Predictive validity in assessment methods.

A comparative analysis of a selection of assessment methods.

For over a hundred years, psychologists has sought to identify the best assessment methods in predicting people's ability to succeed professionally.

Frank L. Schmidt's Meta Analysis

In 2016, University of Iowa Professor *Franck L. Schmidt* published a meta-analytic compilation of the field's findings in his working paper, **"The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years."**

In this meta-analysis, Schmidt rigorously researched and compared 31 assessment methods in order to distinguish those that are of true practical utility in selection procedures, from those that are of little or no use, and which may indeed be detrimental when differentiating candidates based on their ability to succeed professionally.

The following is a summarized account of the findings.

What is predictive validity?

The predictive validity of an assessment method is a mesure representing just how exact the method is in formulating predictions.

It is often determined by correlating the different elements of the method's output (for example, scores obtained on personality tests, technical knowledge tests, or even recommendations resulting from an interview) to the criteria that the method attempts to evaluate (for example, performance levels of new hires in a given role).

This correlation, represented by the letter *r*, can vary from -1 to +1.

When *r* = +1

A correlation value of +1 means that an assessment method can predict individual performance with 100% accuracy.

This would mean that you could "blindly" trust this method. It would allow you to make the right decision 100% of the time and to anticipate without error the future performance of any candidate in a specific role.

It goes without saying that no assessment method provides such accurate results. The level of complexity of the phenomena being predicted, combined with the intricacy of the causes that determine or influence those phenomena, makes it impossible to reach that level of predictive power.

It does not, however, preclude attempts to match it closely, as we will see below.

When *r* = -1

A correlation value of -1 means that an individual's performance is systematically opposed to the prediction issued by the method.

In a nutshell, every time the method anticipates candidate success, the candidate will fail… and vice versa. It would simply require selecting the opposite of whatever the method predicts. It it says black, you chose white. If it says white, you choose black.

Predictive validity in the real world

In practice, most assessment methodologies have a positive correlation ranging from .01 to .70.

A correlation value is considered acceptable starting at .30. Predictions from assessment methods with lower values ($r < .30$) should be considered unreliable.

To what degree can performance be effectively measured using any particular assessment method?

The value of *r* determines the correlation between the output of an assessment method and the criterion that the method attempts to predict.

To obtain an accurate representation of the utility of the predictions of any assessment method, we need to calculate the *shared variance* of the predictions, otherwise known as r^2 or "r squared.

For example:

Imagine we are assessing 100 sales executives, all with varying levels of performance. Some are great, some are good and some are not so good.

The observed differences in performance between these sales executives is what we call *performance variability*. It will be very large if the difference between the best and the worst sales executives is very large. It will be very small if all the executives have a similar level of performance.

If a particular assessment method for sales executive selection has a predictive validity coefficient of .50 ($r = .50$), this means the method can account for 25% of the differences in

performance between the best and worst sales executives (50^2 = 0.25 = 25%).

Since the maximum predictive validity currently possible is about .70, this means that—at most—we can account for approximately 50% of the differences between the best and worst candidates through the most powerful assessment methodologies we have today ($.70^2$ = 0,49 = 49%)

Is this a significant percentage?

This is, indeed, debatable. Being able to predict 50% of a person's ability to succeed in a role, also means not being able to say anything about the remaining 50%…

But critical analysis of the context acknowledges the significant predictive power of any method with an operational validity score of .70.

Consider the following: A person's qualities will be a particularly determinant factor on their ability to succeed and settle into the team… but there are a number of other factors that will also impact an individual's level of performance:

- The coherence between how a position is described and the reality of the job.
- The way the person will be onboarded.
- The style and quality of management of their supervisor.
- The company and team culture.
- The affinity between the new hire and the existing team members.
- The conflicts likely to arise within the team or with other colleagues (internal or external).
- The tools and methodologies that may or may not optimise the work of each employee.
- The company's financial situation at the moment the new hire joins the team.
- The degree of competition the company faces.
- The changes in business strategy at the time of recruitment or six months after their arrival.
- The changes to legislation directly impacting the profession or the specific transactions the new hire is responsible for.

- …

All of these are factors with the potential to impact performance for new hires (and this list is not exhaustive).

Therefore, being able to predict 10, 15, 20, 30 or even 40% of an individual's future performance *before they have ever set foot in your company* seems far from inconsequential. And upon further analysis, it seems rather substantial!

What doesn’t work.

Useless criteria in selection processes.

One of the many benefits of of Schmidt's meta-analysis is that it highlights the criteria that continue to be used regularly in selection processes, in spite of their low or inexistent predictive value.

These criteria do not only fail to contribute value, but their use might also result in an unintended counter-productive effect for decisions which depend on these criteria.

Among them are:

A candidate's age.

The correlation coefficient between a person's age and their performance is .00.

This means that there is strictly no relationship between a person's age and their ability to succeed in any given role. A very young candidate and a very old candidate are just as equally likely to succeed in any role and display equal levels for performance.

Therefore—scientifically speaking—selecting candidates based on their age is ludicrous.

Years of training.

The correlation between a person's years of training and their performance levels is .10.

With an r^2 of 0.01, this approach can only predict 1% of the differences in performance between the best and worst hires.

Continuing to demand that candidates certify at least 3, 4 or 5 years of training is therefore pointless.

This criterion may reassure recruiters or managers (as they may believe this aspects plays a critical role in success) but from a strictly rational standpoint, it is meaningless.

Years of experience.

The correlation between a candidate's years of experience and their performance level is .16.

With an r^2 of 0.03, this approach can only predict 3% of the differences in performance between the best and worst hires.

Likewise, continuing to demand that candidates certify having at a minimum of 1, 2 or 3 years of previous work experience in a field is futile.

NB:
The conclusions of the study are in no way influenced by specific ideologies. They are the logical result of a meticulous analysis of 100 years of practical findings in personnel selection, worldwide.

What really works.

The *g* factor accounts for 42% of future performance.

The best predictor of a person's performance in professional contexts is the *g* factor, assessed through GMA (General Mental Ability) tests (with r = .65, and r^2 = 42%).

The *g* factor is a measure of a person's ability to reason accurately and solve problems with different levels of complexity.

This specific conclusion is perfectly aligned with the observations of HR professionals and certain managers, who describe problem-solving as the most critical skill in achieving success in modern work environments (constantly changing, less certain, more complex, and more ambiguous), and this independent of the technical level required by daily tasks in a specific professional context.

AssessFirst's BRAIN Test

Part of the AssessFirst suite, BRAIN was specifically designed to evaluate the *g* factor through a unique adaptive approach and fully gamified experience.

An adaptive test is one in which each new question presented to the test-taker is based on their answers to previous questions. This results in a more pleasant experience for test-takers, since no person is asked to answer questions of a level of complexity beyond their means.

By following the principles of gaming, we increase engagement and focus their attention during the time it takes to assess their cognitive ability.

Our adaptive approach allows us to cut in half the time required to complete a full assessment of the person's *g* factor.

The more complex the role, the more relevant the *g* factor will be.

The predictive power of the *g factor* increases exponentially with the complexity of the role.

In low-complexity roles, a person who scores high on *g* factor tests like BRAIN will perform X times better than a person who scores low on the same test.

In mid-complexity roles, the difference in performance will be threefold (X * 3).

In high-complexity roles, the difference in performance will be eightfold (X * 8) between those who obtain a high score and those who obtain a low score on the *g* factor test.

Personality: A powerful predictor… if contextualized.

Although the predictive validity of personality assessments (Big5 / OCEAN / FFM) varies according to the traits being considered, it remains highly relevant in the context of candidate selection.

By combining a GMA test like BRAIN and a personality assessment like SHAPE, you can obtain predictive validity scores ranging from .65 to .78, which represents a performance variability ranging from 42% to 61%.

AssessFirst's SHAPE Test

SHAPE—our proprietary personality questionnaire—allows you to assess 20 dimensions of personality which directly influence and determine the behaviors people display in professional contexts.

It's important to note that, on their own, personality traits have a low predictive value for professional success in general.

Each role, each profession and each job is unique in its specificities. Therefore, the personality dimensions required to predict success will vary widely from one job to the next, and from one profession to the next.

This means that, for personality tests to be truly useful, it is indispensable that the personality criteria which determine success for a specific role be well defined in advance.

Motivations: a valuable complement.

Personal interests and motivations, assessed through questionnaires like DRIVE, have predictive validity scores of about .31 on their own. By combining them with *g* factor assessments, operational validity scores can reach an *r* of .71, which translates into an r^2 of 50%.

As with personality, motivations need to be contextualised in order to increase their predictive power for candidate selection.

It goes without saying that each role and occupation demands different motivations.

AssessFirst's DRIVE test

DRIVE—AssessFirst's proprietary motivations questionnaire—was specially designed to consider the widest range possible of individual drivers.

This allows us to assess 20 fundamental needs through one single questionnaire.

As a result, it can be used to predict professional success within a large range of jobs, roles and occupations.

In addition, those 20 needs are particularly useful in defining the specific characteristics of a company's unique culture.

Boosting the predictive power of your hiring process…

Maximizing the predictive power of an assessment method—whether for pre-selecting or selecting candidates—is highly dependent on the recruiter's ability to accurately identify success predictors.

In order to identify the best success predictors for any give role, AssessFirst highly recommends using our Artificial Intelligence module to create customized predictive models based on your current workforce.

This allows you to build and rely on personalized algorithms that demonstrate high predictive power (with an operational validity ranging from .65 to .70, on average).

Why do recruiters get mixed results?

It is no secret that the majority of recruitment processes today turn out to be failures in the short term.

In fact, 19% of all new hires fail within the first six months. Within the first year, 36% of them have quit or are let go. And by the 18-month mark, this number rises to 46%—about half of all new hires.

In 89% of cases, hiring failure is due to an inaccurate assessment of behavioral factors.

This is in part because of the undue importance given to technical ability or "hard skills", when it's soft skills that have a high predictive power for a considerably large number of roles.

Additionally, most recruiters fail to rigorously define the right factors to be used as pre-selection criteria. Often, these criteria are defined based solely on the hiring manager's preference for the profile they seek.

Unfortunately, this method is highly limited, as it only considers the operational manager's representation of that which predicts success—a representation which is often far from reality.

Field observation vs. subjective representations

To identify criteria that can accurately predict success in a role prior to candidate preselection, we recommend conducting a field-oriented study rather than relying on subjective accounts—whether or not these accounts come from people who currently hold the position in question.

The best approach is to assess people in the same role as the open position, and to identify for each person the indicators that accurately measure their performance (be it on a daily, weekly or monthly basis). A meticulous cross-analysis of the data will enable you to distill the criteria that you should focus on during preselection.

AssessFirst's proprietary Artificial Intelligence allows you to automate a significant percentage of the arduous work necessary to arrive at these accurate predictors.

About AssessFirst

AssessFirst has developed a predictive recruitment solution allowing companies to predict how well candidates and employees will succeed and thrive in their job. The AssessFirst solution analyzes data on over 5,000,000 profiles, whether candidates, employees or recruitment professionals. Today, over 3,500 companies use the AssessFirst solution to raise their performance by up to 25%, drive down their recruitment costs by 20% and reduce their employee turnover rate by 50%.

Find out more: www.assessfirst.com