



**Executive summary:**

**The Value of Assessment Tools in Personnel Selection**



# The Value of Assessment Tools in Personnel Selection

Recently, just around new year, [a new meta-analysis](#) by Paul Sackett and colleagues was published that shook up the scientific and HR Community. This meta-analysis concludes that most of the aspects we measure or assess during a selection process do not predict performance as well as we previously thought.

Boom! The main reason why we use assessment tools and tests is of course to get an informed idea of how well a candidate will perform in the job we are hiring for. So, if those tools don't predict as well as we thought they did, can we then still use them and trust their results?

To really understand which lessons to draw from this meta-analysis, Hudson R&D has thoroughly scrutinised it and discussed its results with prof. dr. Filip Lievens, world authority in the field of selection and assessment research and one of the co-authors of the paper. And phew, what a relief to find out that assessment tools remain very relevant! Let's have a look.

## How to interpret the results of the new meta-study?

There are a few important aspects to take into account when looking at the results:

- 1) The authors **use specific statistical techniques to avoid over-correction**. Where the old meta-analyses (like the renowned paper by [Schmidt & Hunter \(1998\)](#)), tend to over-correct for artefacts like unreliability of the criterion or range restriction in their attempts, Sackett et al. (2021) use the principle of conservative estimation: when there is any uncertainty about how to correct, they claim it is better not to correct at all than to overcorrect. This logically results in more modest validity estimates than the ones we know from older meta-analyses.
- 2) The new validity **estimates might seem rather on the low side, but in fact they are not**. For example, the association between cognitive ability tests and work performance (average validity of .31 in the new paper) is much stronger [than the effect of ibuprofen on pain reduction](#) (average validity of .14).



So, if you look at the new estimates from the perspective of a medical researcher, being able to predict human behavior with such high accuracy is rather fantastic.

- 3) The estimates in the study **cover a large variety of studies, tools and work settings**. It is important to realise that the validity figures reported in meta-analyses are not 'exact' numbers. The values reported are averages, calculated from the validities found in a wide variety of primary studies. In fact, the predictive value of any given selection method varies substantially across studies.

Fortunately, to get an idea of the degree of variability, Sackett and colleagues also reported the **credibility intervals** for each of their estimates. For example, while structured interviews have the highest mean validity in the paper, they also have a very wide credibility interval. This means that an interview's predictive power varies strongly across settings, as it will depend on a myriad of factors: the interview design, specific questions asked, the scoring method, question quality, the match (or lack thereof) between questions asked and job characteristics. And of course a lot will depend on the interviewer: their level of experience, how well were they trained, how effective they are in preventing any personal biases from clouding their judgment and so on.

- 4) The study of Sackett et al. (2021) is an international meta-analysis, which draws on a lot of primary studies conducted in the US. But **not all of these studies are relevant in a European/Benelux selection context**.
  - **Highly fakeable tests or questionnaires** (e.g. integrity tests, job interest questionnaires, ...) are more problematic in Europe than in the US. When applying for a job, Europeans will not easily admit that they have once stolen from an employer (a question typically asked in an 'overt' integrity questionnaire) or that they are not really that interested in the vacancy. In the US, however, before taking part in a selection process, candidates are often required to sign an 'honesty contract' which formally states that any answers given are truthful and honest. And, the legal consequences for breaching such a contract can be quite severe.



The consequences of faking or dishonesty are much more serious in the highly legalised US context than in Europe.

Resulting in a high predictive power (knowing that the candidate has once stolen or is actually not interested in the vacancy, of course predicts that there might be some issues) for easily fakeable tests, which then unfortunately simply do not work as well over here. Easily fakeable tests are thus considered less relevant in Europe, as the results cannot be trusted.

- **Sense of privacy:** European candidates will be reluctant to answer questions that ask for private information with no clear link to the vacancy they are applying for. Some of the tools reported in the new meta-analysis fall in that category; for example non-contextualised personality tests with no link to the work context, or empirically keyed biodata questionnaires which probe behaviours and events that occurred early in one's life and that are proven to correlate with work performance. Typical biodata questionnaire items ask about things like the amount of "parental warmth" received when growing up, or the number of hours spent studying as a student. Although these aspects may be predictive of work performance, most candidates in Europe will experience such questions as a serious breach of their privacy.

Therefore, these types of tests are therefore hardly ever used over here. We usually focus on the job-relevant tools only, even if by doing so we might miss out on some highly predictive information.

- 5) A few **other selection tools** reported in the meta-analysis might **need some more background information** on the specifics to be able to understand the results.

- Personality-based Emotional Intelligence (EI) or [mixed Emotional Intelligence](#) should not be confused with much better known ability-based EI tests that consider EI to be an actual ability or facet of intelligence. Personality-based EI is in fact a **personality-based**



**compound that encompasses a constellation of personality traits, affect and self-perceived abilities** rather than aptitude.

It is indeed in the combination of different relevant personality traits (in this case labelled 'Emotional Intelligence') that we can often see whether someone is fit for a job or not.

- The results of the **individual personality dimensions** mentioned in the meta-analysis (Conscientiousness, Emotional Stability, Extraversion, Agreeableness, Openness) indicate the **predictive power** of these dimensions for work performance in general, i.e. **for all types of jobs**. However, we know from other studies that while the validity of any dimension may be very low for some jobs, it is much higher for other jobs. Openness, for example, was shown to be a [good predictor](#) for jobs requiring a high level of creativity and adaptability, while Agreeableness is a [good predictor](#) of success in customer service jobs. It is thus rather striking that the validities of the Big Five personality factors for all jobs are still fairly high in this meta-analysis.
- **Assessment Centers (ACs)** are often only used for a **restricted pool of candidates**. ACs are mostly only used for 'selecting in': meaning that usually only the final candidate takes part in an AC to decide whether they will get the job. ACs are rarely used for 'selecting out': where the entire applicant group takes an AC to filter out (select-out) the weakest candidates. So the fact that the new validity estimates show that even within that pool of top candidates, ACs allow us to differentiate between high performers and less strong performers, makes these results particularly favourable for ACs as a tool for selecting in.



## Overview of the results

The graph here below is a visual representation that includes all aspects discussed above: it shows the degree of variability of each estimate, it includes only tools that are relevant in a European selection context which are termed such that we can fully grasp their meaning.

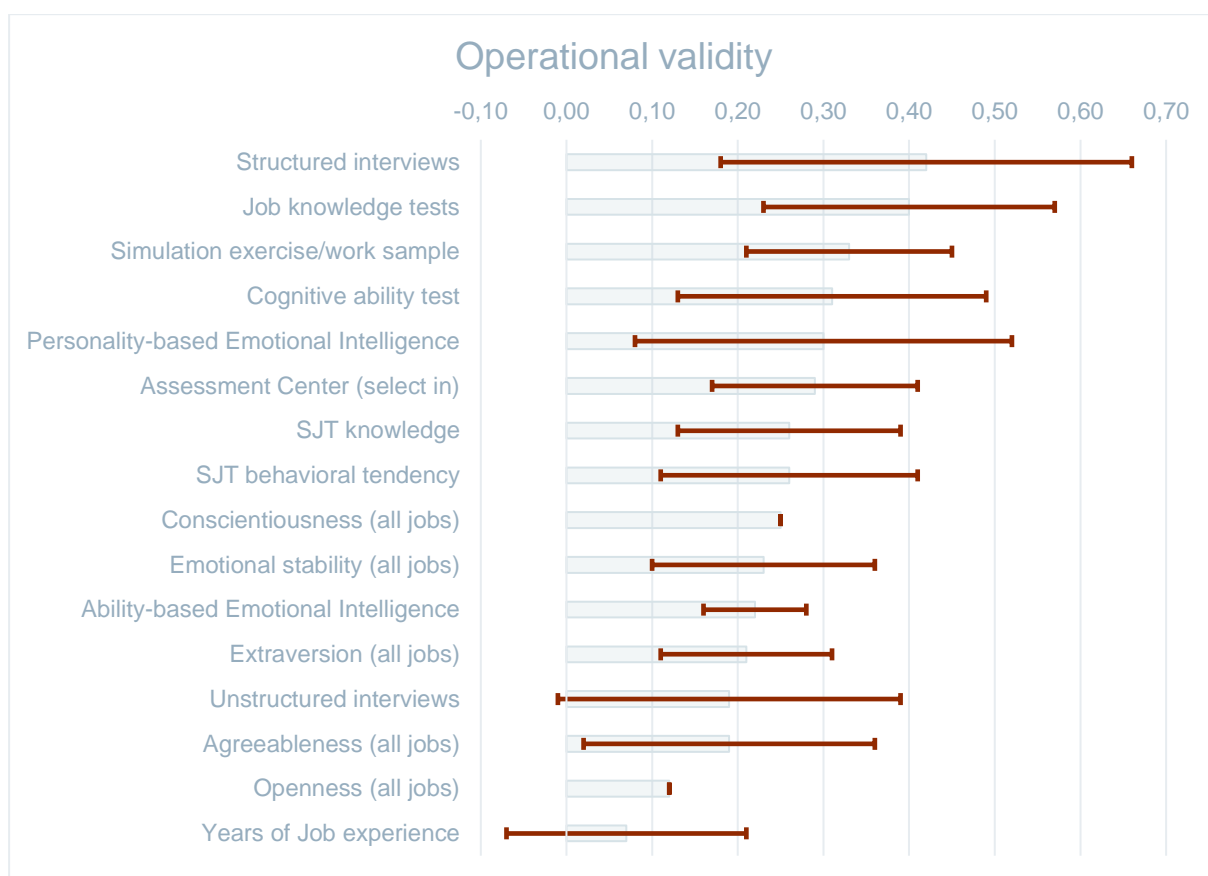


Figure: Operational validity estimates and their 80% credibility intervals taken from Sacket et al. (2021), adapted to the European/Benelux context. The bars represent the 80% credibility intervals around the mean operational validities.



## And now? Which selection method is the best one?

### It depends on the context

The figure above shows that there is some kind of ranking possible: some tools do seem to predict better than others, but the **ranking is not very precise if you look at the high levels of variation in predictive power across settings**. With such a big margin of error it does not seem that useful to focus on exact numbers and rankings.

Most of these estimates are not significantly different from one another, there is a big overlap in the credibility intervals around the estimates. So can we really say that, based on these numbers, structured interviews are better predictors than work samples? Or that personality-based Emotional Intelligence is a better predictor than a Situational Judgement Test (SJT)?

Meta-analyses aggregate the conclusions of many separate studies, that are generalised over all jobs, situations and types of tools within each category. **In order to get an estimate of the predictive validity of a tool in a *specific* context, one should search for more specific studies**. For the frequently used Hudson tools, we do have for example multiple *specific* predictive validity studies available. And we can conduct extra analyses on your own population to see what works best for you.

### It depends on the combination of tools used

Another important aspect to consider is what happens when different techniques are applied together. This is related to **incremental validity**: what is the added value in terms of predictive validity of one technique when you are already using another technique. In the scientific literature, we indeed find that e.g. Structured interviews, personality questionnaires and work samples offer a [significant increase in predictive validity](#) when combined with a cognitive ability test. Moreover, [Assessment Centres](#) are known to explain a sizeable proportion of variance in job performance beyond cognitive ability and personality.



To compose and devise a selection procedure, **HR professionals need to carefully consider the specific context in which they are operating** (organisation, job demands, possible candidate pool, ... ) **to define which combination of methods would be best** to maximise both predictive power and utility.

### And, validity is not all that matters

For many organisations other aspects might be at least as important as predictive validity when setting up a selection procedure. Such as

- **Equal opportunities:** minimising adverse impact to ensure that people from different subgroups (gender, cultural background, age, ..) have an equal chance of being selected on the basis of their test score(s).
- **Cost and practical feasibility:** reducing the resources needed, especially when dealing with a large applicant base. Many organisations will have to balance investments in face-to-face interviews or observations against the administration of computerised tests that can be taken remotely and on a larger scale.
- **Candidate experience:** is important for attracting top talent by building a positive employer brand. The candidates' experience will depend on the perceived face validity (how relevant candidates perceive the tests to be for the job they are applying for) and the perceived fairness of the application and testing procedure.

Up to date, no single selection tool exists with perfect scores on all important aspects: demonstrating high predictive and incremental validity as well as low adverse impact, at a low cost, while being experienced positively by most applicants. All tools have certain advantages and disadvantages and trade-offs exist between the important aspects. For more information on these aspects, we refer interested readers to our paper 'How to choose the most appropriate selection tool'.

In conclusion, to compose and devise a selection procedure, **HR professionals need to carefully and critically consider all these aspects**, taking into account the client's specific context and desires and the requirements of the job they are hiring for.





More detailed information about can be found in our [full white paper](#). Please do not hesitate to contact us if you need additional information.



## Contact

t. +32 497 52 63 86  
amelie.vrijdags@hudsonsolutions.com  
[www.hudsonsolutions.com](http://www.hudsonsolutions.com)

Hudson Belgium N.V./S.A.  
Moutstraat 56 9000 Ghent, Belgium  
t. +32 (0)9 222 26 95  
info@hudsonsolutions.com  
www.hudsonsolutions.com