



## Personality testing in employee selection

“The central issue that led to the disfavour of personality tests 40 years ago - the lack of predictive validity or the extent to which the assessment predicts job performance - still remains an unresolved issue.”

**Peter Capelli, Professor of Management at The Wharton School**

## Challenges, controversies and future directions

# Overview

---

Personality testing is a big business. It has an estimated market value of £4 billion. There are now hundreds of vendors distributing and selling an array of tests (1,319 at the last count), deployed across a range of applications.

Personal and team development, vocational guidance, career counselling, executive coaching and employee selection draw on personality tests. The outcomes of this testing enterprise shape peoples' lives in important ways.

Personality testing is becoming increasingly common in the recruitment process. A report from CEB indicates that 62% of Human Resources departments now use personality tests to vet candidates during the hiring process.

This article is a response to a question asked by a client:

***“Does personality testing work in selection?”***

And what does it mean for personality testing to “work” anyway? A more helpful question would be: “

*“Which specific tests work in which selection scenarios?”*<sup>1</sup>

Given the vast number of available tests and the different ways in which test data can be deployed within a range of selection processes, this doesn't lend itself to a simple evaluation.

This article therefore addresses the broad sweep of findings in this field, summarising the challenges and controversies<sup>2</sup> for personality testing in the high stakes scenario of employee selection. It reviews the evidence base to examine claims for the predictive validity of personality tests and outlines future directions for improved practice.

As with most things in life, to make sense of the future it helps to revisit the past. This article therefore also looks at the origins and evolution of personality testing.

*“The words “I really like most people I meet” flash on to my screen, as the computer asks me how strongly I might agree or disagree with that statement.*

*Welcome to the world of the psychometric test, for which the top prize could be a bank chairmanship and bags of class A drugs.”<sup>3</sup>*

# In a nutshell

---

The proposal is that self report personality measures have now had the best part of a century to demonstrate their practical value in employee selection. But their initial potential - despite some promising signs - has not been translated into the kind of performance that has had a significant organisational impact.

This article outlines five reasons:

- fundamental problems within the **validation research**
- the realities of **self deception** and why we shouldn't expect too much insight from self report measures
- how **faking** in applicant scenarios affects selection decision making, and why solutions to minimise, detect or mitigate faking have largely failed
- the **lack of theory** to connect personality to performance. The complexities of context, cause and consequence make a simple theory unlikely
- the hazards of integrating personality test data within **selection decision making**

There are three responses to the disappointing predictive power of personality tests:

- **abandonment** to accept that the personality testing enterprise should now be dismantled
- **incremental improvement** to search for marginal gains that will establish personality testing's value in future selection
- a **fundamental rethink** to explore alternatives to the instruments provided by the conventional test publishers

And until we shift to:

- **contextualised and customised measures**
- the addition of **objective metrics** rather than rely only on subjective self report measures
- the greater use of image-based assessments for a more **engaging candidate experience** that avoids repetitive tedium
- personality assessments that are **genuinely inclusive**; tests that aren't designed and validated only with WEIRD samples - Western, Educated, Industrialized, Rich, and Democratic

We can anticipate another century of counter-productive debate and confusing claims in which self report personality test data from applicants continue to account for less than 5% of work performance.

Or - at worst - we apply tests that undermine organisational productivity and innovation.

“While the personality measures used in organisations should do better now than years ago, there is not much evidence that they are better.”

**Kevin Murphy**

# The Psychometric James Randi Award

In 1964, the precursor to the “\$1 Million Paranormal Challenge” was announced by James Randi.<sup>4</sup> Randi was a sceptic of the paranormal who had achieved publicity by questioning the spoon bending claims of Uri Geller.

Randi’s offer: \$1 million to any practitioner of the paranormal who could show, under proper observing conditions, evidence of any paranormal, supernatural, or occult power or event. Over a thousand people applied to take it, but none were successful. The challenge was terminated in 2015.

In a variation of the Randi prize, a post appeared on LinkedIn a few years ago. This Psychometric James Randi Award would be given to the test publisher who could provide evidence of the genuine business impact of any personality test in a selection context.

The conditions were demanding but straightforward:

- a **base rate** of current selection success is available. A benchmark is needed to allow comparisons of improvements in impact
- the results of the personality test were **not made available at the time of selection** to influence selection decision making (and avoid any bias in subsequent evaluations)
- there was a decent **sample size** (modestly set at more than 150)
- successful candidates were tracked, and meaningful **performance data** (linked to tangible business outcomes, e.g. sales, productivity, service) was obtained after a year; objective criteria of work outcomes of organisational value



“Uri Geller may have psychic powers by means of which he can bend spoons.”

*If so, he appears to be doing it the hard way.*

**James Randi**

# The Psychometric James Randi Award

The debate was largely informative and good humoured with contributors posting up references to a variety of research studies. The posts also highlighted some of the methodological issues involved in this kind of evaluation. But a handful of posts suggested that the criteria were unreasonable.

One irate test publisher argued that this kind of research design was unethical. Here the argument ran: *“Of course personality testing works, and it would be highly irresponsible not to use the tests to improve selection.”*

No doubt this kind of research methodology is difficult.<sup>5</sup> But given the proliferation of personality tests over the last century, someone, somewhere must have, for example, conducted a randomised control trial in an applicant context.

As the discussion progressed over several weeks it became clear that no one was either able or willing to provide evidence indicative of **any** incremental gain on current selection processes in an applicant context. This is odd.

The verdict on personality testing in the 1950s to 1970s might have been: *“it is difficult to advocate with a clear conscience the use of personality measures in most situations as a basis for making employment decisions about people.”*<sup>6</sup>

But after past challenges and controversies, the narrative became that several meta-analytic studies throughout the 1990s established the validity of personality instruments.

So confident were the personality testers that one of the pioneers of the new wave of personality research announced that researchers and practitioners can now put *“troubling matters and concerns aside.”*

“One day when I was a junior medical student, a very important Boston surgeon visited the school and delivered a great treatise on a large number of patients who had undergone successful operations.

At the end of the lecture, a young student at the back of the room timidly asked, *“Do you have any controls?”*

Well, the great surgeon drew himself up to his full height, hit the desk, and said, *“Do you mean did I not operate on half the patients?”*

The hall grew very quiet then. The voice at the back of the room very hesitantly replied, *“Yes, that’s what I had in mind.”*

Then the visitor’s fist really came down as he thundered, *“Of course not. That would have doomed half of them to their death.”*

God, it was quiet then, and one could scarcely hear the small voice ask, *“Which half?”*<sup>7</sup>

# The Psychometric James Randi Award

---

For the personality testers, the question is not: *“Do personality tests work?”*

Here the response has been an emphatic: *“Yes they do.”*

And the more pressing questions are now: *“which specific tests work best, where and when?”* and how best to optimise their value in selection.

So why after 5 years does the Psychometric James Randi Award still remain open?

To understand why it's useful to go back in time and look at the beginnings of personality testing in employment selection.

“It is no accident that psychologists associated with the test publishers display the most conviction about the value of personality testing in employee selection.”

**Frank Schmidt**



# “Once upon a time”

---

Personality testing, at least in a form recognisable today - self report through a questionnaire format - can be tracked back to 1917 and the design of an instrument to identify soldiers who might be prone to nervous breakdowns. Originally an interview guide to screen out those candidates who might be emotionally unstable and unfit for active combat, this check-list evolved into a self report questionnaire - the Woodworth Personal Data Sheet.

A pragmatic decision was made: a questionnaire format would be much less expensive and time-consuming than an interview from a military psychiatrist.

In a yes-no format, the test included questions of the type:

- *Does it make you uneasy to sit in a small room with the door shut?*
- *Do you feel like jumping off when you are on a high place?*
- and the extraordinary: *Did you ever have St Vitus’s Dance?* with the helpful note, *(Sydenham’s Chorea - you would know)*

After the 1st World War, the author of the test identified opportunities to apply the test for the corporate sector. This test - “the grandfather of all present day personality tests” - met an emerging organisational need to “*root out the undesirable and unstable workers.*”

The marketing claim was that productivity would increase and the danger of workplace radicalism could be reduced if firms could quickly and efficiently screen out the maladjusted and miserable.

The Woodworth Personal Data Sheet was imitated by a series of competing instruments.<sup>8</sup> The most notable rival, the Bernreuter Personality Inventory (BPI) extended the scope of personality beyond adjustment to include other personality factors.

The testing game was well and truly afoot. And the marketing of these new instruments intensified.

As the BPI established itself as the dominant player in the market, commentators noted the irony in which “*the results of so many studies employing the BPI in industrial investigations were negative.*” Repeated findings showed that the BPI was in fact not doing a good job of predicting employee performance.

The dismal results for the BPI opened up an opportunity for better tests and throughout the 1930s and 40s, a new wave of instruments appeared: the Bell Adjustment Inventory, the Guilford-Martin Personality Inventory of Factors, the Nebraska Personality Inventory and the Humm-Wadsworth Temperament Scale (HWTS).

In a rerun of the use of the Woodworth Personal Data Sheet, the Humm-Wadsworth Temperament Scale was then taken up to screen neurotic and psychotic soldiers during the drafts of World War 2.

In their sweep of the history of personality testing, Gibby & Zicker note that much of the success of the HWTS arose from Humm’s aggressive tactics in shaping debate about research and validation.

*“Although it created friction with journal editors and psychologists, these efforts could have been successful in promoting the HWTS as a valid instrument for industry.”*

More personality tests followed, notably the MMPI, the 16PF and the Guilford Zimmerman Temperament Survey.

# Selecting spies

Another key development in testing in the mid 20th century was pioneered by the Office of Strategic Studies - the precursor to the CIA - and a rethink of the assessment process for the selection of secret agents.

Following the template of the British War Office Selection Boards, a team of psychologists and psychiatrists worked with the military to set up an assessment centre - Station S - a three day event incorporating a mix of individual and group tasks, exercises requiring improvisation, interviews, projective tests and conventional personality questionnaires, including the Myers Briggs Type Indicator.<sup>9</sup>

Between 1943 and 1945, the OSS had tested over 5,000 recruits and *“evolved as the most complex and time consuming personality check ever made in history.”*

Although *“validation problems, inherent in all wartime personnel procedures, plagued the programme”*, mysteriously it was generally viewed by key stakeholders as an improvement of the previous selection system. Importantly the methodology of multiple assessments was to make its way into business, initially as part of a research study, then implemented within the US firm AT & T.<sup>10</sup>



*“In light of our knowledge regarding deception, the research question of “Do applicants fake?” is silly. The better question might be: “Why wouldn’t they fake?”*

**R Griffith & M McDaniel**



# Personality testing as corporate conformity

In his 1956 book, “The Organization Man”, William Whyte observed the rapid take up of personality testing by US corporations. By 1954, around 60% of US firms were deploying some kind of personality test or other. He was not impressed.

Whyte’s argument was that corporate America was going down the wrong path. The pioneering and entrepreneurial spirit of its business leaders was being stifled by conformity and conservatism. Personality testing was one of Whyte’s key targets.

He was alarmed at how tests were being deployed in selection, and even more concerned by the way in which tests were being used to “check up” on current employees.

Organisations were now looking for the reliable, well-adjusted individual who would fit in and could be trusted not to rock the boat.

The result: *“a set of yardsticks that reward the conformist, the pedestrian, the unimaginative, at the expense of the exceptional individual without whom no organisation can flourish.”*

Whyte’s argument was that personality testing was doing the opposite of what was claimed. It was eliminating those candidates with distinctive personality.

Whyte went on to provide a detailed critique of the tests, their lack of validity and in particular the way in which candidate personality test data was compared with a norm group.

For Whyte, this norm is *“the result of the instinctive striving of previous test takers to answer as they think everyone else would answer.”*

So exasperated was Whyte with the use of shoddy and inappropriate tests badly implemented within selection, he created a “cheat sheet” with advice to help candidates out-manoeuvre the test publishers.

This advice included the guiding principle:

*“When in doubt about the most beneficial answer to any question, repeat to yourself:*

*I loved my father and my mother, but my father a little bit more.  
I like things pretty much the way they are.  
I never worry much about anything.  
I don’t care for books or music much.  
I love my wife and my children.  
I don’t let them get in the way of company work.”*

*“It’s inherently sexist to view straightforward women as hostile or rude while approving of men who behave the same way.”<sup>11</sup>*

# In the 1960s, things get worse

---

Martin Gross, in the 1962 “The Brain Watchers”, went further in his concerns about personality testing in employee selection. Revisiting the history of personality testing, he questioned the claims of the test publishers.

The Woodworth Personal Data Sheet as applied in the screening of soldiers in World War 1 was *“a failure. It did not satisfactorily select the psycho-neurotics from “good trench material.”*

The value of personality testing in military selection in World War 2: *“our appraisal of screening in World War 2 points to two conclusions: the screen was not very effective and it had little predictive value.”*

The Office of Strategic Studies programme for secret agent selection: *“none of our statistical computations demonstrates that our system of assessment was of great value.”*

Gross also replays George Bennett’s summary of the field in the Buros Mental Measurements Yearbook:

*“Over the past 40 years a great number of self description inventories have been constructed and tried out. This reviewer is unable to recall a well-established instance of useful validity against a criterion of occupational success.”*

Noting the “boundless enthusiasm” for personality testing, Gross reviewed the issues with personality testing, an analysis that was remarkably prescient in identifying the key issues still being played out in today’s debate.

## **Badly designed and inappropriately applied tests**

Gross reviewed the range of tests, a mix of idiosyncratic instruments derived from a clinical context or psycho-analytical theory as well as more systematic attempts to measure personality, and questioned their relevance to employee selection.

## **The lack of validation to demonstrate their practical value in selection**

Here Gross highlighted that validation claims typically rested on studies with small sample sizes, seldom cross validated and never independently replicated.

## **Candidate gamesmanship in selection**

Like Whyte, Gross argues that the savvy candidate should seek to out-wit the testers. *“Keep in mind that you are being statistically compared with a generation of liars before you. Job hunting is hardly the time to upset your career, and the tester’s norms, with honest neurotic replies.”*

## **The lack of theory to guide how personality shapes work outcomes**

Recalling the IBM psychologist, *“If it comes out that our best people are all 5 feet 3 with green eyes, it’s OK with me”*, Gross anticipates the problems with Big Data and predictive analytics that grab patterns without a defensible theory of cause and effect.<sup>12</sup>

## **The smoke screen of complex statistical methodology**

The *“adroit use of statistics - numbers that can and do lie - to create an aura of scientific precision that in fact does not exist.”* Revisiting the classic “How handsome validity can be distilled from raw nothings by clever maths.”<sup>13</sup> Gross highlights how statistical sorcery can conjure a messy data set into an impressive validity coefficient.

# But haven't things moved on since then?

## That was then, and this is now.

The argument from the current generation of test publishers is that yes:

**Badly designed tests were inappropriately applied.** But we now have access to a range of tests, professionally developed against rigorous psychometric standards. We have shifted away from instruments either designed for clinical use or drawing on outdated psychoanalytical thinking to deploy tests of greater occupational relevance.

**Validation was problematic,** but what would you expect in the early days of a new venture? Recent and sophisticated meta-analysis indicate substantial predictive power. The evidence base has been established for personality tests.

The **use of personality tests in selection was largely hit or miss** as researchers and practitioners grappled with a new methodology. We now have complex analytics and decision making models to inform how to weight and integrate personality data into sophisticated formulae to optimise predictive power.

These responses have some merit.

But what remains puzzling is the ongoing controversy in which:

- academic researchers are unable to agree the value of personality tests in employee selection. Some groups indicate the promise of personality testing in selection. Others argue the evidence base hasn't shifted since the 1960s and their application in high stakes situations still can't be justified.
- publishers and consultants are engaged in a game of claim and counter-claim about the superiority of their tests vs their rivals, but with little supporting evidence.
- practitioners are confused in their evaluation of personality tests, which test to use and how test data should be integrated with other assessments in selection.



# Resolving the puzzle of personality testing

The breakthrough findings of the 1990s<sup>14</sup> have, with further analysis, proven to be highly fragile.

Kevin Murphy makes the point: *“empirical support for the use of personality measures in organisational settings has changed very little since the Guion & Gottier review of 1965.”*<sup>15</sup>

And that *“validity estimates for personality measures are often distressingly close to zero.”*

Any number of innovations in psychometric methodology, insights into personality's impact in the work-place and the impact of on line technology **should** have delivered significant predictive gains. And given personality's well established association with a range of significant life outcomes<sup>16</sup> the limitations of self report personality tests in selection are remarkable. Why?

Five themes are outlined:

- problems within the **validation research**; claimed predictive validity is a game of statistical smoke and mirrors
- the realities of **self deception** and its impact in self report measures
- how **faking** in applicant scenarios operates and its impact in selection decision making
- the **lack of theory** to connect personality to performance; and how the interplay of context, cause and consequence makes prediction difficult
- the hazards of integrating personality test data within **selection decision making** and how best to combine the intuition of expert judgement with predictive algorithms

“Measures of broad personality dimensions (e.g., Big Five factors) show levels of validity in predicting performance but rarely all that far from zero.”

Kevin Murphy



# The validation of validity

---

Validity as face validity - it looks like it should work - in personality testing is easy, both in questionnaire completion for candidates as well as in the report back of the results. The helpful Forer effect comes to the rescue to ensure that candidates will for the most part happily agree with their profiles.<sup>17</sup>

Test validation - methodological design, sampling and the interpretation of the findings - is complex. And part of the confusion in the debate for the Psychometric James Randi Award lay in the use of the word “**prediction**”.

How the test publishers use the “prediction” word is very different to the practitioner expectation.

For the practitioner, quite reasonably, a prediction means that a forecast of an outcome in Time 1 is observed in Time 2. There is no expectation of 100% accuracy. But only that we can anticipate that the probabilities have shifted sufficiently to improve predictive power. And *“how predictable something is depends on: what we are trying to predict, how far into the future and under what circumstances.”*

Most test publishers interpret prediction very differently.

Overviewing the evidence base for personality test validity, a few summary points can be made:

Validity coefficients **tend to be higher when reported by test publishers than by academic researchers**. This may be an issue of sample sizes. It may also be the result of cherry picking positive studies for publication, and putting negative findings in the file drawer.<sup>18</sup>

Validity studies are overwhelmingly based on **a concurrent design**. This research draws on current role incumbents rather than use a controlled trial where applicant test data - not used in selection - then tracks performance over time.

For the test publishers, prediction is largely assumed from a correlation between personality test data and work performance observed in Time 1. Given that incumbents represent a different sample group to applicants - applicants can be assumed to have a different response set in completion - this is a problematic claim of predictive power in real life selection.

Concurrent studies also have the potential to confound cause and consequence. Self confidence may be a correlate of performance for a sales group, but self confidence is likely to be more the outcome of higher sales than a predictive input of sales effectiveness.<sup>19</sup>

Most validity studies **use supervisory ratings as the criterion of work performance**. This is understandable. These appraisals are accessible and inexpensive to collect. But this validation enterprise hinges on line management evaluations, evaluations the test publishers dismiss as highly subjective and inadequate metrics of performance.

When validity is established by drawing on flawed criteria, by what logic is validity demonstrated?

# The validation of validity

## What does a validity coefficient practically mean?

Test publishers typically indicate the value of their instruments with correlation coefficients. For the most part this is a confusing enterprise.

As Kevin Murphy argues: *“personality researchers and practitioners seem to have lost track of just what it means for a test to be valid.”*

Validity is not an intrinsic property of a personality test. Validity operates within a context. It only indicates the strength of the inferences that can be drawn from the use of a test within a specific selection scenario. And without an understanding of the base rate of current effectiveness and the selection ratio in recruitment, generalised claims of incremental predictive gains are misleading.

The type of validity coefficients summarised from meta-analyses may be helpful as a high level indicator of the kind of predictive power that might be expected in **principle**. These coefficients do not determine in **practice** what predictive power will be achieved in a given selection context. Wendall Williams makes the point: *“a statistical meta-analysis can suggest: by golly, it might work. But that is a far cry from proclaiming it actually does work.”*

Validity is intended to provide confidence in prediction. The higher the validity, the more confident practitioners can be in the assessments they make.

Life would be so much more simple and straightforward if the test publishers moved away from the abstraction of the correlation coefficient. The recommendation is to shift towards the kind of predictive meaning provided by natural frequencies, expectancy tables or scatter-plots. Even better to follow the lead of the bookmakers when they offer the odds of success.<sup>20</sup>

## The problem of validity generalisation

Meta analyses is the mathematical procedure that pulls together the results of many individual validation studies to control for any number of methodological problems, not least sample size.

From meta-analytical studies, the validity of an assessment method can be generalised. In this scenario we have confidence that a specific method will be valid across different selection scenarios - roles, levels and industries.

And it was validity generalisation in the 1980s that restored the credibility of previously derided measures of general mental ability (GMA). Here there was compelling evidence that for pretty much any role, GMA would be a useful predictor. And in the latest round-up of the evidence, Frank Schmidt<sup>21</sup> confirms that GMA remains the most consistent predictor of work place performance.

The personality test publishers were quick to jump on the bandwagon of validity generalisation.<sup>22</sup> A personality test that can be trusted to work *“any time, any place, anywhere”* is a test with potential for broad appeal in the market place. But here things proved more problematic. Personality test validity does not seem to transport easily to different organisational cultures or roles.<sup>23</sup>

“You can't fix by meta-analysis what investigators bungled by design.”

**Professor James Coyne**

# Self deception: who's kidding who?

“I am a very stable genius.”

**US President Donald Trump**

In the ideal world of personality testing in employee selection, candidates possess full self awareness of themselves and their personality. In this ideal world, candidates are also motivated to disclose fully this information about themselves to recruiters.

We do not operate in this ideal world.

The first challenge is that “*self evaluation is a difficult task.*” Here the evidence base is fairly robust.<sup>24</sup>

*“We are not very good at judging with accuracy our own competence and character.”*<sup>25</sup>

People on average tend to believe themselves to be above average on positive qualities. And on average, to report themselves as below average on negative qualities. To compound the problem, individuals also believe they are more likely than their peers to provide accurate self assessments.

At one level, accuracy in self evaluation has many benefits. Clear thinking about our strengths allows us to optimise our impact. And a knowledge of our shortcomings helps avoid those situations that might expose our weaknesses, weaknesses that if played out, heighten the likelihood of failure.

At another level, as evolutionary psychology suggests, self insight is unlikely to be a straightforward process. A lack of self insight may in fact be important in the game of survival and success.<sup>26</sup>

If self insight is an issue for everyone, the problem is compounded by the Dunning-Kruger Effect.<sup>27</sup> Here the most self insightful individuals report relatively lower competence than the less insightful individuals who claim superior competence.



# Self deception: who's kidding who?

*"What's curious is that, in many cases, incompetence does not leave people disoriented, perplexed, or cautious. Instead, the incompetent are often blessed with an inappropriate confidence, buoyed by something that feels to them like knowledge."*

Self deception is no trivial issue, particularly at senior leadership levels. "Over-confidence" in CEOs, for example, results in the hubris that underpins disastrous mergers and acquisitions."<sup>28</sup>

And it represents a significant hazard in self report measures.

If self awareness, for example, is one element of emotional intelligence, it is difficult to identify the psychological process in which individuals low in self awareness also go on to report themselves as low in emotional intelligence.

*“Don't accept your dog's admiration as conclusive evidence that you are wonderful.”*

**Ann Landers**

"On one fine morning in Pittsburgh (PA), in the year 1995, a man aged 44, known by the name McArthur Wheeler decided to rob a bank. Since he thought he knew a lot about a peculiar chemical property of lemon juice, he decided to smear the juice on his face before executing his plan to rob the bank.

His logic – as lemon juice can be used to write invisible letters that become visible only when the letter is held close to a heat source, he thought, the same thing would work on his face too. By smearing lemon juice all over his face, he thought that his face would become invisible to the security cameras at the bank.

He did not just think that, he was pretty confident about this. He even checked his "trick" by taking a selfie with a polaroid camera. I'm not sure if the film was defective, or the camera wasn't operating properly, but the camera did give him a blank image. The blank image made him absolutely sure that this trick would work. Or he would not have ever dared to rob a bank with lemon juice on his face.

That day, he went on and robbed not one, but two saving banks in Pittsburgh. A few hours after he had done his job, the police got their hands on the surveillance tape and decided to play it on the 11 O'Clock news.

An hour later, an informant identified McArthur in the news video and contacted the police with the man's name. McArthur got arrested on the same day. Ironically, the same surveillance cameras that he was confident would not be able to capture his face, got him behind the bars. During his interaction with the police, he was incredulous on how his ignorance had failed him."<sup>29</sup>



# If you can't make it, just fake it

---

Self deception addresses the relative inability to complete a personality test in which our subjective perceptions correspond to objective reality.

Impression management is more the motivation to project a positive image to others.

We can describe the intentional attempt to distort responses in personality test completion as “impression management”. Alternatively, we can call it “faking”, estimated to occur with around 30% of applicants.<sup>30</sup>

While over 70% of practitioners view faking as a challenge within employee selection, researchers and test publishers are divided about its impact in recruitment decision making.

## **Faking is not an issue.**

Yes, faking does occur, but it doesn't appreciably affect the reliability and validity of personality testing in selection. This is the stance of “Nothing to view here. Move on.” The argument is that intuitively faking might be expected to undermine the value of personality testing in high stakes selection scenarios. Empirically, it doesn't seem to make that much difference.<sup>31</sup>

## **Faking is a good thing.**

This position is that of course impression management occurs<sup>32</sup> and this is a positive. It is part of the social game in which the savvy candidate gets it, and the guileless don't. After all, the ability to understand and play this game is an important predictor of future performance.<sup>33</sup>

This results in the bizarre conclusion from the advocates of faking good that: *“Sometimes telling the literal truth invalidates the assessment process.”* This is an untenable position. The projection of a best self may be an indicator of social competence for some roles.

But if the claim is that the validity of a personality test hinges on who can best fake, we are in choppy professional and ethical waters.

The expectation *“that peoples' responses to personality test items will create the same impression that their behaviour creates in real life”* is extremely optimistic. There is little evidence to indicate that successful fakers will also fake good in the work place. Even worse, organisations encounter the successful faking of the sociopathic candidate.<sup>34</sup>

In times of declining organisational trust, the typical admonition is that: *“there are no right or wrong answers. Please complete honestly”* is a sham that can only contribute to employee cynicism.

One test publisher provides the advice<sup>35</sup> *“Rule 1: when taking a personality test - be yourself.”* Not helpful advice if your results are going to be compared with *“a generation of liars before you.”*

# If you can't make it, just fake it

---

The applicant perception - informed by social media exchanges of "how to pass the personality test" - is that:

*"You play a game.*

*I know you play this game.*

*And you know that I know how to play this game.*

*And that's how the game will work in future for me to succeed as an employee."*

It does nothing for sustainable levels of employee trust.

In the category of "stuff you can't make up" the keenest advocates of the "faking is a good thing" proposal keep pointing to alarming rates of management failure.<sup>36</sup>

*“The best tests assume that people will lie and fake and take that into account, because faking is a sign of competence.”*

**Tomas Chamorro-Premuzic**  
**Chief Talent Scientist, Manpower**  
**Group**

**Faking is a major problem** in employee selection.

This is the challenge facing practitioners in personnel selection. The research base points to significant hazards of faking within personality testing in selection:

- most personality measures can be faked fairly easily
- a significant number of applicants do fake; *"42% of respondents report they had given false impressions of themselves in the completion of personality tests"*
- 74% of applicants believe that other applicants engage in faking activity
- different applicants employ different faking strategies and adopt different tactics from extreme faking to modest faking
- attempts to detect faking are ineffective and may be counter-productive
- applicant samples produce lower validity coefficients than incumbent samples<sup>37</sup>
- faking affects who does and doesn't get hired<sup>38</sup>

# If you can't make it, just fake it

---

## Responses to faking

Faking is not new. And the problem of faking in personality testing has seen a massive research enterprise to identify strategies that will minimise its prevalence, detect its presence and mitigate its impact.

## Preventing faking

The most common strategy is the use of **warning messages** with the implication that faking behaviour can be detected. The ethics are problematic given the reality that faking can't in fact be detected easily.

In addition, there is a paradox in which, for example, highly rule conscious applicants heed the warnings while those low on rule conscious ignore the message. The outcome: low rule conscious candidates end up being profiled as higher on Conscientiousness than those higher on rule conscious.

Inclusion of **bogus statements** within the questionnaire  
This is the design strategy in which applicants are asked to report their knowledge or experience of non-existent items. When applicants respond in the affirmative, the "logic" is that they are over-claiming and attempting to project a positive impression of themselves. The conclusion is that more research is needed and that the ethical issues need to be resolved.

## Covert rather than overt personality test items

The shift in the 1980s and 1990s towards greater occupational relevance in the design of personality tests for employee selection had the virtue of transparency. But it came with a vice: obvious content that was easily fakeable within candidate selection.

Compare the two types of personality test items:

*"I like the feel of furry silk lined slippers"*; taken from the Grygier Dynamic Personality Inventory.

*"I am highly conscientious"*; a typical item from the mainstream popular tests.

Which is more or less fakeable?

Overt items are the personality statements that are clear to candidates; they immediately understand what the statement is intended to measure.

Covert items are less clear to the candidate and more difficult to work out what is being assessed. And there are indications that covert items are less susceptible to faking.<sup>39</sup>

A psychometric balance needs to be struck in the construction of personality tests. This design philosophy steers clear of those statements that are so obvious in their intent they encourage faking. It also avoids items that incorporate such subtlety they lack face validity and are difficult to justify in personnel selection.

Arguably, the shift towards greater occupational relevance in test design has compounded the problem and tilted the balance toward greater fakeability.

# If you can't make it, just fake it

## Detecting faking

**Response distortion scales**, typically known as Lie or Social Desirability Response (SDR) scales, have been used extensively in personality measures. Around 85% of popular personality tests deploy some variation of these scales. The aim is to “*trap dishonest respondents in their deception.*”

Social Desirability Response scales consist of highly unlikely statements along the lines of “*I always keep my promises*” or “*I never tell a lie*” with the assumption that highly positive statements if endorsed, and negative statements aren't endorsed, indicate a candidate who has faked the rest of the personality test.

The give away for the savvy candidate is in the frequency qualifiers.

The research indicates a number of hazards with this approach.

Firstly, the scales themselves tend to be **easily fakeable**. And high SDR scores confuse intentional faking from the naïve candidate - who is unable to rumble the wheeze - with the candidate who is “*blessed with an abundance of socially desirable attributes.*”

Another confounding factor is the candidate interpretation of the statements that make up the scales. Faking may get mixed up with the extent to which some candidates are literal minded and refuse to endorse any “never” or “always” statements versus those candidates who go with the flow of the sentiment behind the statement.

Second different **SDR scales do not correlate well**. “*Different detectors of faking result in different classifications of respondents. Whether an individual is classified as having faked does not necessarily depend on their actual faking behavior but on the method utilised for detection.*”

When the validity of scales designed to detect faking is low, it is difficult to justify their usage.

Third, high SDR scores are also **associated with the traits associated with work performance**.

Even if Social Desirability Response scales did work, the practical issue is then how candidate scores are deployed in selection. The options:

- remove high scoring SDR candidates from the applicant pool
- apply some correction formula to recalibrate the personality test data
- ask high scoring SDR candidates to recomplete with the implication that the first completion was a dishonest response

None of these strategies seem defensible.<sup>40</sup>

The **removal of high SDR candidates** has a negligible effect on both criterion-related validity and job performance.

**Adjustments for high SDR scores** appear to make little difference.

**Retesting candidates**. The retesting of candidates may make a modest difference. Practically it may be difficult to implement, and not without ethical and legal consequences. Non faking candidates who do register high on SDR scales - the false positives - may resent the accusation of deception and seek legal redress if they are unsuccessful in selection.<sup>41</sup>

# If you can't make it, just fake it

---

In the latest systematic review of the research<sup>42</sup> the conclusion:

*“The use of social desirability scales to index faking behaviour is inaccurate, and the use of social desirability scales to correct personality scores may do more harm than good.”*

## Forced choice design

In the 1980-90s, personality test designers found themselves between:

- the rock of occupational relevance but transparent fakeability
- the hard place of non-obvious subtlety but indefensible face validity

Forced choice measurement was seen as one way out of the dilemma. Forced choice formats come in different shapes and sizes, but typically provide pairs, trios or quartets of statements matched for social desirability or undesirability. Candidates are required to select the statement that is more or less descriptive of them. The logic is that since candidates cannot select all the positive statements or avoid all the negative statements, forced choice tests should be less susceptible to faking.

For savvy candidates test taking moves up a level. Not only do they need to avoid detection by keeping their SDR scores low, they also need to work out which statements to rank given their insight into the traits of most job relevance.

Overlooked in this shift towards ipsative forced choice formats for personality testing was the impact on conventional psychometric statistics and measurement properties. Also forgotten was the questionable assumption that ipsative test data can be treated as equivalent to normative data as obtained from the typical rating response task.<sup>43</sup>

But the downsides of ipsative forced choice instruments (problematic psychometric properties and the hazards of candidate comparison) could just about be justified if:

Faking is prevented.

Predictive gains are demonstrated.

## Have forced choice designs mitigated the faking problem?

The jury remains out. Different research and practitioner groups report conflicting findings. In one summary of the evidence base:

*“the empirical literature has not produced persuasive evidence for the elimination of faking amongst highly motivated test takers when forced choice formats are used.”<sup>44</sup>*

## What predictive gains have been demonstrated?

Here again the evidence is inconsistent. The suggestion is that any improvements in validity are less about the personality of candidates than their level of general mental ability.<sup>45</sup>

The conclusion seems to be that when any modest gains in predictive ability in selection are claimed, these are largely the result of candidate cognitive skills to work out the necessary response pattern for the role.

# If you can't make it, just fake it

To summarise the research on faking and its impact within employee selection.

## Faking Performance = Opportunity X Motivation X Ability.

**Opportunity** is largely based on a combination of the transparency of the personality test and the obviousness of any SDR scale that is used in the pretence of detecting faking. Where questionnaire content is apparent and the SDR scale items lack subtlety, we can anticipate successful (undetected) faking. In the scenario where the personality inventory has finesse and nuance in item writing, we can expect greater honesty.

**Motivation** is shaped by the purpose of the personality test. Where results are confidential to the individual and utilised as part of a personal development application, we can assume that faking will be less frequent than in the high stakes scenario of employee selection.

The **Ability** to fake is complicated. In “The “g” in Faking”<sup>46</sup> the evidence indicates that intelligence predicts successful faking ability. The implication is that any incremental validity of personality testing in forced choice formats is due to cognitive aptitude rather than any variance in personality.

Any defensible method to detect faking must:

1. Have a **high detection rate** and avoid false positives. As McCrae & Costa highlight: “*Paradoxically it is the most honest and upstanding citizen that these scales would lead us to accuse of lying.*” When the frequency of false positives is higher than the true positives, there is a selection problem.
2. Scores **should not capture relevant trait variance and eliminate top candidates**. Given that emotional stability, agreeableness and conscientiousness correlate with SDR scores, traits often associated with work performance, the risk is that highly suitable candidates will be ruled out of the selection game.
3. The method should **not be coachable** to allow applicants to spot the detection tactic in a selection situation. In the world of social media and the many guidelines of how to “fake the test”, this is unlikely.

No method meets these three criteria.

“Much like the mythical unicorn, a solution to the faking problem in personality testing has been an exciting but elusive quarry.”

**Michael Zikar & Katherine Sliter**

# In theory, in practice

“What we can say with confidence is that personality is a more complex and versatile phenomenon than is reflected in a personality test score.”

**Luke Smillie**<sup>47</sup>



## Levels of personality

The personality testing industry has been dominated by trait theory. Having overcome the situationalist critique of the 1960s and 70s<sup>48</sup> the trait enterprise in the 1980s rediscovered its mojo to develop a range of new personality tests. More recently, several personality publishers have integrated competency based approaches with the trait perspective.<sup>49</sup>

The trait approach assumes that as individuals we have deep seated temperamental dispositions that shape in important ways what we do and how we do it. These traits are played out in significant life outcomes.<sup>50</sup>

There remains however debate about the best way to map out these personality traits. The Big 5 model (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism) at one point seemed to have achieved reasonable consensus amongst researchers. But this has been recently challenged.<sup>51</sup> The Big Five may be problematic across different cultures, income levels, educational achievements<sup>52</sup> and age groups.<sup>53</sup>

And there is the suggestion that an additional sixth factor (honesty - humility trait) might provide a better alternative.<sup>54</sup> Or even that there is only one single personality dimension as a kind of “p factor” equivalent to the “g factor” in the ability domain.<sup>55</sup>

Perhaps the issue is more fundamental. The assumption that traits can describe personality adequately may be flawed. Dan McAdams<sup>56</sup> makes the point that traits represent only one level of personality. When we apply the language of the Big 5 this is a mental map to help us evaluate others. For example, is this individual more or less agreeable (a low or high threat?) For McAdams, *“a good trait analysis would appear to be little more than a systematic psychology of the stranger.”*

# In theory, in practice

Not trivial, but traits seem to provide only a limited map of personality.

McAdams argues that to understand personality we need to shift to another level. This is to recognise the individual's motivations and aspirations around the range of life's different tasks. Personality becomes the strategies used to navigate through these challenges and to overcome threats. This moves us from the "psychology of the stranger" to a more detailed and nuanced understanding of the individual as an individual and to ground traits in the context of the individual's life goals, plans and tactics.

There is a third level which addresses personal identity and the narratives we use to understand ourselves and others, and the stories we tell to make sense of our past, present and future.<sup>57</sup> This gives a more textured perspective of the individual to understand them as the "story tellers" of their lives.

## Measuring performance

If personality - and how best to conceptualise and measure it - is still in dispute, the challenge is compounded by the issues of performance evaluation.<sup>58</sup> If we can't agree on the criteria of performance - what is being predicted in the first place - it seems difficult to know where to start in the construction of predictors. Here the question is: "*what constitutes job performance?*"

Michael Raynor makes the point that much of the business success genre is looking in the wrong place for genuine metrics of performance.<sup>59</sup> If this is true at an organisational level, it seems particularly applicable for individual performance. Even more awkward: there is reason to believe that some organisations positively reward "bad" behaviour.<sup>60</sup>

*“More care was given to new predictors than the criterion against which they were validated.”*

## Linda Gottfredson

The standard "go to" criterion of work performance has been supervisory ratings. This may well be problematic.<sup>61</sup> When line management ratings of individual effectiveness are only modestly correlated with those of the individual's peers and team members, who is providing the most accurate feedback of performance?

Is validity only in the eye of the beholder?

The solution to the subjectivity and low validity of line management evaluations has been the search for objective performance metrics. And for many roles this has been relatively straightforward. Data about sales, service responsiveness and productivity, etc. can be accessed across a spectrum of jobs.

But not without hazards. When for example, a surgeon's performance is evaluated and rewarded against patient survival rates, an improvement could be anticipated in overall mortality rates. Not necessarily.

Not when surgeons game the system by choosing to operate only on those patients they know have a higher chance of survival and discriminate against those patients they think won't make it through surgery. Here health outcomes are distorted.



# In theory, in practice

Are error rates, for example, within nursing teams an indicator of an under-performing team? No. Paradoxically nursing teams with better manager-nurse relationships report more errors. These teams have the confidence - unlike the less effective teams - to acknowledge any mistakes and review the reasons for future improvements.

In “The Tyranny of the Metrics” Jerry Muller points to the many perverse incentives and negative unintended consequences associated with the introduction of quantifiable measures of performance. And for many roles, objective performance metrics are difficult to access with any confidence.

When Fred Goodwin was selected as Forbes Global's Business Leader of the Year in 2002, was he an excellent CEO? Or a dismal performer when his firm RBS collapsed in 2008, triggering a government bailout to the tune of £45 billion?

Was Jack Welch the greatest business leader of the 20<sup>th</sup> century? Or an exploitative leader whose focus on delivering short term earnings brought about the decline of GE?<sup>62</sup>

“*Measure what you can, evaluate what you measure, and appreciate that you cannot measure the vast majority of what you do.*”

**Ed Catmull, President of Pixar**

## The personality - performance linkage

The standard assumption is that specific attributes of personality increase the likelihood of the behaviours associated with greater effectiveness in the tasks that impact on the outcomes of importance to the organisation.

It is proving difficult however to identify a straightforward dynamic between personality inputs and performance outputs across a range of roles.

There is any interplay of moderating and mediating factors at work in which:

- **context** - the permutations of strategy, structure and culture that set the environment for levels of performance<sup>63</sup>
- **causal** factors - the interactions of experience, cognitive aptitude, motivation and personality and how they combine for different performance outcomes
- **consequence** - the different drivers of short vs long-term metrics of task and contextual outcomes

Context, cause and consequence are proving extremely difficult to disentangle.

# In theory, in practice

---

Any practitioner who assumes a tidy linear one-directional causal effect between personality and performance will be disappointed. Although general mental ability largely operates in a simple way - more input results in more output - there is no reason to assume this is the case for personality attributes.

Instead there are any number of interactive effects. For example, extraversion may lead to higher levels of performance for conscientious employees. But extraversion is associated with lower performance when conscientiousness levels are low.<sup>64</sup>

To add to the complexity, Stephen Woods<sup>65</sup> points to the evidence that it is not simply a case of personality impacting on performance. Work outcomes also shape the development of personality traits. This research insight shifts the debate to explore how work place performance also impacts on employee personality and its development.

“*The concept of validation would collapse to a futile exercise if the criterion measures were idiosyncratic to particular raters or specific alternate indices of performance.*”

**C Viswesvaran & D Ones**

A “good” theory is:

- **simple**; it is easily understandable
- **accurate**; it explains and predicts
- **generalisable**; it works across a range of different contexts

The “theory” to connect the context, causes and consequences of the personality-performance nexus meets none of these criteria. The linkages between personality and performance seem to be:

- highly **complex**, involving any number of mediating and moderating factors
- **imprecise** in making future predictions of work performance
- highly **specific** to the context in which personality affects performance outcomes

The test publishing industry can continue “as is” in the claims it makes of the power of personality in predicting future performance within pretty much any selection scenario. The evidence base indicates that these publishers lack convincing credibility.

As James Clerk Maxwell notes: *“There is nothing more practical than a good theory.”*

To establish the credibility of self report personality testing as a valuable approach in employee selection, a rethink of the interplay of context, cause and consequence is needed for a good theory.

# The so what of decision making

---

“To explain the past we need complex rules.”

To predict the future, less is more.

**Gerd Gigerenzer**



## A thought experiment

For one moment, put aside any concerns about self deception and impression management in candidate selection. Forget the complexities of validity and the messiness of the evidence base. Also ignore the challenges of linking personality inputs to performance outputs.

Imagine that personality is now an established predictor of performance.

## *How practically should personality test data then be utilised in selection decision making?*

Different models can be applied, spanning the spectrum from the intuitive to the empirical.

The most common approach is for personality tests to be used in an informal way within the selection process.

A profile is used to inform interview priorities and coverage to explore how the candidate's personality may shape their work approach and performance. This analysis is interpreted by the assessor (or expert system) <sup>66</sup> to feed into a selection recommendation and/or decision. Here judgement is applied to integrate the results from the personality test with other assessment data to build a better understanding of the candidate.<sup>67</sup>

This seems a low risk strategy. But with a caveat.

As Steve Blinkhorn points out: *“As a candidate I would rather complete a “bad” assessment and be interviewed by a “good” assessor, rather than undertake a “good” assessment with a “bad” assessor.”*

# The so what of decision making

---

Further up the spectrum from the intuitive to the empirical is the **ideal profile**.

A personality profile is created to reflect those personality dimensions that are seen as key to success for a specific role. We know that super-productive software developers share some personality characteristics. This should provide important insights to assess the suitability of candidates for a software developer role.

The challenges are in creating this ideal software developer profile. Software developers come in any variety of shapes and sizes. Here job analysis should be helpful, but often makes unwarranted assumptions about the personality characteristics for a specific role, especially if that role is relatively unusual. It is also a challenge for newly created roles where there is lack of information about who has been more or less successful within the role.

## Danger zone profiles

These profiles map out thresholds of low and high scale scores that may be indicative of risks within a candidate's personality profile. Danger zone profiles operate on the logic of the Goldilocks principle where too little or too much of any personality trait can be counter-productive. These can either be generated by expert judgement or from a local validation study.

At best they provide a risk profile to highlight any potential constraints to the individual's work effectiveness. In practice, these profiles are derived from small sample sizes that fail to replicate in cross validation studies.

Paul Barrett argues that whilst the intuitive approaches should have value, there is little evidence to substantiate these claims.

Or, as Yogi Berra observed: *"In theory there is no difference between theory and practice. In practice there is."*

In **theory**, the deployment of personality testing as part of the selection interview has intuitive appeal. Disentangling test data from expert judgement to evaluate relative predictive gain has been problematic in **practice**.

## Predictive algorithms

In the classic "Clinical versus Statistical Prediction"<sup>68</sup> Paul Meehl studied the successes and failures of predictions in many different settings. He found overwhelming evidence that predictions based on statistical scoring were generally more accurate than those based on expert judgment.

This is reminiscent of the kind of specification equations of Cattell's 16PF in the 1970s in which various traits are combined in a mathematical formula. For example, in the selection of sales people:

*"Success = .44A - .11B - .22E + .11F + .22H....."*

Future performance is predicted by weighting specific personality factors. Largely a failure at the time, this approach has now been revitalised with the recent wave of the predictive analytics that Big Data promises.

# The so what of decision making

There is therefore now a shift away from the informal judgement of the assessor to more systematic ways of utilising personality data in selection. Organisational recruitment practice is moving to the deployment of algorithms that can measure candidate suitability, particularly at the pre-screening phase of the selection process.

There are two concerns with the deployment of Big Data predictive analytics in employee selection:

## Do these analytics actually work?

*“Just use computer-based regression and correlation analysis to find statistically significant influences, then combine them to get a perfect fit to the data. When a data set is left to speak for itself like this, it typically spouts nonsense.”*

Here the suspicion is that often proprietary algorithms derived from Big Data are based on the art of finding spurious correlations<sup>69</sup> to claim predictions that fail to replicate with a different data set. If Big Data is in fact Bad Data, random patterns simply result in highly unstable equations with highly questionable selection outcomes.

Less well understood in the application of analytics is the predictive paradox. When algorithms do improve accuracy and these formulae are used in selection decision making, the model then provides less future predictive power.

Gerd Gigerenzer<sup>70</sup> also makes the point that complex decision algorithms in selection are highly “fragile” and sensitive to change. A complex formula might improve predictive power for a specific role, but any modest shifts in role requirements will undermine its predictive accuracy. Statistical methods for selection may be more robust in low change environments, but shaky in the context of change and uncertainty.

Lyle Ungar argues: *“the bottom line is that if you have lots of data and the world isn’t changing too much, you can use statistical methods. For questions with more uncertainty, human experts become more important.”*<sup>71</sup>

It was Paul Meehl himself, in a not well remembered observation, who noted: *“Shall we use our heads or shall we follow the formula? Mostly we will use our heads because there just isn’t any formula.”*

Or as Robert Matthews puts it: *“in the face of uncertainty about the model, the shiny toys of fancy maths have to give way to experience and judgement.”*

*“With so many reputations resting on regression, it will be a brave researcher who decides to find out how much is baloney.”*

**Professor Robert Matthews**

# The so what of decision making

---

## 2. Are there any negative unintended consequences associated with this approach?

Predictive algorithms are not without their hazards. When Amazon<sup>72</sup> adopted Artificial Intelligence to review job applicant resumes, the intention was to widen the talent pool by scanning the internet for suitable candidates.

*The consequence: the new recruiting engine did not like women.*

Amazon's machine learning was trained to check applicants by observing patterns in resumes submitted over a ten year period. The problem for Amazon was that the training set was overwhelmingly based on a male data set, a reflection of male dominance across the tech industry.

Amazon, after several attempts to fix the problem, abandoned the project.

Jordan Weissmann<sup>73</sup> notes: *“What happened at Amazon really highlights that using such technology without unintended consequences is hard.*

*And if a company like Amazon can't pull it off without problems, it's difficult to imagine that less sophisticated companies can.”*

The argument is not that that Big Data and predictive analytics have no place in employee selection. They do. And we can anticipate more take-up in screening systems in future selection systems.

Instead the recommendation is to encourage greater diligence in the design and implementation of these statistical approaches, as well as attention to monitor their longer-term impact.

Neither is the contention that predictive analytics are inherently discriminatory.<sup>74</sup> But badly implemented, there is much scope for legal challenge and reputational damage.

The trade-offs:

Intuition based on expert judgement applied within personality testing in employee selection is low risk, but its value may well hinge on the wisdom of the assessor.

Empirical approaches that deploy statistical methods should be an improvement on intuition. But there is remarkably little evidence in employee selection to indicate this is the case. Algorithms to predict the future based on past patterns drawn from an out-dated dataset will simply replay irrelevance or bias.

It may be the *“combination of statistical and judgemental methods is associated with the highest predictive accuracy.”*<sup>75</sup>

# Caution and conclusions

In reviewing the research evidence for personality testing within employee selection, the “caution” word is alarmingly frequent.

- caution in the **analysis of research findings** and in the manuals and white papers provided by the test publishers, and the test reviews summarised by professional bodies.
- caution in the **interpretation of validity evidence** and the extent to which the results from a concurrent design can inform practice within a specific selection scenario.
- caution over the **usage of response distortion scales**. The tactics to identify and mitigate faking are ineffective and should be jettisoned.
- caution in how **personality test data should be weighted** and incorporated with a selection decision making process to either screen out or select in candidates.
- caution in responding to the challenges around **privacy and the legal defensibility** of personality test usage.



If after 100 years of effort in the design, validation and implementation of personality tests, “**caution**” remains the dominant recommendation, then it may be that the “game is up”. There are three stances for the future of personality testing in the high stakes scenario of employee selection.

## 1. Abandonment of this enterprise

There remains a lack of consensus within the informed community of researchers and practitioners about the predictive validity of self report personality instruments. Debates that go back to the mid 20<sup>th</sup> century still remain unresolved. And despite a massive research programme that has generated several 100,000 research articles over the last few decades, there is no indication that these issues will ever be resolved.

There are two positions within the abandonment argument. The more **forgiving perspective** is that personality testing adds little value in selection decision making. It probably incorporates unnecessary organisational time and cost, but it is relatively harmless.

The **tougher contention** replays the position of William Whyte and Martin Gross to suggest personality tests are positively detrimental to selection of the talent that drives innovation. Faking, for example, is becoming more prevalent, and the proposed solutions to mitigate its effects, may in fact be undermining validity in selection. Here the efforts of the test publishers are often now the problem not the solution.

Personnel selection advances by discarding those methods that have no predictive value. Graphology, for example, was once marketed as an important tool in employee selection. Despite the best attempts of a few advocates, it has largely disappeared from the assessment landscape.

# Caution and conclusions

---

The experimental and innovative is to be encouraged in exploring how to optimise selection processes. However there comes a time when specific methods have to be thrown aboard. The abandonment perspective suggests that self report personality measures now go into the selection archives of emerging potential that lacked future performance delivery.

## 2. Incremental gains will improve validity

Kevin Murphy asks: *“The more interesting question is not whether personality tests show some validity, but rather why they do not do better.”*

This more up-beat argument claims that there is sufficient promise from the research base for possible improvements in self report personality test validity. This is the argument of marginal gains which, if implemented by test publishers, will establish personality testing as a valuable tool in selection. For example, the shift towards bespoke tools targeted at specific occupational groups may well improve validity.

This strategy also appeals to practitioners to draw on a broader portfolio of tests for different applications rather than fall back on the assumption that every selection problem is a nail that needs to be hammered with one favoured test.

As part of the marginal gains argument, there is also a greater need to work backwards from the criterion of job performance rather than start with the solution of a personality assessment. Too often stereotypes of work success shape the selection blueprint with unwarranted assumptions made about the personality variance that is genuinely associated with effectiveness in the work place.

This stance also argues for greater methodological rigour in building the evidence base for personality test validity. A shift in thinking about the construction of new personality tests will also help.

But personality tests in selection can only become credible as valid instruments when the wider issues facing the psychology profession are addressed.<sup>76</sup> Here there is a move towards the better design of validation studies and an accommodation of the reality of situational specificity rather than the reliance on claims of validity generalisation.

Another key element for practitioners is access to a more informed understanding of how predictive validity is determined and reported. Validity is not the absolute property of a personality test. Any claims by a personality test publisher of 90% predictive accuracy should be a red flag of naiveté or duplicity.<sup>77</sup>

The incremental gains philosophy will also find better ways of integrating personality test data within the sequence of selection decision making. For example in identifying how to optimise personality profiles in pre-screening.

To turn Tolstoy’s famous quote about families on its head:

*“Unsuccessful people are more alike than successful people who are successful in their own way.”*

Failure is easier to predict than exceptional success.<sup>78</sup> Personality testing may be better positioned, for example, as an aid to screen out rather than hire in candidates.



# Future directions: contextualised applications

“Among the competing products developed by psychologists, perhaps the most important are their assessment instruments.”

*Unfortunately, in psychology we have no consumers union to test competing claims and to compare these products on their overall effectiveness.*

*The testing industry provides minor cosmetic successive variants of the same product where only the numbers after the names substantially change.*

*These variants survive because psychologists buy the tests and then loyally defend them.*

**Sternberg & Williams<sup>79</sup>**

## 3. A fundamental rethink

This is the proposal that personality does shape behaviour and performance in the work place. But the off the shelf self report measures available from the conventional test publishers are ineffective in tapping into this dynamic. Despite changes in questionnaire content and response formats, the attempts of quasi-ipsative measures to manage faking, or the deployment of selection decision-making algorithms, conventional self report personality questionnaires have inherent flaws in employee selection that limit their predictive power.

What are the options for practitioners?

### The shift towards contextualised applications

There was a time when the only cost effective option was the deployment of generic instruments. The claim was these personality tests provided tried and tested solutions with known psychometric properties, access to an extensive normative database and established predictive validity. This claim now looks misplaced.

There is reason to think that a move towards more bespoke design targeted at specific occupational groupings and roles represents a more promising way forward.

Contextualisation helps. *“Adding context doubled the validity of personality measures.”<sup>80</sup>*

Asked *“What recommendations would you give about the use of personality tests in selection contexts?”*

Neal Schmitt’s response: *“First avoid published personality measures in most instances. Second, I would construct my own measures that are linked directly to job tasks in a face-valid or relevant fashion.”<sup>81</sup>*

# Future directions: contextualised applications

There is still some debate over whether broad or narrow constructs provide greater predictive power. On the one hand, *“narrow traits seem to add incremental validity to the Big 5 in correlations with managerial success.”*<sup>82</sup>

On the other hand, there is a question mark over narrow measures. Many of the “traits” measured by personality scales may simply not exist.<sup>83</sup> Scales based on items that are near synonyms might be a common test strategy to boost reliability estimates. But these scales may simply show that candidates understand the meaning of words and can respond consistently.

Nonetheless, bespoke development has become a more realistic testing strategy given innovations in open item banks, agile project management and a new generation of on line technology tools.

And the practitioner market is showing signs of impatience with the “take it or leave it” philosophy of the conventional test publishers.

The fundamental shift in the market place is now towards contextualised and customised applications to move away from generic measures that claim to work in any selection scenario.

“*Tailor made, context specific measures may yield superior validity to the Big 5 measures.*”

**Neal Schmitt**



# Future directions: objective metrics

## The utilisation of objective measures

In “Everybody Lies” Seth Stephens-Davidowitz<sup>84</sup> builds on the thesis of Peter Thiel who noted that Facebook and Netflix have built their businesses on the principle: *“Don’t trust people on what they say. Trust what they do.”*

Or, as the Netflix data scientist Xavier Amatrian notes: *“The algorithms know you better than you know yourself.”*

Asking individuals, for example, about their sexual preferences reveals one story. Accessing their browser search history provides a different account. This approach indicates that subjective self report measures need to be replaced by more objective metrics of personality.

There are any number of “facts” that we can access about an individual’s personality. “Snoop”<sup>85</sup> proposes we can obtain important insights *“by looking at stuff. Stuff in offices, bedrooms, cars and bathrooms. What’s there and how it’s arranged can provide clues about who we are and what’s important to us.”*

Not advisable in employee selection. But the concept has some merit provided we avoid stereotypes. And we can also review candidate suitability by “snooping” on social media posts.<sup>86</sup> Again not without its risks.

Alternatively, we can apply the kind of linguistic analysis in personality profiling as pioneered by IBM’s Watson.<sup>87</sup> Here Watson analyses text data from emails, blogs, tweets and other social media to generate a candidate personality profile. At this stage, this is an enterprise based more on trust rather than evidence.

*“Hiring a private detective to shadow a candidate would also gather public information that might be relevant, yet most people would view it as an unacceptable invasion of privacy.”*

**Peter Cappelli<sup>88</sup>**



# Future directions: objective metrics

HireVue<sup>89</sup> in its selection systems, deploys proprietary algorithms to analyse facial movements, word choice and speaking in candidate webcam interviews to generate an employability score.

But it is difficult to evaluate the predictive accuracy of this approach.

*“500,000 data points to provide superhuman precision and impartiality to zero in on the ideal personality”* seems impressive. But as Carl Sagan reminded us: *“Extraordinary claims require extraordinary evidence.”* The extraordinary evidence is unavailable.<sup>90</sup> Here the black box of the predictive formulae is inaccessible for any meaningful evaluation of validity or fairness.

Whatever the level of predictive accuracy is claimed or found, this methodology is not without its critics. Meredith Whittaker at the AI Institute suggests: *“It’s pseudo-science. And a licence to discriminate.”*

And it is clear there are genuine ethical difficulties with a tool that doesn’t explain its decisions or give candidates their assessment scores.

Given the status of the evidence base<sup>91</sup> and concerns that unsuccessful candidates will challenge selection decisions, this approach to personality testing requires a robust risk assessment.

But the principle is sound. Objective and verifiable metrics<sup>92</sup> should outperform subjective self report measures. The challenge is ensuring that the shift towards the capture and utilisation of these kinds of metrics in employee selection is defensible. Defensible as genuine predictors of future performance. And defensible in complying with data protection and privacy legislation.



# Future directions: revisiting projective tests

## The rediscovery of projective tests

Graphology, Rorschach ink blots, the Thematic Apperception Test, the Sentence Completion Test and more seem to have gone into the archives of: “selection stuff that might work but didn’t cut the validity mustard.” Nathan Carter<sup>93</sup> outlines a convincing case why we should revisit the concept of projective testing.

Experiments to translate these kinds of projective measures from clinical applications to organisational contexts in the 1960s had mixed results. But Carter argues there is sufficient promise to suggest that practitioners should continue to explore projective methodologies.

Arguably most projective measures were developed from out-dated psychological theory - often psycho-analytical. A move towards more contemporary perspectives on personality that avoids any intrusive assessment may represent a more promising direction.

Along these lines, the Cambridge Code’s personality test - “An X Ray of the Subconscious Mind” - includes the question: *“Have you ever had an imaginary twin?”* An interesting line of direction, but the claim of: *“proven to uncover the subconscious latent potential”* seems premature.<sup>94</sup>

The rediscovery of projective tests also reflects the trend towards more image based approaches in personality testing.<sup>95</sup> This new wave of methodologies may well improve candidate engagement, reduce user fatigue and shorten completion times. Visually based assessments may also have the potential to be more accessible to those with language or learning difficulties or for candidates from different cultural backgrounds.

A similar design philosophy is also used in the application of gamification<sup>96</sup>, now increasingly deployed in assessments of cognitive aptitude and personality.

No doubt there is much vendor hype<sup>97</sup> about new developments in image based tests. But greater innovation in personality assessment through advances in, for example, Virtual Reality seems likely.<sup>98</sup>

“Traditional tests seem increasingly out of touch. Why?”

*Because the user experience is often incredibly poor. Many tests still involve almost always endless word-crunching.*

*The bigger problem is that they don’t tap into what we humans are best at: visual processing.*

**Heather Myers**

# Future directions

---

It may be that DNA testing is the future for personality assessment in employee selection.<sup>99</sup> For any number of reasons, this is doubtful. We need to look for innovation elsewhere.

In “Pitfalls of Personality Theory” Colin Cooper indicates that complacency is rarely the route to progress.<sup>100</sup> And arguably the personality testing industry has been complacent. In a replay of the mid 20<sup>th</sup> century, today’s publishers have relied too much on the recycling of a previous generation of self report measures.

This strategy has failed. But the gap between the hype of marketing claim and the evidence of predictive validity has widened.

It is unlikely that future innovations - in our understanding of personality and how it shapes performance outcomes, or through the deployment of new technology - will transform the predictive power of personality measurement in employee selection.

Personality - unlike intelligence - is an especially unruly selection method. Modest improvement seems a more realistic goal. But **validity** is not the only issue in how we apply personality measures in selection.

The **engagement of candidates** in the recruitment process is also critical. What does the choice of personality assessments indicate about the organisation? Quick, fast paced and interesting tests in employee selection send out one cultural message. Time-consuming, repetitive and dull tests project a different message about the firm’s culture.

And how does our personality testing strategy engage the candidates we want to attract?

The other key consideration in the choice of assessment methodology is the agenda for **greater inclusion and diversity**.

Much of the rationale behind the use of personality tests in employee selection was the claim of a reduction in adverse impact vis a vis for example cognitive tests.<sup>101</sup> This argument is now challenged.<sup>102</sup>



# Future directions

“Companies that emphasize neuroscience, big data, and gamification may be trying to distract you from the fact that their assessments don't predict workplace performance.”

Ryne A. Sherman<sup>103</sup>



The use of current self report personality tests from the main stream test publishers in the high stakes scenario of employee selection is now highly questionable.

Until we shift to:

- **contextualised and customised measures**
- the addition of **objective metrics** rather than rely only on subjective self report measures
- the greater use of image-based assessments for a more **engaging candidate experience** that avoids repetitive tedium
- personality assessments that are **genuinely inclusive**; tests that aren't designed and validated only with WEIRD samples - Western, Educated, Industrialized, Rich, and Democratic<sup>104</sup>

We can only anticipate another century of counter-productive debate and confusing claims in which self report personality test data from applicants account for less than 5% of work performance.

At worst, the continued usage of conventional personality testing in selection will undermine business productivity and innovation.

# Notes

---

This article is not an academic paper. These notes therefore do not represent a comprehensive set of references.

Instead they provide a mix of additional commentary with a set of links to further research that struck us as interesting as we consolidated key findings in this field.

Many thanks to the academic researchers and expert practitioners who reviewed earlier drafts of the article to point out shortcomings as well as provide constructive feedback and suggestions for improvement.

This piece should be seen as a work in progress iteration to summarise the current state of play. We welcome any feedback from both researchers and practitioners within the field of personality testing in selection. And we will update as additional findings are sent on.





# Notes

---

1. Which personality test is best was once a much loved topic of LinkedIn discussion groups, an exercise in which the vendors made extravagant claims of their specific tests. This is not new. The world of personality testing has been characterised by debate about the best test. But this seems the wrong way to think about personality testing. Different tests seem to “work” better for different purposes. The practitioner search for the “best test” may be a factor in explaining why gains in predictive validity in selection have been limited.

Reviews of personality tests:

A Review and Comparison of 12 personality Inventories on Key Psychometrics, Prewett et al; Handbook of Personality at Work

The Comparative Validity of 11 Modern Personality Inventories: Predictions of Behavioral Acts, Informant Reports, and Clinical Indicators  
[https://projects.ori.org/lrg/PDFs\\_papers/Grucza&Goldberg\\_2007\\_JPA.pdf](https://projects.ori.org/lrg/PDFs_papers/Grucza&Goldberg_2007_JPA.pdf)

2. Controversies. Controversy has been a running theme in personality testing over the years. The recent episode of Cambridge Analytica’s personality test and the harvesting of personal data is the latest in a long line of challenges about how test data is captured and used.  
<https://slate.com/technology/2018/04/how-corporations-convincing-us-that-personality-tests-are-fun.html>

3. “Disgraced former Co-op Bank chairman Paul Flowers did very well in psychometric tests in interviews for the role, a committee of MPs has heard. Mr Flowers, a Methodist minister with little experience in banking, became chairman of the Coop board in April 2010, with a disastrous outcome; “a rescue deal was required with bondholders after it emerged the bank faced a £1.5bn black hole.”  
<https://www.theguardian.com/business/blog/2014/jan/31/paul-flowers-psychometric-testing-bank-chairman>

4. The One Million Dollar Paranormal Challenge;  
<https://www.thevintagenews.com/2016/12/30/james-randi-created-the-one-million-dollar-paranormal-challenge-no-one-ever-claimed-the-prize/>

5. Research design. This kind of longitudinal research requires a combination of methodological savvy and a commitment to the long-term. But this is not new. Over 70 years ago, William Whyte outlined the required design: a rigorous validation would require that a firm test applicants on the new personality test, seal away the results so that the test scores don’t influence either the selection decision or bias managers in their view of the successful recruits. And then match the test data against performance at a later time.

6. The Validity of Personality Inventories in the Selection of Employees, Ghiselli & Barthol, 1953; Validity Of Personality Measures In Personnel Selection; Robert M. Guion Richard F. Gottier, 1965

7. Dr. E. E. Peacock, Jr., quoted in Medical World News (September 1, 1972), p. 45, as quoted in Tufte’s 1974 book Data Analysis for Politics and Policy;  
<http://www.marginalrevolution.com/marginalrevolution/2010/12/the-ethics-of-random-clinical-trials.html>

8. Robert Gibby & Michael Zickar, “A history of the early days of testing in American industry: an obsession with adjustment”, History of Psychology, September 2008  
[https://www.researchgate.net/publication/23562101\\_A\\_history\\_of\\_the\\_early\\_days\\_of\\_personality\\_testing\\_in\\_American\\_industry\\_An\\_obsession\\_with\\_adjustment](https://www.researchgate.net/publication/23562101_A_history_of_the_early_days_of_personality_testing_in_American_industry_An_obsession_with_adjustment)

9. The Tyranny of Personality Testing  
<https://newrepublic.com/article/151098/personality-brokers-book-review-invention-myers-briggs-type-indicator>

10. Assessment Centre validity; <https://amazure.envisialearning.com/wp-content/uploads/2018/09/AssessmentCentresValidityAndOptions.pdf>

# Notes

---

11. Norm groups are problematic in personality testing. A combination of typical scale skew and norms assembled from convenience samples rather than applicant groups can result in the candidate intention in questionnaire completion being distorted in the report outputs.

Are norm groups sexist: <https://qz.com/1201773/we-took-the-worlds-most-scientific-personality-test-and-discovered-unexpectedly-sexist-results/>

12. For an incisive analysis of the challenges (and constraints) of Big Data, "Chancing It: The Laws of Chance and How They Can Work for You", Robert Matthews

13. "How handsome validity can be distilled from raw nothings by clever maths" [https://www.edvul.com/extrapdf/Cureton\\_Baloney.pdf](https://www.edvul.com/extrapdf/Cureton_Baloney.pdf)

14. For example, Barrick, M. R. and Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology* 44(1):

<https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>

15. Why Personality Measures Have Limited Applicability in Personnel Selection, Kevin R. Murphy Jessica L. Dzieweczynski

Morgeson, F., Campion, M., Dipboye, R., Hollenbeck, J., Murphy, K., Schmitt, N. (2007b). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, 60, 1029-1049.

16. The Power of Personality' The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes;

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4499872>

Robust Findings in Personality Psychology

<https://pigeo.wordpress.com/2019/11/12/robust-findings-in-personality-psychology/>

17. The Forer effect; <http://forer.netopti.net/>

Also: Stephen J. Guastello, Denise D. Guastello & Larry L. Craft (1989) Assessment of the Barnum Effect in Computer-Based Test Interpretations, *The Journal of Psychology*,

18 Publication Bias in Test Publishers' Manuals;

<https://pdfs.semanticscholar.org/f169/52884b81aa9b8ee8c1e27585a01bfa05e1e4.pdf>

19. "The halo effect, and other managerial delusions", Phil Rosenzweig; [https://en.wikipedia.org/wiki/The\\_Halo\\_Effect\\_\(business\\_book\)](https://en.wikipedia.org/wiki/The_Halo_Effect_(business_book))

20. On natural frequencies; <https://understandinguncertainty.org/using-expected-frequencies-when-teaching-probability>

Scatter plots and what does a correlation of .5 look like? "To sensibly interpret a correlation coefficient, you need the corresponding scatterplot."

<https://janhove.github.io/teaching/2016/11/21/what-correlations-look-like>

And why we need to rethink prediction; "Polishing The Crystal Ball";

[https://amazure.envisialearning.com/wp-content/uploads/2018/09/PolishingCrystalBall\\_EN.pdf](https://amazure.envisialearning.com/wp-content/uploads/2018/09/PolishingCrystalBall_EN.pdf)

21. The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years" Frank Schmidt;

[https://www.researchgate.net/publication/309203898\\_The\\_VValidity\\_and\\_Utility\\_of\\_Selection\\_Methods\\_in\\_Personnel\\_Psychology\\_Practical\\_and\\_Theoretical\\_Implications\\_of\\_100\\_Years\\_of\\_Research\\_Findings](https://www.researchgate.net/publication/309203898_The_VValidity_and_Utility_of_Selection_Methods_in_Personnel_Psychology_Practical_and_Theoretical_Implications_of_100_Years_of_Research_Findings)

# Notes

---

22. For more on the debate on validity generalisation vis a vis situational specificity, What Personality Does and Does Not Predict and Why: Lessons Learned and Future Directions; [https://www.academia.edu/33781689/What Personality Does and Does Not Predict and Why Lessons Learned and Future Directions?email\\_work\\_card=view-paper](https://www.academia.edu/33781689/What_Personality_Does_and_Does_Not_Predict_and_Why_Lessons_Learned_and_Future_Directions?email_work_card=view-paper)
23. Even Conscientiousness, viewed as the most generalisable of the Big Five traits, looks problematic; the “validity of conscientiousness is moderately overestimated (by around 30%).” [http://www.people.vcu.edu/~mamcdani/Publications/Kepes%20&%20McDaniel%20\(2015\).%20Validity%20of%20conscientiousness%20PLoS%20One.pdf](http://www.people.vcu.edu/~mamcdani/Publications/Kepes%20&%20McDaniel%20(2015).%20Validity%20of%20conscientiousness%20PLoS%20One.pdf)
24. “Flawed Self-Assessment: Implications for Health, Education, and the Workplace”; <https://www.ncbi.nlm.nih.gov/pubmed/26158995>
- Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, 9(2) <http://dx.doi.org/10.1177/1745691613518075>
25. There seem to be two dynamics at play: 1. People lack key information to arrive at an accurate evaluation of themselves. 2. Feedback that might improve accuracy is often difficult to recognise, biased and subject to misinterpretation
26. “If we can lie to ourselves, it helps us lie to others more convincingly.” <https://www.theguardian.com/science/2011/oct/07/deceit-self-deception-robert-trivers>
- “Surprising results that counter the popular idea that knowing yourself is good for you”; <https://www.scientificamerican.com/article/new-insights-into-self-insight-more-may-not-be-better/>

27. On the Dunning-Kruger Effect; <https://theness.com/neurologicablog/index.php/lessons-from-dunning-kruger/>
- <https://www.talyarkoni.org/blog/2010/07/07/what-the-dunning-kruger-effect-is-and-isnt/>
28. CEO Overconfidence and Corporate Investment; Malmendier & Tate <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6261.2005.00813.x>
29. <http://awesci.com/the-astonishingly-funny-story-of-mr-mcarthur-wheeler/>
30. Griffith & Converse, 2011; <https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195387476.001.0001/acprof-9780195387476-chapter-003>
31. Ones, D.S. and Viswesvaran, C. (1997), “A meta-analytic investigation of social desirability influences on integrity test validities: much ado about nothing”
32. A Socioanalytic View of Faking <https://www.advancedpeoplestrategies.co.uk/media/1134/a-socioanalytic-view-of-faking.pdf>
- In “Why the Fake You will Outperform the Authentic You” there is no evidence for the claim that “I fake, and you should do, because it is scientifically proven to make you a higher performer”. [https://www.talentstrategygroup.com/application/third\\_party/ckfinder/userfiles/files/Why%20the%20Fake%20You%20will%20Outperform.pdf](https://www.talentstrategygroup.com/application/third_party/ckfinder/userfiles/files/Why%20the%20Fake%20You%20will%20Outperform.pdf)
33. Here we are back to the William Whyte position that sees personality testing in selection as a kind of filter to check conservative conformity rather than identify distinctive personality. This is to value candidate gamesmanship rather than authenticity.

# Notes

---

34. The Gervais Principle, Or The Office According to “The Office”  
<https://www.ribbonfarm.com/2009/10/07/the-gervais-principle-or-the-office-according-to-the-office/>

35. <https://www.predictiveindex.com/blog/how-to-pass-a-personality-test-and-common-questions-on-faking-assessments/>

36. Are the consequences of these claimed levels of failure a result of the increased usage of personality testing?  
<https://www.hoganassessments.com/bad-management-and-its-consequences/>

37. “A 55% drop in validity for Conscientiousness and a 67% for Agreeableness”; Griffith, R. L., & Converse, P. D. (2012). The rules of evidence and the prevalence of applicant faking. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment*

38. Christiansen, N. D., Rozek, R. F., & Burns, G. (2010). Effects of social desirability scores on hiring judgments. *Journal of Personnel Psychology*, 9(1), 27-39. <http://dx.doi.org/10.1027/1866-5888/a000003>

Faking and selection: Considering the use of personality from select-in and select-out perspectives, *The Journal of applied psychology*;  
[https://www.researchgate.net/publication/10773192\\_Faking\\_and\\_selection\\_Considering\\_the\\_use\\_of\\_personality\\_from\\_select-in\\_and\\_select-out\\_perspectives](https://www.researchgate.net/publication/10773192_Faking_and_selection_Considering_the_use_of_personality_from_select-in_and_select-out_perspectives)

39. “Using Overt and Covert Items in Self-Report Personality Tests: Susceptibility to Faking and Identifiability of Possible Fakers”;  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6037895/>

40. Kiefer, C. and Benit, N. (2016). What is Applicant Faking Behavior? A Review on the Current State of Theory and Modeling Techniques. *Journal of European journal of European Psychology Students*

Schmitt & Oswald, The Impact of Corrections for Faking on the Validity of Noncognitive Measures in Selection  
[https://www.researchgate.net/publication/7046252\\_The\\_Impact\\_of\\_Corrections\\_for\\_Faking\\_on\\_the\\_Validity\\_of\\_Noncognitive\\_Measures\\_in\\_Selection\\_Settings](https://www.researchgate.net/publication/7046252_The_Impact_of_Corrections_for_Faking_on_the_Validity_of_Noncognitive_Measures_in_Selection_Settings)

The Red Herring Needs to be Grilled: Issues of Social Desirability in Personality Testing, Rosanna Miguel;  
[https://www.academia.edu/22338855/The\\_Red\\_Herring\\_Needs\\_to\\_be\\_Grilled\\_Issues\\_of\\_Social\\_Desirability\\_in\\_Personality\\_Testing](https://www.academia.edu/22338855/The_Red_Herring_Needs_to_be_Grilled_Issues_of_Social_Desirability_in_Personality_Testing)

41. Another perspective on retesting. “With so much riding on the outcome of job screening personality tests, it’s probably not surprising that failing candidates who opted to take retake the test would deliberately change their responses.”  
[https://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1019&context=cahrs\\_researchlink](https://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1019&context=cahrs_researchlink)

Sackett PR, Burris LR, Ryan AM. (1989). Coaching and practice effects in personnel selection. *International review of industrial and organizational psychology*

42. Faking in Personality Assessment”, Carolyn Maccann, Matthias Ziegler & Richard Roberts

43. Ipsative measurement has its place, specifically in identifying personal priorities within some domains of individual difference (e.g. vocational preferences, leadership styles) to highlight relative preferences but remain problematic in comparisons across candidates.

The “ipsative debate” from the 1980s continues to the present. Reviews and summaries;

[https://www.pbarrett.net/publications/Ipsative\\_Testing\\_Hammond%20and\\_Barrett\\_1996.pdf](https://www.pbarrett.net/publications/Ipsative_Testing_Hammond%20and_Barrett_1996.pdf).”

# Notes

---

<https://oprablog.wordpress.com/2010/10/27/ipsative-tests-psychometric-properties/>

<https://www.prevuehr.com/resources/insights/ipsative-vs-normative/>

The psychometrician Steve Blinkhorn, reviews recent attempts to recalibrate ipsative data, for example

<https://kar.kent.ac.uk/44775/1/Personality%20assessment%2C%20Forced-choice%20FINAL%20for%20sharing.pdf> and makes the point: “ipsative chalk cannot be turned into normative cheese.”

Also; Fisher, Peter A.; Robie, Chet; Christiansen, Neil D.; Speer, Andrew B.; and Schneider, Leann (2019) "Criterion-related Validity of Forced-Choice Personality Measures: A Cautionary Note Regarding Thurstonian IRT versus Classical Test Theory Scoring," Personnel Assessment and Decisions

44. Application of Preventive Strategies, in “New Perspectives on Faking in Personality Assessment” Dr Matthias Ziegler

45. Neil D. Christiansen, Gary N. Burns & George E. Montgomery (2005) Reconsidering Forced-Choice Item Formats for Applicant Personality Assessment, Human Performance

Forced-choice assessments of personality for selection: evaluating issues of normative assessment and faking resistance. J Appl Psychol. 2006 Jan;91(1):9-24

Effects of Applicant Faking on Forced-Choice and Likert Scores; Goran Pavlov, Alberto Maydeu-Olivares, Amanda J. Fairchild;  
<http://journals.sagepub.com/doi/pdf/10.1177/1094428117753683>

John R. Braun (2018) Forced-Choice Self-Report Devices: A Look at Some Unwarranted Claims, Measurement and Evaluation in Guidance, 2:3

Cao, M., & Western, Educated, Industrialized, Rich, And Democratic, F. (2019, May 9). Does Forcing Reduce Faking? A Meta-Analytic Review of Forced-Choice Personality Measures in High-Stakes Situations. Journal of Applied Psychology.

Advance online publication. <http://dx.doi.org/10.1037/apl0000414>

46. The “g” in Faking: Doublethink the Validity of Personality Self-Report Measures for Applicant Selection

<https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02153/full>

47. There’s more to personality than a test score;

<https://theconversation.com/theres-more-to-personality-than-a-test-score-30898>

48. Ironically it was Walter Mischel, the critic of personality traits, who in the “marshmallow test” observed the importance of delayed gratification. This finding was over-turned by additional research which suggested social and economic background, not the ability to delay gratification, lay behind the children’s’ long-term success.

49. This has had the consequence of building tests that are essentially self report measures of ability rather than personality, compounding the problems of self deception and impression management.

50. Ian Deary also explores the issue of whether traits are simply descriptive or have explanatory value. “The Trait Approach to Personality”, The Cambridge Handbook of Personality Psychology

51. “World’s top personality test doesn’t really work – should we ditch it?”

<https://www.newscientist.com/article/2209360-worlds-top-personality-test-doesnt-really-work-should-we-ditch-it/>

52. Challenges to capture the big five personality traits in non-WEIRD populations; <https://advances.sciencemag.org/content/5/7/eaaw5226>

# Notes

---

53. Beck, E. D., Condon, D. M., & Jackson, J. J. (2019, July 9). Interindividual Age Differences in Personality Structure. <https://doi.org/10.31234/osf.io/857ev>

54. Hexaco; <http://hexaco.org/scaledescriptions>

A alternative framework; Open to experience - closed to intelligence: Why the 'Big Five' are "really the 'Comprehensive Six' Chris Brand; <https://onlinelibrary.wiley.com/doi/abs/10.1002/per.2410080407>

55. The General Factor of Personality: A meta-analysis of Big Five intercorrelations; <https://advances.sciencemag.org/content/5/7/eaaw5226>

56. Dan McAdams and three levels of personality; <https://www.psychologytoday.com/gb/blog/the-personality-analyst/201011/three-levels-knowing-person>

57. For example. "How attitudes to time shape 10 Leadership Patterns; [https://amazure.envisialearning.com/wp-content/uploads/2018/09/ItsAboutTime\\_EN.pdf](https://amazure.envisialearning.com/wp-content/uploads/2018/09/ItsAboutTime_EN.pdf)

58. Perspectives on Models of Job Performance; [https://www.researchgate.net/publication/229645528\\_Perspectives\\_on\\_Models\\_of\\_Job\\_Performance](https://www.researchgate.net/publication/229645528_Perspectives_on_Models_of_Job_Performance)

59. "A Random Search for Excellence  
Why "great company" research delivers fables and not facts; <https://www2.deloitte.com/us/en/insights/topics/operations/a-random-search-for-excellence-why-great-company-research-delivers-fables-not-facts.html>

60. Dark side leadership. "Based on prior theory and the experiences of practitioners, we hypothesized negative relationships between the subclinical disorders and leader development over time. With regard to this hypothesis, our results were certainly more ambivalent. ....we found positive relationships with leader development over time and across multiple dimensions of leadership. The remaining dimensions had few or inconsistent effects. The results require some explanation."  
Harms, P.D., Spain, S., & Hannah, S. (2011). Leader development and the dark side of personality. *The Leadership Quarterly*, 22

61. The test review process of the British Psychological Society has been instrumental in undermining credibility in personality testing in employee selection by lowering standards of predictive validity. By allowing validation studies of small size concurrent design to be acceptable evidence of practical value in selection, several test publishers have been able to claim BPS endorsement of their tools in their marketing.

The EFPA Test-Review Model: When Good Intentions Meet a Methodological Thought Disorder, Paul Barrett  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5791023/>

62. At The Extreme: Leadership and Exceptional Success  
[https://amazure.envisialearning.com/wp-content/uploads/2018/09/Extreme-Leadership\\_EN.pdf](https://amazure.envisialearning.com/wp-content/uploads/2018/09/Extreme-Leadership_EN.pdf)

63. 5 phases to craft talent management strategy;  
[https://amazure.envisialearning.com/wp-content/uploads/2018/09/TheCraftOfTalentManagement\\_EN.pdf](https://amazure.envisialearning.com/wp-content/uploads/2018/09/TheCraftOfTalentManagement_EN.pdf)

64. Witt, L. "The Interactive Effects of Extraversion and Conscientiousness on Performance" *Journal of Vocational Behavior*

65. The influence of work on personality trait development: The demands-affordances TrAnsactional (DATA) model, an integrative review, and research agenda, *Journal of Vocational Behavior*, February 2019

# Notes

---

66. In “Personality testing in personnel selection: Love it? Leave it? Understand it!” Diekman & Konig note: “Especially in case of tests from commercial publishers, it is often difficult to evaluate how these interpretations are generated, which statistical methods and which interpretative algorithms are used to combine test results.”

67. “The holistic approach to judgment and prediction has simply not held up to scientific scrutiny”; Highhouse, S. (2002), ‘Assessing the Candidate as a Whole: A Historical and Critical Analysis of Individual Psychological Assessment for Personnel Decision Making,’ Personnel Psychology

68. “Clinical versus statistical prediction: A theoretical analysis and a review of the evidence”; <https://psycnet.apa.org/record/2006-21565-000>

69. “Correlations are like coincidences. We’d take them less seriously if we were more aware of how easily we find them.” Per capita cheese consumption correlates with the number of people who died by becoming tangled in their bedsheets ( $r = 0.947$ ). More spurious correlations; <https://www.tylervigen.com/spurious-correlations>

70. Gerd Gigerenzer on heuristics; <https://www.annualreviews.org/doi/abs/10.1146/annurev-psych-120709-145346>

“Heuristics are often conceived of as a source of systematic error, whereas logic and statistics are regarded as the sine qua non of good decision making. Yet, this view can be incorrect for decisions made under uncertainty, as opposed to risk. Research on fast and frugal heuristics shows that simple heuristics can be successful in complex, uncertain environments and also when and why this is the case.” <https://pdfs.semanticscholar.org/06b4/c251a2a27abf732ce2730068161f00300c44.pdf>

71. Lyle Ungar on forecasting; <https://intelligence.org/2014/03/26/lyle-ungar/>

72. On Amazon and AI in recruitment; <https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e>

73. <https://slate.com/business/2018/10/amazon-artificial-intelligence-hiring-discrimination-women.html>

74. The opportunities and challenges of mitigating bias in AI; <https://www.thetimes.co.uk/static/ai-bias-job-hunting-ibm-recruitment-sexism-discrimination/>

<https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happens-and-why-its-so-hard-to-fix/>

<https://www.cnbc.com/2018/05/30/silicon-valley-is-stumped-even-a-i-cannot-remove-bias-from-hiring.html>

<https://www.theguardian.com/technology/2019/apr/16/artificial-intelligence-lack-diversity-new-york-university-study>

<https://scholarworks.uark.edu/cgi/viewcontent.cgi?article=1052&context=alr/>

75. The Accuracy of Combining Judgemental and Statistical Forecasts; <https://pubsonline.informs.org/doi/10.1287/mnsc.32.12.1521>

76. “The proliferation of constructs and measures is not a sign of a healthy science. Rather it shows the inability of empirical studies to demonstrate that a measure is not valid or that a construct may not exist. This is mostly due to self-serving biases and motivated reasoning of test developers.” <https://replicationindex.com/2019/02/16/the-validation-crisis-in-psychology/>

# Notes

---

Psychology's Replication Crisis Is Running Out of Excuses

<https://www.theatlantic.com/science/archive/2018/11/psychologys-replication-crisis-real/576223>

77. For example, the suggestion that a well known personality instrument "increases predictive validity to 0.9";

[https://amazure.envisialearning.com/wp-content/uploads/2018/09/FairyTalesFactsFuture\\_EN.pdf](https://amazure.envisialearning.com/wp-content/uploads/2018/09/FairyTalesFactsFuture_EN.pdf)

78. [https://amazure.envisialearning.com/wp-content/uploads/2018/09/PolishingCrystalBall\\_EN.pdf](https://amazure.envisialearning.com/wp-content/uploads/2018/09/PolishingCrystalBall_EN.pdf)

79. Sternberg, R. J., & Williams, W. (1998). Your proved our point better than we did: A reply to our critics. *American Psychologist*, 53 (5), 576-577

80. Fisher, David and Cunningham, Sydnie and Kerr, Alison J. and Allscheid, Steven P., Contextualized Personality Measures in Employee Selection: Extending Frame-of-Reference Research with Job Applicant Samples (March 2017). *International Journal of Selection and Assessment*, Vol. 25, Issue 1, pp. 18-35, 2017

81. Personality and Cognitive Ability as Predictors of Effective Performance at Work Neal Schmitt

<http://people.tamu.edu/~w-arthur/611/Journals/Schmitt%20%282014%29%20AROPOB,%20personality%20and%20cognitive%20ability.pdf>

82. Narrow Reasoning about the Use of Broad Personality Measures for Personnel Selection; Sampo V. Paunonen, Mitchell G. Rothstein and Douglas N. Jackson *Journal of Organizational Behavior*

83. Pitfalls of Personality Theory, Colin Cooper

How Not To Discover A Personality Trait;

<https://www.sciencedirect.com/science/article/abs/pii/S0191886919304830>

Cooper's critique is also a challenge to the test publishers whose manuals incorporate more pages on construct validity than genuine predictive validity in selection.

84. <https://www.theguardian.com/technology/2017/jul/09/everybody-lies-how-google-reveals-darkest-secrets-seth-stephens-davidowitz>

85. On snoopology:

<https://www.theguardian.com/science/2008/jun/28/psychology>

"A Room with a cue: Personality judgments based on offices and bedrooms" *Journal of Personality and Social Psychology*, 2002

86. <https://www.bubble-jobs.co.uk/blog/b2b/snoop-candidates-social-media-profiles/>

87. Watson's Personality Insights

<https://personality-insights-demo.ng.bluemix.net/>

But not without reservations: <https://medium.com/taraaz/https-medium-com-taraaz-human-rights-implications-of-ibm-watsons-personality-insights-942413e81117>

88. "When we turn to hiring employers are missing the forest for the trees: Obsessed with new technologies and driving down costs, they largely ignore the ultimate goal: making the best possible hires.

<https://hbr.org/2019/05/recruiting>

89. <https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/>

90. The link to independent research from HireVue's website is unavailable;

<https://www.prnewswire.com/news-releases/independent-research-reveals-hirevue-dramatically-improves-quality-of-hire-recruiter-productivity-and-candidate-experience-252622521.html>



# Notes

---

91. Emotion recognition algorithms, A question mark;

<https://www.theverge.com/2019/7/25/8929793/emotion-recognition-analysis-ai-machine-learning-facial-expression-review>

92. Here personality tests shift towards the kind of design used in biodata measures; Harold, Crystal M. and McFarland, Lynn A. and Weekley, Jeff A., The Validity of Verifiable and Non-Verifiable Biodata Items: An Examination Across Applicants and Incumbents. International Journal of Selection and Assessment

This is also the re-emergence of Cattell's much neglected T data; Santacreu, J., Rubio, V. J., & Hernández, J. M. (2006). The objective assessment of personality: Cattell's T-data revisited and more. Psychology Science Review 23(3)

93. Projective testing: Historical foundations and uses for human resources management; Human Resource Management

94. But no research evidence to back this claim;

<https://thecambridgecode.com/overview/>

95. More Than Meets the Eye: The Evolution of Personality Testing;

<https://blogs.scientificamerican.com/observations/more-than-meets-the-eye-the-evolution-of-personality-testing/>

Visual methodologies in personality test construction;

[https://d13kjz344z5e1m.cloudfront.net/wp-content/uploads/2014/10/CRWhitepaper\\_web.pdf](https://d13kjz344z5e1m.cloudfront.net/wp-content/uploads/2014/10/CRWhitepaper_web.pdf)

96. On gamification

Game-like personality testing: An emerging mode of personality assessment, Personality and Individual Differences, 2019,

<https://digest.bps.org.uk/2019/05/15/an-exciting-new-approach-to-personality-testing-involves-psychologists-analysing-your-decisions-in-game-scenarios/>

Serious Games for Assessment: Welcome to the Jungle;

<http://www.jattjournal.com/index.php/atp/article/view/118669>

97. The Truth About Game-based Talent Assessments;

<https://www.ere.net/the-truth-about-game-based-talent-assessments/>

Survival Guide to the Bright, Shiny Trends in Talent Acquisition;

<https://aptmetrics.com/wp-content/uploads/2019/11/Survival-Guide-to-the-Bright-Shiny-Trends-in-Talent-Acquisition-2019-SIOP-LEC-1.pdf>

98. Virtual Reality as an Emerging Methodology for Leadership Assessment and Training;

<https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01658/full>

99. <https://www.personneltoday.com/hr/could-dna-be-the-key-to-a-successful-career/>

100. Pitfalls of Personality Theory; <https://doi.org/10.1016/j.paid.2019.109551>

101. Personality Testing in Employment; <https://corporate.findlaw.com/human-resources/personality-testing-in-employment.html>

102. More on personality testing and adverse impact;

Employment Tests and Employment Discrimination: A Dissenting Psychological Opinion;

<https://pdfs.semanticscholar.org/2d38/f22fbcf39c30d7a06fdc590d6c801cfc4b2d.pdf>

Recent EEOC Actions Show Dangers of Using Personality Tests in Hiring

Process; <https://www.jdsupra.com/legalnews/recent-eeoc-actions-show-dangers-of-92292/>

103. Beware These Marketing Trends in Psychological Assessment;

<https://www.psychologytoday.com/gb/blog/the-situation-lab/201910/beware-these-marketing-trends-in-psychological-assessment>

104. Rachid Laajaj et al. Challenges to capture the big five personality traits in non-WEIRD populations, Science Advances (2019)

<https://advances.sciencemag.org/content/5/7/eaaw5226>