Enhancing Assessment and Programme Design in the AI Era

Professor Mark J.P. Kerrigan (SOAS University of London)
Dr Sian Lindsay (University of Reading)
Sarah Sherman (Bloomsbury Learning Exchange)

The challenge (1):

Generative AI has magnified a long-standing assessment problem: what happens when we value the final product too much?

- If students can use AI to produce or polish the final output…the final output (*on its own*) may no longer give us enough **assurance** about what learning has happened, raising questions about the **validity** of our assessments, as well as **integrity** of our qualifications.
- So - what other “evidence” of learning do we need? How can we see this?

The challenge (2):

Generative AI has introduced a scaling problem

We all have a lot of modules, a lot of assessment and not many educational developers or associated roles to support the re-design of assessment

1 degree = 8 modules per year → 24 modules in three years → allow for some option modules → ~30 modules → x number of assesments (x2 per module) --> 60 assessments per degree → x number of degrees → over 1,000. BUT these are linked + integrated

Aim:

To develop a process to enhance programme-level assessment design, build AI literacy and strengthen academic integrity as part of a programme design and validation.

Outline:

Pilot, evaluate and refine a practical Assessment Toolkit that supports programme teams to design secure, inclusive and future-focused assessments across online, blended and face-to-face contexts

Programme idea

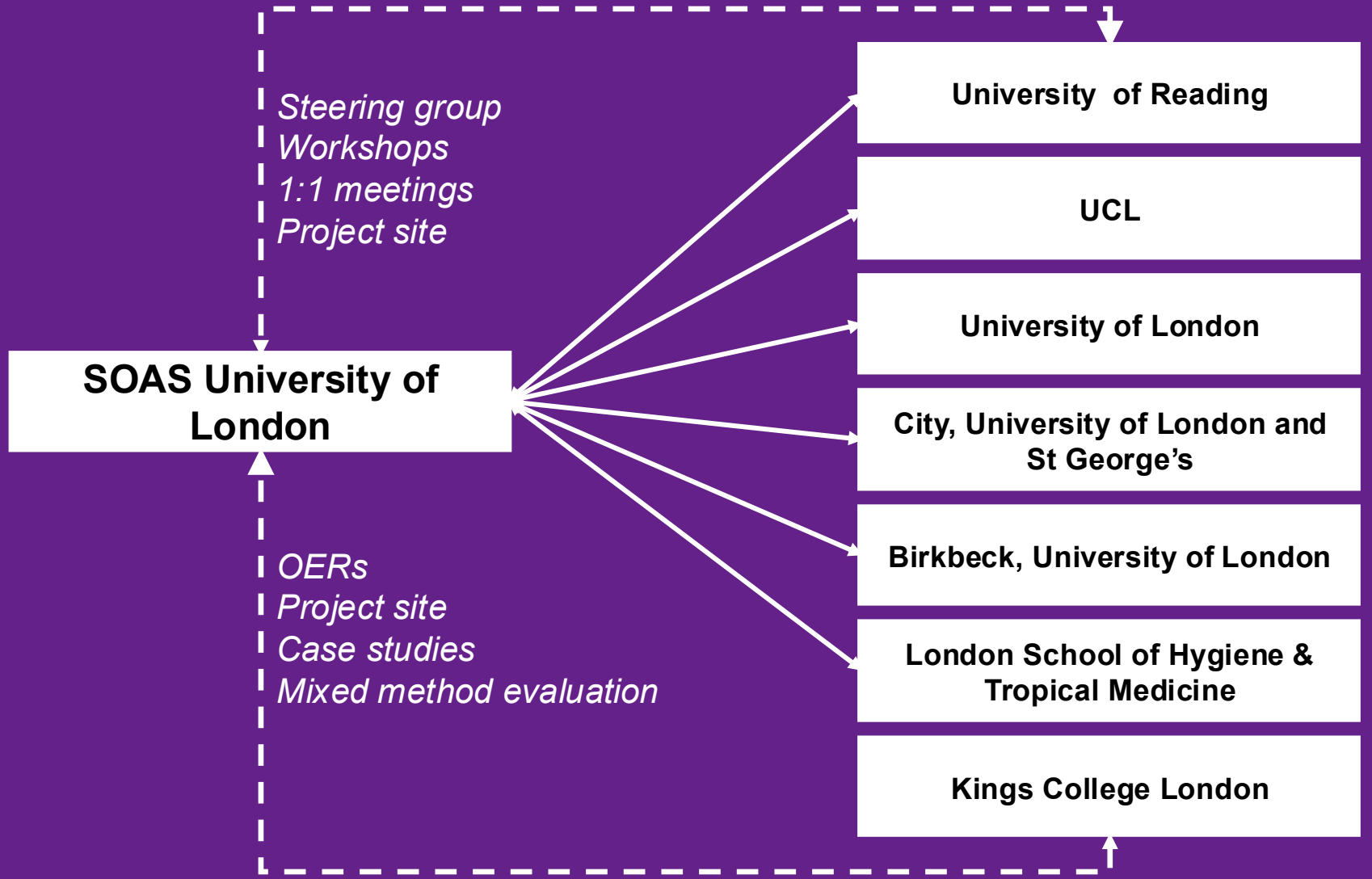
Programme design + development using the AI assessment toolkit

Pilot, evaluate and refine the toolkit that supports programme teams to design secure, inclusive, AI-aware, and future-focused assessment across online, blended and face-to-face contexts.

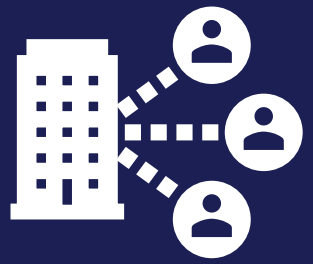
Programme validation

Enhancing Assessment and Programme Design in the AI Era

Co-development



Working in partnership



Co-refined + tested toolkit for holistic assessment in the AI age

A series of support resources and OERs to apply the toolkit in their own context

Key outputs

- Refined AI-assessment toolkit (OER)
- Short case studies on its use + application
- Resource on effective AI + assessment design
- Sector dissemination
- Text + video case studies | Podcast | video
- End of project webinar
- Project site and OER access

Key milestones

Feb-May 2026

PHASE 1

Project setup, confirmation of institutional ethical approval requirements (plus application), and agreement of pilot sites

May-Feb 2027

PHASE 2

Partner institutions pilot the Assessment Toolkit within one or more programme design and/or validation.

Feb-July 2027

PHASES 3-5

Evaluation + reflection. Data collection + analysis. Toolkit refinement. End-of-project dissemination and OER creation.

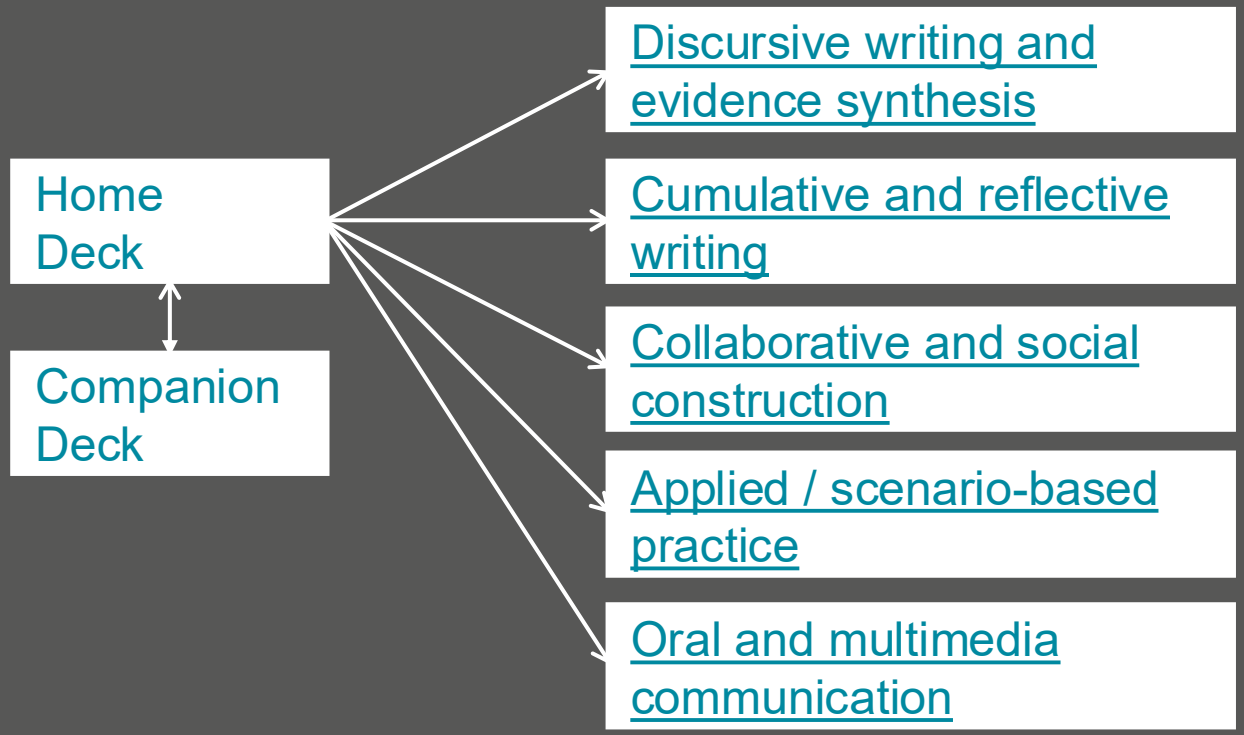
Areas of evidence

Allied to Medicine
Social Sciences
Humanities
Business
Law
Computer Science
Placement
+ others on the way ☺

UG

PG

What is the toolkit?



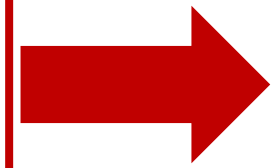
- Overview and rationale
- SWOT analysis
- AI considerations
- Graduate attribute links
- Example assignment briefs
- Rubrics, formative checkpoints and feedback ideas
- Use of institutional e-portfolio to capture process

Home Deck

1. Discursive writing and evidence synthesis	2. Cumulative and reflective writing	3. Collaborative and social construction	4. Applied / scenario-based practice	5. Oral and multimedia communication	6. Live, practical and time-constrained assessment
1A. Essay	2A. Patchwork assignment	3A. Collaborative digital artefact	4A. Scenario-based assessment (case study / role-play / simulation)	5A. Recorded presentation (video / audio)	6A. Assessed workshop / studio critique
1B. Policy / issue analysis brief	2B. Comparative learning journal	3B. Structured debate / dialogue	4B. Data collection task	5B. Live presentation / viva	6B. Laboratory, field or practical task
1C. Annotated bibliography & literature map	2C. Critical incident analysis	3C. Group problem-based learning / consultancy project	4C. Data analysis task	5C. Digital poster / infographic	6C. Lab, field or practice notebook
1D. Research design proposal	2D. Fact-checking dossier		4D. Grant application	5D. Podcast / audio blog	6D. In-class, open-book or time-constrained test
1E. Data analysis report					6E. Structured questions, MCQs, EMIs or coding quizzes
1F. Dissertation / independent research project					

Companion Deck

Click on each box to find out more!



S1: Short explanation check	SM1: Short audio note	SM2: Screencast feedback	SM3: Rubric-linked inline comments
SM4: Office hours / drop-in clinic	M1: Standards comparison session	M2: Acting on feedback	M3: Decision-log / change record
ML1: Whole-class feedback digest	ML2: Annotated exemplar / worked example	ML3: Structured peer checkpoint	ML4: Asynchronous Q&A / clarification thread
ML5: Analytics-driven nudge	L1: Automated self-check quiz	L2: Self-assessment checklist	L3: Tell-Ask-Give (TAG) feedback

KEY

S: < 20 students

S/M: 20-60 students

M: 61-150 students

M/L: 151-499 students

L: 500+ students

2B: Comparative learning journal

A comparative learning journal uses a stable set of lenses to track how understanding changes over time by comparing themes, cases, methods or sources. It may also be called a cross-case journal, dialectical notebook, lens-based journal or synthesis journal.

Students usually complete three or four short entries during a module or teaching block, followed by a brief consolidation or final integrative commentary. The distinctive feature is comparison: students use the same lenses, such as claim, evidence quality, assumptions, ethics, impact or transferability, to explain what changed in their thinking and why one account now seems more convincing.

Return to main menu

2B: Comparative learning journal

SWOT analysis: student lens

STRENGTHS

- Makes change in understanding visible rather than hidden behind a final answer
- Builds confidence in comparison, metacognition and evidence-based judgement
- Encourages regular engagement through short, spaced entries

WEAKNESSES

- Can drift into description if lenses and exemplars are unclear
- Reflection quality may vary unless prompts ask students to judge, not merely narrate
- Multiple entries can feel heavy if the journal is over-designed

OPPORTUNITIES

- Allow text, audio or grid-based entries to widen participation
- Use peer circles and whole-class digests to scale feedback
- Extend into a programme-level longitudinal journal for capstone integration

THREATS

- Allow text, audio or grid-based entries to widen participation
- Use peer circles and whole-class digests to scale feedback
- Extend into a programme-level longitudinal journal for capstone integration

2B: Comparative learning journal

SWOT analysis - educator lens

STRENGTHS

- Strengthens evaluative judgement, metacognition and comparative reasoning
- Captures process evidence and feedback use in a manageable format
- Aligns well with graduate attributes around agility, innovation and global citizenship

WEAKNESSES

- Requires clear lenses, templates and examples to avoid superficial comparison
- Marking load can increase unless entries are sampled or assessed with a streamlined rubric
- It may duplicate other reflective tasks if its comparative purpose is not explicit

OPPORTUNITIES

- Use analytics nudges and checkpoints to sustain participation
- Build a final commentary that asks students to justify what now counts as the better explanation
- Use stable criteria across cases to make disciplinary judgement visible

THREATS

- Too many criteria or entries can create cognitive overload
- Weak alignment between entries and final commentary can reduce synthesis
- Generic reflection prompts may produce formulaic responses

2B: Comparative learning journal

SWOT analysis - GenAI lens

STRENGTHS

- AI can help students brainstorm lenses, test counter-arguments and structure blank comparison grids
- Human-vs-AI entries can reveal omissions, inaccuracies and bias
- Prompt-sensitivity work can make AI reliability and provenance discussable

WEAKNESSES

- AI can generate plausible reflective prose without genuine learning
- Students may accept AI summaries without checking evidence or context
- Confidential or personal data may be exposed if entered into public tools

OPPORTUNITIES

- Include one structured human-vs-AI comparison entry where appropriate
- Ask students to record how feedback became an action plan
- Use AI as a stimulus for critique, not as the author of the journal

THREATS

- Undisclosed AI use may obscure student development over time
- Prompted outputs may narrow comparison if students do not refine the lenses themselves
- Reflection can become performative unless tied to evidence and decisions

2B: Comparative learning journal

Links to graduate attributes

Comparative learning journals can help students develop key attributes by asking them to compare evidence, contexts, and assumptions over time, and to articulate how and why their understanding has changed.

Primary attributes:

- **Reflective:** explains what changed in understanding, how it changed, and what feedback, evidence or comparison prompted that development
- **Critical Thinker:** uses stable lenses to evaluate evidence, assumptions, platform affordances, bias and competing interpretations
- **Self-Directed Learner:** maintains regular entries, manages spaced learning and takes responsibility for developing a final integrative commentary
- **Researcher:** appraises how knowledge is constructed and evaluated in the discipline through repeated comparison of cases, sources or methods

Also possible, depending on design:

- **Globally Engaged:** compares social, cultural and contextual impacts across cases, communities or settings
- **Inclusive:** actively considers diverse perspectives, positionality and whose evidence or experience is represented or missing
- **Digitally Fluent:** uses e-portfolios, online observation, comparison grids or AI logs ethically and critically
- **Confident Communicator:** communicates concise analytical entries and a clear final judgement for a defined audience

2B: Comparative learning journal

🤔 Considerations for GenAI use

Reading GenAI category	Pros / opportunities	Risks / design safeguards	Example assignment brief wording (with reference to I3)
Category 1: Independent work	• Protects learning outcomes where the journal is intended to evidence unaided reflection, comparison and development of judgement over time. • Can be suitable for final consolidation, where the student must articulate their own change in understanding. • Encourages safeguards such as dated entries, observation notes, peer feedback records and explanation checks.	**Main risk:** a ban may be difficult to monitor where entries are completed independently and over time. **Safeguard:** make the rationale explicit, require process evidence, stable comparison lenses, dated notes and a short explanation check or viva where authorship or development needs clarification.	"This is a Category 1 assessment. You must complete the comparative learning journal without GenAI assistance. This includes using GenAI to generate comparison lenses or explanations (**Inquire**), rewrite or polish entries (**Improve**), or incorporate AI-generated analysis, summaries or reflection (**Integrate**). The purpose is to assess your independent comparative judgement, reflection and learning development. You may be asked to explain your lenses, entries or final consolidation."
Category 2: AI-supported	• Can support early brainstorming of comparison lenses, probing questions, counter-positions or blank comparison grids. • Can help students test whether their own entries answer the comparison question and respond to feedback. • Fits where AI supports learning, but the final entries and consolidation remain the student's own work and AI use is not itself graded.	**Main risk:** AI support may become AI-authored reflection or unverified summaries of cases/sources. **Safeguard:** require a GenAI use statement, prompt/output records, verification against original evidence, and a decision log showing how students selected/refined lenses and filled the grid themselves. Do not enter personal, sensitive or confidential data into public tools.	"This is a Category 2 assessment. You may choose to use GenAI to support your learning, for example to brainstorm possible lenses (**Inquire**), generate probing questions or counter-positions on your own draft (**Improve**), or adapt a blank comparison-grid structure (**Integrate**). You must verify all content, fill the journal yourself, keep prompt/output records and include a GenAI use statement. The final consolidation must be in your own words and represent your own judgement."
Category 3: AI-integrated	• Makes AI literacy part of the assessment by asking students to compare their own analysis with an AI summary, audit bias/errors or test prompt sensitivity. • Supports explicit discussion of AI reliability, provenance and limitations in relation to the module content. • Allows students with ethical concerns to critically review publicly available GenAI outputs instead of generating new content.	**Main risk:** students may describe AI use without critically evaluating what changed in their view or why the AI output was reliable, biased or incomplete. **Safeguard:** assess AI literacy explicitly: quality of prompts, critical evaluation of outputs, evidence checking, explanation of accept/reject decisions and articulation of limitations, risks and value.	"This is a Category 3 assessment. You are expected to use or critically evaluate GenAI as part of the comparative learning journal. You may use GenAI to generate a comparison baseline (**Inquire**), refine questions for your own entries (**Improve**), or incorporate selected AI-supported structures where appropriate (**Integrate**). Your submission must critically evaluate AI outputs for accuracy, omissions, bias and relevance, and explain what changed in your understanding. Your AI literacy will form part of the assessment."

👍 Suggested feedback and process-evidence options

Feedback menu option	Why it works
ML3: Structured peer checkpoint	Lets students test lenses and emerging comparisons before the final commentary.
M3: Decision log / change record	Shows how lenses, cases, evidence and conclusions changed across entries.
SM3: Whole-class feedback digest	Scales tutor feedback by addressing common issues in comparison, evidence and reflection quality.

2B: Comparative learning journal

Example assignment brief

- Level 7 ODL module = ***Anthropology of Online Food Practices and Belonging***
- **Type and weighting**: Comparative learning journal, 30% of module mark (minor summative)
- **Word count and structure:** ≈ 1,400–1,500 words total: 3 entries (≈350–400 words each) + 200-word consolidation. References excluded.
- **Submission:** via Moodle, with an accompanying e-portfolio documenting formative checkpoints (drafts, reflections, possible AI log) during the final two-week assessment block.
- **Purpose:** Track how your understanding of diaspora food practices and belonging online develops by comparing cases through a consistent set of analytical lenses, then explain what now counts as the better account and why...

2B: Comparative learning journal

Example assignment brief (cont.)

Task:

1. Choose two contrasting online sites where food, identity and community intersect (e.g., a TikTok recipe hashtag vs a diaspora Facebook group; YouTube cooking channel comments vs an Instagram “heritage dish” thread).
2. Conduct two short micro-fieldwork sessions (30–45 minutes each) of naturalistic observation; optional asynchronous contact (e.g., permissioned DM/email) if appropriate.
3. Write three entries comparing Item A vs Item B using the same lenses each time (suggested: claim/purpose; evidence and representation; platform affordances; ethics/positionality; transferability).
4. Finish with a consolidation note stating what changed in your view and one implication for research or practice.

Entry template:

- Context (100–150 words)
- A vs B through your lenses (core analysis)
- What changed in my view? (2–3 sentences)
- References

Formative checkpoints:

Week 3: lens set + case selection (peer circle)
Week 6: Entry 1 draft (tutor digest)
Week 8: Entry 2 draft (peer QandA)
Week 10: Entry 3 outline + consolidation plan.

AI use (optional):

OK for brainstorming lenses, generating probing questions/counter-positions, or creating a blank comparison grid. You may include one entry contrasting your analysis with an AI summary (log inaccuracies/bias and what changed). Keep an AI-use log; the consolidation is written independently (or may be confirmed via a brief viva).

2B: Comparative learning journal

Example analytic rubric

Criterion	Distinction (70–79; 80+ exemplar)	Merit (60–69)	Pass (50–59)	Fail (<50)
Entry quality (across 3 entries) – relevance; disciplined use of lenses; critical comparison; academic style and structure **Weighting: 45%**	Shows sophisticated, critical understanding; lenses applied consistently to A vs B with explicit reasoning; evaluates evidence/representation and platform affordances; arguments are coherent and independently planned; writing is precise with accurate referencing.	Strong critical insight; lenses used clearly with mostly well-justified contrasts; good balance of detail and theory; organised, fluent writing; referencing correct.	Reasonable understanding with some analysis; lenses used but comparisons sometimes descriptive; structure adequate; referencing mostly correct.	Limited or inaccurate understanding; lenses inconsistently used or missing; largely descriptive; weak structure/style; incomplete/incorrect referencing.
Comparative judgement over time and final integrative commentary – change in understanding; warranted preferences; implications **Weighting: 35%**	Makes a creative, original case for what now counts as the better account and why; traces clear development across entries; weighs trade-offs and states implications/next action with strong justification.	Explains a convincing preference using evidence from entries; shows development with some originality; implications stated and mostly well supported.	Offers a preference with partial support; development noted but uneven; implications general or under-argued.	No clear integrative judgement; little evidence of change; conclusions unsupported or absent.
Process evidence and scholarly practice – e-portfolio trace; micro-fieldwork notes and ethics; use of feedback; verification; AI transparency **Weighting: 20%**	Independent planning/organisation evident; time-stamped notes for ≥2 observations with ethical/positionality reflection; decision-log and response-to-feedback show critical evaluation; sources verified; AI use (if any) transparent and assessed critically.	Planned workflow with appropriate iteration; observation notes present; rationale for choices clear; feedback used meaningfully; accurate referencing.	Basic drafting and observation notes; some rationale for changes; limited reflection; mostly correct referencing; minimal AI/ethics notes.	Sparse or missing process record; absent/poor observation evidence; weak rationale; ethics/verification not evidenced; incomplete/incorrect referencing.

We are happy to answer any questions

Email: mk157@soas.ac.uk