

Battle report from a corporate psychologist: Why inaccurate methods are easier to sell and what we can do about it

Grzegorz Rajca



GRZEGORZ RAJCA

Warsaw, Poland

grzegorz.rajca@gmail.com

About the author

Grzegorz Rajca is a psychologist (MA) with over 10 years of experience in business consulting. He has conducted consulting projects in the area of Human Resources (HR), workplace strategy and change management. He takes an evidence-based approach in his work, and specialises in translating complex business problems into research questions. Grzegorz is a member of the Polish National Awarding Committee, established by the Polish Psychological Association to implement standards for using psychological tests in the business context, developed by the European Federation of Psychologists' Associations (EFPA). He is also an author and co-author of many publications promoting the evidence-based approach among business practitioners.

Abstract

As a psychologist working in consulting, dealing with various types of organizational research is my daily bread. The amount of low-quality methods and tools lacking adequate validation that I encounter in the business world is staggering. Even more worrying is their popularity in business and their resistance to being replaced by validated methods, even if they are available. For some reason business seems to prefer using questionable tools. This paper explores why such tools dominate the market. Several reasons are discussed for why these methods may have an advantage over those designed and validated by academia. Possible solutions to overcome this problem are also discussed.

Keywords: psychological testing, assessment, reliability, validity, business, evidence-based

Introduction

It was 2011. I was still a psychology student and had just started my new job in an HR consulting company. One day my manager told me that there was an HR conference in two weeks, and that it might be a good idea for me to go there and look around. The conference seemed like a good opportunity to explore various assessment tools and meet like-minded professionals, whose fascination with factor analysis and validity metrics equalled mine.

When I arrived at the conference, I was really excited. The room was filled to the brim with exhibitors presenting their solutions. The tools were beautiful. Automated reporting, exports, recommendations, algorithms that calculated job fit in a matter of seconds. Still in awe, I asked one of the exhibitors, *"What is the theory behind this assessment?"*. The exhibitor looked at me and said: *"Huh?"*

As I quickly learned, there was no theory behind the assessment. It was just a collection of questions, mixed into a questionnaire based on basic common sense. I was shocked. They had just sat down one day and wrote a list of questions that seemed like good questions to ask in an assessment!

Still in disbelief, I asked, *"What methods did you use to validate this tool?"*. The exhibitor had an equally surprising answer to my second question. He said, *"We've already sold hundreds of licences for this assessment, our clients love it, so we know it works!"* I spent the next couple of hours talking to all the exhibitors who promoted assessments of any kind. Out of maybe ten different companies that advertised their tools that day, only one had any data on reliability and validity (and we seemed equally happy to find each other). Here's the key issue: it's not even the fact that those other companies didn't have the data on validity or any theory that would support their tools. It's that they didn't understand my questions.

Background

The title of this paper is *"Battle report from a corporate psychologist"*, as I really do believe that in the field of organizational psychology and business assessment there is a battle going on. The key challenge that psychologists face in this battle is that low-quality methods seem to sell, while high-quality methods don't.

One might think that not being able to sell high-quality methods is a problem that should be dealt with by sales representatives and marketing departments. I firmly believe that this is not only a business problem, but also an ethical one for psychologists. What's the point of constantly improving empirically tested, valid tools? Nobody buys them and they are already much better than what's popular on the market. Should we be more active in trying to push low-quality tools out of the market? Should we change the way we design our tools to make them easier to sell? I think that there are many ethical questions to ask here. We should definitely try to understand why evidence-based tools seem to be so difficult to sell.

In this paper, I will focus solely on tests. By tests I mean paper and pencil or computer-based assessment tools, used to measure constructs like personality or aptitude, most often used in employee selection, development or career counselling. The problem described above is relevant to other assessment methods as well. However, tests are well defined tools and as such much easier to compare. Therefore, I will use them as an example to illustrate my point.

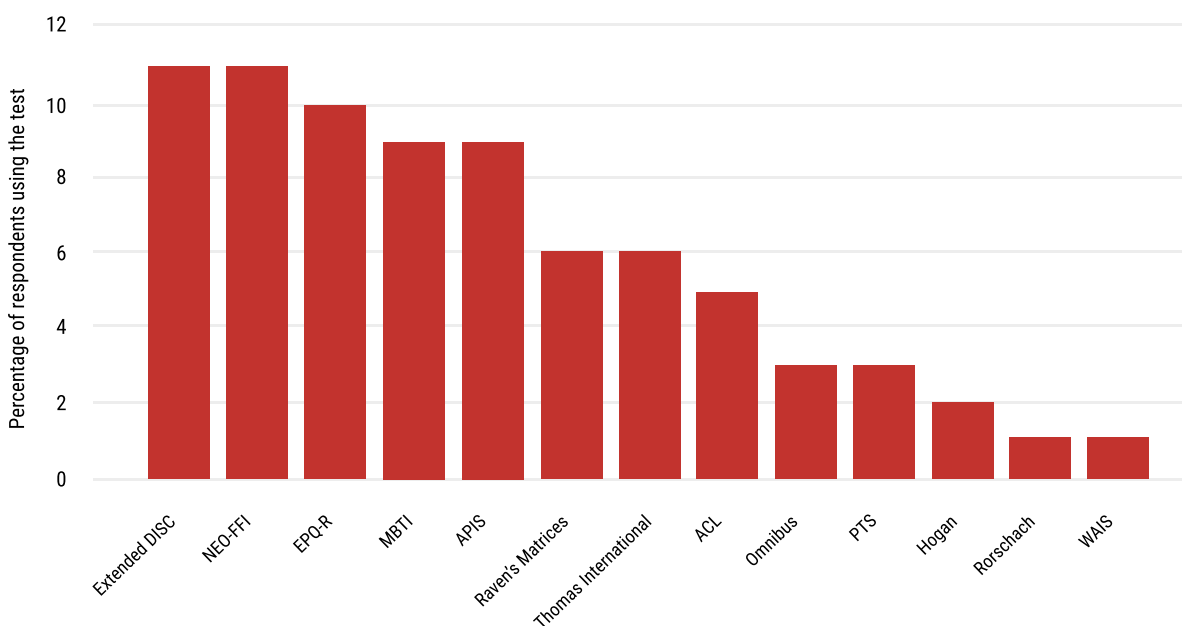
First of all, let's analyse whether there really is a problem. While preparing for my 2019 EAWOP Small Group Meeting presentation in Warsaw, I googled the phrase "most popular assessment tools". The top search result was a list of ten personality tests of which the first two were the Myers-Briggs Type Indicator (MBTI) and the Extended DISC Assessment. The validity of both of these methods either hasn't been confirmed or has been questioned for years both in scientific studies and in the business literature. For instance, see Burnett (2013), Eveleth (2013), Grant (2013), and Stein and Swan (2019), for criticisms of the MBTI. As for the Extended DISC Assessment, it's virtually impossible to find any independent, peer-reviewed research on its validity.

One could of course make a point that googling a phrase is not the best method of searching for valid assessment tools. While this is true, it also seems to be the simplest and most commonly used method in business. It would be really difficult to find a business executive who searches for an answer by performing a systematic literature review.

Fortunately, we don't have to rely solely on web search results. In 2009, a Polish consulting company conducted a survey focused on understanding the practice of selection testing in Poland (Wekselberg, 2011). The survey was completed by 84 participants representing 84 different companies. Some of the tools identified as most commonly used by Polish HR departments were the same ones that came up in my web search results (Figure 1). A fraction of participants even stated that they used the Rorschach inkblot test for selecting managers.

Figure 1

Frequency of using psychological tests in managerial selection in Poland (2009, N=84)



Note: Adapted from Wekselberg (2011), with publisher permission. Full names of tests: NEO-FFI (NEO Five-Factor Inventory); EPQ-R (Eysenck Personality Questionnaire-Revised); MBTI (The Myers-Briggs Type Indicator); APIS (Bateria Testów APIS-Z; eng. APIS-Z Test Battery); Raven's Matrices (Raven's Standard Progressive Matrices); Thomas International (Tests by Thomas International, including Thomas PPA: Personal Profile Analysis and Thomas TST: Tests for Selection and Training); ACL (The Adjective Check List); Omnibus (Test Inteligencji OMNIBUS; eng. OMNIBUS Intelligence Test); PTS (Kwestionariusz Temperamentu PTS; eng. Temperament Questionnaire PTS); Hogan (Tools by Hogan Assessment Systems, including HPI: Hogan Personality Inventory, HDS: Hogan Development Survey and the MVPI: Motives, Values, Preferences Inventory); Rorschach (The Rorschach Inkblot Test); WAIS (Wechsler Adult Intelligence Scale).

Some time ago I participated in a workshop designed around one of these popular, colourful business assessments. The goal of the workshop was to show that people have different psychological profiles and that communicating effectively must take these differences into account. One week before the workshop all participants had to fill out an assessment, so that profiles could be prepared in advance.

During the workshop, I raised questions concerning the validity of the assessment. The trainer – although a psychologist – was not able to provide any information on psychometric properties. She promised, however, to forward my concerns to the test publisher. As agreed, right after the workshop I sent her an email asking for information on the validity of the assessment. I also mentioned that I was interested mainly in criterion or predictive validity. I received a response from the test publisher the next day. The publisher replied that this was a business tool, not a psychological

one, and as such it has not been tested for criterion or predictive validity. The only "validation" method used was calculating the Cronbach's alpha. This approach is really worrying. Apparently, at least some publishers believe that measuring validity is optional in business assessments and that we don't need test results to be related to the real world in any way.

The market is filled with low-quality tests: never-validated and based on outdated research. In fact, such tools seem to be doing very well. What makes them so successful? Perhaps it's their low cost? As it turns out, this is not the case. In Poland, many of the validated tests cost a fraction of what you would have to pay for using a popular business assessment.

If neither quality nor price seems to be the main factor – what is? We have tools that work, have been validated and are often cheaper than business assessments. Why isn't the market dominated by such tools? I believe the problem includes other variables. In one of his lectures Jordan Peterson (2017) suggested that you could make a case that the probability of a company using a test that predicts performance is inversely related to the accuracy of the test, and also – that less accurate tests are easier to sell. He gave two reasons for why this is the case.

Firstly, according to Peterson (2017), such pseudo-tests don't hurt anybody's feelings. One might think that hurting somebody's feelings shouldn't be the main focus of an assessment. However, in today's business reality, recruitment (being able to find and attract candidates) seems to be more important than selection (choosing the right candidate from a pool of applicants). This is due to an increasingly competitive landscape for recruiting and retaining talented employees. As a result, choosing a less accurate tool, based on which everybody is a winner, seems to be a better strategy than using validated assessments and risking a Public relations (PR) crisis, caused by an offended candidate. Secondly, Peterson (2017) claims that most people don't do very well in real, validated assessments, or rather – don't think they do well. The main reason is lack of statistical knowledge and confusing percentages with percentiles. Consequently, being in the 60th percentile is often interpreted as barely passing the test, when in fact it means that you scored better than 60% of the population.

I believe both of these observations to be accurate. However, I also think there are more reasons that help explain why accurate tests are difficult to sell. In this paper, I will describe four such reasons I managed to identify so far in my practice.

Reason 1: “PLEASANT = TRUE”

In HR one of the most prevalent questions, which gets asked after training, workshops, coaching sessions, and even company parties, is a version of “*How did you like it?*” Measuring the quality of an intervention solely by asking participants to rate their overall experience and satisfaction with it is the simplest, most popular way of evaluation. In some cases, assessment tools can also be evaluated in this way, if they’re a part of a workshop or a selection process. As a result, highly rated tools are usually the ones which are fun, short, and give an overall impression that complex phenomena like human behaviour can be explained by splitting people into several distinct and well-defined categories. If measuring overall satisfaction is the only way of evaluating an assessment (and it usually is), then the tools perceived by both HR and employees as the best ones, are those that provide the best overall experience.

Another issue which stems from the simplicity of such assessments is that their interpretations are memorable. The aforementioned workshop provided me with one more interesting insight. After its conclusion, I talked to three colleagues who – prior to completing the assessment for the workshop – had undergone an assessment conducted with a validated psychometric tool. Only one of them told me that this most recent test seemed more like an advanced version of a horoscope than a real assessment. The remaining two didn’t really notice any major differences between the two methodologies. They claimed that both tools gave them a fairly accurate picture of themselves. However, what later caught my attention is the fact that a few months after both assessments, none of the participants could recall any of the results of the evidence-based test. On the other hand, everyone remembered their “dominant personality colour” provided by the business assessment. Some participants also remembered their team members’ colours, and sometimes referenced them in their work. As much as I dislike inaccurate, oversimplified tests, I have to admit that this is something they are doing well. Their results are very memorable and seem to influence participants’ perceptions months, or even years later.

Reason 2: “OLD = WRONG”

If you mention any kind of research to an executive, one of the first questions you will get is probably, “*How old is this research?*”. If it is more than ten years old, there is a fair chance that he or she will immediately lose interest in the findings.

The perceived “expiry date” of research is much shorter in business than in academia. To understand why this is the case, we must first understand how business reports work. In the business world, research reports are published once every quarter, sometimes once a year. The business landscape is very dynamic and provides enough change to fill a report. Therefore, from the business perspective, basing your decisions on last year’s report is a recipe for failure. In a fast-paced, constantly changing world, the key issue is to get relevant information quickly and act on it immediately. In such circumstances, a research paper from 1970 seems prehistoric and completely irrelevant.

In academia, the perception of this topic is very different. Although new findings are interesting, theories that have been developed years ago and replicated multiple times in various contexts are extremely valuable. It is a domain where the value of many concepts and research findings grows with age. In business, the opposite seems to be the norm.

One might wonder why we can’t just provide executives with new, relevant research. At first glance, this seems like a reasonable expectation. Businesses however usually have a very specific problem to solve. It’s not always possible to find relevant, high-quality research, which is also relatively new. The key to effectively use scientific findings in business seems to be rooted in the ability to understand the researched phenomena and apply this understanding to a specific context. For this purpose, insights from research published a year ago can be equally helpful as those from research that is fifty years old.

Reason 3: “NEW (BUT NOT TOO NEW)”

A situation I’ve experienced with one of my clients, with whom I had the opportunity to work as a consultant, may serve as a particularly apt illustration of the “NEW (BUT NOT TOO NEW)” problem. The company hired us to provide recommendations on a particular business problem. One of the priorities and key requirements of the project was that the proposed solution should be innovative. Our team conducted extensive research within the company and suggested a course of action that we believed would help the client achieve their goal.

The solution was a novelty in the Polish market, but already tried and tested in countries like The Netherlands, Australia, Germany or the UK. The management board listened carefully to our presentation. Afterwards, the CEO expressed his interest in trying out the solution. There was only one condition that had to be met before

they decided to proceed with implementation. They wanted us to provide examples of companies that successfully adopted the proposed method. Said companies were expected to be of similar size (more than 1,000 employees), operating in Poland, in the same industry. Only after reviewing these examples the management board would feel confident enough to undertake the project. In the case of this client, innovation meant choosing a safe, well-trodden path.

A common stereotype is that in business you want to be first. In my opinion, this is not exactly true. You don't want to be first, you want to be second. You don't want to be the one experiencing all of the unforeseen difficulties and problems of a novel solution, which is often the fate of pioneers in any field. You want a solution innovative enough that it gives you a competitive advantage, but at the same time established enough to not pose significant and unpredictable threats to your business.

This problem forms a particularly dangerous synergy with the previously described "OLD = WRONG" problem. Why should a business trust a methodology that was developed in academia sixty years ago, has been refined ever since, and still hasn't managed to dominate the market? Why aren't any of our competitors using this idea? Are we really the first company in six decades to think this old concept makes sense? Surely there must be something wrong with it; something that all of our predecessors must have realised, but we are currently unable to see. This combination of distrust towards old concepts and a tendency to use what's already popular on the market is a major obstacle in convincing businesses to use established, validated tools.

Reason 4: "NO THEORY, NO PROBLEM"

A couple of months before I began writing this paper, I had received an email from an HR consulting company. Its aim was to sell me a new test that had recently appeared in the Polish market. The email stated that a large number of commonly used personality assessments are based on social sciences, which rely on observations of human behaviour and ignore chemical reactions of the brain. This claim was followed up with a statement that, according to the latest research, a large part of our behaviour is written into our DNA. Therefore, effective diagnosis should be based on both the observations of behaviour and the measuring of chemical reactions in the brain. In the next paragraph the author of the email claimed that knowledge gained with the use of their tool can help boost performance of teams and entire organizations, prevent conflicts, increase effectiveness of team management and communication, as well as

develop self-awareness, including – among other things – the understanding of one’s own motivators and stressors.

The major obstacle in competing with such tools is that if a test provider doesn’t feel obliged to be true to any theory or research, they can promise whatever they like. Going against such promises with a well-designed, validated test is very difficult. Unless the client is familiar with basic psychometric concepts, they’re unable to compare the quality of various tools and verify unrealistic promises. Therefore a validated test with a clear purpose and a fairly narrow but well-defined range of uses will be perceived as poorly designed, having limited applicability and relying on outdated research. Especially when compared to a tool based on “recent revolutionary research” which – according to the marketing brochure – can diagnose everything from motivation and stressors to chemical reactions in your brain.

Possible solutions

What I suggested above are some of the reasons why evidence-based tests are difficult to sell. The problems I described reflect cultural incompatibilities, rather than differing goals. The key thing to understand: it’s not that companies don’t want accurate tools, it’s that they use a different set of criteria to make decisions. And the important question is, what we can do to increase the use of high-quality methods in business. There are a couple of ways in which we can approach this problem.

In my opinion, the first thing worth considering is the overall assessment experience. Some researchers recommend including participants’ perceptions of the assessment as one of the variables in validity measurements. Moyle and Hackston (2018) suggest we should start measuring something they call *experiential validity* of tools. According to their definition, experiential validity tells us whether the person completing an assessment experienced the assessment process (including feedback) as personally valuable. Additional components of experiential validity include finding out whether the intended development outcomes were achieved, whether key learnings (in the development context) can be recalled months or years later and whether the assessment has an ongoing impact at work. Moyle and Hackston (2018) suggest that the fact that people remember the results, trust them, and are motivated to engage in development activities based on these results should be taken into account when considering the validity of a tool, especially in the development context. I think that experiential validity is important not only in employee development, but also in employee selection.

We could redefine it as “*whether the candidate experienced the assessment process (including feedback) as fair and relevant to the selection goal*”. Paying attention to experiential validity would make tests look more relevant and more believable, which would be helpful in convincing businesses to use these tools.

The other thing worth considering is *simplicity*. Of course, there are some tools and methods that require extensive training and understanding of complex theoretical concepts. However, maybe it’s a good idea to design more basic tools for business purposes. Tools that could be used by an HR professional without psychological background, who has only read the manual. Tools that incorporate all of the positive aspects of popular business assessments, such as the possibility to conduct testing online and automating everything that can be automated without compromising the accuracy of the assessment. The idea behind this is rooted in behavioural economics, which suggests that we should make the right choice an easy choice and the wrong choice a difficult one (Rozin et al., 2011). We can’t make it more difficult to use low-quality methods. We can, however, make it easier to use high-quality ones.

The third thing that, in my opinion, is worth paying more attention to is the *social proof*. The term, coined by Robert Cialdini (2001), states that one of the means people use to determine what is right is to find out what other people think is correct. In my consulting practice, I experienced many situations in which social proof turned out to be critical in making decisions. For instance, clients rarely ask about the methodology behind a tool. If they do, I usually reply that I can go into detail and explain the methodology, if they wish for me to do that. That willingness is usually all that it takes to convince clients that the methodology is in place, and I almost never get asked to explain it. On the other hand, talking about examples of similar projects usually takes a large portion of a business meeting. Clients usually ask many questions about other comparable organizations that we worked with or similar business problems that we solved. Also, social proof seems to explain why some concepts, although old, do very well in business. Maslow’s hierarchy of needs is an example of such a concept, which although a few decades old and questioned by researchers (e.g., Tay & Diener, 2011), is very popular in business and can still be found in brochures and presentations on motivation and well-being. This popularity works like a self-propelling mechanism. After all, this concept has been present in the business world for decades and is still being used by many organizations. Surely it must provide valuable insights. Being aware of this tendency may turn out to be very helpful when planning your conversation with a business executive.

Conclusions

In this paper, I described a problem of business using questionable tests and methods that have not adequately been validated. This problem exists even though, in the case of tests, high-quality methods are usually available and in many cases cheaper than their more popular counterparts. Therefore, the source of this problem seems to lie not in the quality-price ratio, but in differences in communication and in the use of different criteria in decision-making in academia and business. I believe, however, that this problem can be solved or at least mitigated. Some ways to do it include improving experiential validity of assessments, simplifying tools, if possible, and paying more attention to the social proof when communicating with business. The battle is not yet lost. However, perhaps instead of fighting, all we have to do is learn from these popular methods and incorporate what they're currently doing well into our own practice and tools.

References

- Burnett, D. (2013, March 19). Nothing personal: The questionable Myers-Briggs test. *The Guardian*. Retrieved from <https://www.theguardian.com/science/brain-flapping/2013/mar/19/myers-briggs-test-unscientific>
- Cialdini, R. B. (2001). *Influence, science and practice*. Boston, MA: Allyn & Bacon A Pearson Education Company.
- Eveleth, R. (2013, March 26). The Myers-Briggs personality test is pretty much meaningless. *Smithsonian.com*. Retrieved from <https://www.smithsonianmag.com/smart-news/the-myers-briggs-personality-test-is-pretty-much-meaningless-9359770/>
- Grant, A. (2013, September 18). Goodbye to MBTI, the fad that won't die. *Psychology Today*. Retrieved from <https://www.psychologytoday.com/us/blog/give-and-take/201309/goodbye-mbti-the-fad-won-t-die>
- Moyle, P., & Hackston, J. (2018). Personality assessment for employee development: Ivory tower or real world?, *Journal of Personality Assessment*, 100(5), 507-517. doi:10.1080/00223891.2018.1481078
- Peterson, J. B. (2017, June 4). *2017 personality 21: Biology and traits: Performance prediction* [Video]. YouTube. <https://www.youtube.com/watch?v=Q7GKmznagsQ>
- Rozin, P., Scott, S. E., Dingley, M., Urbanek, J. K., Jiang, H., & Kaltenback, M. (2011). Nudge to nobesity I: Minor changes in accessibility decrease food intake. *Judgment and Decision Making*, 6(4), 323-332.
- Stein, R., & Swan, A. B. (2019). Evaluating the validity of Myers-Briggs Type Indicator theory: A teaching tool and window into intuitive psychology. *Social and Personality Psychology*

Battle report from a corporate psychologist: Why inaccurate methods are easier to sell and what we can do about it

Compass, 13(2): e12434. doi: 10.1111/spc3.12434

Tay, L., & Diener, E. (2011). Needs and subjective well-being around the world. *Journal of Personality and Social Psychology*, 101(2), 354-365. doi: 10.1037/a0023779

Wekselberg, V. (2011). Czy testy osobowości mogą być opłacalnym narzędziem w selekcji menedżerów? (Can personality tests be a cost-effective tool for managerial selection?) In T. Szustrowa (Ed.) *Testy w biznesie - standardy i praktyka* (pp. 139-151). Warszawa, Poland: Pracownia Testów Psychologicznych PTP.