HIRING / INTERVIEW SCORING

The Structured-Interview Playbook

Score any structured hiring interview on the six behaviors that separate real evidence from a rehearsed answer, using the words the candidate actually said.

Version 1.0 · August 13, 2026 · talk2bud.com/blog/tsa-hiring-scoring-rubric

INSIDE

USE IN 30 SECONDS

Use this in 30 seconds. Take your last interview. Score each of the six criteria on page 3 a 1 to 5, based only on what the candidate said, and write the line that earned each score. The two lowest criteria are what your next round should probe. A score you cannot attach a quote to is a score you invented.

To have talk2bud do this for you on every interview, see the last page.

Why Two Interviewers Disagree About the Same Candidate

Two people sit through the same structured interview, the same five questions, the same hour, and walk out a point apart. Neither is lying. The gap is not about what the candidate said; it is about what each of them remembered forty minutes later.

Structure is the fix, and the research is not subtle: interview validity rises with structure, from about .20 with none to about .57 at the top (Huffcutt & Arthur, 1994), and a 2022 re-analysis puts structured interviews near .42 against .19 for unstructured ones (Sackett et al., 2022). The mechanism is boring on purpose: the same questions, the same rubric, and a score read from the words rather than the mood.

This playbook scores the six behaviors that separate a real, checkable answer from a fluent one. It is written so a person can run it in ten minutes, and so talk2bud can run it on every interview automatically. The TSA structured interview is the worked example throughout; the TSA does not publish its internal scorecard, so this is a general structured-hiring rubric, not a copy of theirs.

WHAT MAKES THIS DIFFERENT FROM A CHECKLIST

A checklist asks “did the candidate give an example?” and any answer with the word example in it ticks the box. This scores what a 2 sounds like against what a 5 sounds like, with the phrasing, so two interviewers reviewing the same answer land on the same number. That is the only property that makes a rubric worth running twice.

Instructions

Analyze this as a hiring interview. “You” is the interviewer, “Them” is the candidate. Score the interview against the six criteria in the playbook below, giving each a 1 to 5 and quoting the line from the transcript that earned that score. Score each of the interviewer’s questions separately before forming an overall read. Never score a criterion you have no evidence for; mark it n/a and say

what was missing. Judge only what the candidate said out loud, never delivery or likeability you cannot hear in the words. Where a score is 3 or below, write the exact follow-up question the interviewer should have asked to get real evidence.

What this lens is for

A structured hiring interview where the interviewer asks the same questions of every candidate and needs a defensible score, not an impression. A good answer describes a real past situation, the action the candidate personally took, a decision with a tradeoff, and a result you can check. A weak answer runs on maxims (“I always stay calm”) and hypotheticals (“I would handle it professionally”), which cannot be verified. The output should let a second interviewer agree with the score by reading the quotes.

How to read the call

- Read once end to end before scoring. First-pass scoring rewards the loudest story.
- Split the interview by question and score each answer on its own evidence.
- For each answer, mark whether it is a real past event or a hypothetical, because criterion 1 only applies to the first.
- Note every “we” where you needed an “I,” and every result stated as a feeling rather than a fact.
- Check consistency across questions last: one strong answer among four thin ones is a rehearsed story, not a strong candidate.

The criteria

1 A real situation, not a maxim

LISTEN FOR: A specific past event with a when, a where, and who was there, not a general rule about how they behave.

SCORE	SOUNDS LIKE
1-2	A maxim, no event. “I always stay calm and professional under pressure.”
3	A vague event with no anchor. “There was this one time a customer got upset, and I handled it.”
4-5	A dated, located, specific event. “Two weeks ago on the late shift, a customer demanded a refund we stopped giving after 30 days.”

QUOTE: The sentence that establishes the specific situation, or note that none was given.

2 Their action, owned as “I”

LISTEN FOR: What this candidate personally did, in the first person, not what the team did.

SCORE	SOUNDS LIKE
1-2	The team did it. “We handled it as a group and got through it.”
3	Mixed ownership. “We decided, and then I sort of took care of it.”
4-5	Clear personal action. “I checked the receipt date myself and gave him the answer before my manager stepped in.”

QUOTE: The first-person verb that names their action, or the “we” they hid behind.

3 A decision with a tradeoff named

LISTEN FOR: A choice between real options, and why they picked one.

SCORE	SOUNDS LIKE
1-2	Pure compliance, no choice. “I just followed the procedure I was trained on.”
3	A decision with no reasoning. “I decided to give him the credit.”
4-5	Options and a reason. “Two choices: bend the rule or offer store credit. I offered credit, since a chargeback costs us more.”

QUOTE: The line where they weighed one option against another.

4 A result you can check

LISTEN FOR: An outcome with a number or an observable consequence, not a feeling.

SCORE	SOUNDS LIKE
1-2	A feeling, secondhand praise. “It went really well and my manager said it was great.”
3	A soft outcome. “He calmed down and left okay.”
4-5	A checkable result. “He took the credit, and we made that the standard reply for late refunds afterward.”

QUOTE: The outcome sentence, and whether it contains anything you could verify.

5 What they would do differently

LISTEN FOR: Honest reflection, not a rehearsed weakness or a humblebrag.

SCORE	SOUNDS LIKE
1-2	Nothing to improve. “Honestly, I wouldn’t change a thing.”
3	A generic weakness. “I could always communicate a bit better, I guess.”
4-5	A specific, real change. “I’d have brought my manager in earlier so he heard the policy from both of us.”

QUOTE: The reflection line, or note that reflection was absent.

6 On the competency the question asked

LISTEN FOR: Evidence for the thing you asked about, not a nearby virtue.

SCORE	SOUNDS LIKE
1-2	Answers a different question. Asked about conflict, talks about being organized.
3	Drifts toward the competency but never lands on it.
4-5	Directly on target. Asked about conflict, the whole story is the conflict and how it resolved.

QUOTE: The part of the answer that addresses the competency you asked about, or note the drift.

Scoring

Score each criterion 1 to 5. Mark n/a where the transcript gives you nothing, and say what was missing rather than guessing. For a hypothetical question, skip criterion 1 and weight criterion 3. Report the average across the criteria you could

score, to one decimal place, plus the per-question spread.

- **4.0 and up.** Strong evidence. Advance, and note the one competency to confirm in the next round.
- **3.0 to 3.9.** Real but uneven. Name the two lowest criteria; those are the next round's questions.
- **2.0 to 2.9.** Fluent, unproven. Little that was said can be checked.
- **Below 2.0.** No evidence. Treat as a screen that did not happen.

A high average with a 1 on criterion 1 across most questions is not a strong candidate. Say so plainly: the answers were rehearsed, not real.

Red flags

- **All maxims.** Every answer is how they would behave, never a thing that happened.
- **The "we" curtain.** The candidate never once says what they personally did.
- **One rehearsed story.** A single vivid answer, and four thin ones around it.
- **Feelings as results.** Outcomes described only as "it went well," never as a fact you could check.
- **Off-competency.** The strongest story does not answer the question that was asked.
- **No reflection.** Nothing they would change, on any answer.

How to report

- Open with the average, the per-question spread, the strongest criterion, and the weakest. One sentence each.
- Under what went well: each criterion scoring 4 or 5, with its number and the quote that earned it.
- Under what to improve: each criterion scoring 3 or below, with its number, the quote that cost it, and the follow-up the next round should ask.

- Under decisions: only what the panel actually decided, in plain terms (advance, hold, reject).
- Under action items: who owns the next round or the reference check, and by when.
- If instead you are asked for a single performance block, run the six criteria there, item by item, same evidence rule.

Coaching notes

- Score the words, not the delivery. A confident voice is not a checkable result.
- Compare candidates on the same criterion, never on overall gut. That is what makes the scores fair.
- If your whole panel scores soft on criterion 3, your interviews reward likable people who never show judgment. Fix the question, not the scorer.

Calibration: Two Interviewers, One Number

A rubric earns its keep only if two people scoring the same interview land within a point. When they do not, the argument is almost always about an adjective. Calibration forces it down to the quote.

Score criterion 4 on this answer, then read the key: *"So I took the lead on it, and honestly it went really well. The customer was happy, and my manager said it was handled great."* The answer is a 2. It sounds confident and carries a manager's praise, but criterion 4 asks whether there is a result you can check, and "it went really well" is a feeling while "my manager said it was great" is secondhand. No number, no observable consequence, so it scores a 2. A score with no line under it is a number you made up.

Rolling It Out in 30 Days

Week 1. Baseline. Score your last five interviews, or your next five, without changing how you run them. Find your panel's weakest criterion before anyone gets defensive about a number.

Week 2. Calibrate. Two interviewers score the same interview independently, then compare. Anywhere you disagree by more than a point, argue it to the quote until the anchor settles it.

Week 3. Every panel, every candidate. Score every interview on the rubric, and track the gap between the average score and the actual hire decision. When you hire the 3.1 over the 4.2, write down why in one sentence.

Week 4. Read the trend. If the whole panel scores soft on one criterion, that is a question problem, not a scorer problem. A rubric your team has argued with and edited is one they will actually use.

Changelog

- **v1.0 (2026-08-13).** First release. Six criteria with 1-to-5 behavioral anchors, calibration exercise, 30-day rollout, and the importable Interview Lens.

</content>

Run this as a Lens in talk2bud

Scoring one call by hand takes ten minutes. Scoring every call takes a program nobody sustains past week three. talk2bud records the call, transcribes it, and reads it back through this exact playbook, quoting the lines that earned each score.

1. Download **talk2bud-lens-tsa-hiring-scoring-rubric-v1.0.md** from the article this came with.
2. In talk2bud, open **Modes**, switch to **Call lenses**, press **New**.
3. Name it **Structured Interview Rubric**, then paste the file's **Instructions** section into the **Instructions** field.
4. Under **Playbook**, press **Import file...** and pick the same file. Save.
5. Before you analyze your next call, pick this lens. The read comes back scored against these criteria.

Running this with a team? The same file imports into **Channel settings → meeting type → Playbook** with the **Import document** button. Every call of that type then gets the same read, so the scores compare across people instead of across moods.