

STAGE 2 · SCORED REVIEW

A human-centered evaluation tool for educational AI

Most educational AI is judged on accuracy, speed, or features. This tool asks a different question: *what is the learner's experience while interacting with the AI* — and what would make it more fully human?

Ten dimensions across three parts, each scored 1–3, to surface strengths, gaps, and the one change worth making next.

AI FOR HUMAN BECOMING

01 WHO THIS IS FOR

PRIMARY

Learning designers building AI-supported learning — instructional designers, educational technologists, and educators who design or prototype AI tools, reviewing their own work and peers'.

ALSO USEFUL TO

Researchers, programme leads, and adopters who want a shared, human-centered lens for examining an educational AI experience.

NOT A TOOL FOR

Procurement scoring, model benchmarking, or accuracy testing. It reviews the **designed experience**, not the engine behind it.

02 HOW IT WORKS — TWO STAGES

STAGE 1 · DISCOVERY

What the designer intended

The Discovery Prompts capture the design team's stated intent across the same ten dimensions — in their own words.

GAP =
REFINEMENT

STAGE 2 · EVALUATION — YOU ARE HERE

What the build actually does

Score each dimension from observed evidence — transcripts, artifacts, a real session. Where the build falls short of the intent is the clearest next move.

03 THE THREE-POINT SCALE

SCORE 1

Receptive

The learner mostly takes in what the app delivers — listening, receiving, following a path someone else set.

HOW KNOWLEDGE IS TREATED

Transmissive

Knowledge flows one way, app to learner. The app is the source; the exchange is agent-led — the "banking" model.

SCORE 2

Participatory

The learner joins in and acts alongside the app — contributing, choosing, and shaping the exchange.

HOW KNOWLEDGE IS TREATED

Constructive

Knowledge is built, not just delivered — within limits the app still sets, with dips or drift.

SCORE 3

Immersive

The learner is absorbed in the experience and owns it — steering direction, depth, and pace while the app supports their lead.

HOW KNOWLEDGE IS TREATED

Co-creative

Knowledge is co-created and inhabited. The app is a coherent partner that protects the learner's voice and grows with them.

APP / AGENT

LEARNER & CONTEXT

SESSION REVIEWED

REVIEWER & DATE

Does the learner come, return, and go deeper?

The engagement arc — attraction, adoption, and persistence.

SCORE 1

Receptive

transmissive — knowledge delivered

SCORE 2

Participatory

constructive — knowledge built

SCORE 3

Immersive

co-creative — knowledge co-owned

1 Attraction FIRST PULL

Score __ / 3

Does the experience give a first-time learner an immediately clear reason to try it — a felt need or curiosity in the first moments?

STAGE 1 ASKED What pulls a first-time user in, and what do they feel in the first 30 seconds?

WHERE TO LOOK

QUESTIONS TO HOLD

- Is the value or invitation clear within the first 30 seconds?
- Does it meet a real need or spark genuine curiosity?
- Does the first moment feel worth the learner's time?

EVIDENCE TO GATHER

- A recording of a first-time user's opening 30–60 seconds.
- The empty / onboarding state and its first-run copy.
- Time-to-first-"aha" — how long until value lands?

SCORE 1

Receptive

The opening is unclear or generic; little reason to try it, nothing felt in the first moments.

LOOKS LIKE

Opens to a blank prompt beside a feature list. A first-timer skims it, unsure where to start, and leaves.

SCORE 2

Participatory

There's a hook, but the value takes effort to find — or only some learners feel the pull.

LOOKS LIKE

The point is clear, but nothing happens until the learner first supplies their own material.

SCORE 3

Immersive

The need or curiosity lands immediately; within moments it feels worth trying.

LOOKS LIKE

Opens mid-task with a worked example already in motion — in seconds they want to try it with their own.

EVIDENCE & NOTES

REFINEMENT OPPORTUNITY

2 Adoption ROUTINE

Score __ / 3

Does the experience connect to a recurring need so it becomes a go-to the learner folds into how they already work or think?

STAGE 1 ASKED What recurring need makes your app a go-to, so users fold it into how they already work?

WHERE TO LOOK

QUESTIONS TO HOLD

- Is there a recurring moment or need the experience clearly serves?
- Does it fit how the learner already works, rather than adding friction?
- Would the learner return without being prompted?

EVIDENCE TO GATHER

- Return cadence over 2–4 weeks (telemetry, not one demo).
- Where it sits in the learner's existing workflow (interview / diary).
- Unprompted returns vs. nudged ones.

SCORE 1 Receptive

It stays a one-off; nothing brings the learner back or fits it into their routine.

LOOKS LIKE

Tried once for a one-time task, then forgotten; nothing ties it to next week's work.

SCORE 2 Participatory

The learner returns sometimes, but it hasn't become a dependable habit.

LOOKS LIKE

They come back around big moments when they remember, but it hasn't replaced their usual way.

SCORE 3 Immersive

It earns a regular place in the routine and becomes a natural go-to.

LOOKS LIKE

It's where they start every new cycle of work — a standing step they reach for unprompted.

EVIDENCE & NOTES

REFINEMENT OPPORTUNITY

3 Persistence DEEPENING

Score __ / 3

Does the experience get richer the longer the learner stays with it, continuing to develop rather than hitting a plateau?

STAGE 1 ASKED How does the experience get richer the longer someone stays with it?

WHERE TO LOOK

QUESTIONS TO HOLD

- Does the experience deepen with continued use?
- Does the learner keep growing rather than plateauing?
- Does it adapt as the learner's needs and skill change?

EVIDENCE TO GATHER

- Compare an early session with a much later one (e.g. 1 vs. 20).
- Signs of memory / personalization carrying forward.
- Any accumulating artifact — a library, profile, or history.

SCORE 1 Receptive

The experience stays flat or repetitive, and the learner plateaus quickly.

LOOKS LIKE

Session 20 looks like session 1 — same prompts, same scaffolds, no memory of the learner.

SCORE 2 Participatory

It deepens for a while, then levels off or stops adapting to the learner's growth.

LOOKS LIKE

It tailors for a while, then plateaus; the learner's advanced needs outgrow what it offers.

SCORE 3 Immersive

It keeps getting richer and more personalized, growing alongside the learner over time.

LOOKS LIKE

It compounds — recalling style, building on past work, surfacing a growing library of the learner's own.

EVIDENCE & NOTES

REFINEMENT OPPORTUNITY

What is the experience like?

Relational presence, learner agency, engagement, and coherence.

SCORE 1

Receptive

transmissive — knowledge delivered

SCORE 2

Participatory

constructive — knowledge built

SCORE 3

Immersive

co-creative — knowledge co-owned

4 Relational presence CONTINUITY

Score __ / 3

Does the AI feel like a coherent, responsive learning partner over time, rather than a string of disconnected replies?

STAGE 1 ASKED Where does your app feel like one continuous learning partner?

WHERE TO LOOK

QUESTIONS TO HOLD

- Does the agent carry context across turns?
- Does it acknowledge the learner's intent and earlier contributions?
- Does it feel like one relationship or many separate exchanges?

EVIDENCE TO GATHER

- A multi-turn transcript, plus a cross-session recall test.
- Back-references to something the learner said earlier.
- Whether continuity holds under depth or breaks.

SCORE 1

Receptive

Each reply stands alone, and the agent forgets earlier turns.

LOOKS LIKE

Ask about something raised five minutes ago and it has no idea what you mean.

SCORE 2

Participatory

The agent recalls some context, though continuity breaks under depth or across sessions.

LOOKS LIKE

It remembers within a session, but greets you as a stranger the next day.

SCORE 3

Immersive

The agent sustains a coherent thread and builds on what the learner shared, feeling like one partner.

LOOKS LIKE

It picks up last week's thread and builds on what you told it then.

EVIDENCE & NOTES

REFINEMENT OPPORTUNITY

5 Learner agency DIRECTION & PACE

Score __ / 3

How much can the learner shape the direction, depth, and pace of the interaction, exercising judgment rather than following a fixed path?

STAGE 1 ASKED *Where can users steer the experience instead of following a fixed path?*

WHERE TO LOOK

QUESTIONS TO HOLD

- Can the learner steer where the conversation goes?
- Can they slow down, go deeper, or change course?
- Who is leading — the learner or the agent?

EVIDENCE TO GATHER

- Transcript moments where the learner redirects.
- Presence of real choice points, not just a fixed flow.
- Who proposes the next step, turn by turn.

SCORE 1 Receptive

The agent drives a fixed path and the learner mostly follows.

LOOKS LIKE

Every reply ends by telling you the next step; stepping off it breaks the flow.

SCORE 2 Participatory

The learner can make some choices, within limits the agent sets.

LOOKS LIKE

You can pick from its menu, but not set your own direction.

SCORE 3 Immersive

The learner clearly steers direction, depth, and pace, and the agent supports their lead.

LOOKS LIKE

You say "let's go deeper on this part" and it follows your lead.

EVIDENCE & NOTES

REFINEMENT OPPORTUNITY

6 Cognitive–emotional engagement ATTENTION & CHALLENGE

Score __ / 3

Does the interaction hold attention, curiosity, and intellectual engagement without overwhelming or disengaging the learner?

STAGE 1 ASKED *What keeps users engaged and curious, and how do you guard against overload?*

WHERE TO LOOK

QUESTIONS TO HOLD

- Does the experience invite curiosity?
- Is the challenge level right, neither flat nor overwhelming?
- Does the learner stay involved over time?

EVIDENCE TO GATHER

- Session length and where attention drops off.
- Signs of strain or boredom in a think-aloud.
- How well challenge is calibrated to the learner.

SCORE 1 Receptive

The exchange feels flat or draining, or it overwhelms the learner.

LOOKS LIKE

Walls of text; the learner skims, then disengages.

SCORE 2 Participatory

The experience holds attention in parts, with some dips or strain.

LOOKS LIKE

Engaging in bursts, with stretches that drag or briefly overwhelm.

SCORE 3 Immersive

The experience sustains curiosity and engagement at a challenge level that feels right.

LOOKS LIKE

Pitched just past easy — the learner stays curious and leans in.

EVIDENCE & NOTES

REFINEMENT OPPORTUNITY

7 Contextual coherence

ALIGNMENT TO PURPOSE

Score __ / 3

Does the AI stay aligned with the learning purpose, task, and the learner's goals across exchanges?

STAGE 1 ASKED *How does your app stay on track with the learner's purpose across a session?*

WHERE TO LOOK

QUESTIONS TO HOLD

- Does the agent keep the learning purpose in view?
- Does it stay on task from one turn to the next?
- Does it adapt to the learner's goals without drifting?

EVIDENCE TO GATHER

- The stated goal vs. the agent's thread across turns.
- How often it drifts off-task, and whether it recovers.
- Who has to pull it back when it wanders.

SCORE 1
Receptive

The agent drifts off purpose or loses the thread of the task.

LOOKS LIKE

Asked for a focused task, it wanders into generic advice.

SCORE 2
Participatory

The agent mostly stays aligned, with occasional drift the learner must correct.

LOOKS LIKE

Mostly on task, but you re-state the goal to pull it back.

SCORE 3
Immersive

The agent holds the learning purpose and goals steadily across the whole exchange.

LOOKS LIKE

It holds the goal across the whole session without being reminded.

EVIDENCE & NOTES

REFINEMENT OPPORTUNITY

Does it support meaningful learning?

Reflection, creation, and application — how the experience treats the learner's own thinking, making, and judgment.

SCORE 1

Receptive

transmissive — knowledge delivered

SCORE 2

Participatory

constructive — knowledge built

SCORE 3

Immersive

co-creative — knowledge co-owned

8 Reflective depth REFLECTION

Score __ / 3

Does the AI help the learner interpret experience, make sense of feedback, and connect learning to broader understanding?

STAGE 1 ASKED Where does your app help users make sense of their own thinking, not just receive answers?

WHERE TO LOOK

QUESTIONS TO HOLD

- Does the agent invite the learner to notice their own thinking?
- Does it mirror patterns and assumptions back to them?
- Does it deepen interpretation rather than hand over answers?

EVIDENCE TO GATHER

- The ratio of questions-back to answers-given.
- Prompts that ask the learner to interpret or justify.
- The learner's own meaning-making in the transcript.

SCORE 1

Receptive

The agent gives answers, with little room for the learner to reflect.

LOOKS LIKE

It hands over the answer; no invitation to think it through.

SCORE 2

Participatory

The agent prompts some reflection, though it often returns to answers.

LOOKS LIKE

It asks one reflective question, then reverts to answering.

SCORE 3

Immersive

The agent consistently helps the learner interpret, notice patterns, and make meaning.

LOOKS LIKE

It mirrors your reasoning back: "you assumed X here — why?"

EVIDENCE & NOTES

REFINEMENT OPPORTUNITY

9 Creative authorship CREATION

Score __ / 3

Does the learner generate, shape, and own the work, with AI as a collaborator rather than the author?

STAGE 1 ASKED *Where does your app help users make and own something, rather than receive finished output?*

WHERE TO LOOK

QUESTIONS TO HOLD

- Does the learner produce and revise their own ideas?
- Does the agent expand possibilities without taking over?
- Whose voice carries the final work?

EVIDENCE TO GATHER

- The artifact the learner produced.
- Its revision history — who changed what.
- Whose phrasing and voice dominate the output.

SCORE 1 Receptive

The agent produces finished work and the learner mostly selects from it.

LOOKS LIKE

It writes the finished piece; the learner picks option A or B.

SCORE 2 Participatory

The learner shapes the work, though the agent's voice still leads at times.

LOOKS LIKE

The learner edits its draft, but its voice still sets the shape.

SCORE 3 Immersive

The learner authors and owns the work, and the agent expands options while protecting their voice.

LOOKS LIKE

The learner drafts; it offers options and protects their voice.

EVIDENCE & NOTES

REFINEMENT OPPORTUNITY

10 Applied judgment APPLICATION

Score __ / 3

Does the learner carry the learning into a real decision or context, with responsibility staying human?

STAGE 1 ASKED *Where does your app help users connect the learning to a real choice or situation?*

WHERE TO LOOK

QUESTIONS TO HOLD

- Does the agent connect ideas to real situations and choices?
- Does it support weighing options and consequences?
- Does the learner keep responsibility for decisions?

EVIDENCE TO GATHER

- Where a real decision actually gets made.
- Who makes the call — the learner or the agent.
- Whether consequences are weighed with the learner.

SCORE 1 Receptive

The agent makes recommendations or decisions for the learner.

LOOKS LIKE

"Use this one." The decision is effectively made for them.

SCORE 2 Participatory

The agent supports decision-making, though it nudges toward its own answers.

LOOKS LIKE

It helps them decide, but keeps nudging toward its pick.

SCORE 3 Immersive

The agent helps the learner weigh options and consequences while the learner decides.

LOOKS LIKE

It lays out the trade-offs; the learner makes the call.

EVIDENCE & NOTES

REFINEMENT OPPORTUNITY

OVERALL REFLECTION

Where is this experience now — and what moves it?

-- / 30

10-16

Receptive

17-23

Participatory

24-30

Immersive

Across all ten dimensions, where is this experience strongest, where does it feel most receptive, and what one change would move it most toward immersive?

A NOTE ON SCOPE

This tool reviews the designed experience — the work you can see during design and prototyping. Whether the learner's core outcome actually grew is a separate question that belongs to implementation and longitudinal study.