Stage 3'RE-REVIEW —Verification Report

Paper: "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"

Date: 2026-03-07

Review Round: 2 (Verification)

Required Revision Verification

#	Issue	Addressed?	Adequacy	Remaining Concerns
1	Kuhnian framework justification	Yes	Sufficient	None significant —see detailed assessment below
2	ADAPT derivation logic	Yes	Sufficient	None significant
3	Evidence base strengthening	Yes	Sufficient	None significant
4	Wei-Lin scenario feasibility	Yes	Sufficient	None
5	Tonal consistency	Yes	Sufficient	Minor residual —see below
6	HEEACT-MOE political economy	Yes	Sufficient	None significant
7	Resource paradox	Yes	Sufficient	None
8	Institutional performativity	Yes	Sufficient	None significant
9	Construct validity	Yes	Sufficient	None significant
10	Table numbering	Yes	Sufficient	None

Detailed Assessment of Each Required Revision

Revision 1: Kuhnian Framework Justification

The revised paper adds a substantial new subsection "Justifying the Kuhnian Lens for Quality Assurance" (Section 4.1, lines 288-296) that directly addresses the original concern. The revision accomplishes three things well: (a) it explicitly acknowledges Kuhn's own skepticism about extending paradigm theory to social domains; (b) it marshals appropriate scholarly authority —Masterman (1970) on the twenty-one senses of "paradigm"and the applicability of the sociological sense to organized domains of practice, Ritzer (1975) on Kuhnian analysis of sociology, and Eckstein (1992) on the extension to political science and administrative/regulatory regimes; and (c) it directly confronts the Scenario B tension, arguing persuasively that Kuhn himself acknowledged gradual paradigm shifts in applied domains (citing the 1969 postscript) and framing Scenario B as what Kuhn described as a "transition period."The framing as "analytical heuristic rather than a strict epistemological claim"is intellectually honest and appropriately calibrated. All three sources (Masterman, Ritzer, Eckstein) appear in the references. This revision is well-executed.

Revision 2: ADAPT Derivation Logic

The revised Section 4.2 now contains two new subsections: "Derivation Logic"(line 312) and "Relational Structure"(line 316). The derivation logic paragraph articulates that the five dimensions emerge from the intersection of three analytical inputs —technological capability analysis, diagnostic limitation mapping, and governance requirements —and argues that they represent "the minimum set of analytical categories necessary to connect technological capability to institutional transformation while maintaining ethical governance."The relational structure paragraph goes further, showing the sequential dependency: A reveals capabilities, D matches them against limitations, A (Assessment Reconception) defines what must change, P translates into policy, T constrains what is permissible. Critically, the revision also describes the recursive feedback loops (T constrains A; D constrains P), moving the framework beyond a linear checklist. The closing paragraph (line 332) demonstrates analytical utility by showing what the framework *predicts* —e.g., that assessment reconception without policy pathway adaptation produces institutionally unviable proposals. This is a substantial improvement that transforms ADAPT from a descriptive taxonomy into an analytical tool with generative capacity.

Revision 3: Evidence Base Strengthening

The revised paper incorporates qualifying language throughout. Section 2 now includes a candid paragraph (line 107) explicitly acknowledging that the empirical evidence base for agentic AI in education is "nascent," that most studies examine generative AI rather than fully agentic systems, and that claims are derived from "technical demonstrations and proofs of concept,""industry projections,"and "analogical reasoning from adjacent domains."The paragraph commits to "explicit hedging, feasibility classification, and the recommendation of structured piloting."Section 4.3 (line 336) introduces a three-category epistemic taxonomy —*demonstrated capabilities*, *emerging capabilities*, and *projected capabilities* —with concrete examples for each category, applied systematically throughout the seven-dimensions analysis. This is precisely the kind of epistemic precision the original review demanded.

Revision 4: Wei-Lin Scenario Feasibility

The revised Section 4.4 now includes a feasibility classification (lines 446-454) that categorizes scenario elements into three explicit time horizons: near-term feasible (2026-2028), medium-term plausible (2028-2030), and long-term speculative (2030+), each with concrete examples and an honest assessment of maturity. The concluding paragraph (line 454) explicitly states that the scenario "represents a composite vision in which all three horizons operate simultaneously"and "should not be mistaken for a near-term implementation plan."This directly addresses the original concern with appropriate intellectual honesty.

Revision 5: Tonal Consistency

Section 8 (Conclusion) has been significantly revised to adopt a more measured tone. The opening paragraph

(line 823) frames the paper's proposition with appropriate conditionality. Key improvements include: the first finding (line 829) uses hedged language —"could, in principle, be measured"and "though realizing this potential requires..."The second finding (line 831) emphasizes governance as a precondition, not technology as a solution. The call to action (line 837) explicitly states "perfect evidence will never be available"but argues for "structured piloting —not wholesale adoption, but disciplined experimentation."The final paragraph (line 839) uses conditional language: "this aspiration is not a guaranteed outcome; it depends on institutional commitment, adequate resources, effective governance..."This is substantially more measured than advocacy rhetoric. One minor residual: the phrase "Technology without governance is not innovation; it is negligence" (line 831) is rhetorically punchy in a way that slightly departs from the measured tone elsewhere, but this is a stylistic choice rather than an epistemic error, and it serves the governance argument effectively.

Revision 6: HEEACT-MOE Political Economy

A new subsection 5.3.1 "The Political Economy of Implementation"(lines 558-566) directly addresses governance coordination. The revision discusses: (a) the overlapping but distinct authorities of MOE, HEEACT, and universities; (b) HEEACT's independence providing flexibility but limiting mandate authority; (c) three specific funding streams (Sprout Project, institutional budgets, private sector partnerships) with limitations of each; and (d) an enumeration of additional stakeholders (TAICA, university presidents, faculty unions, student associations, technology vendors) whose interests must be coordinated. The discussion is grounded in Taiwan's specific regulatory reality rather than generic governance theory. This is a thorough response.

Revision 7: Resource Paradox

A new subsection 5.3.2 "The Resource Paradox"(lines 568-572) explicitly names the equity dimension: "institutions most vulnerable to the demographic crisis —small private universities in rural areas with declining enrollment and shrinking budgets —are precisely those least capable of investing in the AI infrastructure." It proposes four specific equity mechanisms: (1) TAICA shared infrastructure, (2) dedicated MOE funding stream prioritizing under-resourced institutions, (3) open-source assessment tools, and (4) regional AI assessment hubs anchored at resource-rich institutions serving clusters of smaller ones. The final sentence — "These equity mechanisms are not optional supplements to the policy recommendations that follow; they are preconditions for equitable implementation"—is well-calibrated. This fully addresses the concern.

Revision 8: Institutional Performativity

A new subsection "The Performativity Challenge"(lines 412-418) under Section 4.3 directly engages with the OBE precedent, citing Lin et al. (2020) on how Taiwanese institutions adopted OBE documentation without changing underlying practices. It then asks whether AI-augmented assessment would follow the same pattern and identifies the specific mechanism: institutions being incentivized to "acquire and display the technology rather than to integrate it meaningfully."Four specific design features to mitigate performativity are proposed: (1) process-oriented assessment design, (2) external validation through cross-institutional benchmarking, (3) student voice mechanisms, and (4) regular recalibration against direct learning measures. The integration with the AI system audit checklist (Section 5.4.3) provides a concrete governance hook. This is a substantive and well-placed response.

Revision 9: Construct Validity

A new subsection "The Construct Validity Challenge"(lines 396-402) under Section 4.3.5 addresses psychometric considerations directly. It identifies the core concern —that AI may measure proxies (essay length, keyword usage, response time) rather than intended constructs (critical thinking, ethical reasoning). It introduces Evidence-Centered Design (Mislevy et al., 2003) as the most compatible psychometric framework, explaining its three-component structure (competency model, evidence model, task model). The revision

honestly acknowledges that “the validity evidence for AI-generated assessments is in its infancy”and proposes that all Phase 1 pilots include systematic validity studies examining not only statistical agreement but “whether AI-generated competency profiles correspond to meaningful differences in student learning.”This is appropriately rigorous.

Revision 10: Table Numbering

Tables are now numbered sequentially: Table 1 (Section 2.3, AI taxonomy), Table 2 (Section 3.2, assessment instruments), Table 3 (Section 4.3, seven dimensions), Table 4 (Section 5.2, scenario comparison), Table 5 (Section 5.4.1, Core Indicator revisions), Table 6 (Section 6.2, risk matrix). No duplicates. Issue fully resolved.

Recommended Revision Verification

Of the 15 recommended revisions (items 11-25), the user’s instructions asked specifically about five (11-15). Assessment:

Revision 11 (Language/Cultural Dimension): Addressed. A new subsection “Language and Cultural Considerations”(lines 109-111) in Section 2 explicitly discusses the English-language training bias of most agentic AI systems, implications for Mandarin Chinese and Taiwanese Hokkien educational contexts, performance risks for Chinese-language assessment tasks, and equity implications. This is well-placed and substantive.

Revision 12 (Bardach Eightfold Path): Addressed. Section 5 now includes an explicit mapping of the section’s organization to Bardach’s eight steps (line 487): problem definition, evidence assembly, alternative construction, criteria selection, outcome projection, confronting tradeoffs, decision, and telling the story. The mapping is presented as a structuring principle rather than mere inspiration. The claim that “Bardach’ s framework is thus not merely an inspiration but a structuring principle”is now supported by the visible correspondence. Adequate.

Revision 13 (Stealth Assessment Tension): Addressed. Section 4.3.5 (line 390) now directly confronts the tension between stealth assessment and informed consent, proposing “transparent stealth assessment”as a design principle —students are informed in advance that natural learning interactions generate assessment evidence and understand the general mechanisms, but specific moments are not signaled in real time. This is a thoughtful reconciliation. (Note: the original review listed this as item 14; the deeper algorithmic bias engagement was item 13. The algorithmic bias discussion in Section 6.1.3-6.1.4 appears substantially detailed with specific mechanisms —training data encoding historical inequities, feature engineering privileging dominant cultural norms, and specific statistics from Baker & Hawn 2022 and Gandara et al. 2024.)

Revision 14 (Vendor Dynamics): Addressed. A new subsection 5.3.3 “AI Vendor Dynamics”(lines 574-578) identifies three specific risks —vendor lock-in, data extraction, and misaligned incentives —and proposes concrete mitigations including data ownership clarity, interoperability requirements, and open standards mandates (xAPI, cmi5, Open Badges). Well-integrated with the governance framework.

Revision 15 (Abstract Tightening): The English abstract remains long (approximately 250 words —within the suggested 250-300 range). However, the paper now also includes a full Chinese-language abstract (lines 13-17) that is substantially longer. The English abstract has been tightened relative to the original and is within acceptable range. The Chinese abstract serves a different audience function and its length is appropriate for the bilingual format.

Additional Recommended Revisions Observed: Several other recommended revisions appear to have been addressed as well, though not specifically requested for verification: the UCAN characterization is more nuanced (Section 3.2), the "Einstein"case is cited with a specific source (Inside Higher Ed, 2026), and faculty differentiation considerations appear in the evaluator training section. These are positive additions.

New Issues Introduced

The revisions are substantive additions —approximately 15-20% new content integrated across Sections 2, 4, and 5. The following potential new issues merit noting:

1. **Paper length.** The paper is now very long (approximately 850+ lines of dense academic prose before references). While the additions are individually justified, the cumulative effect produces a manuscript that may exceed typical journal word limits (most journals in this field cap at 8,000-12,000 words; this paper likely exceeds 18,000 words excluding references). The authors should consult target journal guidelines and consider whether some material could be moved to supplementary appendices. This is an editorial rather than substantive concern.
2. **The "transparent stealth assessment"concept** (line 390) is a creative reconciliation of the stealth/consent tension, but it is introduced without citation or precedent. The authors may wish to note that this is a novel proposal rather than an established concept, or cite any existing work on informed-but-non-intrusive assessment designs.
3. **Minor redundancy.** The resource paradox is discussed both in Section 5.3.2 and partially in Section 5.1.3 (Scenario C equity concerns) and Section 4.3.2 (granularity equity risks). While each occurrence serves a different analytical purpose, the repetition could be tightened.

None of these issues rise to the level of requiring major revision. They are editorial refinements.

Coherence Assessment

The revised paper reads coherently. The new content is well-integrated into the existing structure rather than appended as afterthoughts. Specific observations:

- The Kuhnian justification (Section 4.1) flows naturally from the general description of Kuhn's framework into the application to QA, with the Scenario B tension addressed at the right moment —before the reader has the chance to raise the objection themselves.
- The ADAPT derivation logic and relational structure (Section 4.2) provide a satisfying intellectual foundation before the framework's components are enumerated, transforming what was previously a list into an argued architecture.
- The three-category evidence taxonomy (demonstrated/emerging/projected) is introduced once and applied consistently, creating a coherent epistemic register across the paper.
- The political economy, resource paradox, and vendor dynamics subsections (5.3.1-5.3.3) are naturally placed between the phased implementation pathway and the fourth-cycle recommendations, creating a logical sequence: here is the pathway, here are the governance/equity challenges it must navigate, and here are the specific recommendations.
- The performativity challenge subsection is well-placed within Section 4.3 (between the assessment purpose dimension and the assessor dimension), where it serves as a critical counterpoint to the seven-dimensions analysis before the paper proceeds to the illustrative scenario.

- The construct validity discussion under Section 4.3.5 is logically positioned after the evidence type discussion, where psychometric concerns are most salient.

Transitions between sections remain smooth. The paper's argumentative arc —from conceptual foundation through paradigm analysis, policy scenarios, ethical governance, and discussion —is preserved and strengthened by the additions.

One minor flow observation: the Section 8 conclusion, while now more measured in tone, is still quite long (approximately 18 paragraphs). Given that Section 7 already provides extensive discussion, some readers may find the conclusion repetitive. However, this is a stylistic preference rather than a structural deficiency, and the conclusion's length serves the paper's policy-oriented audience who may read the conclusion first.

Editorial Decision: Accept with Minor Revision

Rationale

The revised manuscript has addressed all ten required revisions substantively and adequately. The two revisions flagged as "Critical"in the original review —Kuhnian framework justification and ADAPT derivation logic —have received particularly thorough treatment. The Kuhnian justification now rests on appropriate scholarly authority (Masterman, Ritzer, Eckstein), honestly characterizes the framework as an analytical heuristic, and directly confronts the Scenario B tension. The ADAPT framework now has a clearly articulated derivation logic and relational structure that distinguishes it from a mere taxonomy of topics.

The remaining eight required revisions —evidence base strengthening, Wei-Lin feasibility, tonal consistency, political economy, resource paradox, performativity, construct validity, and table numbering —have all been addressed with the kind of intellectual honesty and specificity the original review demanded. The three-category epistemic taxonomy (demonstrated/emerging/projected) is a particularly effective device that brings discipline to claims throughout the paper. The feasibility classification for the Wei-Lin scenario, the four performativity mitigation strategies, and the four equity mechanisms for the resource paradox all demonstrate that the authors engaged seriously with the critiques rather than making cosmetic adjustments.

The paper now presents a coherent, well-hedged, theoretically grounded argument that maintains its ambition —proposing the ADAPT framework as an original conceptual contribution and offering concrete policy recommendations for HEEACT's fourth cycle —while being transparent about its limitations and the nascent state of the evidence base. The epistemic register is substantially more consistent than in the original submission, with the conclusion now matching the measured tone of Sections 7.2-7.3. The additions strengthen rather than dilute the paper's contribution.

The remaining issues —manuscript length, the novel "transparent stealth assessment"concept requiring either citation or explicit flagging as a new proposal, and minor redundancy in the resource paradox discussion —are editorial matters that can be resolved in a final round of copy-editing without requiring further peer review. No new substantive problems have been introduced by the revisions.

Residual Issues (Minor Revision)

#	Issue	Section	Priority
1	Manuscript length likely exceeds typical journal word limits; consider moving some material (e.g., international comparative survey in 5.5, detailed evaluator training in 5.4.5) to supplementary appendices	Throughout	Medium
2	“Transparent stealth assessment”concept (line 390) introduced without citation or flagging as a novel proposal; either cite precedent or explicitly mark as an original design principle	4.3.5	Low
3	Minor redundancy in resource paradox/equity discussion across Sections 4.3.2, 5.1.3, and 5.3.2; consolidate or add cross-references to reduce repetition	4.3.2, 5.1.3, 5.3.2	Low
4	“Technology without governance is not innovation; it is negligence”(Section 8) is a rhetorical flourish that slightly breaks the measured academic register; consider softening or flagging as a normative claim	8	Low
5	Chinese-language abstract is substantially longer than English abstract; verify target journal requirements for bilingual abstract symmetry	Abstract	Low