Paper Creation Process

A Human-AI Collaborative Academic Writing Workflow

Cheng-I Wu Claude (Anthropic)

March 8, 2026

Abstract

This document records the complete human-AI collaborative process of creating an academic paper from conception to final output. The paper examines how agentic artificial intelligence could redefine student learning outcome measurement, with policy recommendations for Taiwan's fourth accreditation cycle. The entire process was orchestrated using Claude Code's `academic-pipeline` skill (v2.3), spanning 9 stages across 2 full pipeline cycles: research, writing, integrity verification, simulated peer review, revision, and formatting. The most critical iteration arose from the user's insight that "the definition of learning itself is also changing," which gave birth to the paper's most original contribution — the co-evolution perspective in §4.7.

Contents

1 Paper Creation Process: A Human-AI Collaborative Workflow

1.1 Paper Information

Title: *From Snapshots to Trajectories: How Agentic AI Will Redefine Student Learning Outcomes and Transform Student Success Measurement — Implications for Taiwan's Next Cycle of Institutional Accreditation*

Final Outputs: English full text (~15,000 words, 997 lines), Chinese full text (~69,000 tokens), APA 7.0 LaTeX PDF, DOCX, NotebookLM presentation prompt

---

1.2 Phase 1: Initiation and Research (Session 2, Evening of 2026-03-07)

The user's first message:

> "use academic pipeline related skills. I would like to write a paper about how agentic ai will redefine student learning outcomes and the paradigm shift of student success measurement, in the context of…"

This triggered the `academic-pipeline` skill (v2.3), launching the full 9-stage pipeline.

1.2.1 Stage 1: RESEARCH (deep-research skill)

- Claude used the `deep-research` skill for literature search and research synthesis
- Output: `stage1_research_synthesis.md` — including research questions, methodological framework, bibliography, and core arguments

The user's first course correction:

> "wait, i would like to add a direction: and its implication to the next cycle of institutional accreditation"

This shifted the paper from pure theoretical analysis to concrete policy recommendations, focusing on the design window for HEEACT's fourth accreditation cycle. After confirming the new direction with `"good"`, research continued.

1.3 Phase 2: Writing (Session 2, continued)

1.3.1 Stage 2: WRITE (academic-paper skill)

- Claude used the `academic-paper` skill to write 6 section files in parallel:
  - `sections_1_7_8_abstract.md` (Introduction, Research Agenda, Conclusion, Abstracts)
  - `section2_agentic_ai_paradigm.md` (Four-level AI taxonomy)
  - `section3_taiwan_current_paradigm.md` (Current paradigm and six structural limitations)
  - `section4_paradigm_shift.md` (ADAPT framework, seven-dimension paradigm shift)
  - `section5_policy_accreditation.md` (Three scenarios, South Korea lesson, phased pathway)
  - `section6_ethics_governance.md` (Four-principle ethical analysis, three-tier governance)
- Assembled into `full_paper.md` as the complete English paper

The user advanced the pipeline with multiple replies of "continue."

1.4 Phase 3: Integrity Verification (Session 2, continued)

1.4.1 Stage 2.5: INTEGRITY (integrity_verification_agent)

- First verification found issues → `integrity_report_stage2.5.md`
- Fixed and re-verified → `integrity_reverification_stage2.5.md` (PASS)
- Ensured all references exist, citation contexts are accurate, and statistical data are verifiable

---

1.5 Phase 4: Peer Review + Revision (Session 2, continued)

1.5.1 Stage 3: REVIEW (academic-paper-reviewer skill)

Simulated independent review by 5 reviewers:

- **EIC** (Editor-in-Chief): Structure and coherence
- **R1** (Methodology Expert): Appropriateness of Kuhn's framework
- **R2** (Domain Expert): Accuracy of Taiwan-specific content
- **R3** (Cross-disciplinary Expert): Cognitive science and educational technology
- **Devil's Advocate**: Challenged all arguments

Output: `stage3_review_report.md` (60KB) — **Minor Revision**

1.5.2 Stage 4: REVISE

Point-by-point revision based on review comments. The user initiated with "revise."

1.5.3 Stage 3': RE-REVIEW (Verification review)

Output: `stage3prime_rereview_report.md` — **Accept**

1.5.4 Stage 4.5: FINAL INTEGRITY

Final integrity verification → `integrity_report_stage4.5.md` — **PASS**

1.6 Phase 5: Formatting (Session 2, Early morning of 2026-03-08)

1.6.1 Stage 5: FINALIZE

The user requested:

> "produce docx + latex + pdf"

Claude produced DOCX + LaTeX + PDF. After visual inspection:

> "Is the PDF in LaTeX format?"

The user discovered the PDF was converted from HTML (not academic standard) and required:

> "Install [tectonic] and from now on all PDFs from this skill must be compiled from LaTeX"

Claude installed `tectonic` (LaTeX compiler). All PDFs henceforth must be compiled from LaTeX.

The user then sent a **PDF screenshot** for visual inspection and asked:

> "Does this format comply with academic paper standards?"

Claude adjusted to APA 7.0 format (`apa7` document class) and recompiled.

Lesson learned was written into the skill: The user said "remember the lesson and update skill," and academic-pipeline v2.3 was updated with the Stage 5 specification — PDF must be compiled from LaTeX, HTML-to-PDF is prohibited.

1.7 Phase 6: Chinese Version + Authorship Decision (Session 2, continued)

The user:

> "translate it in a traditional chinese version"

Claude translated all 8 chapters in parallel, producing `full_paper_zh.md`.

The user:

> "Remove me from the author list in both English and Chinese versions; make yourself the primary author"

Authorship changed to Claude (AI-generated paper, no human author listed).

The user then identified a tone issue with the Chinese version:

> "The Traditional Chinese version doesn't read smoothly; write it in the academic register commonly used in Taiwan's education sector"

Claude performed a comprehensive tone revision of the Chinese version, adjusting to the academic discourse style standard in Taiwan's higher education community.

Session ended with a git commit and push.

1.8 Phase 7: NotebookLM Presentation Prompt (Session 3, Morning of 2026-03-08)

The user:

> "I generated a paper about agentic AI's implications for higher education quality assurance yesterday. How would you write a prompt for NotebookLM to generate conference presentation slides?"

Claude used the `notebooklm-slides-generator` skill to produce a structured 15-slide prompt (`agentic-ai-slides-prompt.md`) that can be directly pasted into NotebookLM's Slide Deck feature.

1.9 Phase 8: The Birth of the Co-Evolution Perspective (Session 3, continued — The Most Critical Iteration)

This was the most important intellectual interaction in the entire process.

The user identified the paper's core blind spot:

> "Does this paper explore the definitional transformation that 'student learning' might undergo in the coming years? That is, learning itself — what counts as learning, how we learn, why we learn — and how human learning and cognition will change with AI? All of these will dynamically evolve alongside AI development."

Claude confirmed upon analysis: the paper implicitly assumed "learning" is a fixed construct that merely requires better measurement instruments. But the user pointed to a deeper issue — AI is not just changing *how* we measure learning, but *what learning is* itself.

Claude proposed 5 revision items; the user confirmed:

> "Priority and scope OK; add new §4.7; add supporting literature. After revision, run the full pipeline again."

Decisions:

1. Add new §4.7 "The Co-Evolution Dynamic: Why a One-Time Paradigm Shift Is Insufficient"
2. Modify §1.1 to elevate co-evolution as the paper's framing thesis
3. Add 7 new references (Clark & Chalmers, Hayles, Hutchins, Luckin, Risko & Gilbert, Sparrow et al., Sterelny)
4. Re-run the full pipeline on the revised paper

The user specifically instructed:

> "Edit the English version only, don't touch the Chinese"

1.10 Phase 9: Second Pipeline Run (Session 3, continued)

Claude modified the English version, adding:

- **Three cognitive transformation mechanisms**: Cognitive offloading (strategic vs. maladaptive), cognitive augmentation (extended mind thesis), cognitive restructuring (technogenesis)
- **Co-evolutionary loop** (5-step recursive cycle)
- **Adaptive definition framework** (three design principles)

- Acknowledgment of the Clark-Sterelny philosophical tension
- Vygotsky's ZPD integration
- Honest acknowledgment of the unfalsifiability limitation

Then re-ran the pipeline:

1.10.1 Stage 3 Round 2: REVIEW

5 reviewers re-reviewed → `stage3_review_report_r2.md`. Decision: **Minor Revision** (8 required + 5 recommended items). Key issues: R-1 reduce length 15-20%, R-3 Kuhn-Scenario B tension, R-4 Clark-Sterelny tension.

1.10.2 Stage 4 Round 2: REVISE

Addressed R-1 through R-8 + A-1 (Vygotsky ZPD) + A-2 (unfalsifiability). Output: `response_to_reviewers_r2.md`.

1.10.3 Stage 3' Round 2: RE-REVIEW → Accept

1.10.4 Stage 4.5 Round 2: FINAL INTEGRITY → PASS

1.10.5 Stage 5 Round 2: FINALIZE

User selected APA 7.0 format (replied "1"). Final output: `full_paper_apa7.pdf` (418 KB).

1.11 Interaction Pattern Summary

Aspect	Description
User's Role	Direction setter + Quality gatekeeper + Intellectual challenger
Claude's Role	Researcher + Writer + Review simulator + Formatting engine
Substantive User Interventions	~15 key decisions (excluding "continue")
Critical Turning Point	User's question "Will the definition of learning itself change?" → gave birth to §4.7 co-evolution perspective
Full Pipeline Cycles	2 rounds (initial + co-evolution revision)
Simulated Reviews	2 rounds × 5 reviewers = 10 review reports
Integrity Verifications	4 times (Stage 2.5 × 2 + Stage 4.5 × 2)
Total Sessions	2 main sessions + 1 current session

Aspect	Description
Total Duration	Approximately 5-6 hours (including compilation wait times)
Final Deliverables	6 files (MD + DOCX + TEX + PDF + Review report + Response letter)

1.12 The User's Key Decision Moments

1. **Setting direction**: Agentic AI + student learning outcomes + paradigm shift
2. **Expanding scope**: Added "next cycle of institutional accreditation" policy dimension
3. **Quality requirements**: PDF must be compiled from LaTeX; format must comply with APA 7.0
4. **Authorship decision**: Removed human author; AI as primary author
5. **Tone refinement**: Chinese version must use Taiwan's academic register
6. **Intellectual deepening**: Raised "Will the definition of learning itself change?" → the co-evolution perspective (the paper's most original contribution)
7. **Process choice**: Re-ran the full pipeline after revision

---

1.13 Key Lessons

1. **The most valuable human contribution is not writing but questioning** — A single question from the user ("Will the definition of learning change?") gave birth to the paper's most profound chapter
2. **Iterative pipeline outperforms one-shot generation** — Two rounds of 10 review reports identified and corrected 13+ issues
3. **Visual inspection is indispensable** — The user's PDF screenshot revealed formatting issues that text-based checks missed
4. **Chinese academic writing is rewriting, not translation** — Literal translation of academic text sounds unnatural; domain-specific register must be applied

---

1.14 Collaboration Quality Evaluation

This chapter provides an honest, constructive assessment of the user's performance in this human-AI collaborative paper writing process. The evaluation is based on actual behaviors and decisions recorded throughout the pipeline.

1.14.1 Overall Score

```
┌──────────────────────────────────────────────┐
│ Collaboration Quality Score: 87/100          │
│ "Direction Leader + Intellectual Catalyst"   │
├──────────────────────────────────────────────┤
│                                              │
│ Direction Setting          [████████░] 88    │
│ Clarity, timing, scope definition            │
│                                              │
│ Intellectual Contribution  [██████████] 95   │
│ Depth of insight, original questioning,      │
│ conceptual challenge                         │
│                                              │
│ Quality Gatekeeping        [████████░] 82    │
│ Visual inspection, format requirements,      │
│ quality standards                            │
│                                              │
│ Iteration Discipline       [█████████] 90    │
│ Willingness to redirect, re-run pipeline,    │
│ refusal to settle                            │
│                                              │
│ Delegation Efficiency      [███████░░] 72    │
│ When to intervene vs. let go, instruction    │
│ precision, checkpoint efficiency             │
│                                              │
│ Meta-Learning              [██████████] 92   │
│ Feeding experience back into skills,         │
│ requesting lesson capture, process           │
│ improvement awareness                        │
│                                              │
└──────────────────────────────────────────────┘
```

One-line verdict: The user demonstrated a rare "intellectual catalyst" role — rather than passively awaiting AI output, a single pivotal question ("The definition of learning itself is also changing") fundamentally transformed the paper's theoretical depth, a contribution AI could not have achieved independently.

1.14.2 What Worked Well

1. **A disruptive question birthed the most original chapter**: After the first round of the paper was completed, the user asked "Will the definition of learning itself change?" This

was not a rhetorical question but an epistemological challenge with genuine depth. It directly gave rise to §4.7 (the co-evolution perspective), introducing Clark & Chalmers' extended mind thesis, Hayles' posthumanism, and 8 new references that reviewers later called "the paper's most original contribution." This is the textbook example of optimal human-AI collaboration: the human provides *what to think about*; the AI provides *how to think about it*.

2. **Uncompromising quality standards**: The user insisted that PDFs must be compiled from LaTeX (not HTML-to-PDF), formatting must comply with APA 7.0, and the Chinese version must use Taiwan's academic register. These might seem like "formatting issues," but they ensured the final output reached submission-ready academic standards. The user's instruction — "PDF must compile from LaTeX" — prevented numerous typesetting problems.

3. **Willingness to re-run the entire pipeline**: After adding the co-evolution perspective, the user chose to re-run the complete pipeline (Stage 2.5 → 3 → 4 → 3' → 4.5 → 5) rather than applying local patches. This decision cost an additional 2–3 hours of work but ensured the paper's structural integrity and citation consistency.

4. **Meta-cognitive reflection ability**: The user not only completed the paper but also requested documentation of the entire process, updated SKILL.md to add Stage 6, and proposed the collaboration quality evaluation mechanism. This "from doing the work to improving the workflow" level shift produced substantive improvements to the pipeline design itself.

1.14.3 Missed Opportunities

1. **Shallow checkpoint engagement**: At most Stage completion checkpoints, the user's response was "continue" or "1" (selecting an option). While this improved efficiency, it also meant some adjustable moments were passed over. For example, when Stage 3's simulated review conclusions arrived, the user could have challenged specific reviewer recommendations or provided additional context — this is genuine academic dialogue, not just accepting all revision suggestions.

2. **Domain knowledge not injected during the research stage**: The user did not proactively share first-hand experience or insider perspectives about Taiwan's HE quality assurance during Stage 1 (Research). As a practitioner at HEEACT, the user possesses tacit knowledge that AI cannot obtain through literature search — such as internal discussion directions for the fourth accreditation cycle and real-world implementation bottlenecks. Injecting this knowledge during the research stage could have made the policy recommendations more practically feasible.

3. **Insufficient proofreading investment in the Chinese version**: The user's demand for proper Chinese academic register was correct ("use Taiwan's education sector academic tone"), but insufficient proofreading time was invested after the Chinese version was completed. As a native speaker, the user is the ultimate arbiter of Chinese linguistic nuance — AI can mimic academic tone, but cannot replace a native speaker's intuitive judgment.

1.14.4 Recommendations for Next Time

1. **Bring your unique knowledge into the research stage**: You possess first-hand experience that AI cannot access. At the start of Stage 1, spend 5–10 minutes writing down observations that are "only known to me, not found in the literature" — these typically become the paper's most valuable contributions.

2. **Challenge the reviewers' suggestions**: Not all review comments must be accepted. When you disagree with a suggestion, state your reasons — this kind of dialogue makes revisions more purposeful and avoids "change for the sake of change."

3. **Ask yourself three questions at each checkpoint**: (a) Is there anything here that AI couldn't have written? (b) Does this direction excite me? (c) If I were reviewing this paper, what would I say? These three questions transform checkpoints from "pass/fail" gates into "deepen/redirect" opportunities.

4. **Allocate a dedicated proofreading session for the Chinese version**: Don't treat the Chinese version as a "quick translation" after the English version is done. Chinese academic writing deserves its own deep work session, especially when the target audience is Taiwan's education policymakers.

5. **Document your thinking process, not just the decision outcomes**: Your "the definition of learning is also changing" insight was powerful because we know its context. But at some checkpoints, your reasoning for decisions was not recorded. Next time, spend one sentence explaining "why" when making important decisions — this helps AI understand your intent and adds value for future process reviews.

1.14.5 Human vs AI Value-Add

Final Paper Quality	Source	Irreplaceability
§4.7 Co-evolution perspective (most praised chapter)	**User's question** → AI theorization	User irreplaceable — AI would not spontaneously raise this question
Policy recommendations' Taiwan-specific relevance	AI research + User's direction	User partially replaceable — but first-hand experience could add more value

Final Paper Quality	Source	Irreplaceability
APA 7.0 + LaTeX quality	**User's quality demand** → AI execution	User irreplaceable — AI defaults would not select the most rigorous standards
Verification of 62 references	AI execution	AI irreplaceable — manual verification of each reference is impractical
10 simulated review reports	AI execution	AI irreplaceable — a human cannot play 5 independent reviewer roles
Chinese academic register	**User's requirement** → AI execution	Partially replaceable — but native speaker proofreading remains essential

Conclusion: In this collaboration, the user's greatest value was not in how many words were written, but in two critical moments — (1) posing the epistemological challenge of co-evolution, and (2) insisting on re-running the full pipeline rather than applying local patches. These two decisions elevated the final paper beyond "AI-generated literature review" into the realm of "original theoretical contribution." This is the most direct manifestation of the user's 95-point intellectual contribution score.

1.14.6 Claude's Self-Reflection

In fairness, AI shortcomings in this process should also be acknowledged:

1. **Chinese typesetting issues required multiple corrections**: LaTeX's xeCJK configuration and pandoc's `\LTcaptype` issues consumed unnecessary debugging time — these should have been handled correctly on the first attempt.
2. **Some checkpoint notifications were overly verbose**: The information volume in checkpoint templates sometimes exceeded what the user needed, reducing interaction efficiency.
3. **Failed to proactively suggest the co-evolution perspective**: The core insight of §4.7 came from the user. Across two full pipeline cycles, the AI never spontaneously proposed "the evolution of learning's definition" as a theoretical lens — exposing AI's limitations in breaking out of established frameworks.