

AI Engagement Quality Framework Developmental Evaluation Report — Session 2

22 March 2026 · Framework version 1.0 · Evidence base: This conversation only

Evaluation Summary

Dimension	Score	Band	Movement
1. Deliberate Planning	8/10	Capable	+1 from Session 1
2. Effective Use and Iteration	8/10	Capable	Unchanged
3. Self-Evaluation of AI Use	8/10	Capable	+1 from Session 1
4. Critical Thinking	9/10	Exemplary	Unchanged
Practice Coherence	8/10	Capable	Unchanged

Dimensional Evaluations

Dimension 1: Deliberate Planning — 8/10 — Capable (+1 from Session 1)

Strengths

The full conversation arc confirms that the opening prompt was the start of a planned sequence rather than an improvised exploration. The forward signal mid-conversation — 'we will pack that up after this exercise' — is the clearest evidence: the user had a destination in mind before the current step was complete. The correction of the PDF omission was also planned rather than reactive: the user named the specific gap, located it to the development conversation, and directed a targeted fix. The close of the conversation — running the meta review as an explicit final step — confirms the session had a structured shape throughout.

Weaknesses

The opening prompt contained several typos, suggesting it was drafted quickly. The combined document request emerged late rather than being specified at the outset — a minor planning gap given that document format is a production decision that could have been named earlier.

Visibility Note

The planning process itself remains unobservable. The rating reflects what the conversation arc and forward signalling suggest about prior intent — a stronger inference than was available at the first evaluation point.

Area for Development

Name the intended deliverables as a set at the outset. For a session like this one, a single sentence listing the intended outputs would have made the planning visible from the start and may have surfaced the combined document requirement earlier.

Dimension 2: Effective Use and Iteration — 8/10 — Capable (Unchanged)

Strengths

Iteration throughout the session was targeted and produced material improvements. The three-challenge sequence remains the strongest example. The PDF omission correction is equally precise: the user identified a specific design decision that had been agreed but not implemented, and directed a fix to that exact point. The decision to wait for desktop access for the skills upload — rather than pursuing impractical alternatives — reflects good judgement about when to stop iterating and accept a constraint.

Weaknesses

Several outputs across the session were accepted without visible interrogation — the cross-framework synthesis section, the record block format, the PDF report layout, and the combined document structure all passed without challenge. Whether this reflects active acceptance of strong proposals or passive acceptance remains ambiguous where no signal was given.

Visibility Note

Direct observation is strong for iteration that produced visible challenge or correction. The ambiguity remains around uncontested acceptance. The additional session material confirms the pattern rather than changing it.

Area for Development

Where a significant structural decision is accepted without challenge, a brief active acceptance signal would strengthen the iteration record. The combined document in particular — a substantial production decision — passed without comment on its structure or completeness.

Dimension 3: Self-Evaluation of AI Use — 8/10 — Capable (+1 from Session 1)

Strengths

This dimension has strengthened since the first evaluation. The PDF omission correction is the key piece of new evidence: the user identified that a design decision had been agreed but not carried through, named it as an omission, and directed a correction. That is self-evaluation applied to the outputs of the session rather than to its stated subject, and it emerged

unprompted rather than being scaffolded by the framework. The decision to run a second meta review at the close, in developmental mode, also demonstrates an ongoing monitoring habit.

Weaknesses

Self-evaluation is still most visible when applied to structured outputs or explicit quality checks. There is limited evidence of unprompted reflection on the engagement process itself mid-session — reflection surfaces at defined points rather than continuously.

Visibility Note

This is the dimension that has moved most clearly between the two evaluations. The PDF correction provides direct evidence of self-evaluation operating outside a structured framework context. The rating increase from 7 to 8 is warranted on this evidence.

Area for Development

The next evidence threshold is self-evaluation applied to an engagement where quality checking is not the stated purpose. That would confirm whether the habit is fully independent of the scaffold.

Dimension 4: Critical Thinking — 9/10 — Exemplary (Unchanged)

Strengths

The strongest dimension throughout. The iPad/PDF constraint exchange is worth noting: when offered three options, the user assessed each practically and honestly — identifying that two presented alternatives did not actually work rather than accepting a plausible-sounding solution. The PDF omission correction also reflects discriminating judgement: the user distinguished between what had been agreed and what had been delivered, and named the gap precisely. Throughout the session, challenge has been purposeful and well-calibrated.

Weaknesses

The potential for the two frameworks to conflict — rather than simply complement — remains unexplored. This was identified in the first evaluation and has not been returned to. At this level of practice, identifying which available threads are worth pursuing is itself a critical thinking question.

Visibility Note

This dimension continues to have the strongest direct evidence base in this conversation. The rating reflects what is observable with confidence.

Area for Development

The single remaining gap is the conflict scenario — whether the rubric and the engagement framework can produce incompatible readings of the same practice. Worth one focused exchange to resolve.

Holistic Measure: Practice Coherence

Practice Coherence — 8/10 — Capable (Unchanged)

The four dimensions are reinforcing each other more visibly across the full conversation than was apparent at the first evaluation point. The planning produced a structured session with a clear arc; the iteration was driven by critical thinking rather than following it; the self-evaluation surfaced a specific gap mid-session and directed a correction. These are not parallel competencies — they are operating as a connected practice.

The coherence limit identified in the first evaluation — that self-evaluation was scaffolded by the framework — has partially resolved. The PDF correction provides evidence of self-evaluation operating outside the explicit evaluation context. What remains is confirmation across a different type of engagement.

Area for Development

A single evaluation against a non-evaluation conversation would close the remaining coherence question. The practice visible here is integrated and deliberate — the missing evidence is not a gap in practice but a gap in the record.

Development Priorities — Developmental Mode | Session 2

Where to focus first

Planning — specifically, naming the intended deliverables as a set at the outset of sessions where the destination is already known.

Suggested next step

Before the next structured session, write one sentence that lists the intended outputs. Not the questions, not the process — the outputs. For a session like this one, that sentence would have read: 'Intended outputs: skill file, project instructions, combined Word document, PDF evaluation report.' That takes thirty seconds and changes the architecture of everything that follows.

Longer-term trajectory

The practice visible across this conversation is operating at Capable across three dimensions and Exemplary on one. The move toward Exemplary on the remaining dimensions is not about correcting weaknesses — the weaknesses here are marginal. It is about generating evidence in contexts where the practice is active but not the subject. Run this framework against a bid or briefing conversation and the coherence picture will be complete. The conflict question between the two frameworks is also worth one focused exchange — it is the single remaining analytical thread that has been identified but not pursued.

Evaluation Record

Date	22 March 2026
Evaluation mode	Developmental
Evidence base	This conversation only
Framework version	1.0
Session	Session 2 of 2
1. Deliberate Planning	8/10 — Capable (+1 from Session 1)
2. Effective Use and Iteration	8/10 — Capable (Unchanged)
3. Self-Evaluation of AI Use	8/10 — Capable (+1 from Session 1)
4. Critical Thinking	9/10 — Exemplary (Unchanged)
Practice Coherence	8/10 — Capable (Unchanged)
Highest-priority development area	Name intended deliverables as a set at the outset of structured sessions.
Notes	Movement on Dimensions 1 and 3 reflects additional evidence from the extended conversation — particularly the PDF omission correction as unprompted self-evaluation, and forward signalling as planning evidence. Dimension 4 and Practice Coherence are stable. Published with the practitioner's full approval.

This evaluation was produced using the Dynes Insights AI Engagement Quality Framework v1.0, in Developmental mode. Published with the practitioner's full approval. No content has been altered from the original evaluation.

AI Engagement Quality Framework · Version 1.0 · Field Testing · 22 March 2026