Behavioral Assessment Selection Guide

How to choose an assessment for hiring: what separates the categories, what the validity evidence says, and eight questions to put to a vendor.

Most organizations choose an assessment the way they choose a conference venue: someone recommends one, it looks professional, and it becomes the standard. Here is what to ask instead.

First, decide what you want it to do

The single most common mistake is using a tool built for one purpose to make a decision of another kind.

IF YOU WANT TO…	YOU NEED…
Help a team understand each other	A development tool. Basic personality assessments can work, as can a behavioral assessment like PATH, and measures of psychological safety or accountability preferences have value here too. They give you a way to see how a team works through a particular lens.
Predict who will perform well in a specific role	A selection tool, scored against a profile of that role. The essential element of behavior hiring is building a behavior profile that is not based on a job description but on day-to-day modes of work: purpose, approach, thinking, habits and so on. Early versions approximate the profile and evolve as specific behaviors prove more predictive of particular workplace outcomes.
Decide where to invest in someone after hiring	A development tool, which follows the same process as hiring but focuses on the essential areas for development, creating the opportunity for specific and measurable growth in key roles.

A development tool used for selection is the problem case. It describes without comparing, so whatever decision you make with it, you made for other reasons and attributed to the report.

The categories

Personality inventories. Measure broad traits. MBTI's publisher advises against using it for hiring. Big Five instruments are better grounded scientifically, but general conscientiousness measures came in at .19 validity in the 2022 Sackett estimates, roughly the same as an unstructured interview.

Cognitive ability tests. Measure reasoning. Revised to .31. Legitimate predictive power, with the largest subgroup differences of any common method, which is why they need care and legal

review.

Situational judgment tests. Present scenarios and ask what you would do. Cheap to run, easy to coach for, and increasingly easy to answer with AI assistance.

Work samples. Ask the candidate to do a piece of the actual job. Revised to .33, and among the most defensible methods you can use, though they take real effort to build and score.

Work-context behavioral assessments. Measure how someone prefers to work: what motivates and drains them, how they approach problems, collaboration, thinking, habits and organizing their work. Their value depends almost entirely on whether the results are compared to a defined profile of the role, and whether they are verified afterward. On their own they describe. Inside a structure they predict.

GoodJob's PATH assessment is an example of the last category used inside a structure: results are read against a RoleDNA profile built for the specific role, then tested in a structured behavioral interview where interviewers rank the answers. GoodJob reports predictive validity of .65 to .75 for the full process from its action research, a figure that comes from practitioner data rather than from independent meta-analysis, and that describes the whole process rather than the questionnaire.

Eight questions for a vendor

1. What exactly does this predict, and what is your evidence? Ask for the technical manual, not the brochure.
2. Is the validity evidence for the questionnaire alone, or for a process that includes interviews? These are very different claims.
3. Was the evidence produced independently, or by you?
4. What are the subgroup differences, and what adverse impact analysis has been done?
5. How do candidates game it, and what stops them?
6. What happens when a candidate's results and their interview evidence disagree? A vendor whose answer is "trust the assessment" is selling something you shouldn't buy.
7. Does it compare candidates to the role, or only describe them?
8. What does the candidate get to see about their own results?

That last one is not a formality. An assessment the candidate never sees, used to make a decision they are never told about, is a reputational problem waiting to happen, and in several jurisdictions a legal one.

How to weigh the answers

Be wary of any single validity figure quoted without its source, its method, and what it was measured against. The honest version of a claim sounds like: this number, from this kind of

study, for this part of the process. Vendors who can produce that are telling you something. Vendors who cannot are telling you something too.

Be equally wary of your own enthusiasm. An assessment that confirms what you already thought about a candidate feels like validation and is usually just an echo. The one that disagrees with you is where the value is, which is why the step after the assessment has to be an interview designed to test the disagreement rather than smooth it over.

Where it sits in the process

An assessment is a middle step, never a gate on its own. The sequence that works: define the role's behaviors, screen for the non-negotiables, assess, then interview against both the role profile and whatever the assessment raised. Teams that put the assessment first use it to reject people, which is where most of the legal exposure and most of the missed candidates live.