Behavioral Interview Guide

How to design and run a structured behavioral interview, calibrate interviewers, and avoid the errors that pull a good process back toward a conversation.

Structured interviews are the strongest single predictor of job performance we have evidence for, at an average validity of .42 in the 2022 Sackett re-analysis. Unstructured ones sit at .19. Same hour, same candidate, more than double the predictive value.

But .42 is an average with a wide band around it, roughly .18 to .66. Structure alone doesn't get you to the top of that range. What gets you there is how closely the interview tracks the specific job. This guide is about that gap.

What structure actually means

Four things, all decided before the first candidate:

1. Every candidate gets the same core questions.
2. Those questions come from the job's behaviors, not from interviewer preference.
3. Every answer is scored against criteria written in advance.
4. Scores are recorded before anyone discusses the candidate.

Drop any one and you are back toward .19. The fourth is the one teams skip most often and the cheapest to keep.

Designing the interview

Split the behaviors across interviewers. Five behaviors, two interviewers, is better than both covering all five shallowly. Overlap one behavior deliberately so you have a calibration point.

Budget the time. A 45-minute conversation holds about five behavioral questions with real probing, plus ten minutes for the candidate's own questions. If you have planned eight, you will rush all of them.

Decide who asks what, in writing. Two interviewers both asking about accountability, and nobody asking about ambiguity, is the most common failure in a well-intentioned process.

Running it

Open by explaining the format. Thirty seconds: "I'll ask about four or five specific situations from your past. I'll take notes. Take a moment before answering if you need it." This improves

answers, particularly from people who think more slowly than they speak, and it costs you nothing.

Ask the question, then stop talking. The silence after a behavioral question is uncomfortable and productive. Interviewers who fill it tend to supply the answer they were hoping for.

Probe for the specifics. A first answer is usually a summary. The evidence is one or two layers down.

IF THEY SAY	ASK
"We decided to..."	What did you personally do?
"I usually..."	Tell me about the last time
"It worked out well"	How did you know? What changed?
"It was a difficult situation"	Difficult how? What specifically was at stake?
A hypothetical	Has that come up? Tell me about that time

Take notes on what they did, not how it felt. "Reconciliation error, told CFO same day, added a day-two check" is usable in a debrief. "Seemed honest" is not.

Score immediately. Before the next meeting, before you speak to anyone. The scorecard is built so this takes two minutes.

Calibrating interviewers

The difference between a .30 interview and a .60 one is mostly here.

Run one practice round before interviewing real candidates: two interviewers score the same recorded or role-played answer independently, then compare. Where they diverge, the criteria are ambiguous, and you fix the wording rather than the people.

After the first two or three candidates in a search, check the grids. If one interviewer is consistently a point above everyone else, that is worth a conversation. They may be lenient, or they may be asking a better follow-up question than anyone else.

Common errors

Talking too much. Interviewers who speak more than a third of the time learn less and like the candidate more. These two facts are related.

Selling before you've assessed. Sell at the end. Candidates who have been sold to give more polished answers to everything that follows.

Letting the résumé drive. Walking through someone's career chronologically feels natural and produces no comparable data. The behaviors drive the interview; the résumé is context.

Mistaking fluency for competence. Some people narrate their work well. Some do the work well. The overlap is smaller than it feels in the room, and AI interview coaching has widened the gap between the two.

behaviorworking.com Free to copy and share.

Panel interviews by default. McDaniel's meta-analysis found individual structured interviews predicted better than panels. Panels save the candidate's time and cost you independence of judgment. Use them when scheduling demands it, not as the standard.

What to do when an answer isn't there

Sometimes a candidate genuinely has no example, because the situation has never arisen in their work. That is information about fit, not a character flaw, and it should be scored as "didn't get to it" rather than as a failure.

The exception worth probing: if someone has no example of being wrong, no example of disagreeing with a manager, and no example of a mistake being discussed openly, you may be looking at someone who came from an environment where none of that was survivable. Ask what they expected would happen. The answer tells you where they have been, which shapes how they will behave in your team on day one.