AI Psychometrics Inc.

Theta AI™

Judgment Evidence Report

5-option situational judgment · Best+Worst responses · Individual

What this person judged most effective and most harmful across 20 work situations, where their judgment is solid or shaky by competency, and what to check in the interview.

Prepared for SYNTHETIC-DEMO (not customer data)

Candidate Sample 001 (pseudonymous)

20/20 answered · 41 pages · v6.1

Generated October 8, 2026

Pre-calibration · descriptive | key_version theta-key-2026-10-05 | scoring bw-s1-v1 | No composite rank · no verdict | Missing stays missing

02 Executive summary
03 Visual dashboard · 30-second view
04 overview dashboard
05 difficulty & candidate scores
06 candidate vs means
07 candidate harmful & inversions
10 discrimination & candidate group
11 candidate stability
12 candidate alternations
13 candidate bands
14 candidate nine-metric synthesis
15 Competency profile
16 Risky picks · harm areas
17 Interview · development guide
18 Interviewer deep questions
21 Development 30/60/90
22 Two-axis judgment map
24 Choice fingerprint 20×5
25 Choice quality · gap
26 Attractive distractors
29 Difficulty-adjusted story
31 Consistency · type stability
32 Response pattern
33 Response time in depth
34 Values fingerprint · decision style
35 Measurement notes · evidence
37 Competency interpretation
38 Appendix R

Verify this report
TH-8MNP-5RKS-C

© 2026 AI Psychometrics Inc. · ai-psychometrics.com

Descriptive · pre-calibration

The one-minute conclusion before the interview. Every number comes from a chart or table later in the report.

The conclusion in one sentence

Mean 6.2/8, above the cohort mean (4.6), with spotting effective action (0.80) clearer than harm spotting (Worst credit 0.75). Check CB-03-A, CB-08-A, CB-03-C first in the interview.

3 situations above the cohort

20 **CB-12-C** · 8/8 · cohort 5.0 (+3.0)
09 **CB-05-D** · 7/8 · cohort 4.0 (+3.0)
02 **CB-01-C** · 7/8 · cohort 4.2 (+2.8)

3 points to check

05 **CB-03-A** · risky pick · 3/8 · cohort 4.6
14 **CB-08-A** · risky pick · 3/8 · cohort 4.8
06 **CB-03-C** · furthest below the cohort · 5/8 · cohort 4.3

For follow-up, not reasons to reject.

3 questions to ask in the interview

05 “Have you acted like ‘Quietly spread complaints about that colleague to other teammates so you have people on your side before the next meeting’ in a situation like CB-03-A? What did you weigh first?”
14 “Have you acted like ‘Push back on each point one by one right there in the meeting’ in a situation like CB-08-A? What did you weigh first?”
06 “In CB-03-C, why did ‘It is personal, not work, so let the two of them sort it out’ look like the worst choice?”

Follow-ups and signs of strong and weak answers are on the ‘Interviewer deep questions’ page.

30-minute interview plan

0–5 min · open on the most solid competency, ‘Customer & stakeholders’
5–20 min · check the 2 risk picks with the deep-question page (reason → alternative → outcome)
20–27 min · ask for one real example in ‘People & collaboration’
27–30 min · candidate questions · write down the reasoning for the decision meeting

Is the evidence enough?

- **20/20** answered · nothing missing
- Mean item score 90% interval **5.7–6.7** (out of 8)
- Type stability **99%** (1,000 item resamples)
- Response integrity: **no signal**
- Comparison: the 24-person cohort in this campaign · descriptive, pre-calibration (key theta-key-2026-10-05)

One line per competency · Best credit + Worst credit (0–8)

Customer & stakeholders · **6.8/8** · above the cohort median (4.8) · lowest situation 01 CB-01-B (6/8)
Integrity & risk · **6.7/8** · above the cohort median (4.4) · lowest situation 18 CB-11-C (5/8)
People & collaboration · **4.8/8** · in line with the cohort median (4.8) · lowest situation 05 CB-03-A (3/8)
Execution under pressure · **6.3/8** · above the cohort median (5.0) · lowest situation 13 CB-07-C (5/8)

What this report says

- How well effective and harmful actions were spotted in 20 situations, with intervals
- Position against the cohort (not a rank)
- Specific situations and questions to check in the interview

What it does not say

- Whether to hire · composite score · rank
- A calibrated ability score (before key and calibration approval)
- Personality · misconduct

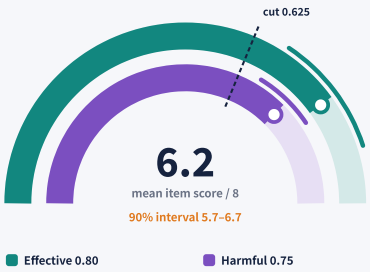
Pages to read before deciding

- **p16** Risky picks · harm areas
- **p18** Interviewer deep questions · rubric
- **p21** Development 30/60/90 days
- **p35** Measurement notes · evidence

VISUAL DASHBOARD · 30-SECOND VIEW

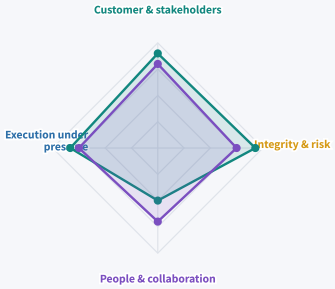
Mean item score 6.2/8 across 20 situations. See at a glance whether harm spotting or effective-action spotting is clearer.

Two-sided judgment gauge



Outer = spotting effective action, inner = spotting harmful action (0–1). Thin arc = 90% interval. Descriptive before calibration, not θ.

Competency profile · Best vs Worst



Teal = Best credit, purple = Worst credit (mean 0–4 per competency) · competency tags are a sample mapping.

Item score mix



7–8 solid 10
5–6 mostly right 8
0–3 needs a check 2
Under S1, a score of 4 cannot occur.

All 20 situations at a glance · Best / Worst / total

Item score = Best credit key(B)−1 + Worst credit 5−key(W). Orange cell = 3 or below.

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Best credit (0–4)	2	4	4	4	0	2	4	4	4	3	4	2	4	1	4	4	4	2	4	4
Worst credit (0–4)	4	3	2	4	3	3	3	3	3	3	3	4	1	2	3	4	3	3	2	4
Item score (0–8)	6	7	6	8	3	5	7	7	7	6	7	6	5	3	7	8	7	5	6	8

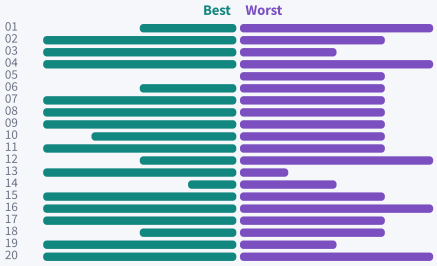
Position in the 24-person cohort · mean item score



Blue shape = cohort spread (dark = middle 50%) · orange = this candidate with 90% interval. A position on the spread, not a rank or percentile.

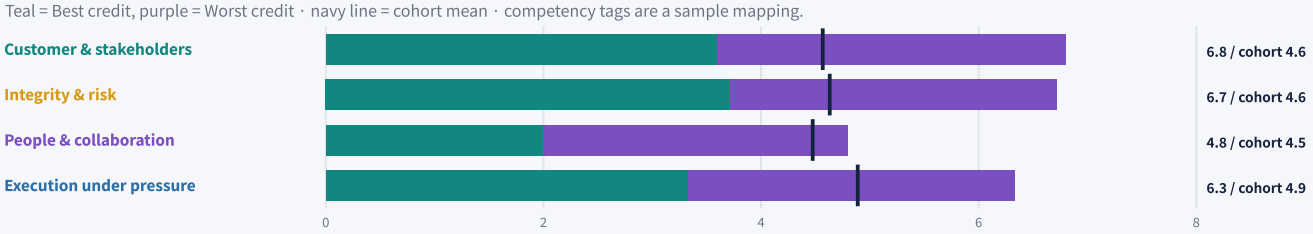
Best · Worst credit by situation

Left teal = Best credit, right purple = Worst credit (0–4 each).



Competency snapshot · Best credit + Worst credit (0–8)

Teal = Best credit, purple = Worst credit · navy line = cohort mean · competency tags are a sample mapping.



RELATIVE-COMPARISON EVIDENCE · OVERVIEW (CANDIDATE MARKED)

The candidate's mean is 6.2, above the cohort mean of 4.61.
Type: balanced, stability 98.6%.

Candidate Sample 001 at a glance

Candidate mean	Judgment type	Axis alternations
6.2	balanced	9
cohort mean 4.61 · 0-8	resample stability 98.6%	cohort mean 5.5

Harmful picks	Inversions · near	Answered
2	2 · 3	20/20
none	situations with gap < 0 · gap = 1	all situations answered

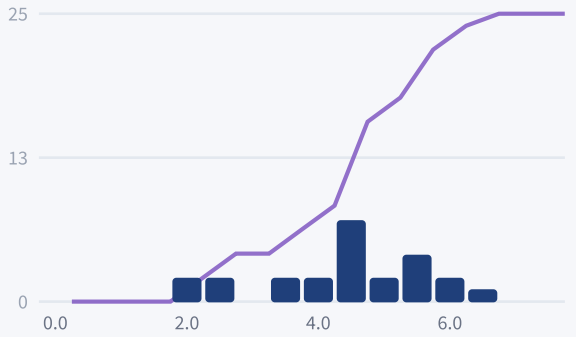
Cohort distributions (comparison base)

Type and score distributions

Judgment-type distribution (N · share)

Balanced	2 · 8.3%
Undifferentiated	7 · 29.2%
Effective-leaning	0 · 0.0%
Borderline	15 · 62.5%

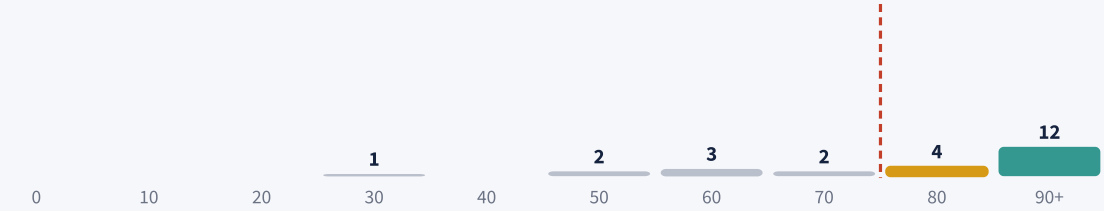
Mean-score distribution · purple line = cumulative share



Median 4.9 · mean 4.61 · range 0-8

Stability bands

Type-stability bands (1,000 resamples) · 80% bar



RELATIVE-COMPARISON EVIDENCE · DIFFICULTY (CANDIDATE MARKED)

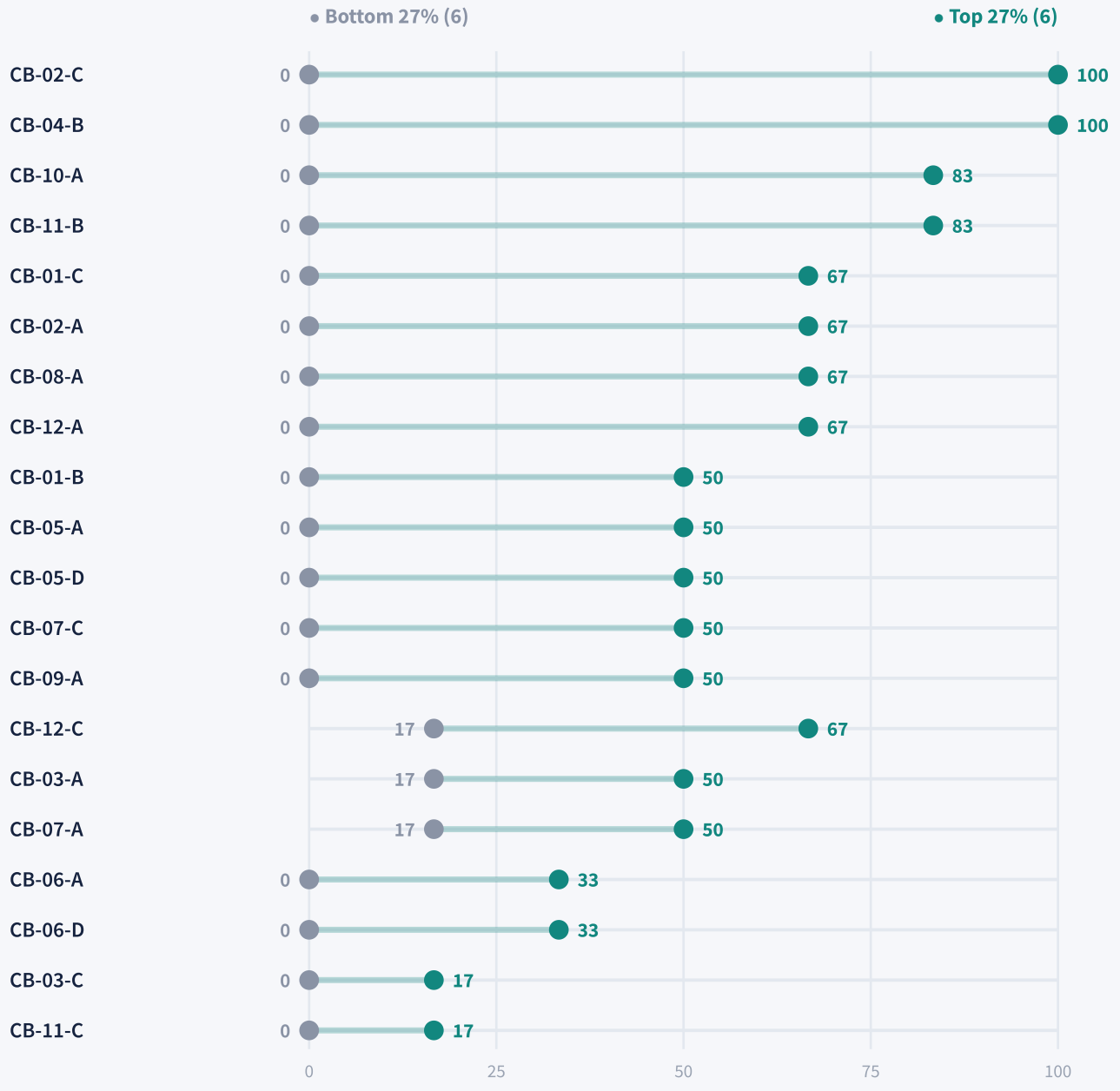
Highest for the candidate: CB-02-C (8.0); lowest: CB-03-A (3.0).

Candidate per-situation scores

Situation	Score	Pick	Situation	Score	Pick
CB-01-B	6.0	best rank 3	CB-06-D	7.0	model pick
CB-01-C	7.0	model pick	CB-07-A	6.0	best rank 3
CB-02-A	6.0	model pick	CB-07-C	5.0	model pick
CB-02-C	8.0	model pick	CB-08-A	3.0	best rank 2
CB-03-A	3.0	best rank 1	CB-09-A	7.0	model pick
CB-03-C	5.0	best rank 3	CB-10-A	8.0	model pick
CB-04-B	7.0	model pick	CB-11-B	7.0	model pick
CB-05-A	7.0	model pick	CB-11-C	5.0	best rank 3
CB-05-D	7.0	model pick	CB-12-A	6.0	model pick
CB-06-A	6.0	best rank 4	CB-12-C	8.0	model pick

Cohort difficulty dumbbells (comparison base)

Model-pick dumbbells · by situation



RELATIVE-COMPARISON EVIDENCE · MEANS (CANDIDATE MARKED)

The widest gap from the cohort mean is CB-12-C (+3.0).

Candidate vs cohort means

Situation	Cohort mean	Candidate	Delta	Situation	Cohort mean	Candidate	Delta
CB-01-B	4.54	6.0	+1.5	CB-06-D	4.42	7.0	+2.6
CB-01-C	4.21	7.0	+2.8	CB-07-A	5.00	6.0	+1.0
CB-02-A	4.88	6.0	+1.1	CB-07-C	4.33	5.0	+0.7
CB-02-C	5.21	8.0	+2.8	CB-08-A	4.83	3.0	-1.8
CB-03-A	4.58	3.0	-1.6	CB-09-A	4.42	7.0	+2.6
CB-03-C	4.29	5.0	+0.7	CB-10-A	5.33	8.0	+2.7
CB-04-B	4.46	7.0	+2.5	CB-11-B	5.38	7.0	+1.6
CB-05-A	4.42	7.0	+2.6	CB-11-C	4.08	5.0	+0.9
CB-05-D	4.04	7.0	+3.0	CB-12-A	4.67	6.0	+1.3
CB-06-A	4.25	6.0	+1.8	CB-12-C	4.96	8.0	+3.0

Cohort mean back-trace (comparison base)

Score distribution by situation · mean ±SD · median



Descriptive · pre-calibration

The candidate shows harmful picks in 2, inversions in 2, and near-inversions in 3 situations.

Candidate harmful & inverted situations

Harmful picks: 2

Situation	Detail	Score
CB-03-A	harmful option as Best	3.0
CB-08-A	harmful option as Best	3.0

Inversions: 2 · near: 3

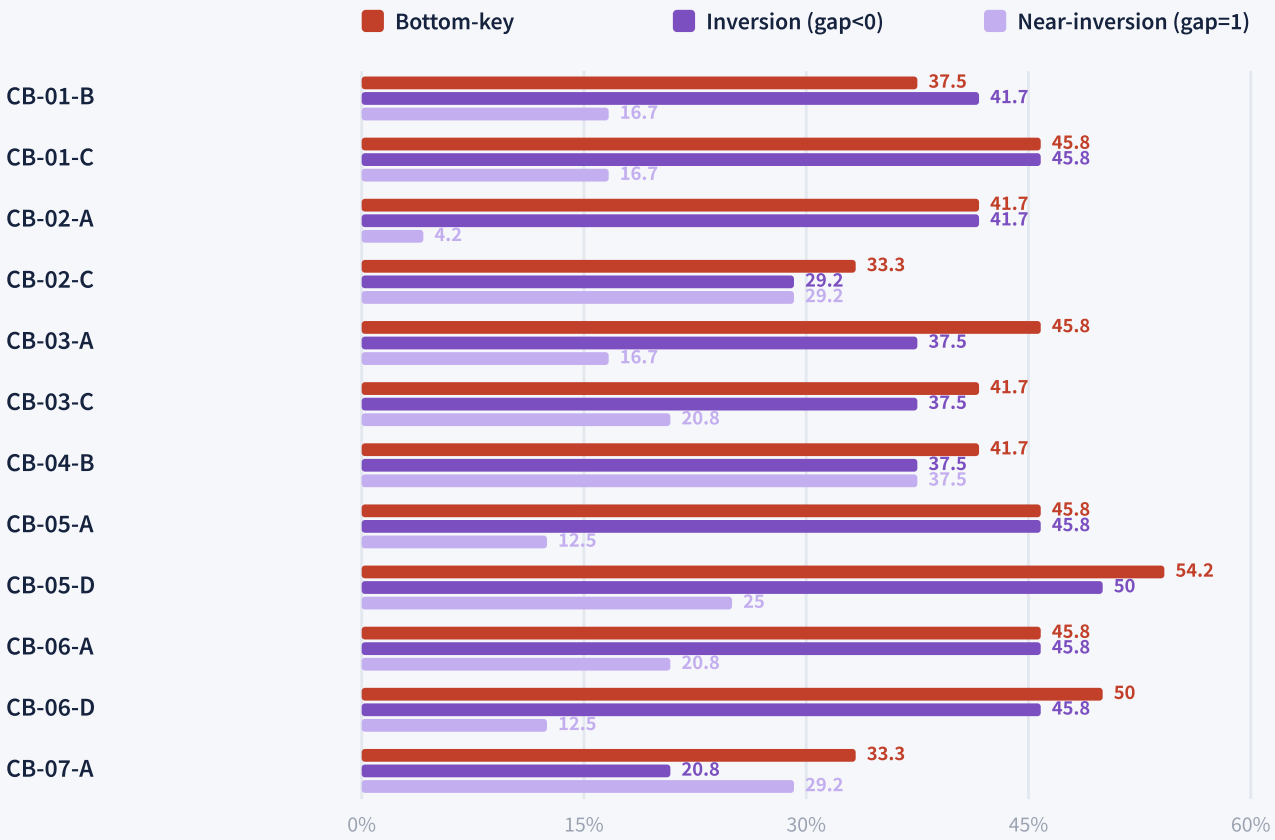
Situation	gap	Score
CB-03-A	gap -1	3.0
CB-08-A	gap -1	3.0
CB-03-C	gap 1	5.0
CB-07-C	gap 1	5.0
CB-11-C	gap 1	5.0

Different signals, different questions

A rank-based follow-up flag marks Best key rank 1–2. An inversion means the selected Best has a lower key rank than the selected Worst. A near-inversion has a rank gap of 1. These counts overlap: they are not distinct groups to add together. Check the original option choice before asking for a real example.

Bottom-key Best picks exceed 10% in 20 situations, peaking at CB-05-D (54.2%). Inversions concentrate on CB-05-D (50%).

Bottom-key picks · inversions · near-inversions by situation

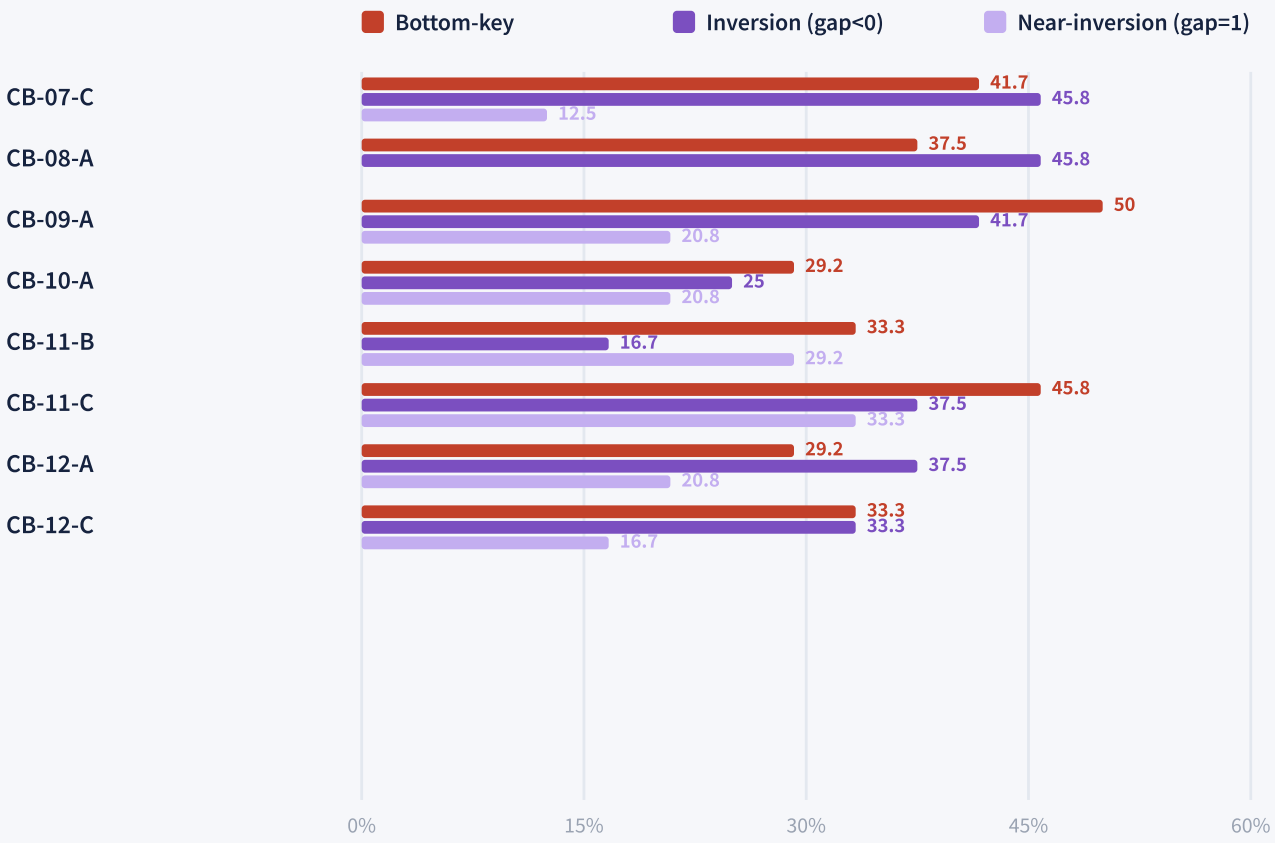


Figures

Situation	Bottom-key	Inversion	Near-inversion	Model Best
CB-01-B	37.5%	41.7%	16.7%	29.2%
CB-01-C	45.8%	45.8%	16.7%	29.2%
CB-02-A	41.7%	41.7%	4.2%	29.2%
CB-02-C	33.3%	29.2%	29.2%	37.5%
CB-03-A	45.8%	37.5%	16.7%	29.2%
CB-03-C	41.7%	37.5%	20.8%	4.2%
CB-04-B	41.7%	37.5%	37.5%	25%
CB-05-A	45.8%	45.8%	12.5%	25%
CB-05-D	54.2%	50%	25%	20.8%
CB-06-A	45.8%	45.8%	20.8%	12.5%
CB-06-D	50%	45.8%	12.5%	25%
CB-07-A	33.3%	20.8%	29.2%	25%

Bottom-key picks continued · situations 13-20.

Bottom-key picks · inversions · near-inversions by situation



Figures

Situation	Bottom-key	Inversion	Near-inversion	Model Best
CB-07-C	41.7%	45.8%	12.5%	25%
CB-08-A	37.5%	45.8%	0%	33.3%
CB-09-A	50%	41.7%	20.8%	20.8%
CB-10-A	29.2%	25%	20.8%	29.2%
CB-11-B	33.3%	16.7%	29.2%	37.5%
CB-11-C	45.8%	37.5%	33.3%	20.8%
CB-12-A	29.2%	37.5%	20.8%	33.3%
CB-12-C	33.3%	33.3%	16.7%	41.7%

How to read, and what shows

• Bottom-key Best = a Best pick on one of the two lowest-ranked key options (key ranks 1–2). Inversion = the Best pick's key rank is below the Worst pick's (gap<0). Near-inversion = ranks one step apart (gap=1) - a near-flipped judgment.
• Of 480 response pairs: 183 inversions (38.1%), 95 near-inversions (19.8%). 24 members (100.0%) picked at least one bottom-key Best.
• Situations high on both inversion and near-inversion are key/option review candidates; bottom-key-only situations are cohort coaching targets.
• Severity-tagged risky picks (fatal/high) are counted on the dashboard, item health and interview bank pages.

RELATIVE-COMPARISON EVIDENCE · DISCRIMINATION (CANDIDATE MARKED)

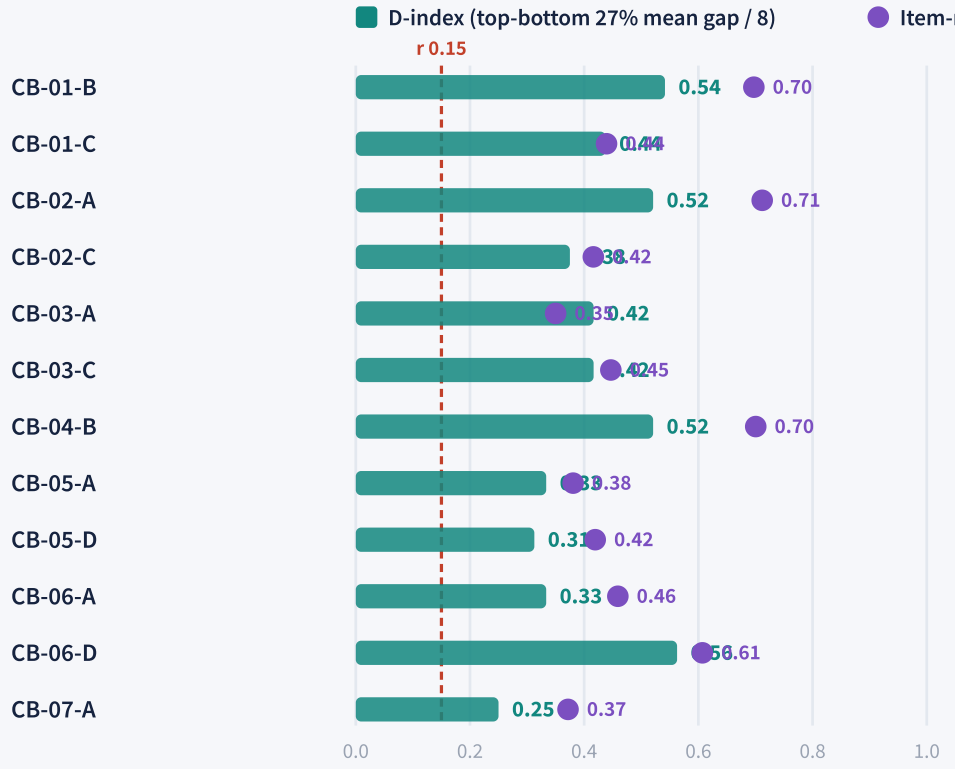
On the discrimination split the candidate falls in the top 27% group.

Where the candidate sits in the discrimination math

In the top/bottom 27% split the candidate falls in the **top 27% group** (mean 6.2; groups are cut by mean score). This metric is a property of the items, not an evaluation of the candidate.

Cohort discrimination (comparison base)

Discrimination by situation · D-index & item-rest r



RELATIVE-COMPARISON EVIDENCE · STABILITY (CANDIDATE MARKED)

Candidate type stability is 98.6%, on the firm side.

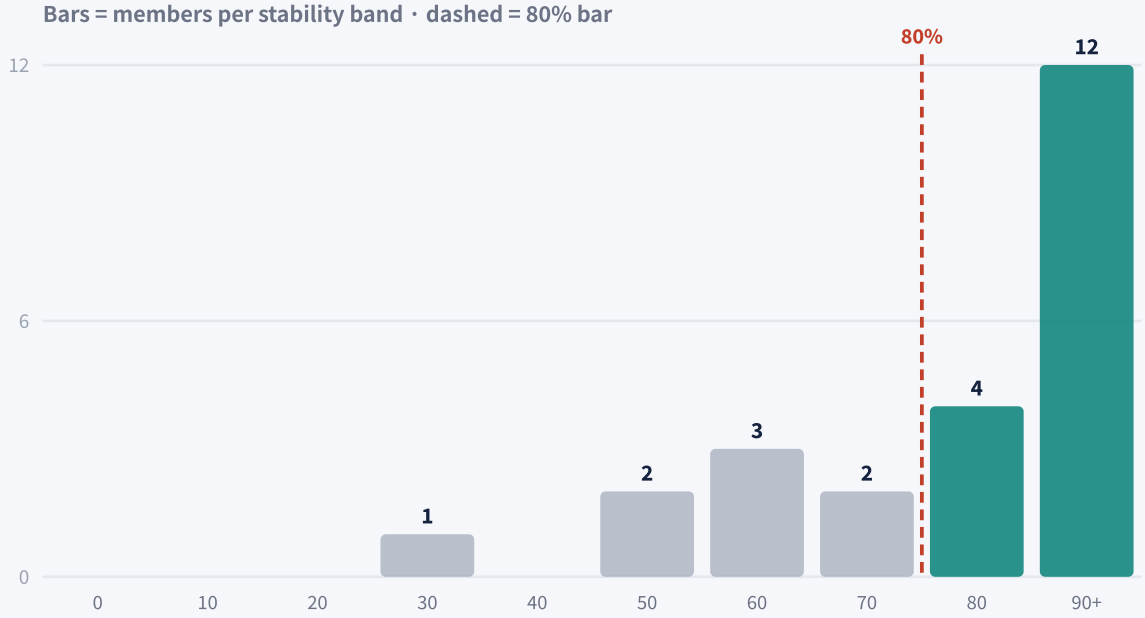
Candidate type stability

Candidate stability	Type	Position in cohort
98.6%	**balanced**	**62.5% of members below**
1,000 deterministic resamples	cohort borderline: 15	pseudo-standardized (descriptive, not an evaluation)

Cohort stability distribution (comparison base)

Type stability distribution · 1,000 resamples

Bars = members per stability band · dashed = 80% bar



RELATIVE-COMPARISON EVIDENCE · ALTERNATION (CANDIDATE MARKED)

The candidate alternates axes 9 times, at CB-01-B→CB-01-C, CB-02-C→CB-03-A, CB-03-C→CB-04-B, CB-06-D→CB-07-A, CB-07-A→CB-07-C, CB-07-C→CB-08-A, CB-08-A→CB-09-A, CB-11-B→CB-11-C, CB-11-C→CB-12-A.

Candidate axis alternations

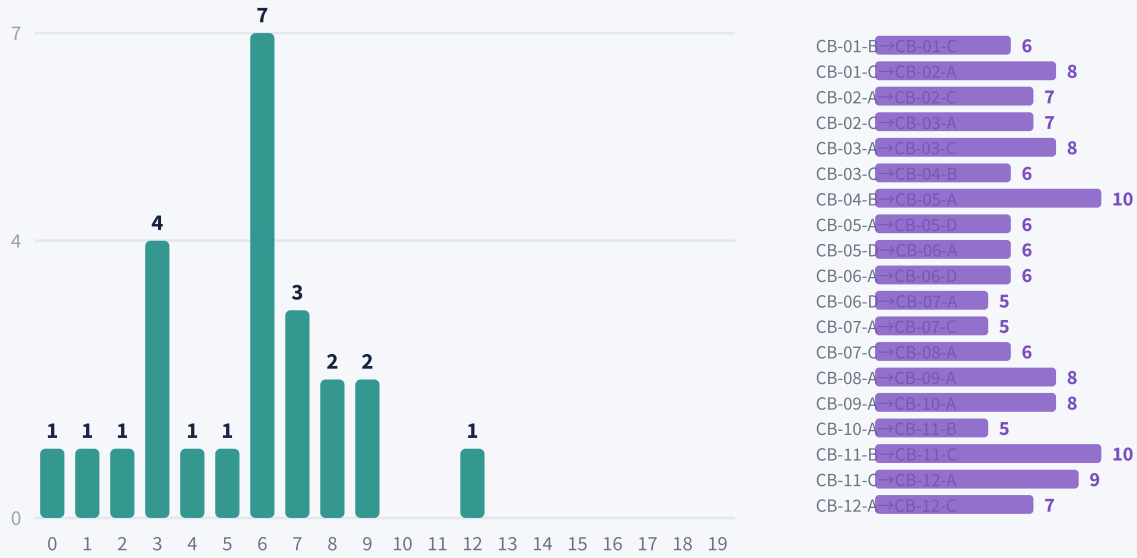
Alternations
9
cohort mean 5.5 · max 12

Boundaries
9
CB-01-B→CB-01-C, CB-02-C→CB-03-A, CB-03-C→CB-04-B, CB-06-D→CB-07-A, CB-07-A→CB-07-C, CB-07-C→CB-08-A, CB-08-A→CB-09-A, CB-11-B→CB-11-C, CB-11-C→CB-12-A

Cohort alternation distribution (comparison base)

Axis alternation · by member and boundary

Left = alternations per member · right = boundaries where flips happen



RELATIVE-COMPARISON EVIDENCE · BANDS (CANDIDATE MARKED)

15 of the candidate's responses sit in the 6-8 band; 2 fall below 4.

Candidate band placement

Of 20 situations: 6-8 band 15, 4-6 3, 2-4 2, 0-2 0. Below 4: CB-03-A, CB-08-A.

Situation	Score	Band
CB-01-B	6.0	**6-8**
CB-01-C	7.0	**6-8**
CB-02-A	6.0	**6-8**
CB-02-C	8.0	**6-8**
CB-03-A	3.0	**2-4**
CB-03-C	5.0	**4-6**
CB-04-B	7.0	**6-8**
CB-05-A	7.0	**6-8**
CB-05-D	7.0	**6-8**
CB-06-A	6.0	**6-8**
CB-06-D	7.0	**6-8**
CB-07-A	6.0	**6-8**
CB-07-C	5.0	**4-6**
CB-08-A	3.0	**2-4**
CB-09-A	7.0	**6-8**
CB-10-A	8.0	**6-8**
CB-11-B	7.0	**6-8**
CB-11-C	5.0	**4-6**
CB-12-A	6.0	**6-8**
CB-12-C	8.0	**6-8**

Cohort band matrix (comparison base)

Situation × band matrix

Cell = members in the band (N) · bottom row = band totals

	0-2	2-4	4-6	6-8
CB-01-B	3	7	4	10
CB-01-C	4	7	4	9
CB-02-A	2	8	1	13
CB-02-C	0	7	7	10
CB-03-A	5	4	4	11
CB-03-C	5	4	5	10
CB-04-B	1	8	9	6
CB-05-A	5	6	3	10
CB-05-D	2	10	6	6
CB-06-A	3	8	5	8
CB-06-D	4	7	3	10
CB-07-A	4	1	7	12
CB-07-C	3	8	3	10
CB-08-A	1	10	0	13
CB-09-A	4	6	5	9
CB-10-A	1	5	5	13
CB-11-B	2	2	7	13
CB-11-C	6	3	8	7
CB-12-A	1	8	5	10
CB-12-C	5	3	4	12
Band totals	**61**	**122**	**95**	**202**

RELATIVE-COMPARISON EVIDENCE · SYNTHESIS (CANDIDATE MARKED)

Candidate: mean 6.2, stability 98.6%, 9 alternations, harmful picks in 2 situations.

Candidate summary (nine metrics)

1. Mean score	2. Best · worst situation	3. Inversions
6.2 cohort 4.61	8.0 / 3.0 0-8 scale	2 38.1% of cohort pairs
4. Model picks	**5. Stability**	**6. Alternations**
13 Best rank 5	98.6% 80% bar	9 cohort mean 5.5
7. 6-8 band	**8. Type**	**9. Harmful picks**
15 42.1% of cohort pairs	balanced cohort borderline 15	2 100.0% of cohort

Cohort nine-graph synthesis (comparison base)

Nine pieces of evidence at a glance

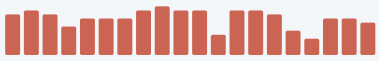
1. Difficulty gap (top - bottom 27%)
100pp / 17pp
CB-02-C / CB-03-C



2. Per-situation mean range
4.0 – 5.4
CB-05-D – CB-11-B



3. Inverted picks (of all pairs)
38.1%
gap < 0 of 480 pairs



4. Discrimination D
mean 0.41
lowest CB-11-C 0.25



5. Type stability ≥80%
16 · 67%
borderline 15 least stable



6. Axis alternations (mean)
5.5
CB-04-B→CB-05-A 10



7. 6-8 band share
42.1%
0-2 band 12.7%



8. Reliability (α · split-half)
0.88 · 0.88
20/20 items at or above r = 0.15



9. Members with a harmful pick
100.0%
24 of 24



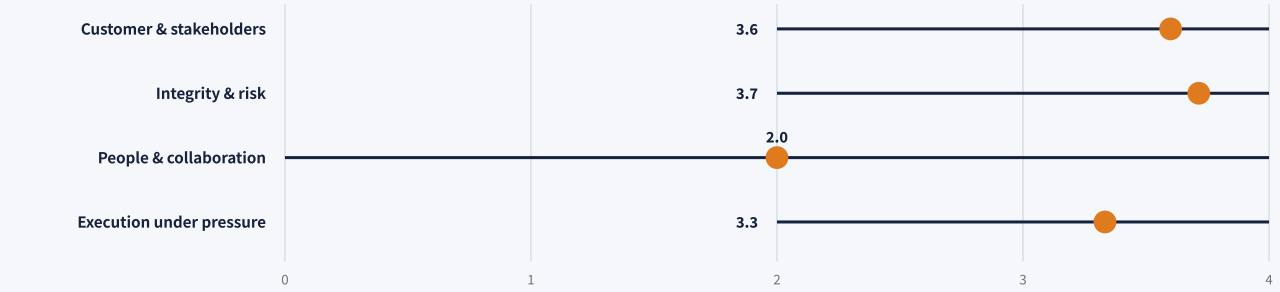
Synthesis in three sentences

① Across 20 situations, mean D is 0.41; 42% of responses sit in the 6-8 band. ② Separation varies widely by situation: CB-02-C shows a 100pp top-bottom gap while CB-03-C manages only 17pp, leaving it nearly flat. ③ Types are stable for 67%, but the 15 borderline members can shift with the item mix, and the 100% harmful-pick rate with 38.1% inversions marks interview territory.

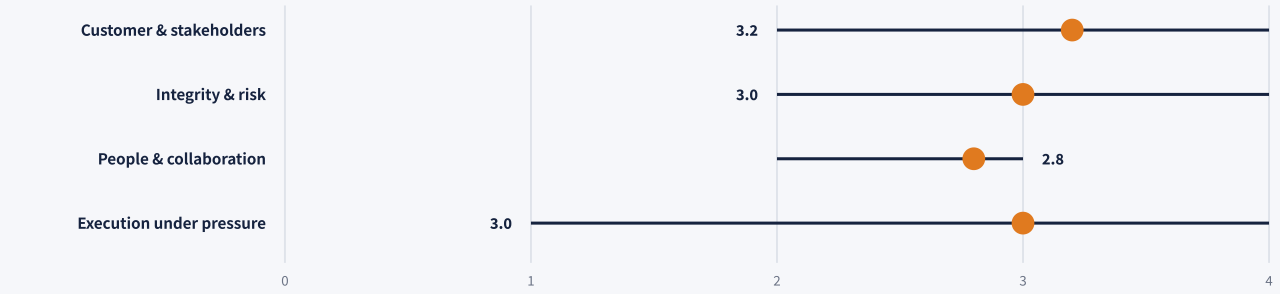
COMPETENCY PROFILE

Judgment is most solid in ‘Customer & stakeholders’; ‘People & collaboration’ has a wide interval and is the point to check in the interview.

Best credit by competency · this candidate on the cohort spread



Worst credit by competency · this candidate on the cohort spread



Most solid competency

Customer & stakeholders

highest Best+Worst mean

Check in the interview

People & collaboration

lowest mean, widest interval

How to read

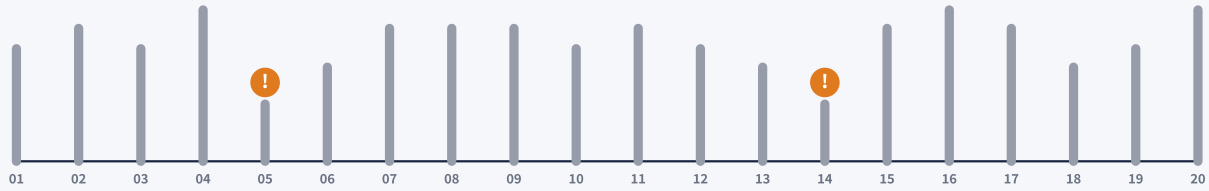
Orange dot = this candidate, navy line = item range. With few situations per competency the ranges are wide. A starting point for the conversation, not a ranking.

RISKY PICKS · HARM AREAS

There are 2 risky picks and no repeated pattern.

Risky-pick timeline

bar = item score (0–8) · circle = risky pick (red critical · orange high · gold model action as Worst)



Judgment by harm area

	Harmful options shown	Rejected as Worst	Endorsed as Best
Ethics	6	3	0
Law & policy	4	2	0
Safety	2	1	0
Data protection	4	2	0
Client trust	5	2	0
People	8	3	1
Quality	6	0	0

Numbers = count of picks. Deeper purple = harmful actions rejected; orange = a harmful action endorsed as Best.

Risky picks to follow up

05 CB-03-A · harmful option (severity 2) picked as Best

14 CB-08-A · harmful option (severity 2) picked as Best

Risky picks are for interview follow-up, not reasons to reject.

Rank and tag evidence

The rank-based follow-up rule flags CB-03-A and CB-08-A. The supplied form metadata assigns a high-severity People tag to the first action. The second action is untagged by design. The area table counts only supplied tags, so its endorsed-Best count is 1, not 2. Review both original choices without assigning an invented severity to the untagged action.

INTERVIEW · DEVELOPMENT GUIDE

3 things to check in the interview, and development notes ready to use after hire.

Structured interview guide · questions built from this candidate's evidence

Question 1 · High risky pick 05

"In 'CB-03-A', you picked 'Quietly spread complaints about that colleague to other teammates so you have people on your side before the next meeting' as the best action. Walk me through what you weighed."

Follow-up: Model action: 'Raise the moment one-on-one after the meeting and propose a way forward'. How do you see the difference?

Strong answer	Concern	Evidence
Weighs the situation's priorities and risks, and reaches the model action's principle	Defends the harmful option or reads the priorities backwards	CB-03-A: B key 1 · W key 2

Question 2 · High risky pick 14

"In 'CB-08-A', you picked 'Push back on each point one by one right there in the meeting' as the best action. Walk me through what you weighed."

Follow-up: Model action: 'Accept the fair points; show evidence on the wrong ones one-on-one after'. How do you see the difference?

Strong answer	Concern	Evidence
Weighs the situation's priorities and risks, and reaches the model action's principle	Defends the harmful option or reads the priorities backwards	CB-08-A: B key 2 · W key 3

Question 3 · Lowest-scoring item 06

"In 'CB-03-C', you picked 'Split tasks so the two do not work directly for now, and collect updates from each separately' as the best action. Walk me through what you weighed."

Follow-up: Model action: 'Hear each separately, then meet as three to agree roles and communication'. How do you see the difference?

Strong answer	Concern	Evidence
Weighs the situation's priorities and risks, and reaches the model action's principle	Defends the harmful option or reads the priorities backwards	CB-03-C: B key 3 · W key 2

Development notes (for onboarding after hire)

People & collaboration judgment

Lowest competency at 4.8/8. Review one similar real case with a senior colleague in month one.

Effective-action axis

5 picks were 'safe' options (key 3–4) instead of the model action. Stretch tasks that require bringing stakeholders to agreement would help.

Strength to use

Customer & stakeholders is most solid (6.8/8). Start with work that uses this strength.

Interview priority · items from lowest score



Descriptive · pre-calibration

Interview evidence record · 1 of 3

05 CB-03-A · high · severity 2

“Have you acted like ‘Quietly spread complaints about that colleague to other teammates so you have people on your side before the next meeting’ in a situation like CB-03-A? What did you weigh first?”

Follow-ups

“Who was affected by that decision?”
“What would you change if you did it again?”
“How did you check the rules or approval process?”

Signs of a strong answer

Same direction as the model action: “Raise the moment one-on-one after the meeting and propose a way forward”
Names the risk and the stakeholders first
A concrete case · outcome · lesson

Signs of a weak answer

Defends the pick as is: “Quietly spread complaints about that colleague to other teammates so you have people on your side before the next meeting”
Gives only speed or convenience as the reason
General statements with no real case

Independent interview record

Record what the candidate actually says. The expected signals above are prompts for checking evidence, not an extra score.

Evidence to capture	Candidate account / interviewer note
Context and people affected	
Options considered and reason	
Rules / approval checked	
Action and observed outcome	
What changed after reflection	

Use one concrete case. Keep observed evidence separate from inference. Discuss the interview record alongside the original response, without changing the item score.

Interview evidence record · 2 of 3

14 CB-08-A · rank-based follow-up · untagged

“Have you acted like ‘Push back on each point one by one right there in the meeting’ in a situation like CB-08-A? What did you weigh first?”

Follow-ups

“Who was affected by that decision?”
“What would you change if you did it again?”
“How did you check the rules or approval process?”

Signs of a strong answer

Same direction as the model action: “Accept the fair points; show evidence on the wrong ones one-on-one after”
Names the risk and the stakeholders first
A concrete case · outcome · lesson

Signs of a weak answer

Defends the pick as is: “Push back on each point one by one right there in the meeting”
Gives only speed or convenience as the reason
General statements with no real case

Independent interview record

Record what the candidate actually says. The expected signals above are prompts for checking evidence, not an extra score.

Evidence to capture	Candidate account / interviewer note
Context and people affected	
Options considered and reason	
Rules / approval checked	
Action and observed outcome	
What changed after reflection	

Use one concrete case. Keep observed evidence separate from inference. Discuss the interview record alongside the original response, without changing the item score.

Interview evidence record · 3 of 3

06 CB-03-C · lowest-scoring situation

“In CB-03-C, why did ‘It is personal, not work, so let the two of them sort it out’ look like the worst choice?”

Follow-ups

“Who was affected by that decision?”
“What would you change if you did it again?”
“How did you check the rules or approval process?”

Signs of a strong answer

Same direction as the model action: “Hear each separately, then meet as three to agree roles and communication”
Names the risk and the stakeholders first
A concrete case · outcome · lesson

Signs of a weak answer

Defends the pick as is: “It is personal, not work, so let the two of them sort it out”
Gives only speed or convenience as the reason
General statements with no real case

How to record the answer

- For each question, tick the signs you heard and write one sentence of what the candidate said.
- Do not create a new score. Bring this record and the report to the decision meeting together.

Independent interview record

Record what the candidate actually says. The expected signals above are prompts for checking evidence, not an extra score.

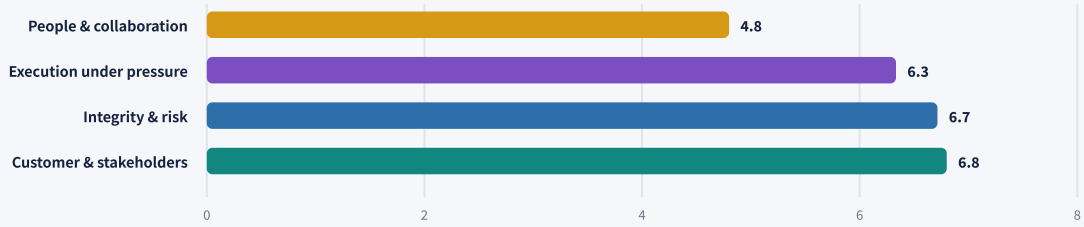
Evidence to capture	Candidate account / interviewer note
Context and people affected	
Options considered and reason	
Rules / approval checked	
Action and observed outcome	
What changed after reflection	

Use one concrete case. Keep observed evidence separate from inference. Discuss the interview record alongside the original response, without changing the item score.

DEVELOPMENT SUGGESTIONS · 30/60/90 DAYS

If hired, start the first 90 days with ‘People & collaboration’. Every suggestion comes from choices this person actually made.

Priority by competency · lowest first



People & collaboration · 4.8/8

Evidence · CB-03-A (3/8) · Best “Quietly spread complaints about that colleague to other teammates so you have people on your side before the next meeting”
30 days · review one similar real case with a senior colleague, against the model action “Raise the moment one-on-one after the meeting and propose a way forward”
60 days · before a ‘People & collaboration’ decision, write down the most harmful option first (think Worst first)
90 days · check two cases with the manager · see whether choices changed in the same kind of situation

Execution under pressure · 6.3/8

Evidence · CB-07-C (5/8)
30 days · review one similar real case with a senior colleague, against the model action “Line up backup venues fast and, once set, tell attendees the change and why”
60 days · before an ‘Execution under pressure’ decision, write down the most harmful option first (think Worst first)
90 days · check two cases with the manager · see whether choices changed in the same kind of situation

Integrity & risk · 6.7/8

Evidence · CB-11-C (5/8) · Best “Lock the screen so no one else sees it and move to another PC”
30 days · review one similar real case with a senior colleague, against the model action “Log out right away without looking and tell security”
60 days · before an ‘Integrity & risk’ decision, write down the most harmful option first (think Worst first)
90 days · check two cases with the manager · see whether choices changed in the same kind of situation

Customer & stakeholders · 6.8/8

Evidence · CB-01-B (6/8) · Best “Explain that compensation is outside your authority and give them the official contact”
30 days · review one similar real case with a senior colleague, against the model action “Acknowledge the impact, explain the investigation and review steps, and connect them to the right team”
60 days · before a ‘Customer & stakeholders’ decision, write down the most harmful option first (think Worst first)
90 days · check two cases with the manager · see whether choices changed in the same kind of situation

Basis and limits of these suggestions

Every suggestion comes from choices this candidate actually made. They are starting points for a development conversation, not an evaluation.

JUDGMENT MAP · 2 AXES

Spotting effective and harmful actions is shown separately.
Effective-action ID: 0.80 (0.69-0.90), the full interval is above the cut.
Harmful-action ID: 0.75 (0.68-0.81), the full interval is above the cut.



Effective-action ID · contributor ①

CB-01-C

Best: Contact the customer to check the issue, then post a factual public reply (key 5)

Effective-action ID · contributor ②

CB-02-A

Best: Tell your manager and send a corrected version to everyone who received it (key 5)

Harmful-action ID · contributor ①

CB-07-C

Worst: Lay out two or three options, including going online, present them to your manager, and get a decision (key 4)

Harmful-action ID · contributor ②

CB-02-A

Worst: Adjust the numbers in next month's report and explain in a footnote (key 3)

Pulled the action axis down ①

CB-03-A

Picked key 1 as Best: Quietly spread complaints about that colleague to other teammates so you have people on your side before the next meeting · model (key 5): Raise the moment one-on-one after the meeting and propose a way forward

Pulled the action axis down ②

CB-08-A

Picked key 2 as Best: Push back on each point one by one right there in the meeting · model (key 5): Accept the fair points; show evidence on the wrong ones one-on-one after

Descriptive · pre-calibration

Four types, borderline and cohort spreads

Four types and borderline

Balanced
Both axes at or above the cut

Strong on effective action, risk identification to develop
Only the effective-action axis at or above the cut

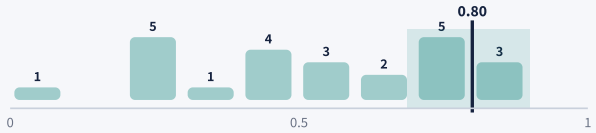
Strong on risk identification, effective action to develop
Only the harmful-action axis at or above the cut

Insufficient evidence on both axes
Both axes below the cut

Borderline
Interval crosses the cut · type not set

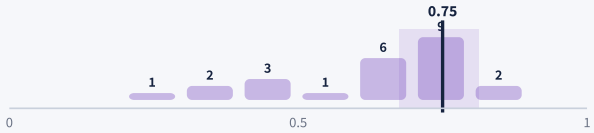
x = mean best_credit = (key(B)−1)/4 · y = mean worst_credit = (5−key(W))/4 · interval = 90% over 1,000 item resamples · before calibration this is descriptive, not θ · the cut is a criterion reference, not the cohort median.

Effective-action ID · cohort spread



Bars = the 24-completer spread on this axis · shaded band = this person's 90% interval · line = their center 0.80. Descriptive, pre-calibration.

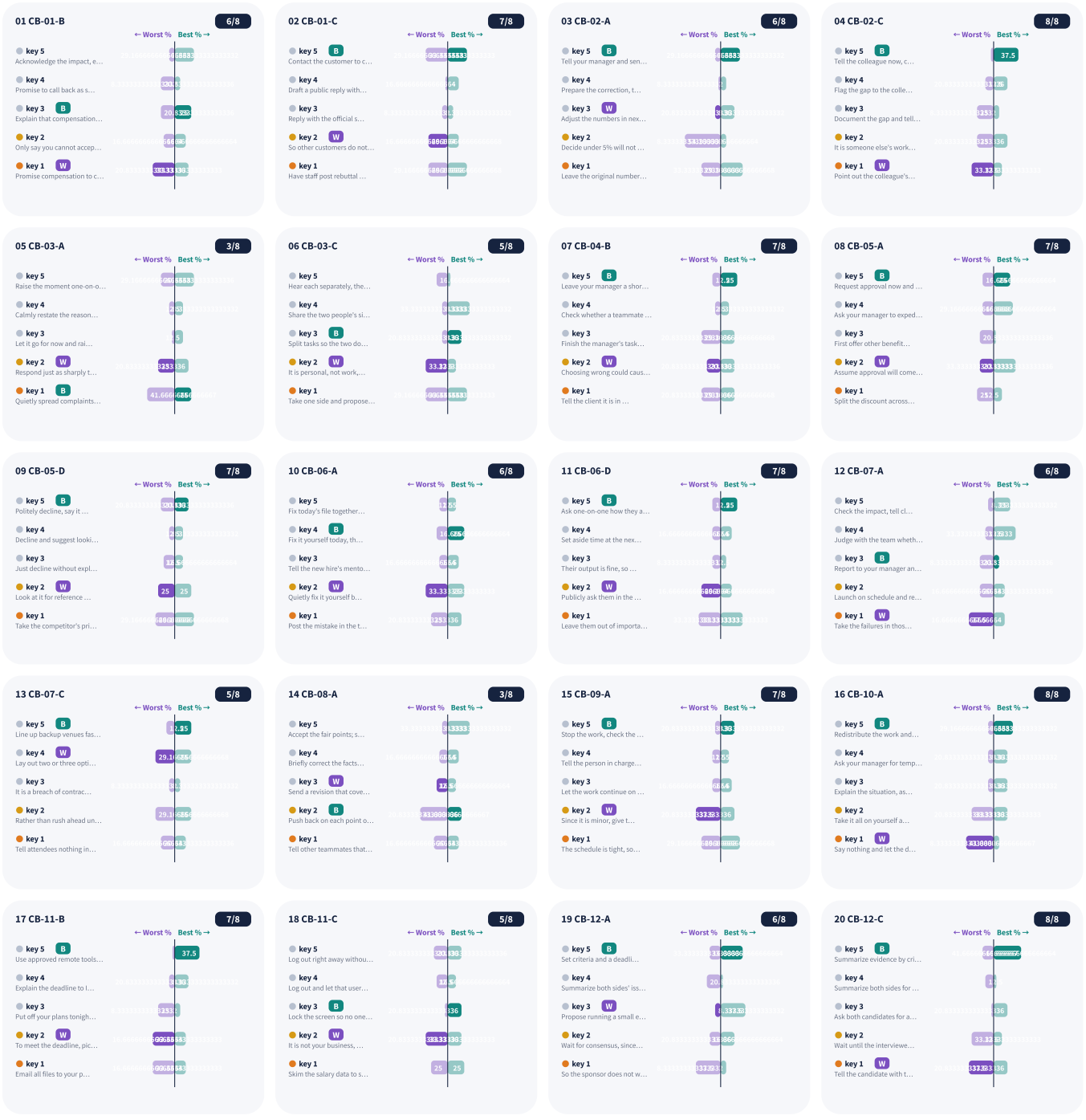
Harmful-action ID · cohort spread



Bars = the 24-completer spread on this axis · shaded band = this person's 90% interval · line = their center 0.75. Descriptive, pre-calibration.

CHOICE FINGERPRINT · 20 SITUATIONS × 5 OPTIONS

Model action picked as Best in 13 of 20; most harmful action picked as Worst in 5. Where this person differs from the cohort shows at a glance.



Rare choices (picked by under 15% of the cohort)

- W **03 CB-02-A** · key 3 · Adjust the numbers in next month's report and explain in … · cohort 8%
- B **12 CB-07-A** · key 3 · Report to your manager and leave the launch decision to them · cohort 8%
- W **19 CB-12-A** · key 3 · Propose running a small experiment and deciding on the data · cohort 8%

A rare choice is not a wrong choice. It is a good place to ask for the reasoning.

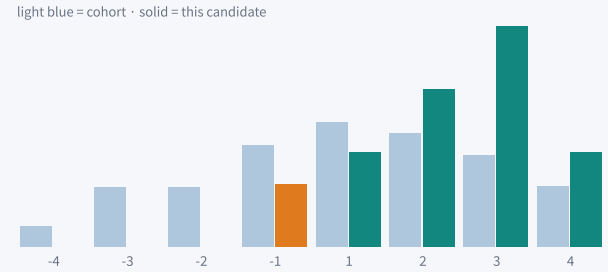
CHOICE QUALITY · JUDGMENT GAP

The gap is +3 or more in 10 of 20 situations. The 2 inverted picks are the first thing to check.

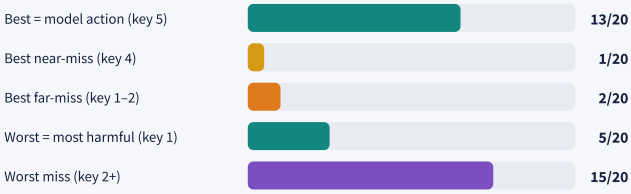
Judgment gap by situation · key(Best) − key(Worst)



Gap distribution vs the cohort



Near-miss vs far-miss



2 inverted picks · the Worst pick has a higher key than the Best pick

05 **CB-03-A** · Best key 1 “Quietly spread complaints about that colleague to other teammates so you have people on your side before the next meeting” · Worst key 2 “Respond just as sharply to that colleague's proposal at the next meeting”

14 **CB-08-A** · Best key 2 “Push back on each point one by one right there in the meeting” · Worst key 3 “Send a revision that covers only the fair points, and let the wrong ones go without bringing them up again”

An inversion means the priorities in that situation were read in reverse order. Ask first how the person understood the situation. Cohort inversion rates appear with a reference group.

Where the 20 picks land on the S1 score table

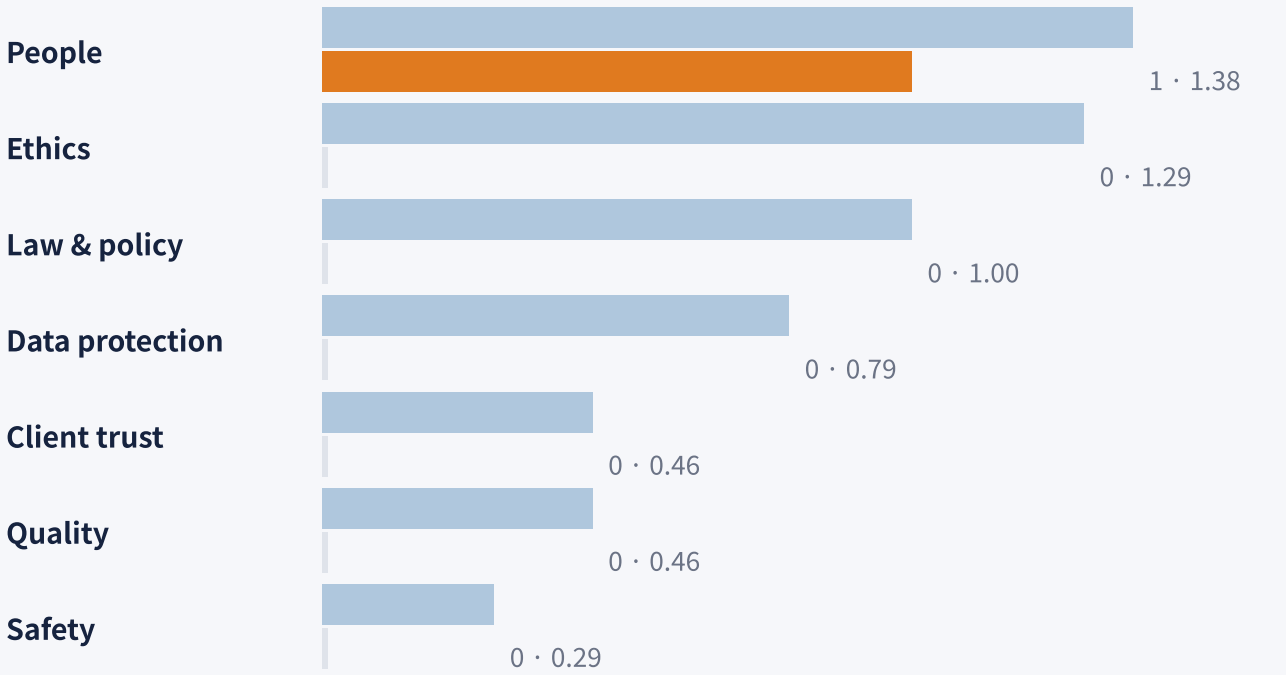


Cell = Best key × Worst key and its score (0–8) · circle = candidate picks (situation no., or count) · cell shading and cohort % appear with a reference group.

ATTRACTIVE DISTRACTORS · HARM AREAS

Of 7 Best picks that were not the model action, 1 was the second-best action. The only area where a harmful option was picked as Best is People.

Pull toward harmful options by area · this candidate vs cohort



Which wrong options were attractive

- Of 7 Best picks that were not the model action: 1 near-miss (key 4), 2 far-miss (key 1–2)
- Areas where a harmful option (severity 2+) was picked as Best: People
- Cohort mean 5.67 per person; this candidate 1
- If much of the cohort made the same wrong pick, the option itself may be an attractive distractor, not just this person.

Two different follow-up counts

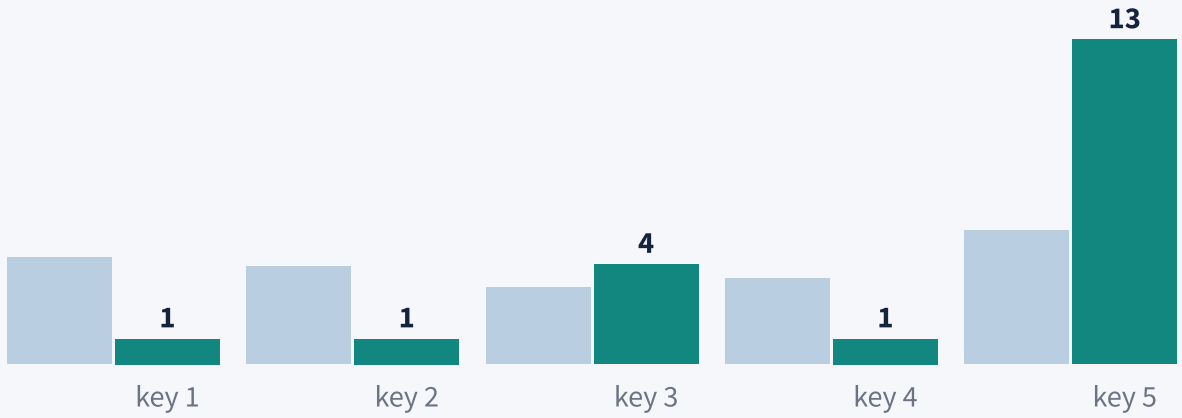
The key-rank rule flags 2 picks (Best rank 1–2). The supplied area/severity tags identify 1 high-severity pick in People. CB-08-A's rank-2 action is untagged by design. An untagged option has no assigned area or panel severity; it does not erase the rank-based follow-up signal.

Descriptive · pre-calibration

Best/Worst key distribution

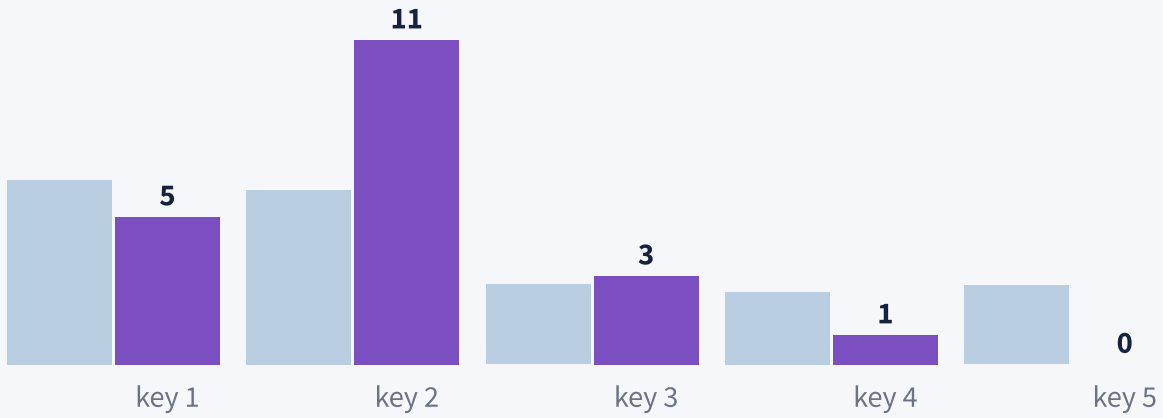
key of Best picks

light = cohort · solid = this candidate (number = count)



key of Worst picks

light = cohort · solid = this candidate (number = count)



Counts and scale

Each distribution sums to 20 responses. The solid bars count this person's picks at each key rank. Light bars sum cohort choice shares to expected counts on the same 20 situations. Rank 5 is the model action; rank 1 is the lowest-ranked action under the scoring key. Best and Worst distributions are separate views, not a composite rank.

Descriptive · pre-calibration

Harmful choice evidence · continued

Every Best pick that was not the model action

Situation	key	Picked as Best	Area · severity	Cohort same pick	model action
01 CB-01-B	key 3	Explain that compensation is outside your authority and give them the official contact	Untagged · no area assigned	25%	Acknowledge the impact, explain the investigation and review steps, and connect them to the right team
05 CB-03-A	key 1	Quietly spread complaints about that colleague to other teammates so you have people on your side before the next meeting	People · sev 2	25%	Raise the moment one-on-one after the meeting and propose a way forward
06 CB-03-C	key 3	Split tasks so the two do not work directly for now, and collect updates from each separately	Untagged · no area assigned	21%	Hear each separately, then meet as three to agree roles and communication
10 CB-06-A	key 4	Fix it yourself today, then walk the new hire through the error pattern tomorrow morning	Untagged · no area assigned	25%	Fix today's file together, then find the cause and build a checklist with them
12 CB-07-A	key 3	Report to your manager and leave the launch decision to them	Untagged · no area assigned	8%	Check the impact, tell clients first, and offer launch options
14 CB-08-A	key 2	Push back on each point one by one right there in the meeting	Untagged · no area assigned	21%	Accept the fair points; show evidence on the wrong ones one-on-one after
18 CB-11-C	key 3	Lock the screen so no one else sees it and move to another PC	Untagged · no area assigned	21%	Log out right away without looking and tell security

Full option text is in the authorized appendix. Tags are supplied for this form; untagged options stay unassigned.

DIFFICULTY-ADJUSTED ITEM STORY

Ahead of the cohort on 18 of 20 hard items. Missed 0 easy item(s).

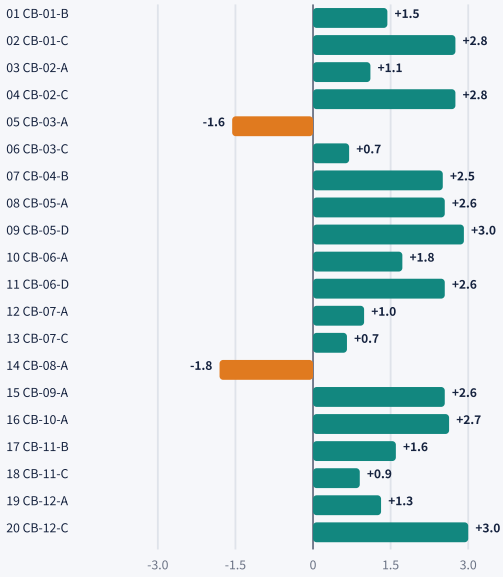
Hard items mean	Easier items mean	Furthest ahead	Furthest behind
6.2	—	20	14
cohort 4.6 · 20 items	cohort — · 0 items	+3.0 · CB-12-C	-1.8 · CB-08-A

Item difficulty × discrimination · circle = this candidate vs cohort



Circle size = size of the difference from the cohort mean · teal = 0.5+ above · orange = 0.5+ below · gold = similar.

Difference from the cohort mean



Ahead on hard items

20 CB-12-C · cohort mean 5.0 → 8
09 CB-05-D · cohort mean 4.0 → 7
02 CB-01-C · cohort mean 4.2 → 7
04 CB-02-C · cohort mean 5.2 → 8
16 CB-10-A · cohort mean 5.3 → 8
08 CB-05-A · cohort mean 4.4 → 7
11 CB-06-D · cohort mean 4.4 → 7
15 CB-09-A · cohort mean 4.4 → 7
07 CB-04-B · cohort mean 4.5 → 7
10 CB-06-A · cohort mean 4.3 → 6
17 CB-11-B · cohort mean 5.4 → 7
01 CB-01-B · cohort mean 4.5 → 6
19 CB-12-A · cohort mean 4.7 → 6
03 CB-02-A · cohort mean 4.9 → 6
12 CB-07-A · cohort mean 5.0 → 6
18 CB-11-C · cohort mean 4.1 → 5
06 CB-03-C · cohort mean 4.3 → 5
13 CB-07-C · cohort mean 4.3 → 5

Missed on easy items

None.

How to read

Being ahead on a hard item is stronger evidence than being ahead on an easy one. Differences on high-discrimination items (top) are more trustworthy. Based on the 24-person cohort.

DIFFICULTY-ADJUSTED ITEM STORY

Ahead of the cohort on 18 of 20 hard items. Missed 0 easy item(s). · continued

Item figures · the source of every dot and bar on this page

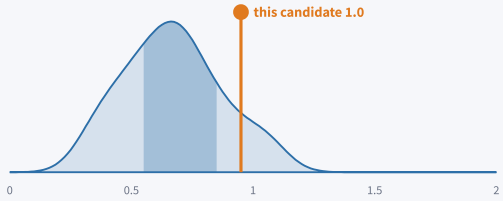
Situation	Competency	Cohort mean	Ease p	D	Item-rest r	This person	Diff	Hard
01 CB-01-B	Customer & stakeholders	4.54	0.57	0.54	0.70	**6**	**+1.46**	✓
02 CB-01-C	Customer & stakeholders	4.21	0.53	0.44	0.44	**7**	**+2.79**	✓
03 CB-02-A	Integrity & risk	4.88	0.61	0.52	0.71	**6**	**+1.13**	✓
04 CB-02-C	Integrity & risk	5.21	0.65	0.38	0.42	**8**	**+2.79**	✓
05 CB-03-A	People & collaboration	4.58	0.57	0.42	0.35	**3**	**-1.58**	✓
06 CB-03-C	People & collaboration	4.29	0.54	0.42	0.45	**5**	**+0.71**	✓
07 CB-04-B	Customer & stakeholders	4.46	0.56	0.52	0.70	**7**	**+2.54**	✓
08 CB-05-A	Integrity & risk	4.42	0.55	0.33	0.38	**7**	**+2.58**	✓
09 CB-05-D	Integrity & risk	4.04	0.51	0.31	0.42	**7**	**+2.96**	✓
10 CB-06-A	People & collaboration	4.25	0.53	0.33	0.46	**6**	**+1.75**	✓
11 CB-06-D	People & collaboration	4.42	0.55	0.56	0.61	**7**	**+2.58**	✓
12 CB-07-A	Execution under pressure	5.00	0.63	0.25	0.37	**6**	**+1.00**	✓
13 CB-07-C	Execution under pressure	4.33	0.54	0.25	0.36	**5**	**+0.67**	✓
14 CB-08-A	People & collaboration	4.83	0.60	0.35	0.45	**3**	**-1.83**	✓
15 CB-09-A	Integrity & risk	4.42	0.55	0.44	0.58	**7**	**+2.58**	✓
16 CB-10-A	Execution under pressure	5.33	0.67	0.46	0.59	**8**	**+2.67**	✓
17 CB-11-B	Integrity & risk	5.38	0.67	0.33	0.46	**7**	**+1.63**	✓
18 CB-11-C	Integrity & risk	4.08	0.51	0.25	0.26	**5**	**+0.92**	✓
19 CB-12-A	Customer & stakeholders	4.67	0.58	0.42	0.50	**6**	**+1.33**	✓
20 CB-12-C	Customer & stakeholders	4.96	0.62	0.60	0.55	**8**	**+3.04**	✓

p = cohort mean / 8 · D = (upper 27% mean − lower 27% mean) / 8 · hard = p < 0.76 · 24-person cohort.

CONSISTENCY · TYPE STABILITY

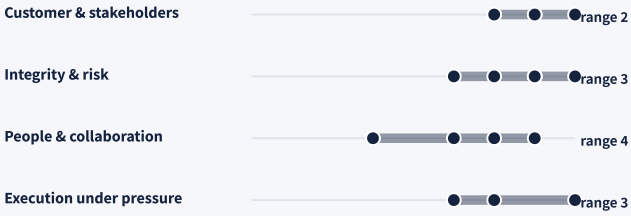
‘Balanced’ in 99% of resamples.

Agreement with the cohort's most common pick



Best agreement 8/20 + Worst agreement 11/20. Blue shape = cohort spread. Differing from the majority is not the same as being wrong.

Consistency within each competency



Type stability · 1,000 item resamples



Distance to each type centre



Diamond = cohort centre for each type · number = distance to this candidate (orange) · dashed = 0.625 cut.

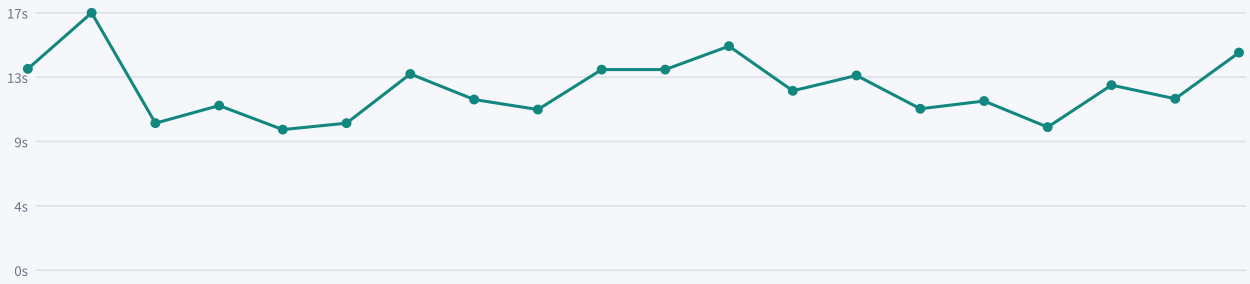
Agreement by situation · ✓ = same as the cohort's most common pick, % = cohort share making the same pick

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Best agrees	25%	✓	29%	✓	25%	21%	✓	25%	21%	✓	25%	8%	✓	21%	21%	✓	✓	21%	33%	✓
Worst agrees	✓	29%	8%	✓	25%	✓	21%	21%	25%	✓	29%	✓	✓	17%	✓	✓	✓	✓	8%	✓

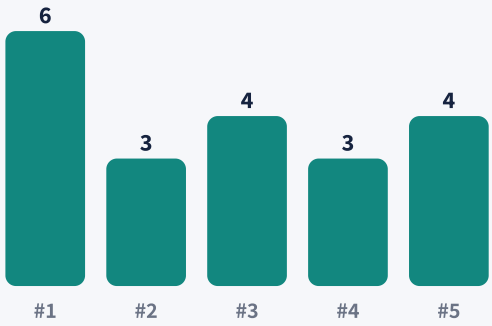
RESPONSE PATTERN · TIME · POSITION

No review signals in response timing or pick patterns.

Response time per item · on the cohort band



Screen position of Best picks



Positions follow the order this candidate was shown. The live form shuffles option order per candidate.

Response process summary

- Median response time 12s
- Longest deliberation: 02 CB-01-C (17.1s)
- Best = Worst answers: 0 (rejected by the server)
- Integrity signals: none
- Signals are for review, not proof of misconduct. Median time < 8 s, Best at the same screen position ≥ 15/20. Never combined with scores.

RESPONSE TIME IN DEPTH

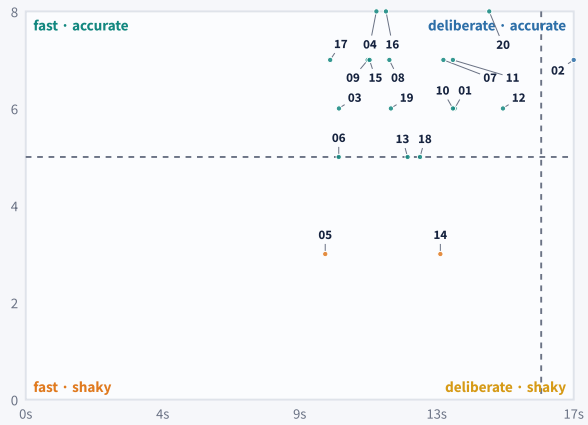
Most answers were faster than the cohort, and deliberation time and scores were weakly linked (r = 0.18). These are the hard moments to hear about in the interview.

Response time per item · vs cohort (log z)



Number = item score · teal = 6+ · orange = 5 or below · up = longer than the cohort.

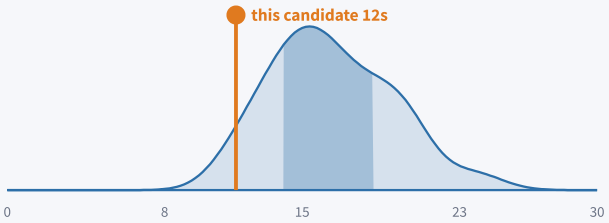
Deliberation × accuracy quadrant



All IDs, including overlapping points · exact coordinates (s / score)

01 13.355s / 6	**02** 17.063s / 7	**03** 9.751s / 6	**04** 10.92s / 8
05 9.327s / 3	**06** 9.747s / 5	**07** 13.006s / 7	**08** 11.323s / 7
09 10.649s / 7	**10** 13.302s / 6	**11** 13.3s / 7	**12** 14.858s / 6
13 11.89s / 5	**14** 12.907s / 3	**15** 10.703s / 7	**16** 11.216s / 8
17 9.481s / 7	**18** 12.275s / 5	**19** 11.367s / 6	**20** 14.432s / 8

Median response time · this candidate on the cohort spread

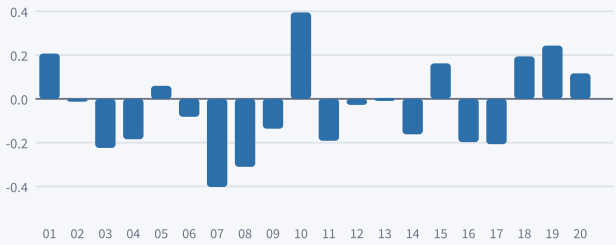


Review threshold: median below 8s.

How time relates to judgment

- This candidate's time–score correlation r = 0.18 (cohort median 0.06)
- The correlation is too weak to read a direction.
- Fast but shaky items: 2
- Fast response (cohort bottom 25%) + harmful Best: 1
- A correlation over 20 points is unstable. Use it only as a conversation opener.
- Time signals are for review, not proof of misconduct.

Time–score correlation by item · 24-person cohort

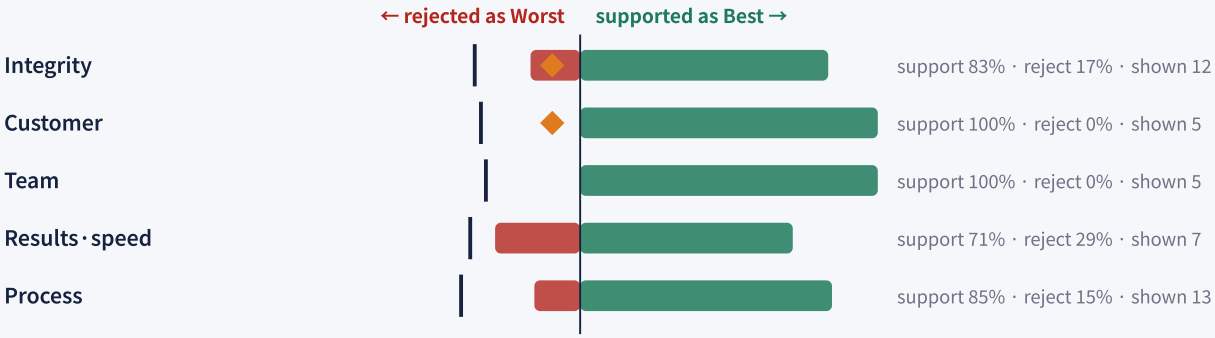


Whether people who took longer scored higher on that item (r). The strongest link is item 07 (r = -0.40).

Descriptive · pre-calibration

2 picks clash with company priority values. The value supported most often is ‘Customer’.

Value priorities shown in effectiveness judgments (in-house value framework draft, R4)

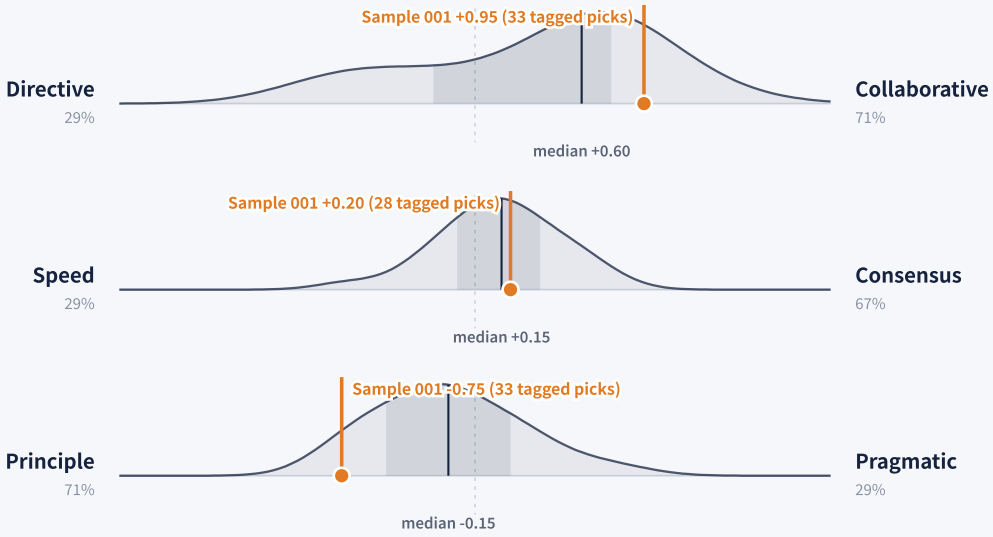


◆ = top value in the company culture profile · vertical line = cohort median reject rate · harmful options excluded · no total fit score · these are priorities shown in “what the candidate judges effective,” not personal values.

Tension with priority values 2 picks clash with company priority values — check in the interview (not a rejection reason). · **CB-02-A** · priority value (Integrity) option as Worst · **CB-08-A** · priority value (Integrity) option as Worst

Not shown Shown fewer than 4 times: Autonomy

Decision style from actual picks (3 axes, R6)



Score = mean Best tag − mean Worst tag (−2 to +2) · orange line = this candidate · curve = cohort distribution (darker = middle 50%) · knowledge-type instructions, so this means “how the candidate thinks one should act” · faking resistance still to be validated.

Interview questions from this page

Directive ↔ Collaborative · score +0.95

“Tell me about a real case where you chose a more collaborative approach. When would the opposite approach have worked better?”

Principle ↔ Pragmatic · score -0.75

“Tell me about a real case where you chose a more principle approach. When would the opposite approach have worked better?”

The two axes furthest from 0 are shown · there is no right answer · use the answers to see whether the approach adapts to the situation.

Descriptive · pre-calibration

Every number is computed from the 20 response records and can be traced on this page.

What this result says

- What this candidate picked as Best/Worst in the 20 situations shown, and how that differs from the key
- Concrete points to check in the interview
- A starting point for a development conversation after hire

What it does not say

- It does not predict on-the-job performance
- It is not a verdict on personality or attitude
- It is not a ranking against other candidates (cohort comparisons land with a reference group, with percentile disclosures)
- Pre-calibration descriptive figures, not θ

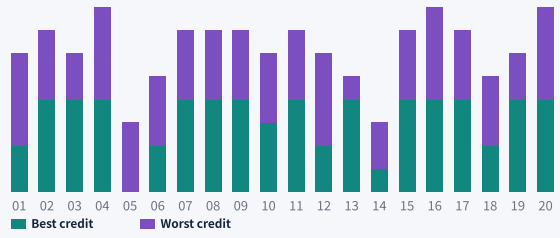
Evidence trail · where every number comes from

Scenario	Response ID	Picks and key	Screen position B/W	Time	Key · scoring version
01 CB-01-B	Sample 001-01	B=key 3 · W=key 1	4 / 1	13.4s	theta-key-2026-10-05 · bw-s1-v1
02 CB-01-C	Sample 001-02	B=key 5 · W=key 2	1 / 5	17.1s	theta-key-2026-10-05 · bw-s1-v1
03 CB-02-A	Sample 001-03	B=key 5 · W=key 3	1 / 3	9.8s	theta-key-2026-10-05 · bw-s1-v1
04 CB-02-C	Sample 001-04	B=key 5 · W=key 1	3 / 4	10.9s	theta-key-2026-10-05 · bw-s1-v1
05 CB-03-A	Sample 001-05	B=key 1 · W=key 2	3 / 5	9.3s	theta-key-2026-10-05 · bw-s1-v1
06 CB-03-C	Sample 001-06	B=key 3 · W=key 2	1 / 4	9.7s	theta-key-2026-10-05 · bw-s1-v1
07 CB-04-B	Sample 001-07	B=key 5 · W=key 2	1 / 2	13.0s	theta-key-2026-10-05 · bw-s1-v1
08 CB-05-A	Sample 001-08	B=key 5 · W=key 2	5 / 4	11.3s	theta-key-2026-10-05 · bw-s1-v1
09 CB-05-D	Sample 001-09	B=key 5 · W=key 2	5 / 1	10.6s	theta-key-2026-10-05 · bw-s1-v1
10 CB-06-A	Sample 001-10	B=key 4 · W=key 2	3 / 2	13.3s	theta-key-2026-10-05 · bw-s1-v1

Response IDs use candidate-ref + item order. Each maps 1:1 to the raw response recorded by the server at form submission.

Pre-calibration · descriptive | key_version theta-key-2026-10-05 | Reference cut 0.625 | No composite · no ranking | Missing stays missing

Score build-up · Best + Worst credit over 20 situations



Share of the cohort making the same pick · Best / Worst



Teal = share of the 24-person cohort making the same Best pick as this candidate, purple = the same Worst pick. Lower means rarer, not wrong.

Every number is computed from the 20 response records and can be traced on this page. · continued

Reading this response record

Each row is one scored response from Sample 001. Best and Worst refer to options under the same item key. Item credit is (Best rank − 1) + (5 − Worst rank), from 0 to 8. Response time is the recorded duration, not a screening score. Display position follows the recorded option order. The table continues the source response list from the previous page.

Evidence trail · where every number comes from

Scenario	Response ID	Picks and key	Screen position B/W	Time	Key · scoring version
11 CB-06-D	Sample 001-11	B=key 5 · W=key 2	5 / 3	13.3s	theta-key-2026-10-05 · bw-s1-v1
12 CB-07-A	Sample 001-12	B=key 3 · W=key 1	5 / 2	14.9s	theta-key-2026-10-05 · bw-s1-v1
13 CB-07-C	Sample 001-13	B=key 5 · W=key 4	2 / 5	11.9s	theta-key-2026-10-05 · bw-s1-v1
14 CB-08-A	Sample 001-14	B=key 2 · W=key 3	2 / 3	12.9s	theta-key-2026-10-05 · bw-s1-v1
15 CB-09-A	Sample 001-15	B=key 5 · W=key 2	1 / 3	10.7s	theta-key-2026-10-05 · bw-s1-v1
16 CB-10-A	Sample 001-16	B=key 5 · W=key 1	3 / 2	11.2s	theta-key-2026-10-05 · bw-s1-v1
17 CB-11-B	Sample 001-17	B=key 5 · W=key 2	4 / 3	9.5s	theta-key-2026-10-05 · bw-s1-v1
18 CB-11-C	Sample 001-18	B=key 3 · W=key 2	1 / 5	12.3s	theta-key-2026-10-05 · bw-s1-v1
19 CB-12-A	Sample 001-19	B=key 5 · W=key 3	2 / 3	11.4s	theta-key-2026-10-05 · bw-s1-v1
20 CB-12-C	Sample 001-20	B=key 5 · W=key 1	4 / 2	14.4s	theta-key-2026-10-05 · bw-s1-v1

Response IDs use candidate-ref + item order. Each maps 1:1 to the raw response recorded by the server at form submission.

Descriptive · pre-calibration

Response scores are not a hiring verdict

Individual competency raw scores · pre-calibration

Best + Worst credits are descriptive under the item key, not a hiring verdict or IRT ability percentile.

Four competency scores

Customer & stakeholders · 6.8/8 · cohort median 4.8 (n=24)

Integrity & risk · 6.7/8 · cohort median 4.4 (n=24)

People & collaboration · 4.8/8 · cohort median 4.8 (n=24)

Execution under pressure · 6.3/8 · cohort median 5.0 (n=24)

Comparison cohort of 24 is calculated from this request's source responses. See Appendix R for ranks.

Use each score with its evidence

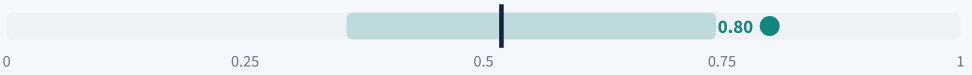
A competency score is the mean raw item credit for situations assigned to that competency. The comparison median comes from this same 24-person cohort and the same form. It does not estimate the share of a wider population below the candidate. Review lower-scoring situations and the selected actions, then ask for concrete interview evidence. Do not turn the four scores into an automatic hiring decision.

Appendix R · Relative standing

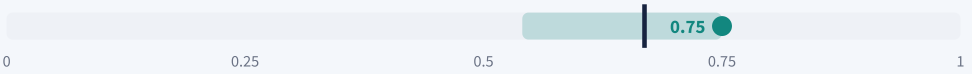
Where the candidate stands in the 24-person cohort.

Two axis scores · ranks computed from this cohort

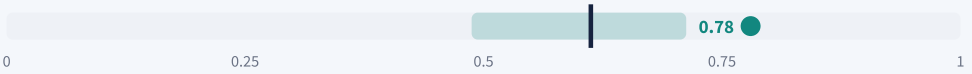
Effective-action ID
1 of 24 · 3 tied · top 4.2%



Risk ID
6 of 24 · 2 tied · top 25.0%



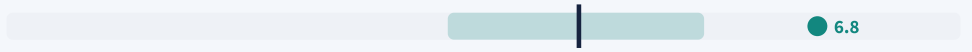
Axis mean
2 of 24 · 1 tied · top 8.3%



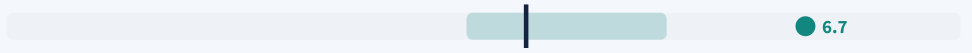
Overall mean 6.2/8 · above cohort median 4.9

Competency standing · vs cohort quartiles (P25 · median · P75)

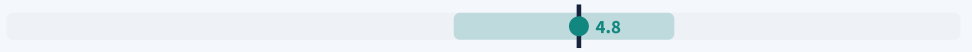
Customer & stakeholders
Top quartile (above P75 5.8)



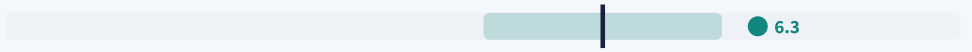
Integrity & risk
Top quartile (above P75 5.5)



People & collaboration
At the cohort median



Execution under pressure
Top quartile (above P75 6.0)



Appendix R · Relative standing · continued

Situation standing · 01–10

Situation standing · all 20 situations

Situation	Cand.	Cohort med.	Q1–Q3	Same-band count	Position
CB-01-B	**6.0**	5.0	3.0-6.0	10 of 24	Top-25% edge
CB-01-C	**7.0**	5.0	2.8-6.3	9 of 24	Top 25%
CB-02-A	**6.0**	6.0	3.0-7.0	13 of 24	Middle 50%
CB-02-C	**8.0**	5.0	3.0-7.0	10 of 24	Top 25%
CB-03-A	**3.0**	5.0	2.8-6.3	3 of 24	Middle 50%
CB-03-C	**5.0**	5.0	2.0-6.0	5 of 24	Middle 50%
CB-04-B	**7.0**	5.0	3.0-5.3	6 of 24	Top 25%
CB-05-A	**7.0**	5.0	2.8-7.0	10 of 24	Top-25% edge
CB-05-D	**7.0**	4.0	2.0-5.3	6 of 24	Top 25%
CB-06-A	**6.0**	5.0	2.8-6.0	8 of 24	Top-25% edge

Reading relative standing

Cand. is this person's raw 0–8 item credit. Cohort med. is the median of the same 24 people. Q1–Q3 is the middle half of their item credits. A position at a quartile edge is marked as an edge rather than treated as a strict above-cut rank. Same-band count includes this person. These are within-cohort comparisons, not population norms or a hiring grade.

Appendix R · Relative standing · continued

Situation standing · 11–20

Situation standing · all 20 situations

Situation	Cand.	Cohort med.	Q1–Q3	Same-band count	Position
CB-06-D	**7.0**	5.0	2.8-7.0	10 of 24	Top-25% edge
CB-07-A	**6.0**	5.5	5.0-6.3	12 of 24	Middle 50%
CB-07-C	**5.0**	5.0	2.0-6.0	3 of 24	Middle 50%
CB-08-A	**3.0**	6.0	3.0-7.0	6 of 24	Middle 50%
CB-09-A	**7.0**	5.0	3.0-6.3	9 of 24	Top 25%
CB-10-A	**8.0**	6.0	4.5-7.0	13 of 24	Top 25%
CB-11-B	**7.0**	6.0	5.0-7.0	13 of 24	Top-25% edge
CB-11-C	**5.0**	5.0	2.5-6.0	8 of 24	Middle 50%
CB-12-A	**6.0**	5.0	2.8-6.0	10 of 24	Top-25% edge
CB-12-C	**8.0**	5.5	3.0-7.3	12 of 24	Top 25%

Reading relative standing

Cand. is this person's raw 0–8 item credit. Cohort med. is the median of the same 24 people. Q1–Q3 is the middle half of their item credits. A position at a quartile edge is marked as an edge rather than treated as a strict above-cut rank. Same-band count includes this person. These are within-cohort comparisons, not population norms or a hiring grade.

Appendix R · Relative standing · continued

Rank definitions and calculation basis

Rank definition · ties share a rank and the next rank skips (competition ranking) · denominator = all 24 members incl. borderline · top % = rank ÷ 24

Axis ranks and competency/situation positions are computed directly from this cohort's responses · quartiles use linear interpolation

Overall mean 6.2/8, cohort median 4.9; 3 of 24 members are in the 6–8 band (12.5%)

Judgment axes and raw scores

Measure	Calculation	Scale
Effective-action ID	(effective-action identification mean − 1) ÷ 4	0–1
Risk ID	(harm identification mean − 1) ÷ 4	0–1
Axis mean	((effective-action mean − 1) + (harm identification mean − 1)) ÷ 8	0–1
Overall mean	Mean of scored Best + Worst item credits	0–8

Ranks and ties

Axis ranks use scores rounded to two decimals. Rank = 1 + the number of cohort members with a higher rounded score. Equal rounded scores share a rank; the next rank skips the tied positions. Top percentage = rank ÷ 24 × 100. This is relative standing within this cohort, not a population percentile.

Quartiles and the middle 50%

P25, the median and P75 use linear interpolation over the sorted scores from this cohort. The shaded band runs from P25 to P75. Competency comparisons use each member's mean for that competency; situation comparisons use the same item across members.

Cohort and denominator

This appendix uses 24 members from the report's supplied cohort and the same assessment form. Borderline members remain in the denominator. Every situation is compared with the same item ID. The cohort count and source answers are shared with the individual report's comparison panels.

Score bands and missing comparisons

Item bands are 0–2, above 2 to below 4, 4 to below 6, and 6–8. Same-band count is the number of cohort members in the candidate's band for that item. The scorer skips a missing response, identical Best and Worst choices, and choices absent from the item key. No zero is inserted for those skipped items. Situation comparisons use available scores for the matching item; missing answers are not evidence of lower performance.

How to use this appendix

Read the candidate dot beside the cohort median and middle-50% band. Use the item table to trace the comparison to a specific situation. These ranks describe this cohort's answers. They are not a hiring verdict, a pass mark or a norm for the wider population.