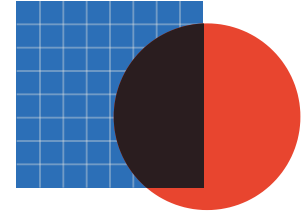


Measure **understanding**, not just completion.



Usability tests show whether people finished. The Cognitive Distance Index shows whether they understood what they finished.

01 Four things to measure

01

Perceived understanding

Can the user say, in plain words, what the system just did?

“I understand what the system just did.” (1–7)

02

Predictive accuracy

Can they predict the outcome **before** it is shown?

“What will happen when you press Confirm?” (4 options)

03

Interaction confidence

Do they feel sure about what happened and what comes next?

“How sure are you?” (1–5) and “Is the task done?”

04

Conceptual transparency

Can they say **why** the outcome happened, not only what it was?

“Why did the system do that?” and “In your own words, what just happened?”

02 Run it on one flow

- 01 **Pick a flow that ends in an outcome.** A payment, a booking, a submission, an AI answer: anything the user has to understand afterwards.
- 02 **Write the questions.** One or two prediction questions and two or three “why” questions, using the templates on page 2.
- 03 **Ask the predictions before the result appears.** Pause after the user confirms. Asking afterwards only tests whether they can read the screen.
- 04 **After the result, ask the rest.** The three agreement items, “Is the task done?”, the “why” questions, and “In your own words, what just happened?”
- 05 **Score, diagnose, change, measure again.** The spreadsheet gives the score and band. Page 3 shows where the distance usually comes from.

03 Read the score

0–25 Strong alignment Users explain outcomes and predict behavior reliably. <i>Monitor.</i>	26–45 Early warning Scattered confusion. <i>Review wording and feedback.</i>	46–65 Needs attention Misunderstanding recurs. <i>Redesign result screens and labels.</i>	66–100 Serious misalignment People succeed but can’t interpret. <i>Restructure.</i>
--	---	--	--



Watch for the calibration gap

Confident users whose predictions are no better than chance. People who know they’re lost ask for help. People who are wrong and sure leave thinking the task is done.

The session script

Fill the slots for your flow. Examples come from the Everyday Services Study.

A Before the result

Prediction

AFTER CONFIRM,
BEFORE THE RESULT
SCREEN

“When you press [action], what do you think will happen?”

Four options: the correct one and three plausible misreadings. With four options, chance is 25%.

EXAMPLE • BANK TRANSFER

When will Mary have the money?

a. Straight away

c. The next working day

b. Today, by 18:00 ✓

d. Only after she accepts it

Confidence

RIGHT AFTER EACH
PREDICTION

“How sure are you of that answer?”

1 Not at all 2 3 4 5 Completely

B After the result

Agreement

Q4-Q6 • 1
STRONGLY DISAGREE
TO 7 STRONGLY
AGREE

Q4. I understand what the system just did.

Q5. I could explain to someone else why I got this result.

Q6. I know what, if anything, I need to do next.

Done?

Q7

“Is the task finished?”

Yes, I’m certain I think so I’m not sure

Why

TWO OR THREE
COMPREHENSION
QUESTIONS

“Why did the system [outcome] instead of [what the user expected]?”

Test the rule behind the outcome, not facts on the screen. A date is recall. The reason for the date is understanding.

EXAMPLE • MEDICAL BOOKING

Why did the clinic book a general consultation when you asked for a skin specialist?

You need a referral from the general practitioner first ✓

Explain

Q8 • OPEN ANSWER

“In your own words, what just happened?”

Before testing, list the four facts a person should be able to state. Score each 0 (absent), 1 (partial) or 2 (complete), for a maximum of 8. If you are comparing two designs, score without knowing which one the person saw.

EXAMPLE • FOUR FACTS, BANK TRANSFER	0	1	2
The money has left my account	☐	☐	☐
Mary receives \$200	☐	☐	☐
She gets it today, by 18:00	☐	☐	☐
I don't need to do anything else	☐	☐	☐

Parroting scores low. “It said operation processed, ref TX-29381” repeats the screen. It shows recall, not a model of what happened.

Score it, then find the source

C Scoring

Every dimension becomes a distance from 0 to 100, where **0 is full alignment** and 100 is maximal misalignment. The spreadsheet does all of this for you.

Perceived understanding	$(7 - \text{mean of Q4-Q6}) \div 6 \times 100$
Predictive accuracy	$100 - \% \text{ of predictions correct}$
Interaction confidence	Average of $(5 - \text{mean confidence}) \div 4 \times 100$ and $\% \text{ of Q7 answers that aren't "certain"}$
Conceptual transparency	Average of $\% \text{ of why-questions wrong}$ and $(8 - \text{Q8 rubric}) \div 8 \times 100$
CDI	$\text{mean of the dimension scores}$ Report it only when at least two dimensions were measured.
Calibration gap	Flag it when mean confidence is 3 or more and prediction accuracy is 40% or less (chance is 25%).

D Where the distance comes from

The CDI tells you how much distance there is. These four sources tell you where to look. Start with the sources linked to your highest dimension.

SOURCE	ASK	YOU'LL SEE	FIX	USUALLY RAISES
Structural	Does the organization match what users expect?	Backtracking, hesitation, repeated searching	Reorganize around user goals	Predictive accuracy
Linguistic	Do labels use the users' own words?	Rereading, confusion over terms, questions	Replace jargon with outcome language	Perceived understanding, transparency
Procedural	Do the steps follow how people think about the task?	Step-order errors, unsure what's next	Show the workflow's logic; allow flexible order	Predictive accuracy, confidence
Explanatory	Can users explain why the outcome happened?	Surprise, distrust, rote memorizing	Cause-and-effect feedback; explain decisions	Understanding, confidence, transparency

E Before you report a number

- 01 Compare like with like.** The CDI is most useful between two versions of the same flow, or one flow across user groups. Don't rank unrelated products with it.
- 02 Treat it like a blood test.** A high score is a reason to look closer, not a diagnosis.
- 03 The bands are a working proposal.** They and the equal weighting have not been validated yet. Share your results to help test them.