

In the midst of the razzmatazz surrounding AI, we hear a lot about its potential for the recruitment process. But is it really the silver bullet that recruitment-tech start-ups claim? Adrian Furnham advises caution.

WHY MANY MODERN PSYCHOLOGY TEST PUBLISHERS FAIL

by Adrian Furnham



I am a Professor of Psychology with a long-time interest in psychometric testing. I have published books on the topic (the latest last year, *Twenty Ways to Assess Personnel*, Cambridge University Press), and developed a number of tests sold to test publishers (e.g., *High Potential Trait Indicator*).

Over the past three years, a number of young entrepreneurs (perhaps half a dozen) have contacted me to help them develop new tests. They wanted to be part of the new, and potentially very lucrative, AI-inspired wave of talent management and person profiling. They were small business start-ups — some with good backing, others not.

They had a lot in common and, as far as I can tell, they all failed financially, like so many other testing-tech start-ups. One main issue was that they competed with the traditional tests on scalability, candidate experience, and time and cost, but neglected both evidence of validity, surely the most central feature, and clients' willingness to invest in training the algorithms to suit the business needs.



DISRUPTION

For well over 50 years, the test-publishing “model” went something like this. Authors and academics with ideas and theories devised tests (usually of personality, motivation, and values). Think of the famous MBTI, which apparently is taken by somebody every 10 seconds somewhere in the world. They then sold these tests to the few test publishers that existed. These publishers, who might or might not be very sophisticated or technologically inclined, sold the published tests to institutions and consultants, who used them in various ways, often in the process of selection. Paper-based tests started to be replaced by electronic tests about 20 years ago, but the model persisted.

It was a “nice little earner” for a few, and the test-publishing world was a small, quiet oligopoly, happy to potter on. There were rises and falls in the enthusiasm for tests but social media often made them more popular. Big American publishing firms dominated the world. Many still exist after 50 to 70 years.

Then the AI revolution sprang into action with its ready-made solution for this field, as well as numerous others. AI comes in two formats: a simple scoring engine to dynamically improve the algorithm (as opposed to a fixed syntax), and also the more experimental approaches that don't rely on self-report assessment, but

rather scrape internal or external big data, mine interview voice and video, use natural language processing, etc.

Companies are attracted to using AI resumé-scraping tools to “search and match” candidates based on their hard skills, just parsing CVs and looking at similarities with past candidates or job vacancies.

The aim is to give people a test online, and the results are fed into a smart algorithm which spits out an accurate, reliable, unbiased, and valid score which can easily and very simply be used in the accept / reject or rank-ordering of candidates.

SMALL BUSINESS START-UPS

Young entrepreneurs, in particular, were excited by the potential. There were three aspects to this. The first was the still-prevalent belief and hope that selection could be much more efficient and “scientific” with the application of AI. So, the aim is to give people a test online, and the results are fed into a smart algorithm which spits out an accurate, reliable, unbiased, and valid score which can easily and very simply be used in the accept/reject or rank-ordering of candidates. The “white heat” of the new digital world!

The second was that one could “cut out the middle men”. You don't need test publishers taking a big cut and controlling access to their products (e.g., for the “necessary” training). Consultants can now devise tests and sell them to other consultants or to clients directly (B2B, B2C). All sorts of people may want to assess themselves and their friends—a potentially huge new market.

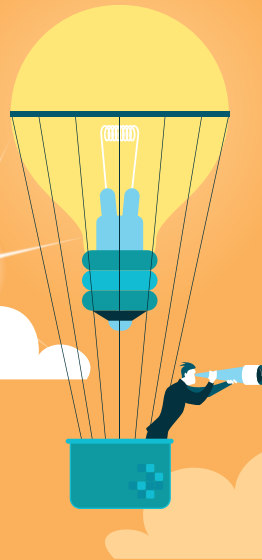
The third was that tests could be done on a mobile device, phone, or even a smart watch. Indeed, all have the advantage that they can also be used to gather behavioural data, which could supplement what the testee has to say about themselves.

Indeed, five years ago, a friend of mine who started and ran the biggest test-publishing company in Britain sagely made the following predictions: *Smartphones will replace computers for employee assessment. High-quality psychometric testing services will be sold direct to consumers. Advances in the neuroscience of personality will reveal which are the most valid individual differences to measure and how best to measure them. The digital badging movement, coupled to the use of big data and new forms of digital CV, will render many of the current applications for high-stakes testing redundant. The basis for employee development will in the near future be derived from the data yielded by wearable devices and not from psychometric tests.* The Brave New World was just around the corner.

AI-based methods have expanded the market for “assessment”, and traditional methods remain a very niche market. Things are still a bit better for traditional assessments when it comes to feedback for development, self-awareness, coaching, and leader selection, but this could change very quickly. AI-based scraping of Zoom coaching sessions could produce a dynamic bright-and-dark-side profile that feeds into the coaching session, and you could replace the 360 with simulations.

Various smart young entrepreneurs suddenly became aware of a very big opportunity to create value and to make money. The testing world seemed absolutely ripe for disruption and digitalisation. And it seemed so easy.

The basis for employee development will in the near future be derived from the data yielded by wearable devices and not from psychometric tests.



MY EXPERIENCE

The various groups who approached me, from four different countries, had many features in common. All were groups of three or four young, hungry, and very clever mainly (but not always) men. They had met in business school or consultancy businesses. They were smart, super tech-savvy, hungry, and not risk-averse. For most, it was not their first venture or success. Indeed, they knew a lot about venture capitalists and they shared a very similar goal. Start-up to billionaire in a few years, because the total addressable market (“TAM”, in the VC world) is “huge”.

Some had done their homework more than others, but all were shocked by the expensive, lazy, and inefficient world of HR and recruitment – something they had seen themselves along their professional careers in large organisations. They found that there were a surprisingly limited number of tests that had been around for 30 to 50 years and that people wanted some new ones, as well as to know which of the more established tests was the most accurate.

What they wanted to say (and often did, without much evidence) was that their new tests had better psychometrics, particularly predictive validity, because the research sample size would be significantly larger, allowing for some iterations while being used

to change the weighting of questions, correlate answers between tests, and add data that was never collected before. They argued that their tests reduced or avoided “older method” artefacts or problems like impression management (i.e., lying) because of the way the tests were administered, often over time.

They said that testees like them; there was a good reaction from them, which meant a better candidate experience. They tried to persuade HR buyers that sexy new tests were good PR for

the tester, being up to date and fairer, in a society where discrimination bias had become more widely acknowledged and challenged. And, of course, they argued that there were many savings in terms of time, effort, and cost. What more could the HR world ask for?

They were certainly all clear about the “schtick”. You hear (without any evidence) claims like “next-generation technology”, “twenty-first-century generation”, “digs deeper”, “powered by neuroscience”, “state-of-the-art”, “has less adverse effect”, “leads to more diverse choices”, “authentic” and, not to be left out, “disruptive”.

They all came to me for the same reason: they had come across my name because of academic papers, popular articles, conference talks, and workshops that I had run. They wanted to know if I could develop tests for them or advise on how existing ones could be applied in business for the specific purpose of selecting the people who “fit” the organisation, in terms of both predispositions and preferences.

WHAT THEY GOT WRONG

They all failed as far as I know, one spending \$2 million of their investor’s money without generating sufficient traction to continue their journey. And they did so for all the classic reasons.

First, they were not that clear about the true cost of getting their potential clients to start using a solution. The concept of on-boarding a new technology in businesses that have developed over decades to achieve a structured recruiting process today is not easy to accept, when the risk of breaking what they have is so high – regardless of how many benefits it might possibly bring.

Second, most seemed to know very little about the multiplicity of decision makers in the

The concept of on-boarding a new technology in businesses that have developed over decades to achieve a structured recruiting process today is not easy to accept, when the risk of breaking what they have is so high – regardless of how many benefits it might possibly have.

process of selection and none had any real idea of how the market for HR solutions functions. They became obsessed by their PR and what their super-designers could do, but they “forgot” to find out what “their clients” wanted.

Third, and most important, in their rush to get sexy-looking feedback, they forgot the hardest bit: collecting real evidence that the test works, namely by predicting behaviour. Asserting validity is not enough. Collecting anecdotes from old professors and happy clients is not enough. Clients want proof-as-you-go as much as hype. The candidates deserve it,

and the courts may demand it.

When I told them about validity and how you achieve it, some seemed surprised. I explained that of the two tests I have sold to a





Asserting validity is not enough. Collecting anecdotes from old professors and happy clients is not enough. Clients want proof-as-you-go as much as hype. The candidates deserve it, and the courts may demand it.

customer, it was easy enough to devise the tests, but it took three to five years to collect enough “real-world” data to get the evidence to make the claims that clients wanted, and needed, to hear. It’s called “the criterion problem” and it involves looking at the relationship between test scores and actual job performance over time. As much as the validation process of such relationships can be done across all potential clients of such a start-up, for the client the validity for their employees is all that matters.

All were blinded by their own propaganda – how the AI revolution has, could, and will change testing. They all thought that testing was a sleepy backwater where “swashbuckling” disruptive experts could bring some sexy science and “make a killing”.

So, the lesson is not “Don’t try to disrupt or don’t try to innovate.” The real problem is that entrepreneurs often don’t try to learn from others’ experiences. They thought that

the reason for not using the tests was only the channel to reaching the testees and they often failed at the same point. In the development of their businesses, they did not use the so-called AI that would have

learnt over time from the many other entrepreneurs who faced the same challenges and failed. [EBR](#)

I would like to thank a number of entrepreneurial friends for their insights, comments, and critique: Simmy Grover, Robert McHenry, and Alexandre Gruca.

ABOUT THE AUTHOR



Adrian Furnham is a Professor at the Norwegian Business School in Oslo.