

3 shapes, costed, and the yearly check

A technical assessment measures whether the candidate can do the work this team does. The 3 shapes below are costed on the same hire. Only one tests the job. Page 2 is the check that tells you, once a year, whether yours does.

1. The 3 shapes on one hire

SHAPE	CANDIDATE'S TIME	YOUR TIME	WHAT IT MEASURES	WHAT IT LOSES
Algorithm test	60 min, plus the revision it takes	60 min	Memory of a balanced tree. A job you do not have.	Engineers whose last tree was at university. Validity of unstructured conversation: about 0.20.
Take-home	6 hours (you believe 3)	45 min of marking, per submission	Writing alone, over hours. The job in 1 case only.	Everyone with options, and the candidate with 2 children and a current job. You selected for free time.
Pair on a broken service	60 min	60 min, no marking	The work: a retry that duplicates writes, a cache that lies under load.	An exercise can leak: rotate 2 or 3 versions. Validity of a structured interview: about 0.42 (Sackett and colleagues, 2022).

2. The pair, in 5 lines

1. A small working service with a subtle, behavioural bug. An off-by-one under concurrency. A retry that duplicates writes.
2. Forty to 60 minutes, every tool allowed, AI included. The candidate narrates, the interviewer watches.
3. After the fix, 4 questions: Why does it work? What might it break? What did the tool suggest that you rejected? Why 3 retries?
4. Someone who understands the fix answers all 4. Someone who pasted it runs out around the second.
5. The same service with a heavier bar per level. Senior names the second bug the fix exposes. Staff says what they would not fix, and why.

Build 2 or 3 versions and rotate them. Refresh once a year. If one leaks, the 4 questions still work on someone who read the answer.

3. The take-home, only when writing alone is the job

- ☐ Capped at 3 hours in writing, and checked against what people submit. If the strongest submissions took a weekend, the cap is fiction.
- ☐ Sent after the 30-minute screen, never as the first step. A stranger sent a link will not spend 3 hours on it.

- ☐ Followed by a conversation about the code.
- ☐ Nothing submitted is used, and the brief says so. A task that resembles your roadmap reads as exactly that, and the candidate tells the next one.

4. The yearly check

Run once a year, in one afternoon. Three numbers. If the first one is "nothing", cut the stage this quarter and put the pair in its place.

CHECK	HOW	YOUR NUMBER	READ IT
Did the score predict anything about the last 10 hires?	Assessment score next to the day-90 answer for each of the last 10 hires.	_____ of 10 agree	Under 6 of 10: the stage is measuring something other than the job.
Are you measuring the marker?	Two markers score the same 3 submissions blind. Count the lines more than 1 level apart.	_____ lines apart	A wide spread means the score is the marker's, and the marking sheet needs strong, mixed and weak written out.
Who never starts it?	Completion rate by stage, from the ATS export.	_____ % start, _____ % submit	A take-home half the candidates never start is a filter you did not choose.

Where the assessment sits in the loop. Monday: 30-minute screen. Thursday: the pair, then the design round, then the hiring manager, on the same day. Scores in writing, alone, before the debrief.

Every scorecard line the assessment feeds describes something the hire does in month 2. A line it cannot feed is a line for another round.