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

Claude (Anthropic)

Anthropic

Author Note

This paper was authored by Claude, an AI assistant developed by Anthropic. The research design, analytical frameworks, policy recommendations, and ethical arguments were generated by Claude using the Academic Research Skills pipeline. All claims are grounded in cited literature; no original empirical data were collected.

This work was commissioned and directed by a human researcher who provided the research topic, scope, and iterative feedback throughout the pipeline.

Abstract

Taiwan's higher education quality assurance system relies on periodic accreditation cycles, document-based self-assessment, and indirect measurement instruments that capture artifacts of learning rather than its substance. This paper examines how agentic AI — systems capable of autonomous planning, persistent memory, and adaptive reasoning — could enable a paradigm shift from retrospective institutional snapshots to continuous individual learning trajectories. Crucially, it argues that AI is not only transforming *how* we measure learning but *what learning is*: the co-evolution of AI and human cognition creates a "double moving target" requiring adaptive, not fixed, assessment architectures. The analysis employs the ADAPT framework (Assessment-Design for Agentic Paradigm Transformation), an original conceptual model integrating five dimensions — Agency Architecture, Diagnostic Mapping, Assessment Reconception, Policy Pathways, and Trust & Ethics Safeguards — grounded in Kuhnian paradigm shift analysis, Bardach's eightfold policy path, and principlist ethics. Six structural limitations in Taiwan's current quality assurance architecture function as Kuhnian anomalies; three policy scenarios are evaluated, with a phased implementation pathway (2026–2030+) recommended for HEEACT's fourth accreditation cycle. A four-principle ethical analysis identifies eight systemic and five agentic-AI-specific risks, addressed through a three-tier governance framework. Taiwan's mature quality assurance ecosystem, AI Basic Act (2025), and INQAAHE standing position it to lead — provided it builds assessment architectures capable of evolving alongside the learners they measure.

摘要

台灣高等教育品質保證體系仰賴週期性評鑑、文件式自評及間接測量工具，所捕捉的是學習的表象產出而非實質內涵。本文探討代理型人工智慧（agentic AI）如何促成學生學習成效測量的典範轉移——從回溯性的機構快照邁向連續性的個人學習軌跡。本文進一步主張，AI 不僅改變測量方式，更改變「學習」本身的定義：AI 與人類認知的共同演化形成「雙重移動靶」（double moving target），要求評量架構必須具備持續適應的能力，而非固定不變。本研究提出 ADAPT 架構（Assessment-Design for Agentic Paradigm Transformation）作為原創性概念貢獻，整合五大面向——能動性架構、診斷對應、評量再概念化、政策路徑、信任與倫理防護——並以孔恩典範轉移分析、Bardach 政策分析法及原則主義倫理評估為基礎。分析辨識出台灣現行品保架構六項結構性限制作為孔恩「異例」，評估三種政策情境後建議採分階段實施路徑（2026–2030+），配合 HEEACT 第四週期校務評鑑之設計窗口。四原則倫理分析辨識出八項系統性與五項代理型 AI 特有風險，提出三層級治理架構因應。台灣成熟的品保生態系統、2025 年《人工智慧基本法》及 INQAAHE 國際地位，使其有條件引領 AI 增強之品質保證——前提是建構能隨學習者共同演化的評量架構。

Keywords: agentic AI, student learning outcomes, paradigm shift, higher education accreditation, HEEACT, quality assurance, Taiwan

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

Introduction

The Measurement Problem

In Taiwan's higher education system, the primary external evidence used to determine whether approximately 900,000 students are learning effectively is collected once every six years. An institution prepares a self-assessment report, assembles documentary evidence of its educational practices, and submits both to a panel of peer reviewers who visit the campus for one to two days. On the basis of this evidence — curated by the institution itself, compiled months before the visit, and evaluated over a period measured in hours — the institution receives an accreditation judgment that will stand for the next six years (HEEACT, 2023a). During those six years, curricula will change, faculty will turn over, labor markets will shift, and entire fields will be reshaped by technological disruption. The accreditation judgment, once rendered, will capture none of it.

This temporal architecture would be unremarkable if learning itself were static — if the competencies students needed in 2024 were the same ones they would need in 2030, and if the methods for developing those competencies changed only incrementally. But we live in an era of radical discontinuity. Generative artificial intelligence emerged as a mass technology in late 2022; within eighteen months, it had reshaped professional practice in law, medicine, software engineering, journalism, and education itself (UNESCO, 2023). Agentic AI — systems that can autonomously plan, execute, adapt, and remember — arrived shortly after, introducing a category of technology that does not merely respond to human prompts but pursues goals, orchestrates tools, and operates across extended time horizons with minimal human direction (Arunkumar et al., 2026; Masterman et al., 2024). The world in which Taiwan's accreditation framework was designed — a world of incremental change, stable competency requirements, and human-only assessment — is receding.

The paradox is stark. We now possess technologies capable of continuous, real-time, individual-

ized measurement of human learning across multiple modalities — technologies that can track a student's developing competencies not as a six-year retrospective summary but as a living, evolving trajectory. Yet we continue to measure learning with instruments that produce snapshots: point-in-time surveys, aggregate statistics, and curated documentary narratives. The gap between what is measurable and what is measured constitutes not merely a technical limitation but a structural failure — one that carries consequences for students whose learning goes unrecognized, for institutions whose quality goes unmeasured between accreditation cycles, and for a nation whose economic competitiveness depends on a higher education system that can adapt faster than its assessment architecture currently permits.

Yet this paradox has a deeper structural layer that constitutes the central thesis of this paper. The framing above implicitly treats "learning" as a stable construct — a fixed target that merely requires better measurement instruments. In reality, AI is not only transforming *how* we measure learning; it is transforming *what learning is*. When students collaborate with AI systems to think through problems, cognitive processes extend beyond the biological brain into technological systems (Clark & Chalmers, 1998). When information retrieval is offloaded to AI, human memory shifts from storing facts to managing knowledge sources and evaluating epistemic reliability (Sparrow et al., 2011). When AI assumes routine cognitive tasks, "knowing" migrates from individual possession of knowledge to the capacity to effectively mobilize human-AI collaborative networks (Hutchins, 1995). The measurement problem is therefore a *double moving target*: not only do the instruments need updating, but the object of measurement — the very definition of learning, the nature of knowledge, the boundaries of cognition — is itself being continuously reconstructed by the technologies we deploy. This co-evolutionary dynamic — AI changes learners, changed learners reshape how AI should be adopted, which further changes learners — is the thread that runs through the entire analysis that follows, from the Kuhnian paradigm shift theorized in Section 4 through the policy recommendations in Section 5 to the adaptive assessment architecture proposed in Section 4.7. Any reform agenda that addresses only "how to measure better" while ignoring "what we are measuring is changing" remains structurally incomplete.

Taiwan's Higher Education Context

Taiwan's higher education system encompasses more than 140 institutions — national universities, private universities, and universities of technology — serving a student population that has been declining steadily as a consequence of one of the world's lowest birth rates (MOE, 2025). The system is governed by the Ministry of Education (MOE) and quality-assured externally by the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT), an independent foundation established in 2005 that holds full compliance with the International Network for Quality Assurance Agencies in Higher Education (INQAAHE) International Standards and Guidelines and recognition by the Council for Higher Education Accreditation (CHEA) in the United States (HEEACT, 2023a; Lin et al., 2021). HEEACT administers both institutional accreditation, mandatory for all higher education institutions on a six-year cycle, and program accreditation, voluntary since a 2017 policy shift toward institutional autonomy (HEEACT, 2024).

The urgency of the present analysis is driven by three converging forces. First, the demographic crisis has reached existential proportions: in 2024 alone, seven universities ceased operations, and the MOE projects that up to 40 private universities may close by 2028 as the cohort of students entering higher education continues to shrink (Sharma, 2024; Taiwan News, 2024; MOE, 2025). In a contracting system, the quality assurance question transforms from "does this institution meet minimum standards?" to "can the assessment system identify, quickly enough to protect students, which institutions are delivering sufficient learning value to justify continued operation?" A six-year accreditation cycle is structurally incapable of providing that kind of early warning.

Second, Taiwan has moved decisively into the AI governance arena. The AI Basic Act (人工智慧基本法), enacted in December 2025, articulates seven governing principles — human autonomy, privacy protection, transparency, fairness, safety, accountability, and sustainability — that provide a legal foundation for responsible AI deployment across sectors, including education (Legislative Yuan, 2025). The Taiwan AI College Alliance (TAICA), launched in 2024 with 55 member universities, has begun developing shared AI curricula. The Taiwan Adaptive Learning Platform (TALP,

因材網) has integrated generative AI capabilities into its nationwide adaptive learning infrastructure (MOE, 2025). These developments signal that Taiwan's policy environment is increasingly receptive to AI integration in education — but the quality assurance framework has not yet adapted to evaluate, much less leverage, AI-driven assessment.

Third, and most consequentially, the design window for the fourth cycle of institutional accreditation is imminent. The third cycle concludes in academic year 2025; the decisions made in the next two to three years about the fourth cycle's standards, indicators, evidence requirements, and evaluation protocols will determine whether Taiwan's quality assurance architecture enters the AI era by design or by default. This paper argues that a deliberate, evidence-informed, and ethically grounded approach to integrating agentic AI into the fourth-cycle framework can position Taiwan as a regional leader in AI-augmented quality assurance — while a passive approach risks leaving the framework increasingly disconnected from the educational reality it purports to assure.

Agentic AI: Beyond Chatbots

The public discourse on artificial intelligence in education has been dominated since late 2022 by generative AI — specifically, large language models such as ChatGPT that can produce human-like text, answer questions, and assist with writing tasks. This discourse, while important, has obscured a more consequential technological development: the emergence of agentic AI systems that differ from generative AI not merely in degree but in kind.

A generative AI tool responds to prompts. It produces text, code, or images when asked and ceases activity when the user disengages. Its relationship to education is instrumental: it is a tool that educators and students use, for better or worse, within existing pedagogical structures. An agentic AI system, by contrast, pursues goals autonomously. It can decompose a complex objective into sub-tasks, select and orchestrate tools to accomplish those sub-tasks, monitor its own progress, adapt its strategy when initial approaches fail, and maintain persistent memory across interactions that enables it to build cumulative models of the entities it works with (Arunkumar et al., 2026; Masterman et al., 2024; Bandi et al., 2025). When applied to educational assessment, this constellation

of capabilities does not merely automate existing assessment tasks — grading faster, scoring more consistently — but enables assessment modalities that were previously impossible: continuous competency tracking across courses and semesters, personalized assessment strategies calibrated to individual learner profiles, multi-modal evidence integration that captures learning processes as well as products, and longitudinal developmental trajectories that reveal how competencies emerge, consolidate, and transfer over time.

The distinction matters for policy. The governance frameworks, ethical guidelines, and quality assurance standards appropriate for a generative AI chatbot that helps students draft essays are wholly inadequate for an agentic AI system that autonomously designs assessment strategies, evaluates student competencies, and triggers interventions based on its own judgment. If Taiwan's fourth accreditation cycle is designed with only generative AI in mind, it will be obsolete before it is implemented.

Research Questions and Scope

This paper addresses five research questions, each corresponding to a layer of the analytical framework:

RQ1 (Conceptual Foundation): What constitutes "agentic AI" in higher education, and how do its autonomous planning, execution, and adaptation capabilities create qualitatively new possibilities for learning outcome assessment that were impossible with prior AI paradigms?

RQ2 (Problem Diagnosis): What structural limitations in Taiwan's current learning outcome measurement ecosystem — spanning MOE indicators, HEEACT accreditation criteria, and institutional assessment practices — create the demand conditions for a paradigm shift, and which of these limitations are tractable through agentic AI intervention?

RQ3 (Paradigm Shift): Through what mechanisms can agentic AI transform learning outcome measurement from periodic, standardized, and summative approaches to continuous, personalized, and formative paradigms — and what would a conceptual framework for this transformation look

like?

RQ4 (Policy Implications): What regulatory and accreditation adaptations are necessary for Taiwan's MOE and HEEACT to accommodate agentic AI-driven learning outcome measurement, and what implementation pathway would balance innovation with quality assurance?

RQ5 (Ethics and Governance): What ethical risks are unique to agentic AI in student success measurement — particularly regarding learner autonomy, algorithmic accountability, equity, and data sovereignty — and what governance frameworks can mitigate these risks within Taiwan's legal and cultural context?

The scope of the analysis is bounded in three ways. First, it focuses on Taiwan's higher education system specifically, drawing on international comparators for context but grounding all policy recommendations in Taiwan's regulatory, institutional, and cultural environment. Second, it is a theoretical and policy-analytical paper, not an empirical study: it constructs conceptual frameworks, evaluates policy scenarios, and proposes governance structures on the basis of existing evidence and reasoned argument, but it does not present original data from field implementations. Third, while it discusses AI capabilities that are currently emerging, it treats these capabilities as analytically established possibilities whose institutional implications warrant examination now — before, rather than after, they become embedded in educational practice without deliberate governance.

Agentic AI — A New Paradigm in Educational Technology

The rapid maturation of artificial intelligence has produced a category of system that differs not merely in degree but in kind from the AI tools that higher education institutions have deployed over the past two decades. Where earlier generations of educational AI operated as passive instruments — responding to discrete queries, scoring individual items, or flagging anomalies — a new class of *agentic* AI systems can autonomously plan multi-step assessment strategies, execute complex evaluation workflows, maintain persistent models of individual learners, and adapt

their approaches on the basis of accumulated evidence (Arunkumar et al., 2026; Masterman et al., 2024). This section establishes a rigorous taxonomy for understanding this evolution, articulates the specific capabilities that distinguish agentic AI from its predecessors, and proposes a definitional boundary that separates AI-assisted assessment from genuinely AI-agentic assessment. The taxonomy developed here provides the conceptual scaffolding for subsequent sections, which examine how these capabilities intersect with the structural limitations of Taiwan's current quality assurance regime.

From Tools to Agents: A Taxonomy of AI in Education

Any serious attempt to integrate AI into educational assessment must begin with a clear-eyed classification of what different AI systems can and cannot do. The field has suffered from imprecise language: university administrators, accreditation bodies, and policymakers often conflate a simple plagiarism detector with a sophisticated adaptive learning platform, treating both as instances of "AI in education." This terminological looseness obscures critical differences in capability, risk, and governance requirements. Drawing on Russell and Norvig's (2021) canonical agent typologies and Ouyang and Jiao's (2021) three-paradigm framework for AI in education, we propose a four-level taxonomy that maps the evolution of AI in educational settings with increasing specificity.

Level 0: Static Tools. At the most rudimentary level, educational institutions employ rule-based software that performs fixed, deterministic operations on student work. Spell-checkers, grammar correctors, and first-generation plagiarism detection systems (e.g., early versions of Turnitin) exemplify this category. These tools apply predefined rules or string-matching algorithms without any model of the student, the learning context, or the broader educational objectives. In Russell and Norvig's (2021) taxonomy, they correspond to *simple reflex agents* — systems that map percepts directly to actions through condition-action rules, with no internal state and no capacity for learning. Their contribution to assessment is narrow but well-understood: they automate mechanical checks that would otherwise consume instructor time. Critically, they make no evaluative judgments about learning quality.

Level 1: Reactive AI. The next tier encompasses systems that maintain an internal model of the student and adjust their behavior accordingly, but only within a single interaction session or a narrowly bounded task. Computerized adaptive testing (CAT) platforms, such as those underlying the GRE and GMAT, exemplify Level 1: they select subsequent test items based on a running estimate of examinee ability, adjusting item difficulty in real time using Item Response Theory models (van der Linden & Glas, 2010). Intelligent tutoring systems like Carnegie Learning's MATHia similarly track student performance within a lesson and modify hints or problem sequences in response to observed errors. In Ouyang and Jiao's (2021) framework, these systems represent the *AI-directed* paradigm, in which the AI assumes a controlling role — determining what the learner sees and when — but operates within a fixed pedagogical script. The AI reacts to student input; it does not reflect on its own strategies or reconsider its goals.

Level 2: Deliberative AI. A qualitative shift occurs when AI systems move beyond reactive item selection to deliberate reasoning about pedagogical strategy. Learning analytics dashboards that synthesize data from multiple sources — LMS engagement logs, assignment submissions, discussion forum participation, assessment scores — and generate actionable recommendations for instructors represent this level. Systems like Civitas Learning or Brightspace Insights aggregate longitudinal data, identify at-risk students, and suggest interventions. More advanced instantiations include AI-powered essay scoring engines (e.g., ETS's e-rater) that apply multi-dimensional rubrics encompassing content, organization, and language use, and recommendation engines that suggest personalized learning pathways based on competency gap analyses. These systems correspond to Ouyang and Jiao's (2021) *AI-supported* paradigm, in which AI augments human decision-making without supplanting it, and to Russell and Norvig's (2021) *model-based* and *goal-based agents* that maintain representations of how the world works and select actions to achieve specified objectives. The crucial limitation remains: a human must interpret the system's outputs and decide what to do. The AI deliberates, but it does not act autonomously.

Level 3: Agentic AI. The most recent and consequential development transcends the tool-user rela-

tionship entirely. Agentic AI systems — corresponding to Ouyang and Jiao's (2021) *AI-empowered* paradigm and Russell and Norvig's (2021) *learning agents* — possess the capacity to autonomously formulate goals, devise multi-step plans to achieve those goals, execute those plans through tool use and environmental interaction, monitor their own performance, and revise their strategies based on outcomes (Arunkumar et al., 2026). In educational assessment, a Level 3 system does not merely score an essay or recommend a next item; it can design an entire assessment strategy for a student, orchestrate the delivery of multiple assessment modalities (formative quizzes, peer review tasks, reflective prompts, portfolio evaluations), evaluate the resulting evidence against a competency framework, identify gaps, and autonomously initiate remediation sequences — all while maintaining a persistent, evolving model of the learner that spans courses and semesters. Yan (2025) describes this transition as the movement from "passive tool" to "socio-cognitive teammate," a characterization that captures the fundamentally collaborative — rather than instrumental — relationship between agentic AI and human educators.

This four-level taxonomy is not merely academic. It has direct implications for governance: the regulatory frameworks, ethical guidelines, and quality assurance mechanisms appropriate for a Level 0 spell-checker are wholly inadequate for a Level 3 agentic system that autonomously designs and administers assessments. As subsequent sections will argue, Taiwan's current accreditation framework was designed for a world in which AI occupied Levels 0 and 1 at most.

What Agentic AI Can Do That Traditional AI Cannot

The distinction between Level 2 and Level 3 — between deliberative AI and agentic AI — warrants further elaboration, because it is precisely at this boundary that the most consequential implications for educational assessment emerge. Building on the unified taxonomy proposed by Arunkumar et al. (2026) and the analysis of opportunities and challenges of large language models in education by Kasneci et al. (2023), we identify six capabilities that collectively define agentic AI and differentiate it from all prior paradigms. Each capability is illustrated with a concrete assessment scenario to ground the theoretical discussion in institutional practice.

Capability 1: Autonomous Planning. Agentic AI can decompose complex, high-level learning objectives into structured sequences of assessment activities without human specification of each step. Consider the objective "demonstrate mastery of linear algebra." A Level 2 system might recommend a pre-built sequence of quizzes from an item bank. A Level 3 agentic system can analyze the competency's constituent sub-skills (vector operations, matrix transformations, eigenvalue decomposition, applications to systems of equations), assess the student's current proficiency in each through diagnostic probing, and construct a personalized assessment roadmap — including formative checkpoints, a collaborative problem-solving task, and a culminating performance assessment — calibrated to the student's demonstrated strengths and weaknesses (Kasneci et al., 2023). The plan is not retrieved from a template; it is *generated* through reasoning about the specific learner and the specific competency structure.

Capability 2: Dynamic Adaptation. While Level 1 reactive systems adjust item difficulty within a testing session, agentic AI maintains a continuously updating student model that informs adaptation across assessment contexts, time scales, and modalities. If a student demonstrates strong procedural fluency in matrix operations but struggles with conceptual transfer to real-world applications, the agentic system does not simply present harder matrix problems; it shifts the assessment modality to case-based scenarios that probe transfer, adjusts the scaffolding level of subsequent prompts, and recalibrates its confidence estimates about the student's competency profile (Agent4EDU, 2024). This adaptation is not pre-programmed; it emerges from the system's reasoning about discrepancies between expected and observed performance.

Capability 3: Tool Use. A defining feature of agentic AI, as distinguished from monolithic AI models, is the capacity to autonomously invoke external tools, APIs, and data sources as needed to accomplish its objectives (Arunkumar et al., 2026). In an assessment context, an agentic system might query the institution's learning management system to retrieve a student's submission history, invoke a rubric engine to evaluate a written artifact against program-level learning outcomes, call a statistical analysis module to compute inter-rater reliability across peer assessments, and trigger a

notification to the course instructor when a student's performance crosses a predefined threshold — all as coordinated steps within a single assessment workflow. This capacity for tool orchestration transforms the AI from a standalone application into an *infrastructure layer* that integrates disparate educational systems into a coherent assessment pipeline.

Capability 4: Multi-Step Reasoning. Traditional AI assessment tools typically perform a single evaluative operation: score this essay, classify this response, predict this student's risk level. Agentic AI can execute iterative, multi-step evaluation processes that mirror the reasoning of expert human assessors. Consider essay evaluation: rather than producing a single holistic score, an agentic system can first analyze content accuracy against domain knowledge bases, then evaluate argumentative structure and logical coherence, then assess evidence integration and citation quality, and finally synthesize these dimensions into a nuanced evaluative narrative that identifies specific strengths and targeted areas for improvement (Masterman et al., 2024). Each step can inform the next — if the content analysis reveals factual errors, the subsequent argumentation analysis is contextualized by that finding. This iterative, self-referential evaluation process produces assessment feedback of a qualitative richness that single-pass automated scoring cannot achieve.

Capability 5: Persistent Memory. Perhaps the most transformative capability for educational assessment is the agentic system's maintenance of longitudinal learner models that persist across courses, semesters, and even degree programs. Current assessment practices in higher education are overwhelmingly episodic: each course's assessments are designed, administered, and graded in isolation, with minimal systematic connection to assessments in prerequisite or subsequent courses. An agentic system with persistent memory can track a student's development of critical thinking skills from a first-year general education course through disciplinary methods courses to a senior capstone project, identifying trajectories of growth, persistent misconceptions, and emergent competencies that no single-course assessment could capture (Bandi et al., 2025). This longitudinal perspective is precisely what accreditation bodies aspire to when they mandate program-level learning outcome assessment — yet it has remained largely aspirational because the data integration and

analytical labor required to achieve it manually are prohibitive.

Capability 6: Multi-Agent Collaboration. The most sophisticated agentic AI architectures deploy multiple specialized agents that collaborate to achieve assessment objectives no single agent could accomplish alone. Andrew Ng's four agentic design patterns — reflection, planning, tool use, and multi-agent collaboration — identify this last pattern as the most powerful and the most complex (Ng, 2024). In an educational assessment context, one can envision a *Tutor Agent* that manages instructional interactions and identifies assessment opportunities, an *Assessment Agent* that designs and administers evaluative tasks, a *Feedback Agent* that generates personalized formative feedback, and an *Analytics Agent* that aggregates data across students to identify program-level patterns (Agent4EDU, 2024). These agents share information through structured communication protocols, coordinate their actions to avoid redundancy or conflict, and collectively produce an assessment ecosystem that is more coherent, more responsive, and more comprehensive than any single system could provide.

The market trajectory suggests that these capabilities are moving from prototype to deployment. Gartner projects that by 2028, 15% of day-to-day work decisions will be made autonomously by agentic AI, and that by 2026, 40% of enterprise applications will embed AI agents as core components (Gartner, 2025). The education sector, while typically lagging enterprise adoption curves, is unlikely to remain insulated from a transformation of this magnitude. Yan's (2025) Agentic-Profiling-Collaborative-Personalized (APCP) model provides an initial blueprint for how these capabilities might be harnessed in educational contexts while preserving meaningful human oversight.

A candid assessment of the evidence landscape is warranted here. The empirical evidence base for agentic AI in education is nascent. Most existing studies examine generative AI tools — chatbots, automated essay scoring, adaptive testing — rather than fully agentic systems that plan, adapt, and act autonomously over extended periods. The claims advanced in this section regarding agentic AI capabilities are derived primarily from technical demonstrations and proofs of concept (Arunkumar et al., 2026; Masterman et al., 2024), industry projections and analyst reports (Gartner, 2025), and

analogical reasoning from adjacent domains — healthcare, software engineering, and enterprise automation — where agentic AI deployment is more advanced. While these sources establish the technological *possibility* of the capabilities described, they do not constitute the kind of rigorous, replicated, domain-specific empirical evidence that would ordinarily ground claims of paradigm-level transformation. This evidentiary limitation is acknowledged here and addressed throughout the paper through explicit hedging, feasibility classification, and the recommendation of structured piloting (Section 5.3) designed to generate the empirical evidence that the current literature does not yet provide.

Language and Cultural Considerations

A practical barrier that warrants explicit acknowledgment is the language dimension of AI deployment. Most agentic AI systems — including the large language models that underpin their reasoning capabilities — are trained predominantly on English-language data. Taiwan's higher education operates primarily in Mandarin Chinese, with significant use of Taiwanese Hokkien in some institutional contexts and specialized academic terminology that may not be well-represented in English-centric training corpora. For assessment tasks requiring nuanced evaluation of argumentative writing, critical thinking, cultural competency, or disciplinary expertise expressed in Chinese, AI systems trained on English-dominant data may perform significantly less accurately than their English-language benchmarks suggest. This language gap has implications for both the feasibility and the equity of AI-augmented assessment: institutions serving primarily Chinese-language student populations may find that AI assessment tools perform less reliably than advertised, and students whose academic work is conducted in Chinese may be systematically disadvantaged by assessment systems whose linguistic competence is strongest in English. Any pilot implementation in Taiwan must include rigorous evaluation of AI system performance in Chinese-language educational contexts.

From "AI-Assisted" to "AI-Agentic" Assessment: The Definitional Boundary

The preceding taxonomy and capability analysis prepare the ground for a question of considerable practical importance: when does an AI system cross the threshold from tool to agent in the context

of educational assessment? This is not a purely theoretical question. Accreditation standards, institutional policies, academic integrity regulations, and faculty governance structures all presuppose a particular model of how assessment is designed and conducted. If that model changes — if the AI shifts from instrument to collaborator, from tool to agent — then the entire governance apparatus must be reconsidered.

Table 1 synthesizes the key differentiators across three paradigmatic categories: Traditional AI (Levels 0-1), Generative AI (Level 2, exemplified by large language models used as tools), and Agentic AI (Level 3).

Dimension	Traditional AI	Generative AI	Agentic AI
Interaction mode	Single query-response	Conversational, session-based	Autonomous, goal-directed over extended periods
Task scope	Narrow, predefined tasks	Broad but bounded by prompt	Open-ended, self-decomposed into sub-tasks
Autonomy	None; fully human-directed	Low; requires human prompting for each action	High; operates independently toward specified goals
Inference pattern	Rule-based or statistical	Single-pass generative	Iterative, self-correcting, multi-step reasoning
Adaptation	Fixed or session-bounded	Within conversation context	Persistent across sessions, courses, and time
Output type	Scores, classifications, flags	Text, code, multimedia	Actions, decisions, orchestrated workflows

Dimension	Traditional AI	Generative AI	Agentic AI
Tool use	None	Limited (plugins, function calls)	Autonomous orchestration of multiple tools and systems

This typology reveals that the transition from generative AI to agentic AI is not merely an incremental improvement in language model capability; it represents a categorical shift in the system's relationship to the educational process. A generative AI tool like ChatGPT, when used in assessment, remains fundamentally passive: it responds to human prompts, generates text or evaluations on demand, and ceases activity when the human disengages. An agentic AI system, by contrast, maintains its own goals, monitors its own progress, and continues operating — checking student submissions, updating learner models, triggering interventions — without moment-to-moment human direction.

The implications of this shift are thrown into sharp relief by a recent case that has prompted considerable discussion in the higher education community. In early 2026, researchers demonstrated that an agentic AI system — colloquially named "Einstein" by its developers — could autonomously complete entire online university courses, including all assessments, achieving passing or above-average grades without any human intervention (Inside Higher Ed, 2026). The system navigated course management platforms, read assigned materials, completed quizzes, wrote essays, participated in discussion forums, and submitted final projects. This demonstration is not merely a curiosity or a stunt; it constitutes a rigorous *validity test* of current assessment practices. If an AI agent with no genuine understanding of the material, no lived experience, and no authentic learning trajectory can satisfy all assessment requirements for a university course, then those assessment requirements are measuring something other than — or in addition to — the learning outcomes they purport to measure. The "Einstein" case does not indict AI; it indicts assessment designs that fail to capture the dimensions of human learning that AI cannot replicate: embodied experience, ethical

reasoning grounded in personal values, collaborative meaning-making in authentic communities, and the capacity for genuine intellectual transformation.

But the significance of the "Einstein" case extends well beyond assessment validity. It forces a more fundamental question: *what counts as learning?* If an AI agent can produce assessment performances indistinguishable from those of a human learner, what criteria distinguish "genuine learning" from "simulated performance"? The traditional answer — understanding, internalization, transformation — presupposes a cognitive model enclosed within the individual brain. Yet the extended mind thesis in cognitive science argues that cognition does not stop at the skull but extends into the tools and environments that individuals use (Clark & Chalmers, 1998). If cognition itself is distributed, then in the AI era "learning" may need to be reconceived not as an event occurring purely within the individual but as a co-constitutive process between human and technological systems. Hayles (2012) captures this dynamic with the concept of *technogenesis* — the co-evolution of humans and technics, where humans reshape their tools and tools in turn reshape human cognition. Under this lens, the "Einstein" case reveals not merely a blind spot in assessment design but a *conceptual lag*: we are evaluating a digital-age cognitive phenomenon with an analog-age definition of learning. This deeper challenge — that AI changes what learning *is*, not just how we measure it — will be elaborated in Section 4.7.

This provocation clarifies the stakes of the present analysis. Agentic AI does not merely create new tools for assessment; it simultaneously reveals the fragility of assessment architectures that were never designed to withstand agents capable of autonomous, multi-step, tool-using, adaptive behavior. More fundamentally, it exposes a structural problem: when AI transforms how humans cognize, learn, and know, the definitions that assessment systems are built upon become moving targets — and any measurement framework anchored to a fixed definition will inevitably drift from the reality it purports to capture. Any quality assurance framework that aspires to relevance in the agentic AI era must grapple with both dimensions: harnessing the constructive potential of agentic AI for assessment (the six capabilities detailed in Section 2.2) while fortifying assessment designs

against the validity threats that these same capabilities introduce.

We therefore propose the following working definition, which will anchor the analysis in subsequent sections:

> **Agentic AI in educational assessment** refers to AI systems that autonomously plan assessment strategies, execute multi-step evaluation workflows, maintain persistent learner models, and adapt their approach based on accumulated evidence — operating as collaborative assessment partners rather than passive tools. Such systems are characterized by goal-directed behavior, tool orchestration capability, longitudinal memory, and the capacity for multi-agent collaboration, and they require governance frameworks that address their autonomous decision-making authority rather than merely their technical outputs.

This definition deliberately foregrounds the governance implications of agentic AI, because it is at the governance level — accreditation standards, quality assurance criteria, institutional policies — that the most urgent adaptation is required. A system that autonomously designs assessments raises fundamentally different questions about academic authority, faculty prerogatives, and institutional accountability than a system that scores essays on demand. The definition also emphasizes the collaborative framing — "assessment partners rather than passive tools" — to signal that the most productive institutional response is neither uncritical adoption nor blanket prohibition, but a thoughtful renegotiation of the division of labor between human educators and AI agents.

The taxonomy and definitional framework established in this section provide the analytical vocabulary for examining a concrete case: how does Taiwan's current higher education quality assurance system — with its specific accreditation criteria, assessment expectations, and institutional practices — measure up against the capabilities and challenges of agentic AI? Section 3 turns to this question, analyzing the structural limitations of Taiwan's assessment landscape that agentic AI both

exposes and has the potential to remedy. ## 3. The Current Paradigm: Student Learning Outcome Measurement in Taiwan

Any attempt to theorize a paradigm shift must first establish a rigorous account of the paradigm it proposes to supersede. This section undertakes that task by mapping the institutional architecture through which student learning outcomes are currently defined, measured, and acted upon in Taiwan's higher education system. The analysis proceeds in four stages. First, it examines the accreditation framework administered by the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT), which constitutes the primary external quality assurance mechanism and the structural backbone of learning outcome assessment at the system level. Second, it surveys the complementary policy instruments deployed by the Ministry of Education (MOE), including competency platforms, graduate tracking surveys, and institutional research infrastructure. Third, it identifies six structural limitations inherent in the current measurement paradigm – limitations that function, in Kuhnian terms, not as minor imperfections but as systematic anomalies that the reigning paradigm cannot resolve from within its own logic. Fourth, it traces the accumulation of external pressures – demographic, economic, technological, and comparative – that are intensifying these anomalies to crisis proportions. Together, these analyses establish the demand conditions for the paradigm shift theorized in Section 4.

The HEEACT Framework: Standards, Indicators, and Assessment Logic

Taiwan's contemporary quality assurance architecture for higher education is anchored by the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT), established in 2005 as an independent, non-governmental foundation commissioned by the MOE to conduct third-party accreditation (HEEACT, 2023a). HEEACT's creation marked Taiwan's formal entry into professionalized external quality assurance, aligning domestic practice with the accountability and quality improvement movements that had reshaped higher education governance across OECD nations since the 1990s (Lin et al., 2021). By 2026, HEEACT holds full compliance with the International Network for Quality Assurance Agencies in Higher Education (INQAAHE) International Standards

and Guidelines (ISGs), maintains recognition by the Council for Higher Education Accreditation (CHEA) in the United States, and serves as an active member of the Asia-Pacific Quality Network (APQN) (HEEACT, 2023a; Lin et al., 2021).

HEEACT operates a dual-track system that has been in place since 2012: mandatory external institutional accreditation for all higher education institutions (HEIs), and voluntary program accreditation for individual departments and degree-granting units (HEEACT, 2024). In 2017, the MOE introduced a significant policy adjustment, rendering program accreditation voluntary rather than mandatory, provided that institutions can demonstrate "alternative mechanisms for ensuring teaching quality" (HEEACT, 2024, p. 2). This policy shift toward institutional autonomy and self-accountability represents an important evolution in Taiwan's quality assurance philosophy, moving from compliance-driven to enhancement-oriented accreditation.

Institutional Accreditation: The Third Cycle (2023–2025)

The third cycle of institutional accreditation, implemented from 2023 to 2025, is scheduled to evaluate a total of 83 HEIs, including 67 public and private universities, 8 religious schools, 6 military schools, and 2 open universities (HEEACT, 2023a). The framework is organized around four standards, each operationalized through core indicators that institutions must address in their self-assessment reports (SARs) and demonstrate during on-site visits:

- **Standard 1: Institutional Governance and Management** – encompasses mission clarity, organizational structure, resource planning, decision-making mechanisms, internal quality assurance, institutional research (IR), and stakeholder engagement (Core Indicators 1-1 through 1-4).

- **Standard 2: Teaching and Academic Professionalism** – addresses faculty performance, evaluation, and reward systems; recruitment quality; curriculum planning and review; and quality assessment in teaching (Core Indicators 2-1 through 2-4).

- **Standard 3: Student Learning and Outcomes** – the standard most directly relevant to this

paper's inquiry – examines undergraduate education and outcomes (3-1), graduate education and outcomes (3-2), evaluation mechanisms for general and interdisciplinary education (3-3), and evaluation mechanisms for inter-collegiate and cross-border education (3-4) (HEEACT, 2023a).

- **Standard 4: Social Responsibility and Sustainable Development** – covers equal educational opportunities, social responsibility practices, and financial sustainability (Core Indicators 4-1 through 4-3).

Two conceptual pillars undergird this framework: the Plan-Do-Check-Act (PDCA) cycle, requiring evidence of continuous improvement, and the "empowerment model" (賦權模式) that encourages institutions to add distinctive indicators beyond the required core set (HEEACT, 2023a). The accreditation process follows a five-stage sequence (preparation, self-assessment, document review, on-site visit, decision-making) spanning approximately 18 months, with tiered outcomes: 6-year accreditation, 3-year accreditation with follow-up, or re-accreditation (HEEACT, 2023a).

Program Accreditation: The Current Cycle (2024 Edition)

The HEEACT Program Accreditation framework (2024 edition) operates in parallel at the departmental level, organized around three standards with 12 core indicators covering program governance, faculty and teaching, and students and learning (HEEACT, 2024). Accreditation results mirror the institutional tier: 6 years, 3 years, or re-accreditation required.

What is analytically significant about both tracks is how they operationalize "learning outcomes." The evidence required under Standard 3 reveals the paradigm's epistemological commitments: lists of students' learning performance (research, creations, exhibitions, certificates, awards); lists of graduates' performance (within 3 years); and analysis of student learning and graduate survey data (HEEACT, 2023a, pp. 32–36). These evidence categories privilege countable outputs and institutional self-reports – a pattern whose limitations are examined in Section 3.3.

MOE Policy Instruments

Beyond the HEEACT accreditation framework, the MOE deploys several complementary policy instruments that shape how student learning outcomes are defined, measured, and incentivized across Taiwan's higher education system.

Higher Education Sprout Project Phase II (2023–2027)

The Higher Education Sprout Project (高教深耕計畫), Taiwan's flagship funding mechanism for higher education quality enhancement, entered its second phase in 2023 with a total budget of approximately NT$83.6 billion (roughly USD 2.6 billion) over five years, following Phase I's NT$86.85 billion allocation in 2018–2022 (MOE, 2025). The project is structured around four focus areas: (a) strengthening teaching quality and learning outcomes, (b) developing institutional features and research excellence, (c) enhancing social responsibility and lifelong learning, and (d) promoting international competitiveness. Funding is allocated through a competitive application process, with institutional performance assessed against self-defined key performance indicators (KPIs), many of which relate directly to student learning outcomes, graduate employability, and employer satisfaction.

The Sprout Project represents the MOE's primary lever for translating quality assurance standards into institutional behavior through financial incentives. However, as Hou et al. (2020) have noted, the project's KPI-driven evaluation model can inadvertently encourage metric optimization over genuine pedagogical transformation – institutions may gravitate toward easily measurable indicators that satisfy reporting requirements rather than pursuing deeper, harder-to-quantify changes in learning quality.

UCAN Platform

The University Curriculum and Career Mapping (UCAN) platform (大專校院就業職能平台), operational since 2010, maps occupational competencies to curricular structures through professional competency diagnostics (66 career clusters) and general competency assessments (eight workplace-ready skills) (MOE, 2025). However, UCAN measures students' *perceptions* of their competencies

through self-reported surveys rather than their *demonstrated* competencies – a distinction critical to evaluating measurement paradigm adequacy.

Graduate Tracking Surveys

Taiwan mandates graduate tracking surveys at 1-year, 3-year, and 5-year intervals post-graduation, collecting data on employment, salary, satisfaction, and perceived relevance of education (MOE, 2025). Results feed into institutional self-assessment and MOE policy evaluations.

Institutional Research (IR) Infrastructure

Taiwan's IR capacity varies dramatically: research universities have developed sophisticated analytics, while many smaller private institutions lack dedicated IR offices. This unevenness creates a significant equity dimension – institutions with the most vulnerable student populations often have the weakest capacity to measure and improve learning outcomes.

Assessment Instruments Summary

Table 2 synthesizes the primary instruments currently deployed for student learning outcome measurement in Taiwan, organized by administering body, measurement type, and temporal characteristics.

Table 2

Summary of Student Learning Outcome Assessment Instruments in Taiwan's Higher Education System

Instrument	Administering Body	Measurement Type	Temporal Frequency	Primary Evidence Mode
Institutional Accreditation (Standard 3)	HEEACT	External peer review of institutional self-assessment	6-year cycle	Document-based (SAR + on-site visit)

Instrument	Administering Body	Measurement Type	Temporal Frequency	Primary Evidence Mode
Program Accreditation (Standard 3)	HEEACT	External peer review of program self-assessment	6-year cycle (voluntary)	Document-based (SAR + on-site visit)
UCAN Competency Assessment	MOE	Self-reported competency diagnostics	Annual (optional for students)	Survey-based (indirect)
Graduate Tracking Surveys	MOE	Self-reported employment and satisfaction	1, 3, 5 years post-graduation	Survey-based (indirect)
Higher Education Sprout Project KPIs	MOE	Institutional self-reported KPI achievement	Annual review, 5-year cycle	Quantitative indicators + narrative
Course Evaluation Surveys	Individual HEIs	Student satisfaction and perceived learning	Semester-based	Survey-based (indirect)
Capstone/Thesis Assessment	Individual HEIs	Direct assessment of student work	At program completion	Performance-based (direct, but unstandardized)

Note. Adapted from HEEACT (2023a, 2024), MOE (2025), and Coates and Zlatkin-Troitschanskaia (2019).

The table reveals a striking pattern: of the seven primary instruments, five rely on indirect measurement (self-reports, surveys, document review), while only capstone and thesis assessment provides

direct evidence of learning – and this is administered at the institutional level without standardization or cross-institutional comparability. This dominance of indirect measurement constitutes a foundational characteristic of the current paradigm.

Structural Limitations of the Current Paradigm

Having mapped the institutional architecture, we can now identify the structural limitations embedded within it. These limitations are not incidental deficiencies that can be remedied through incremental reform; they are, in Kuhn's (1962/2012) terminology, *anomalies* – systematic discrepancies between what the paradigm promises to measure and what it actually captures. Six such anomalies are identifiable.

The temporal limitation. The current paradigm operates on periodic snapshot logic. Institutional accreditation follows a 6-year cycle; program accreditation similarly evaluates performance over multi-year windows (HEEACT, 2023a, 2024). Self-assessment reports look backward, synthesizing 3.5 to 4 years of historical data into a document submitted months before the on-site visit. Graduate tracking surveys capture employment outcomes 1 to 5 years after graduation. In a higher education environment increasingly shaped by rapid technological change, labor market volatility, and pedagogical innovation, this temporal architecture means that the assessment system is structurally backward-looking in a world that demands forward-looking agility (Coates & Zlatkin-Troitschanskaia, 2019). A program can receive full accreditation based on outcomes from a cohort that graduated before the emergence of generative AI, with no mechanism to assess whether current students are being prepared for the post-AI labor market.

The granularity limitation. The assessment architecture operates at the institutional and program levels, producing judgments about collective entities – "this institution is accredited," "this program meets standards." These aggregate judgments obscure individual learning trajectories. A program can be accredited while individual students within it struggle invisibly; a university can demonstrate satisfactory learning outcomes in its SAR while significant sub-populations (first-generation students, students with disabilities, students from disadvantaged socioeconomic backgrounds) ex-

perience markedly different educational realities. The current paradigm has no mechanism to disaggregate learning outcomes below the program level in any systematic, real-time fashion (Coates & Zlatkin-Troitschanskaia, 2019). This granularity gap is not merely a measurement inconvenience; it represents a fundamental misalignment between the unit of analysis (programs and institutions) and the unit of educational concern (individual learners).

The modality limitation. The dominant evidence modality in the current paradigm is documentary. Self-assessment reports constitute the primary basis for accreditation judgments, supplemented by on-site visits of 1 to 2 days' duration (HEEACT, 2023a). This document-based approach cannot capture real-time learning processes, moment-to-moment pedagogical interactions, or the dynamic evolution of student competencies across a semester or a degree program. Assessment relies fundamentally on what institutions *report*, not on what students *experience*. The SAR, by its nature, is a curated narrative – institutions select evidence that supports favorable evaluation, organize it according to the prescribed framework, and present it in a format optimized for reviewer assessment. This is not a critique of institutional integrity; it is a recognition that the documentary modality itself introduces systematic selection bias into the evidence base (Tam, 2001).

The agency limitation. In the current paradigm, the assessed entity is simultaneously the reporter. Institutions prepare their own SARs, select their own evidence, define their own distinctive indicators, and present their own narrative of educational quality. While the on-site visit provides a partial external check, and while HEEACT's review panels exercise professional judgment in evaluating self-reported claims, the fundamental information asymmetry remains: institutions possess far more information about their actual performance than can be captured in any document review or 1-to-2-day visit. This structural feature creates what organizational theorists term "impression management" incentives (Goffman, 1959) – not necessarily dishonest, but systematically tilted toward favorable self-presentation. The agency limitation is particularly consequential for learning outcome measurement, where the gap between reported outcomes (lists of awards, certifications, and employment rates) and actual learning (critical thinking development, deep understanding,

competency growth) is widest.

The competency capture limitation. The evidence categories specified in HEEACT's accreditation standards reveal a systematic privileging of countable outputs over complex competencies. The supporting documents required under Standard 3 of institutional accreditation emphasize lists of students' learning performance – research output, creations and exhibitions, hands-on work, certificates, and awards from national and international competitions – alongside graduate survey data (HEEACT, 2023a, pp. 32–36). While these outputs are legitimate evidence of learning, they represent only the visible tip of Spencer and Spencer's (1993) iceberg: the knowledge and skills that are readily observable and measurable. Below the waterline – the self-concept, traits, and motives that Spencer and Spencer identified as the deeper, more predictive layers of competency – the current paradigm has no systematic measurement approach. Complex competencies such as critical thinking, creative problem-solving, intercultural sensitivity, ethical reasoning, and collaborative intelligence are precisely the competencies most valued by employers and most resistant to the output-counting logic of the current framework (Association of American Colleges and Universities [AAC&U], 2018).

The indirect measurement dominance. As Table 2 documents, the current paradigm relies overwhelmingly on indirect evidence of learning – surveys, self-reports, document reviews – rather than direct assessment of student performance against defined learning outcomes. UCAN measures perceived competencies; graduate tracking surveys measure self-reported employment and satisfaction; course evaluation surveys measure student perceptions of teaching; and accreditation itself evaluates institutional *mechanisms* for assessing learning rather than learning itself. The only consistently direct assessment instruments – capstone projects, theses, and dissertations – operate at the institutional level without standardization, cross-institutional comparability, or systematic analysis of learning progression over time. This pattern mirrors a broader challenge identified in the international literature: higher education systems worldwide have been more successful at building quality assurance *processes* than at developing robust *measures* of student learning (Coates &

Zlatkin-Troitschanskaia, 2019; Shavelson, 2010).

Anomalies Accumulating: Why the Current Paradigm Is Under Strain

The six structural limitations identified above have been present since the inception of Taiwan's modern quality assurance architecture. What has changed is the external environment, which has intensified these limitations from tolerable imperfections into urgent anomalies. Several converging pressures are straining the current paradigm to its breaking point.

The Demographic Crisis

Taiwan's birth rate, among the lowest in the world, has produced an enrollment crisis of existential proportions for the higher education sector. Total higher education enrollment stands at approximately 900,000, and the MOE projects that first-year university enrollment will decline to approximately 173,000 by the early 2030s – a figure that many institutions cannot survive (MOE, 2025). In 2024 alone, seven universities ceased operations, and projections suggest that up to 40 private universities may close by 2028 (Sharma, 2024; Taiwan News, 2024). This demographic reality transforms the quality assurance problem: in a contracting system, the question is no longer merely "are institutions meeting minimum standards?" but rather "can the assessment system identify which institutions are delivering sufficient learning value to justify continued operation – and can it do so quickly enough to protect students at failing institutions?" The 6-year accreditation cycle is structurally incapable of providing that kind of early warning.

The Employment Mismatch

Despite a university enrollment rate exceeding 95% – making Taiwan's higher education system effectively universal – graduate unemployment hovers at approximately 4.5%, and more consequentially, there is growing evidence of qualitative mismatch between graduate competencies and employer needs (MOE, 2025). Recent research has found that 63% of AI-related positions in Taiwan require application-level skills, while only 37% of higher education curricula address these competencies – a structural gap that current assessment mechanisms have neither detected nor corrected in a timely manner. The employment mismatch anomaly is particularly damaging to the current paradigm's legitimacy because graduate employment outcomes are among the few indi-

cators the paradigm *does* measure, yet even here the measurement is too delayed (1–5 years post-graduation) and too coarse (aggregate employment rates) to drive meaningful curricular adaptation.

Faculty Resistance to Outcomes-Based Education

The international shift toward outcomes-based education (OBE), which HEEACT's frameworks formally endorse through their emphasis on learning outcomes across all standards, faces significant implementation resistance at the faculty level. Research on Taiwanese higher education has documented widespread faculty skepticism toward OBE, rooted in concerns about academic freedom, the reductionism of competency frameworks, the administrative burden of outcomes documentation, and fundamental philosophical disagreements about whether education's purposes can or should be captured in measurable outcomes (Lin et al., 2021). This resistance creates a gap between the formal architecture (which demands outcomes evidence) and the ground-level reality (where teaching and assessment practices may remain input- and process-focused). The current paradigm has no mechanism to detect or address this implementation gap beyond the stylized evidence presented in SARs.

The Digital Divide

Taiwan's higher education system encompasses enormous institutional diversity – from research-intensive national universities with sophisticated digital infrastructure to small private institutions with limited technology capacity. The current assessment paradigm, being document-based, is agnostic to this digital divide; it evaluates all institutions through the same SAR-and-site-visit process regardless of their technological maturity. However, as digital learning environments become increasingly central to educational delivery (accelerated by the COVID-19 pandemic), the inability of the assessment framework to evaluate digital pedagogy, online learning quality, or educational technology effectiveness represents a growing blind spot.

International Comparators Moving Faster

Taiwan's quality assurance framework does not exist in isolation; it operates within an increasingly competitive international landscape where peer systems are evolving rapidly. The European Standards and Guidelines (ESG) underwent significant revision in 2015 and are currently undergoing

further updates that increasingly integrate digital competency and learning analytics considerations (ENQA, 2015). Australia's Tertiary Education Quality and Standards Agency (TEQSA) has developed explicit digital delivery standards and is exploring AI-augmented quality assessment processes. Singapore's EdTech Masterplan 2030, announced in 2023, positions technology-enhanced learning assessment as a national strategic priority (MOE Singapore, 2023). Against these comparators, Taiwan's document-based, 6-year-cycle, largely survey-dependent assessment paradigm risks falling behind – not because it lacks rigor, but because the rigor it offers is of a kind that is becoming structurally insufficient for the emerging higher education environment.

The AI Competency Gap

Perhaps the most consequential anomaly is also the newest: the current assessment framework has no mechanism to evaluate AI literacy as a student learning outcome. In a world where generative AI is rapidly transforming professional practice across virtually every field, the absence of AI competency from accreditation standards, UCAN assessment modules, and graduate tracking instruments means that the paradigm is structurally unable to assess whether Taiwanese graduates are prepared for the professional reality they will enter upon graduation. This is not merely a content gap that can be filled by adding a new indicator; it represents a paradigm-level challenge, because AI competency is not a static body of knowledge that can be measured through traditional output counting but a dynamic, evolving capability that requires new assessment modalities altogether.

Taken together, these accumulated anomalies create what Kuhn (1962/2012) identified as the "crisis" conditions that precede a paradigm shift. The current paradigm – built on periodic cycles, document-based evidence, institutional self-reporting, aggregate judgments, and indirect measurement – was adequate for an era when higher education was expanding, change was incremental, and the primary quality assurance question was whether institutions met minimum thresholds. That era is ending. Taiwan now faces a contracting system with accelerating change, a labor market demanding competencies that current instruments cannot measure, a demographic crisis requiring

early-warning capabilities that 6-year cycles cannot provide, and an AI revolution that renders the entire assessment modality potentially obsolete. The question is not whether the current paradigm will change, but how – and whether Taiwan can shape that change proactively rather than reactively.

Section 4 proposes an answer: the integration of agentic AI systems into the learning outcome measurement architecture, constituting not merely a technological upgrade but a paradigmatic transformation in how student learning is defined, captured, analyzed, and acted upon.

The Paradigm Shift: From Static Measurement to Dynamic Learning Evidence

The preceding sections have established two foundational arguments: that agentic AI possesses a constellation of capabilities — planning, adaptation, tool use, multi-step reasoning, memory, and multi-agent collaboration — that create qualitatively new possibilities for educational assessment (Section 2), and that Taiwan's current quality assurance paradigm exhibits six structural limitations that cannot be resolved through incremental improvement (Section 3). This section synthesizes these arguments to address the paper's central research question: through what mechanisms can agentic AI transform learning outcome measurement from periodic, standardized, and summative approaches to continuous, personalized, and formative paradigms? Drawing on Kuhn's (1962/2012) theory of scientific revolutions, this section proposes the ADAPT framework — Assessment-Design for Agentic Paradigm Transformation — as an original conceptual model for understanding this transformation, articulates seven dimensions along which the paradigm shift operates, and illustrates the new paradigm through a grounded scenario set in a Taiwanese university of technology. Critically, this section also confronts the limits of technological transformation by examining the irreducible role of human judgment and the ontological implications of redefining what counts as evidence of learning.

Kuhn's Framework Applied to Assessment Theory

Thomas Kuhn's (1962/2012) account of scientific revolutions provides a powerful lens for understanding the current moment in educational assessment. Kuhn argued that scientific fields operate

within paradigms — shared frameworks of assumptions, methods, and exemplars that define legitimate problems and acceptable solutions. Progress within a paradigm constitutes "normal science," but when anomalies accumulate that the paradigm cannot resolve, a crisis emerges, eventually catalyzing a revolutionary shift to a new paradigm that is incommensurable with the old.

Justifying the Kuhnian Lens for Quality Assurance

Kuhn's framework was developed for the natural sciences, and he himself expressed skepticism about extending it to social domains. This skepticism must be taken seriously. However, a substantial body of scholarship has demonstrated that Kuhn's core concepts — paradigm, normal science, anomaly, crisis, and revolution — possess analytical utility well beyond the natural sciences when applied with appropriate qualification. Masterman (1970), in her influential analysis, identified twenty-one distinct senses in which Kuhn used "paradigm" and argued that the sociological sense — a shared set of exemplars, practices, and standards that define a professional community's work — is applicable to any organized domain of practice, not merely to scientific disciplines. Ritzer (1975) applied Kuhnian analysis to sociology, demonstrating that the concept of paradigm could illuminate how professional communities adopt, defend, and eventually abandon shared frameworks of practice. Eckstein (1992) extended the framework to political science, showing that administrative and regulatory regimes can function as paradigms when they exhibit the key characteristics Kuhn identified: shared exemplars, a "disciplinary matrix" of accepted methods, and a professional community that defines legitimate problems and solutions.

We employ Kuhn's framework as an analytical heuristic rather than a strict epistemological claim. The argument is not that quality assurance is a natural science undergoing a scientific revolution in the precise sense Kuhn described, but rather that the Kuhnian vocabulary is analytically productive for understanding the current moment in educational assessment. Taiwan's quality assurance system exhibits the structural features that make the Kuhnian analogy illuminating: it operates through shared exemplars (the Self-Assessment Report template, the on-site visit protocol, the Core Indicator framework), a disciplinary matrix (the Plan-Do-Check-Act cycle that structures all accreditation

activities), a professional community with defined norms (HEEACT's evaluator corps, trained in specific methodologies and sharing standards of evidence), and a set of agreed-upon "puzzles" (how to measure learning outcomes, how to ensure institutional accountability) that practitioners solve within the paradigm's logic. These features are not metaphorical; they are observable institutional structures that function analogously to Kuhn's paradigmatic components.

The apparent tension between the "paradigm revolution" framing and this paper's recommendation of Scenario B (incremental framework evolution) requires direct engagement. This tension is less paradoxical than it appears. Kuhn himself acknowledged, particularly in the 1969 postscript to *The Structure of Scientific Revolutions*, that paradigm shifts are not always sudden ruptures; in applied domains, they can unfold gradually as practitioners accumulate evidence that the old framework is inadequate and progressively adopt new methods and standards. Scenario B represents what Kuhn described as the "transition period" — a phase during which elements of the new paradigm coexist with the old, and practitioners begin to reconceptualize their work even before the full gestalt shift occurs. The three-phase implementation pathway proposed in Section 5.3 is designed to manage this transition deliberately, neither forcing premature revolution nor foreclosing the possibility of paradigm-level change as the evidence base matures. Moreover, as Section 4.7 will argue, the co-evolution of AI and human cognition means that a single revolutionary rupture followed by stable "normal science" is structurally impossible in this domain — the paradigm shift is itself continuous, making incremental adaptation not a compromise but the only epistemically honest response to a moving target.

More broadly, scholars have productively applied the Kuhnian framework to education (Shepard, 2000) and, more recently, to the specific question of how artificial intelligence reshapes educational paradigms (Zhong & Zhao, 2025).

Normal science in assessment. The current paradigm of higher education quality assurance — periodic, standardized, summative, document-based, and institution-reported — constitutes what Kuhn would recognize as normal science. Its methods are well-established: accreditation cycles

of five to seven years, standardized self-assessment reports, statistical indicators aggregated at the program or institutional level, and peer review by human panels (HEEACT, 2023). Within this paradigm, quality assurance professionals refine instruments, adjust indicator weights, and develop rubrics — the "puzzle-solving" activity that characterizes normal science (Kuhn, 1962/2012, p. 35). The paradigm has produced genuine achievements: Taiwan's HEEACT has built one of Asia's most respected quality assurance systems, and periodic accreditation has demonstrably improved institutional attention to teaching quality and student outcomes (Hou et al., 2012).

Anomalies. Yet the six structural limitations identified in Section 3 — temporal rigidity, granularity collapse, modality restriction, agency asymmetry, competency capture failure, and indirect measurement dominance — function as Kuhnian anomalies. They represent problems that the current paradigm recognizes as important but cannot solve within its own logic. The temporal limitation, for instance, is not a bug that can be fixed by shortening accreditation cycles from six years to three; even annual assessment would remain fundamentally periodic rather than continuous. The granularity limitation cannot be resolved by collecting more aggregate data; it requires an entirely different unit of analysis. These are not puzzles waiting for cleverer solutions within the existing paradigm but structural contradictions inherent in the paradigm itself.

Crisis. Three external pressures intensify these anomalies into what Kuhn (1962/2012) termed a crisis. First, Taiwan's demographic decline — from approximately 270,000 first-year university students in 2015 to a projected 173,000 by 2028 (Ministry of Education, 2024) — makes every student's learning trajectory a matter of institutional survival, rendering aggregate cohort measures dangerously inadequate. Second, persistent employer dissatisfaction with graduate competencies, despite decades of accreditation, suggests that the current paradigm measures institutional compliance rather than authentic learning outcomes (MOE, 2025). Third, the rapid emergence of AI competencies as essential graduate attributes creates assessment demands — for creativity, human-AI collaboration, ethical reasoning in algorithmic contexts — that existing instruments were never designed to capture (UNESCO, 2025). The accumulated weight of these anomalies, as Kuhn's

framework predicts, generates a sense that something is fundamentally wrong with the paradigm, not merely with its implementation.

Revolution. Agentic AI does not merely offer incremental improvements to the existing paradigm — faster data processing, more efficient report generation, or automated indicator calculation. Rather, it enables what Kuhn (1962/2012) described as a gestalt shift: a fundamentally different way of seeing assessment. Where the current paradigm sees learning as an endpoint to be measured at intervals, the agentic paradigm sees learning as a trajectory to be continuously observed. Where the current paradigm treats evidence as documents that institutions produce, the agentic paradigm treats evidence as multimodal traces that emerge from the learning process itself. Where the current paradigm positions the assessor as an external auditor, the agentic paradigm envisions assessment as a collaborative activity among students, faculty, AI agents, and quality assurance bodies. These are not quantitative differences — more data, faster processing — but qualitative differences in the ontology of assessment: what learning is, what evidence counts, and who assesses. As Zhong and Zhao (2025) argue, the AI age demands a fundamental rethinking of teaching, learning, and assessment paradigms.

Incommensurability. Kuhn's (1962/2012) most controversial claim — that successive paradigms are incommensurable, meaning they cannot be fully translated into each other's terms — applies with particular force here. A graduation rate and a learning trajectory are not different measures of the same thing; they reflect fundamentally different conceptions of what "student success" means. An accreditation panel's judgment and an AI agent's continuous competency mapping are not different methods for answering the same question; they answer different questions entirely. This incommensurability does not mean that the new paradigm renders the old one worthless — periodic human review retains essential value, as Section 4.5 will argue — but it does mean that the two paradigms cannot be smoothly blended. The transition requires deliberate architectural choices, which the ADAPT framework is designed to support.

The ADAPT Framework: A Conceptual Model for Transformation

To provide a structured analytical lens for understanding how agentic AI transforms learning outcome measurement, this paper proposes the ADAPT framework: Assessment-Design for Agentic Paradigm Transformation. The framework is not a prescriptive implementation guide but a conceptual model that enables systematic analysis of both opportunities and risks.

Derivation Logic

The five dimensions of the ADAPT framework emerge from the intersection of three analytical inputs developed in the preceding sections: the technological capability analysis (Section 2), which identifies what agentic AI can do; the diagnostic limitation mapping (Section 3), which identifies what the current paradigm cannot do; and the governance requirements identified in the literature on responsible AI deployment in education (UNESCO, 2023; Beauchamp & Childress, 2019). The framework's dimensions are not an arbitrary selection or a relabeling of the paper's research questions; they represent the minimum set of analytical categories necessary to connect technological capability to institutional transformation while maintaining ethical governance. Understanding what agentic AI can do (Agency Architecture) is a precondition for identifying where the current paradigm fails (Diagnostic Mapping); these diagnostic gaps define what must change in assessment practice (Assessment Reconception); reconceived assessment requires institutional and regulatory adaptation (Policy Pathways); and all of the preceding must be constrained by ethical safeguards (Trust & Ethics Safeguards).

Relational Structure

Unlike a checklist, these five dimensions are sequentially dependent and mutually constraining. Agency Architecture (A) reveals what new assessment modalities are possible; Diagnostic Mapping (D) matches those capabilities against the current paradigm's limitations; Assessment Reconception (A) defines what must change (Section 4.3); Policy Pathways (P) translates possibilities into regulatory realities (Section 5); and Trust & Ethics Safeguards (T) constrains what is permissible (Section 6). Critically, the framework is recursive: trust considerations constrain permissible agency architectures (T → A), and diagnostic mapping shapes policy design (D → P). The ADAPT framework

thus functions as both a linear analytical progression and a feedback loop of mutual constraints — generating analytical predictions, such as that assessment reconception without corresponding policy adaptation will produce institutionally unviable proposals, or that deploying agentic capabilities without trust safeguards will produce backlash analogous to South Korea's AI textbook rollback (Section 5.2).

Seven Dimensions of the Paradigm Shift

The Assessment Reconception component of the ADAPT framework identifies seven dimensions along which the paradigm shift from static measurement to dynamic learning evidence operates. Table 3 summarizes these dimensions; the analysis that follows examines each in detail. To maintain epistemic precision, the discussion distinguishes among three categories of capability: *demonstrated capabilities* that exist in current educational technology implementations (e.g., automated essay scoring, adaptive testing, learning analytics dashboards); *emerging capabilities* that have been demonstrated in prototype or limited deployment (e.g., multi-modal assessment, cross-course competency tracking); and *projected capabilities* that are technically plausible but not yet demonstrated in educational contexts at scale (e.g., autonomous multi-agent assessment ecosystems, real-time institutional quality dashboards aggregating student-level data across programs).

Table 3

Seven Dimensions of the Assessment Paradigm Shift

Dimension	Current Paradigm	Agentic AI Paradigm	Mechanism of Change
Temporality	Periodic (5–7-year cycle)	Continuous	Persistent memory + always-on monitoring
Granularity	Aggregate (program/cohort)	Individual trajectory	Dynamic student modeling + personalized assessment

Dimension	Current Paradigm	Agentic AI Paradigm	Mechanism of Change
Agency	Institution-reported	AI-observed + student-co-created	Autonomous data collection + multi-agent evidence synthesis
Feedback latency	Retrospective (months/years)	Real-time	Iterative reasoning + immediate feedback loops
Evidence type	Documents, statistics, surveys	Multimodal learning traces	Tool use + LMS integration + multimodal analytics
Assessment purpose	Accountability/compliance	Improvement/personalization	Goal-directed planning + adaptive intervention
Assessor	Human reviewers	Human-AI collaborative assessment	Multi-agent collaboration + human-in-the-loop oversight

Temporality: From Periodic to Continuous

Under the current paradigm, learning outcomes are measured at fixed intervals — typically aligned with five-to-seven-year accreditation cycles — producing temporal snapshots that may be outdated before they are analyzed (Section 3.1). An agentic AI system, by contrast, could potentially leverage persistent memory to maintain continuously updated models of learning processes. Unlike stateless AI tools that respond to individual prompts, agentic systems retain context across interactions, building cumulative representations of student learning that evolve over semesters and years (Arunkumar et al., 2026; Masterman et al., 2024).

In a Taiwanese university context, this would mean that an AI agent embedded within a program's learning management system continuously monitors student performance across courses, tracking not merely grades but patterns of engagement, conceptual development, and competency acquisition. The agent's persistent memory enables it to detect longitudinal trends — a gradual decline in

analytical writing quality across a cohort, for instance — that periodic assessment would capture only years after the decline began. OpenAI's (2025) learning outcomes measurement tools demonstrate early prototypes of such continuous monitoring, though current implementations remain far from the seamless integration this dimension envisions.

The challenge accompanying this shift is surveillance creep. Continuous monitoring, even when benign in intent, creates pervasive observation environments that may inhibit risk-taking, creativity, and the productive failure that is essential to deep learning (Bearman & Ajjawi, 2023). The temporal dimension of the paradigm shift thus requires careful calibration: continuous in capacity but judicious in application, with clear boundaries around what is monitored, when, and for what purpose.

Granularity: From Aggregate to Individual

Current quality assurance operates at the level of programs and cohorts: average graduation rates, aggregate employment statistics, mean satisfaction scores. These aggregate measures, while useful for institutional benchmarking, obscure the variance within populations that is often more informative than the mean (Pellegrino et al., 2001). An agentic AI paradigm shifts the unit of analysis to the individual learning trajectory, using dynamic student modeling to construct personalized competency profiles.

This shift is created by the technical preconditions of adaptive behavior: the AI agent could adjust its assessment strategies based on individual student responses, learning patterns, and demonstrated competencies, much as an expert tutor would (Black & Wiliam, 1998). For a Taiwanese university, this means moving beyond the question "What percentage of graduates in this program are employed within one year?" to the far richer question "How did this student's problem-solving competency develop across their four years, and what learning experiences were most consequential?"

The risk here is equity. Individualized assessment powered by AI could exacerbate existing in-

equalities if the models encode biases from training data, if students with greater digital fluency generate richer learning traces, or if resource-poor institutions cannot implement sophisticated AI systems. Banihashem et al. (2025) caution that learning analytics optimized for formative assessment can inadvertently create "feedback-rich" and "feedback-poor" environments that mirror and amplify existing socioeconomic disparities.

Agency: From Institution-Reported to AI-Observed and Student-Co-Created

Perhaps the most profound dimensional shift concerns who generates assessment evidence. Under the current paradigm, institutions curate self-assessment reports, selecting which data to present and how to frame it — an arrangement that creates inherent conflicts of interest (Section 3.4). Agentic AI introduces two new sources of evidence: autonomous AI observation and student co-creation.

AI agents equipped with tool use capabilities can independently access learning management systems, assignment repositories, and institutional databases, generating evidence that is not filtered through institutional self-reporting (Ng et al., 2021). Simultaneously, students become active co-creators of their assessment evidence, working with AI agents to build competency portfolios that reflect their own understanding of their learning growth. This dual shift — from institutional monopoly to a triangulated evidence ecosystem — addresses the agency asymmetry identified in Section 3 while aligning with constructivist principles that position learners as active meaning-makers rather than passive subjects of measurement (Biggs, 1999).

The accompanying challenge is accountability. When evidence is generated by AI agents rather than human-authored reports, questions arise about responsibility for errors, the interpretability of AI-generated evidence, and the legal and regulatory status of algorithmically produced assessments. Who is accountable when an AI agent's competency assessment contradicts a faculty member's professional judgment?

Feedback Latency: From Retrospective to Real-Time

The current paradigm's feedback loop operates on a timescale of months to years: accreditation reports are submitted, reviewed, deliberated upon, and communicated back to institutions long

after the assessed learning has occurred (Section 3.1). Agentic AI's capacity for iterative reasoning — decomposing complex assessments into manageable steps, executing them sequentially, and refining conclusions based on intermediate results — enables assessment feedback that approaches real time (Masterman et al., 2024).

The mechanism here is not simply faster computation but fundamentally different assessment architecture. Rather than batch-processing assessment at the end of a cycle, an agentic system continuously generates formative feedback that can be acted upon immediately. Black and Wiliam's (1998) landmark meta-analysis demonstrated that formative assessment with timely feedback is among the most powerful interventions for improving learning; agentic AI provides the infrastructure to deliver formative assessment at scale, continuously, and personalized to individual learners.

In a Taiwanese context, this could mean that a program director receives an alert — not in the next accreditation report but next Tuesday — that students in a particular course section are struggling with a threshold concept, accompanied by the AI agent's analysis of likely causes and suggested interventions. The shift from retrospective reporting to real-time intelligence transforms quality assurance from a backward-looking accountability exercise into a forward-looking improvement engine.

The risk is information overload. Real-time feedback systems can generate more data than humans can meaningfully process, leading to alert fatigue and, paradoxically, less responsive institutions (Swiecki et al., 2022). Effective implementation requires AI agents that exercise judgment about when to escalate — a capability that depends on the planning and adaptation functions of agentic systems.

Evidence Type: From Documents to Multimodal Learning Traces

Current quality assurance relies overwhelmingly on textual documents — self-assessment reports, course syllabi, meeting minutes — supplemented by structured statistical data (Section 3.3). The agentic paradigm expands the evidence base to encompass multimodal learning traces: digital ar-

tifacts generated through the learning process itself, including code repositories, design portfolios, laboratory notebooks, video presentations, collaborative documents, peer review exchanges, and reflective journals.

This expansion is made technically feasible by agentic AI's tool use capability — the capacity to interface with diverse digital systems, retrieve heterogeneous data, and synthesize it into coherent evidence narratives (Ng et al., 2021). Rather than relying on students to demonstrate their learning through standardized examinations, the AI agent observes learning as it unfolds across multiple modalities and environments. Shute and Ventura (2013) termed this approach "stealth assessment" — the extraction of evidence about student competencies from their natural interactions with learning environments, without interrupting those interactions with formal testing. A tension exists between stealth assessment's value and the informed consent requirements articulated in Section 6.1.1: assessing students without their awareness, even when pedagogically motivated, risks violating the autonomy principles this paper endorses. A design principle of "transparent stealth assessment" could reconcile this tension — students are informed in advance that their natural learning interactions will generate assessment evidence, and they understand the general mechanisms involved, but the specific moments and modalities of evidence extraction are not signaled in real time, preserving the ecological validity that makes stealth assessment diagnostically valuable.

For Taiwanese universities, multimodal evidence collection could address the persistent criticism that standardized assessments fail to capture the competencies that matter most for professional practice. A nursing student's competency, for example, is better evidenced by her performance in simulated clinical encounters, her reflective practice journals, and her collaborative problem-solving in team-based care scenarios than by her score on a written examination. Agentic AI provides the integrative infrastructure to synthesize these diverse evidence streams into a coherent competency profile.

The challenge is validity. Multimodal learning traces are noisy, contextual, and difficult to standardize. Establishing that an AI agent's synthesis of diverse evidence streams constitutes valid

assessment — that it actually measures what it claims to measure — requires new frameworks for assessment validity that go beyond classical test theory (Mislevy et al., 2003). Evidence-Centered Design, with its explicit modeling of evidence claims, task situations, and scoring rules, provides a promising foundation but must be substantially extended to accommodate agentic AI's dynamic, multimodal evidence generation.

The Construct Validity Challenge

The expansion of evidence types raises a fundamental psychometric concern that must be addressed directly. A core challenge for AI-generated competency assessments is construct validity: the extent to which the AI system is measuring the intended construct (e.g., critical thinking, ethical reasoning, collaborative problem-solving) rather than proxies correlated with that construct (e.g., essay length, keyword usage, response time, frequency of LMS logins). This concern is not unique to AI — human assessors also sometimes confuse proxies with constructs — but it is amplified in AI systems that learn statistical associations from training data without the contextual understanding that enables human assessors to distinguish genuine competency from superficial performance.

The psychometric framework most compatible with AI-augmented assessment is Evidence-Centered Design (ECD), as articulated by Mislevy et al. (2003). ECD's explicit modeling of three interconnected components — the competency model (what is being assessed), the evidence model (what observable behaviors constitute evidence of the competency), and the task model (what tasks elicit those observable behaviors) — provides a principled basis for designing AI assessment systems that can articulate *why* a particular observation counts as evidence of a particular competency. Without such a framework, AI systems risk producing assessments that are statistically reliable (consistent across administrations) but not valid (actually measuring what they claim to measure).

The validity evidence for AI-generated assessments is in its infancy. Automated essay scoring systems have accumulated validity evidence over two decades, demonstrating acceptable agreement with human raters for certain assessment contexts (e.g., standardized writing assessments with well-defined rubrics). However, the extension of AI assessment to complex, ill-structured

competencies — the very competencies that the agentic paradigm aspires to assess — lacks comparable validation. This paper therefore proposes that all pilot implementations of AI-augmented assessment (Phase 1, Section 5.3) include systematic validity studies as a mandatory component. These studies should examine not only statistical agreement between AI and human assessments but also the deeper question of whether AI-generated competency profiles correspond to meaningful differences in student learning — a question that can only be answered through longitudinal follow-up and convergent validation against multiple independent measures.

Assessment Purpose: From Accountability to Improvement

The current paradigm is overwhelmingly summative: its primary purpose is to render judgments about institutional quality for accountability purposes — accreditation decisions, rankings, funding allocations (HEEACT, 2023). While summative assessment serves legitimate social functions, its dominance crowds out formative uses of assessment that directly improve learning (Shepard, 2000). The agentic paradigm shifts the center of gravity from accountability to improvement, from judgment to development, from summative verdicts to formative intelligence.

This shift is made possible by the goal-directed planning capability of agentic AI systems. Rather than passively recording outcomes for later judgment, an agentic assessment system could actively plan interventions: identifying at-risk students, recommending pedagogical adjustments, and generating resources tailored to specific learning gaps (Banihashem et al., 2025). The AI agent does not merely measure; it acts on what it measures, with the explicit goal of improving learning outcomes.

The tension this creates is between assessment for learning and assessment of learning — a tension that Shepard (2000) identified as fundamental to assessment reform. If the same AI system that supports student learning also generates evidence for accreditation decisions, conflicts of interest emerge: students and institutions may strategically manage their interactions with the AI to optimize their accreditation outcomes rather than their learning. Maintaining the integrity of both formative and summative functions within an integrated agentic system requires architectural sepa-

ration — an assessment design challenge that the ADAPT framework's Trust and Ethics component (Section 6) addresses directly.

The Performativity Challenge

Before proceeding to the final dimension, an important counter-argument must be addressed. Taiwan's adoption of outcomes-based education (OBE) provides a documented case study in institutional performativity — the phenomenon whereby institutions adopt the language and documentation of a reform without necessarily changing the underlying practices the reform was designed to improve. As Lin et al. (2020) observed, many Taiwanese institutions adopted OBE frameworks, produced impressive learning outcome documentation, created curriculum maps aligned with competency frameworks, and generated the evidence artifacts required for accreditation — while continuing teaching and assessment practices that remained substantially unchanged. The documentation was genuine; the transformation it documented was, in many cases, superficial.

There is no reason to assume that AI-augmented assessment would be immune to performativity dynamics. Institutions might adopt AI assessment tools for compliance purposes — deploying a learning analytics dashboard, generating AI-processed competency reports, and presenting these artifacts in accreditation documentation — while continuing traditional practices that the AI tools nominally supplement. The risk is particularly acute if accreditation standards reward the *presence* of AI tools rather than evidence of their *impact* on learning: institutions would be incentivized to acquire and display the technology rather than to integrate it meaningfully into their assessment ecology. Worse, institutions might learn to optimize AI-generated metrics without achieving genuine learning improvement — a Goodhart's Law dynamic in which the measure becomes the target.

Several design features could mitigate performativity in AI-augmented assessment. First, process-oriented assessment design: rather than evaluating whether institutions have AI tools, accreditation standards should evaluate how those tools are used — whether they track learning behaviors and processes (not just outputs), whether their outputs inform actual pedagogical adjustments, and whether those adjustments produce measurable learning improvement. Second, external validation

through cross-institutional benchmarking: AI-generated competency profiles from individual institutions should be periodically validated against external measures — standardized assessments, employer evaluations, capstone project reviews by external examiners — to prevent institutional gaming of internal metrics. Third, student voice mechanisms: students' own accounts of how AI-mediated assessment affects their learning experience provide a check on institutional self-reporting that is difficult to fabricate. Fourth, regular recalibration: AI assessment systems should be periodically recalibrated against direct measures of learning to ensure that the competencies they track correspond to genuine educational outcomes rather than to artifacts of the measurement process itself. The proposed AI system audit checklist (Section 5.4.3) should explicitly include performativity indicators as part of its evaluation criteria.

Assessor: From Human Reviewers to Human-AI Collaborative Assessment

The final dimension concerns the identity of the assessor. The current paradigm vests assessment authority in human experts — accreditation panel members, external reviewers, disciplinary specialists — whose professional judgment constitutes the gold standard of quality assurance (HEEACT, 2023). The agentic paradigm does not eliminate human assessors but fundamentally reconfigures their role through multi-agent collaboration and human-in-the-loop oversight.

In this reconfigured model, AI agents handle the labor-intensive work of evidence collection, pattern detection, and preliminary analysis, while human reviewers focus on the interpretive, evaluative, and deliberative tasks that require professional expertise, contextual knowledge, and ethical judgment. This division of labor reflects what Zhong and Zhao (2025) characterize as the necessary complementarity between artificial and human intelligence in the AI-age educational paradigm: AI excels at processing scale, consistency, and pattern recognition; humans excel at meaning-making, contextual judgment, and value-laden deliberation.

The risk is deskilling. If human reviewers increasingly defer to AI-generated analyses, their own assessment expertise may atrophy, creating a dependency that undermines the very human judgment the system is designed to preserve (Bearman & Ajjawi, 2023). Section 4.5 addresses this risk in

detail.

Illustrative Scenario: A Student at a Taiwanese University of Technology

To ground the seven dimensions in concrete experience, consider the following scenario. It is deliberately realistic — set within existing institutional structures and near-term technological capabilities — rather than speculative.

Wei-Lin is a third-year student in the Department of Mechanical Engineering at a Taiwanese university of technology. Under the current paradigm, Wei-Lin's learning would be measured through course grades, a capstone project evaluation, and — years after graduation — through employment surveys that ask whether her job relates to her field of study. These data points would be aggregated with those of her cohort and reported in the department's next self-assessment report.

Under the agentic paradigm, Wei-Lin's learning journey is observed continuously and multimodally. An AI agent integrated with the university's learning management system tracks her engagement with course materials, noting not merely completion but patterns of interaction: which problems she revisits, where she seeks additional resources, how her approach to engineering design problems evolves across semesters. When Wei-Lin submits code for a computational mechanics assignment, the agent analyzes not only correctness but problem-solving strategy — whether she decomposes problems systematically, whether she tests boundary conditions, whether her documentation reflects metacognitive awareness of her reasoning process. This analysis draws on Evidence-Centered Design principles (Mislevy et al., 2003), maintaining explicit models of the competency claims being assessed and the evidence warranting those claims.

In her second semester, the AI agent detects a pattern: Wei-Lin's performance in courses requiring spatial reasoning — computer-aided design, materials science visualization — has plateaued, while her analytical and computational skills continue to develop. Rather than waiting for this pattern to manifest as a poor grade or, worse, to go undetected entirely, the agent generates a formative alert for Wei-Lin's academic advisor, including a preliminary analysis of possible causes and a set

of recommended interventions — additional visualization exercises, a peer tutoring match with a student whose spatial reasoning is strong, a meeting with the course instructor to discuss alternative approaches to design problems.

Wei-Lin herself interacts with the AI agent as a collaborator in her learning. She uses the agent to build a competency portfolio — a dynamic, evidence-rich representation of her developing expertise that she curates alongside the AI's observations. When the agent identifies a strength she had not recognized — an unusual facility for integrating multidisciplinary perspectives in team projects — she adds it to her portfolio with the AI's supporting evidence. When she disagrees with the agent's assessment of her technical writing — she believes it undervalues her ability to communicate complex ideas to non-specialist audiences — she annotates the disagreement, providing counter-evidence from an industry internship report. The portfolio thus becomes a site of constructive alignment (Biggs, 1999), not between learning objectives and assessment tasks, but between AI observation and human self-understanding.

During Wei-Lin's capstone project — designing a component for a local manufacturer's production line — a multi-agent system coordinates assessment across multiple dimensions. One agent evaluates the technical soundness of her engineering design, drawing on domain-specific knowledge bases. Another analyzes her project management behaviors: how she allocates tasks within her team, how she responds to setbacks, how she integrates feedback from the industry mentor. A third synthesizes evidence from her entire trajectory, placing the capstone performance within the context of her four-year development. The faculty evaluator receives not a grade recommendation but an evidence narrative — a rich, contextualized account of what Wei-Lin knows, can do, and has become as an engineer, supported by multimodal evidence and longitudinal trajectory analysis.

For HEEACT accreditation, Wei-Lin's learning evidence — anonymized and aggregated with appropriate privacy protections — contributes to a continuously updated picture of the program's learning outcomes. Accreditation reviewers do not wait for a self-assessment report to learn whether the program's graduates develop the competencies its curriculum promises; they can access real-

time dashboards showing competency trajectories across cohorts, identify emerging gaps before they become systemic problems, and focus their limited site visit time on the interpretive and evaluative questions that require human judgment.

This scenario is not science fiction, but neither is it an accurate description of current capability. Each of its individual elements — learning analytics, adaptive assessment, competency portfolios, multi-agent systems — exists in some form today (Swiecki et al., 2022; Shute & Ventura, 2013; OpenAI, 2025). What the agentic paradigm adds is integration: the capacity to connect these elements into a coherent, continuously operating, goal-directed assessment system. The gap between current capability and this integrated vision remains significant.

Intellectual honesty requires classifying the scenario's elements by their feasibility horizon:

- **Near-term feasible (2026-2028):** Automated feedback on written assignments using large language models; basic learning analytics dashboards tracking student engagement and performance across a learning management system; competency-tagged rubrics linked to program learning outcomes. These capabilities exist today and are deployed, in varying degrees of maturity, in institutions worldwide.

- **Medium-term plausible (2028-2030):** Cross-course competency tracking with persistent memory, enabling an AI system to maintain a learner model across semesters; adaptive assessment sequencing that adjusts evaluation strategies based on accumulated evidence; multi-modal evidence integration that synthesizes written work, project artifacts, and participation data into coherent competency profiles. These capabilities have been demonstrated in prototype or limited deployment but face significant engineering, data infrastructure, and institutional integration challenges.

- **Long-term speculative (2030+):** Fully autonomous multi-agent assessment ecosystems in which specialized AI agents coordinate assessment across multiple dimensions of a capstone project; real-time institutional quality dashboards that aggregate student-level trajec-

tory data into program- and institution-level quality indicators for accreditation purposes; AI-generated "evidence narratives" that synthesize four years of longitudinal data into nuanced accounts of individual learning development. These capabilities are technically conceivable but have not been demonstrated at scale in educational settings.

The Wei-Lin scenario as presented represents a composite vision in which all three horizons operate simultaneously — a state that, while analytically useful for illustrating the paradigm shift's full implications, should not be mistaken for a near-term implementation plan. The phased implementation pathway proposed in Section 5.3 is designed to move through these horizons sequentially, grounding each phase in the empirical evidence generated by its predecessor.

The Role of Human Judgment in the New Paradigm

The enthusiasm generated by agentic AI's capabilities must be tempered by a clear-eyed assessment of what it cannot do. Bearman and Ajjawi (2023) caution that it is "still too early to say whether GenAI is or will be the paradigm shift that people predict" (p. 1), and their caution is well-taken. The ADAPT framework insists that the paradigm shift is not from human assessment to AI assessment but from human-only assessment to human-AI collaborative assessment — a distinction that is conceptually crucial and practically consequential.

The centaur model. In competitive chess, the most formidable players are neither humans nor AI alone but "centaur" teams that combine human strategic intuition with AI computational power (Kasparov, 2017). The assessment centaur pairs AI's capacity for continuous, large-scale, multimodal evidence processing with human capacity for contextual interpretation, ethical judgment, and meaning-making. Neither alone is sufficient. An AI agent can detect that a student's collaborative behavior has changed over a semester; a human assessor is needed to interpret whether that change reflects deepening leadership skills, social withdrawal due to personal difficulties, or strategic free-riding in a team environment.

Faculty as meta-assessors. In the agentic paradigm, faculty do not abdicate assessment respon-

sibility; they elevate it. Rather than spending hours grading assignments — a task at which AI is increasingly competent — faculty become meta-assessors who evaluate the quality of AI-generated evidence and the validity of AI-generated competency claims. This role requires new forms of professional expertise: assessment literacy that encompasses understanding of how AI systems generate evidence, what biases they may encode, and where their judgments require human override. Shepard (2000) argued that assessment reform requires corresponding transformation in teacher preparation; the agentic paradigm makes this requirement acute.

HEEACT reviewers as system auditors. The role of accreditation reviewers likewise transforms. Rather than evaluating whether an institution's self-assessment report accurately reflects its practices — a task limited by information asymmetry and impression management — reviewers evaluate whether the institution's AI assessment system produces valid, reliable, equitable, and transparent evidence (Pellegrino et al., 2001). This is a more demanding role, requiring understanding of both disciplinary content and AI system behavior, but it is also a more consequential one: auditing the system that produces evidence is more valuable than auditing individual evidence artifacts.

The irreducible human domain. Certain dimensions of learning resist algorithmic assessment not because AI is not yet sophisticated enough but because they are constitutively human. Creativity — not the generation of novel combinations, at which AI excels, but the expression of authentic personal vision. Ethical reasoning — not the application of rules to cases, at which AI is competent, but the navigation of genuine moral dilemmas where values conflict and no algorithm can arbitrate. Interpersonal growth — the development of empathy, cultural sensitivity, and the capacity for authentic human connection. These dimensions of learning are not edge cases to be addressed later; they are central to higher education's mission, and their assessment requires human judgment that no agentic system can replicate (Pellegrino et al., 2001). The ADAPT framework's insistence on human-in-the-loop design is not a concession to current technological limitations but a principled commitment to the irreducibly human dimensions of education.

From "Student Success" to "Learning Trajectories": Redefining What We Measure

The paradigm shift described in this section is not merely methodological — a change in how we measure — but ontological: a change in what we measure and, by extension, in what we value. This final subsection argues that the most consequential implication of the agentic paradigm is the redefinition of educational quality itself.

The poverty of endpoint measures. Current "student success" metrics — graduation rates, employment rates, licensure examination pass rates, grade point averages — share a common characteristic: they are endpoint measures that capture destinations without illuminating journeys. A 90% graduation rate tells us nothing about what graduates learned, how they learned it, or whether the institution's educational practices contributed to their development or merely credentialed prior achievement. UNESCO's (2025) provocative question — "What's worth measuring?" — challenges the assessment community to move beyond easily quantifiable endpoints toward the richer, messier, more consequential question of how students grow.

Learning trajectories as the new unit of analysis. The agentic paradigm enables a fundamental shift in the unit of assessment from endpoints to trajectories — the dynamic paths through which learners develop competencies, dispositions, and identities over time. Zhong and Zhao's (2025) analysis of educational paradigm shifts in the AI age exemplifies this reorientation, arguing that continuous formative assessment aligned across temporal and spatial dimensions — tracking learning development across semesters, courses, and programs — must replace periodic summative snapshots. Agentic AI provides the infrastructure to operationalize such temporally aligned assessment at scale, maintaining persistent models of individual learning trajectories that capture not only what students know at any given moment but how their knowledge, skills, and dispositions are developing.

Process over product. A trajectory-oriented paradigm privileges process over product. How does a student approach an unfamiliar problem? Does she decompose it systematically or proceed by trial

and error? How does she respond to failure — with persistence, strategic adjustment, or avoidance? How does she integrate feedback from peers, instructors, and AI agents into her subsequent work? These process-oriented questions are more diagnostically valuable than product-oriented questions ("Did the student get the right answer?") and more predictive of long-term professional success (Pellegrino et al., 2001). Agentic AI's capacity for continuous observation and multi-step reasoning makes process-oriented assessment feasible at a scale that was previously impossible.

Competency trajectories and metacognitive development. Beyond content knowledge and technical skills, the agentic paradigm enables tracking of metacognitive development — students' growing awareness of their own learning processes, their ability to self-regulate, and their capacity to transfer learning across contexts. Banihashem et al. (2025) demonstrate that learning analytics optimized for formative purposes can provide rich evidence of metacognitive processes, including self-monitoring, strategy selection, and reflective adjustment. When combined with agentic AI's persistent memory and adaptive behavior, such analytics can generate longitudinal metacognitive profiles that capture what may be higher education's most important outcome: the capacity to learn how to learn.

The ontological shift. Ultimately, the move from static measurement to dynamic learning evidence entails a redefinition of educational quality. Quality is no longer a property of institutions — something they possess and demonstrate through self-assessment — but a property of learning experiences: dynamic, relational, and emergent. This ontological shift aligns with constructivist learning theory (Biggs, 1999), which holds that learning is not a transfer of knowledge from institution to student but an active process of meaning construction. The agentic paradigm, for the first time, provides assessment infrastructure commensurate with this theoretical commitment.

Yet this ontological shift is itself nested within a deeper dynamic. The move from endpoints to trajectories presupposes that "trajectory" is a stable unit of analysis — that we need merely shift from measuring destinations to tracking paths. But the co-evolution of AI and human cognition means that the nature of the path itself is changing: how learners interact with AI is evolving,

the cognitive ecology constituted by human-AI collaboration is being reshaped, and the answer to "what constitutes a meaningful learning trajectory" is not fixed. The following subsection examines this co-evolutionary dynamic and its fundamental implications for assessment design.

The Co-Evolution Dynamic: Why a One-Time Paradigm Shift Is Insufficient

The Kuhnian framework employed throughout this paper provides a powerful analytical lens: it identifies an old paradigm beset by anomalies, a crisis precipitated by converging pressures, and a new paradigm that resolves those anomalies. But Kuhn's model, designed for scientific revolutions, carries an implicit assumption that fits uneasily with the phenomenon under analysis: it presupposes that the new paradigm, once established, will achieve a period of "normal science" — a stable equilibrium that persists until the next round of anomalies accumulates. The argument of this subsection is that no such equilibrium is available in the domain of AI-augmented learning assessment, because the relationship between AI and human learning is not static but *co-evolutionary* — and the definitions of learning that any assessment framework must operationalize are themselves continuously being reconstructed by the technologies embedded in educational practice.

Three mechanisms of cognitive transformation. Cognitive science identifies at least three mechanisms through which AI is transforming human learning and cognition, each with distinct implications for assessment.

Cognitive offloading refers to the process by which individuals delegate cognitive tasks to external tools, freeing internal cognitive resources for other purposes (Risko & Gilbert, 2016). When students use AI to retrieve factual information, perform routine calculations, or generate first drafts, they offload tasks that once constituted core learning activities. Critically, offloading is not monolithic: it can be *strategic* — deliberately freeing cognitive resources for higher-order synthesis, creativity, and ethical reasoning — or *maladaptive* — creating over-reliance that atrophies the very capacities education aims to develop. Sparrow et al.'s (2011) seminal finding that internet access changes what people remember — the "Google effect" shifting memory toward source locations rather than content — is being amplified by AI tools that are more capable and more deeply

integrated into learning practice than any prior technology. The assessment implication is that measurement systems must distinguish between students who offload strategically (demonstrating metacognitive resource management) and those who offload dependently (masking underdeveloped competencies) — a distinction that Vygotsky's (1978) zone of proximal development helps illuminate, as AI effectively creates a dynamic, personalized scaffolding zone that continuously recalibrates what learners can accomplish with and without assistance.

Cognitive augmentation describes the enhancement of human cognitive capabilities through AI partnership. When a student uses an AI system to visualize complex data, explore alternative framings of a problem, or receive real-time feedback on reasoning quality, the human-AI ensemble achieves cognitive performances that neither human nor AI could accomplish independently. Clark and Chalmers' (1998) extended mind thesis provides the philosophical foundation: if a tool is reliably available, consistently used, and its outputs are automatically endorsed by the user, then the tool constitutes part of the user's cognitive system. In educational terms, this means that "what a student knows" may be inseparable from "what tools a student can effectively mobilize" — a redefinition with profound implications for assessment validity.

Cognitive restructuring denotes the deeper, slower process by which sustained interaction with AI tools reorganizes internal cognitive architecture. Hayles (2012) theorizes this as *technogenesis*: the recursive process through which humans and technics co-evolve, with each reshaping the other. Sterelny (2010) describes the construction of *cognitive niches* — environments that humans actively engineer to scaffold their own cognitive development. A philosophical tension should be acknowledged here: Clark and Chalmers' (1998) extended mind thesis, cited above, holds that tools *constitute* part of the cognitive system, while Sterelny's scaffolding view treats tools as environmental supports *to* cognition without being constitutive of it. This paper need not resolve this debate; for quality assurance purposes, both positions converge on the same practical implication: assessment must account for the cognitive ecosystem — the full ecology of human-AI interaction — rather than treating the isolated biological learner as the sole unit of analysis. In the AI era,

students do not merely use AI tools; they grow up in cognitive environments saturated by AI, developing reasoning patterns, attention structures, and epistemic habits that are qualitatively different from those of pre-AI cohorts. The assessment implication is temporal: what counts as evidence of "critical thinking" or "creative problem-solving" for a generation that has always had AI as a cognitive partner may differ fundamentally from what counted for prior generations — and will differ again for subsequent ones.

The co-evolutionary loop. These three mechanisms — offloading, augmentation, and restructuring — do not operate in isolation; they constitute a recursive loop between technology adoption and cognitive transformation:

1. AI technologies are deployed in educational settings, changing what cognitive tasks students perform independently and what they delegate to AI.
2. Students' cognitive practices, epistemic habits, and learning strategies adapt in response — what they remember, how they reason, what they value as knowledge.
3. These cognitive changes alter what "learning" means in practice — the competencies that matter, the processes that constitute genuine understanding, the performances that demonstrate mastery.
4. The changed definitions of learning create new requirements for assessment systems and new expectations for how AI should be adopted in education.
5. These new adoption patterns further transform the cognitive ecology, and the cycle continues.

This loop means that the relationship between AI and learning is not unidirectional — AI as a tool applied to a fixed educational target — but bidirectional and recursive: *AI changes how humans learn, and the changes in how humans learn reshape how AI should be adopted, which further changes how humans learn.* Hutchins' (1995) distributed cognition framework, which demonstrates that cognitive processes are distributed across people, tools, and environments rather than contained within individual minds, provides the theoretical architecture for understanding this dynamic. In the agentic AI context, cognition is distributed across students, AI agents, learning man-

agement systems, and the assessment infrastructure itself — and the distribution is continuously being reconfigured.

Implications for assessment design: the adaptive definition framework. If learning definitions are co-evolving with technology, then assessment frameworks cannot be designed around fixed definitions and expect to remain valid. The practical implication is that the paradigm shift theorized in this paper should not be conceived as a one-time move from Paradigm A (snapshots) to Paradigm B (trajectories) that then stabilizes. Rather, it should be conceived as the establishment of an *adaptive assessment architecture* — a system designed to track and respond to the ongoing evolution of what learning means.

Concretely, this entails three design principles that extend the ADAPT framework:

First, *definitional monitoring*: the assessment system should include mechanisms for periodically revisiting and revising the operational definitions of learning outcomes in light of technological and cognitive change. Just as the Kuhnian framework requires vigilance for new anomalies, an adaptive assessment architecture requires vigilance for definitional drift — the gradual divergence between what the system measures and what learning has become. Operationally, this means HEEACT's fourth-cycle indicators could include a required annual review, at the program level, of what counts as "independent student work" and "demonstrated competency" — definitions that must be updated as AI capabilities evolve rather than fixed for a six-year cycle.

Second, *human-AI boundary tracking*: as cognitive offloading and augmentation shift the boundaries of what students do independently versus what they accomplish through AI partnership, assessment must continuously recalibrate what it attributes to the learner versus what it attributes to the tool. This is not a one-time calibration but an ongoing negotiation as tools evolve and students' relationships with them change. Luckin's (2018) "interwoven intelligence" framework, which identifies seven elements of human intelligence that AI cannot replicate and maps how human and machine capabilities relate, provides a useful starting point. Operationally, accreditation

self-assessment reports could include a "human-AI contribution map" for each assessed learning outcome, documenting which competency demonstrations involve AI partnership and how the institution distinguishes learner capability from tool capability.

Third, *generational sensitivity*: assessment criteria should be developed with awareness that successive cohorts of students will have qualitatively different cognitive relationships with AI. The "digital natives" concept (now widely critiqued) captured an important intuition — that growing up with technology shapes cognition — but applied it too coarsely. The co-evolutionary framework suggests that even within a single decade, the cognitive profiles of entering students may shift meaningfully as AI tools become more capable and more pervasive. Operationally, longitudinal comparison of cohort learning profiles should adjust baselines for AI-native versus pre-AI cohorts, and accreditation criteria should be reviewed on a shorter cycle (e.g., triennially) than the full accreditation period to accommodate this generational variation.

Why this matters for Taiwan's fourth cycle. The co-evolutionary perspective reinforces the case for Scenario B (framework evolution) over Scenario A (conservative integration) or Scenario C (paradigm replacement). Scenario A, by maintaining fixed assessment definitions, is structurally unable to accommodate definitional evolution. Scenario C, by committing to a comprehensive new paradigm, risks crystallizing a new set of definitions that will be overtaken by the next round of co-evolutionary change. Scenario B, with its incremental revision mechanisms and structured piloting, provides the institutional flexibility needed to treat assessment definitions as living constructs rather than permanent fixtures — provided the design principles articulated above are built into the fourth-cycle architecture from the outset.

The seven dimensions of the paradigm shift — temporality, granularity, agency, feedback latency, evidence type, assessment purpose, and assessor identity — collectively describe a transformation that is neither purely technological nor purely conceptual but both simultaneously. Agentic AI

provides the enabling infrastructure; the ADAPT framework provides the analytical structure; and the ontological shift from endpoints to trajectories provides the conceptual foundation. Yet frameworks and technologies do not transform educational systems; policies do. Section 5 turns to the question of how Taiwan's quality assurance policies — specifically HEEACT's accreditation standards — might evolve to enable, regulate, and govern this paradigm shift, translating theoretical possibilities into institutional realities. ## 5. Policy and Accreditation Implications

The preceding sections have established that agentic AI possesses the theoretical capacity to transform learning outcome measurement from periodic, retrospective snapshots into continuous, adaptive evidence streams. Yet theoretical potential alone does not reshape institutional practice. The critical mediating variable is policy – specifically, how Taiwan's Ministry of Education (MOE) and the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT) choose to position agentic AI within the nation's quality assurance (QA) architecture. This section addresses RQ4 by analyzing three policy scenarios for integrating agentic AI into Taiwan's accreditation framework, conducting a comparative evaluation across multiple dimensions, and proposing a phased implementation pathway. Most critically, it offers concrete recommendations for the fourth cycle of institutional accreditation (第四週期大學校院校務評鑑), whose design window represents an immediate and consequential opportunity. The analysis draws on Bardach and Patashnik's (2019) eightfold path for policy analysis while remaining grounded in Taiwan's specific regulatory context, institutional culture, and the cautionary lessons emerging from peer nations' early experiments with AI in education. To make the eightfold path's analytical structure visible, the section's organization maps to Bardach's steps as follows: problem definition (the structural limitations identified in Section 3 and the policy challenge of integrating agentic AI); evidence assembly (the technological capability analysis of Section 2 and the South Korean cautionary tale in Section 5.2); alternative construction (the three policy scenarios in Section 5.1); criteria selection (the multi-criteria evaluation matrix in Table 4); outcome projection (the comparative analysis across feasibility, equity, cost, and timeline); confronting tradeoffs (the acknowledgment that no scenario dominates across all criteria); decision (the recommendation of a phased approach beginning with Scenario

B); and telling the story (the concrete fourth-cycle recommendations in Section 5.4 that translate the analysis into actionable guidance). Bardach's framework is thus not merely an inspiration but a structuring principle for the policy analysis that follows.

Three Policy Scenarios

The integration of agentic AI into higher education quality assurance is not a binary proposition. Following the scenario planning methodology widely employed in technology governance (OECD, 2023), this subsection presents three distinct scenarios representing points along a continuum from minimal disruption to paradigm replacement. Each scenario carries different assumptions about institutional readiness, regulatory appetite, and the pace of technological maturation.

Scenario A – Conservative Integration

Under Scenario A, agentic AI functions as a supplementary analytics tool operating entirely within the existing HEEACT framework. The third cycle's four standards and their associated Core Indicators (核心指標) remain unchanged. Institutions that wish to deploy AI-driven assessment tools – intelligent tutoring systems, automated rubric scoring, predictive learning analytics – do so voluntarily, at their own discretion and expense. HEEACT evaluators continue to examine traditional evidence artifacts: course syllabi, grade distributions, student surveys, capstone portfolios, and employer feedback. AI-generated evidence, if presented, is treated as supplementary documentation rather than primary accreditation evidence.

This scenario's principal advantage is its low regulatory risk. No legislative amendments, no revisions to accreditation handbooks, and no new evaluator competencies are required. It respects what Lin et al. (2020) identified as the foundational principle of Taiwan's QA system: institutional autonomy (大學自主) balanced with public accountability. Institutions experimenting with AI tools can do so without fear that their innovations will be penalized by evaluators unfamiliar with the technology.

However, conservative integration carries significant costs. Most critically, it risks creating what Coburn (2003) termed a "two-tier system" – a growing divide between well-resourced institutions

that leverage AI to enhance learning outcomes and those that lack the capacity or incentive to do so. When accreditation frameworks fail to recognize innovative practices, they implicitly devalue them, sending a signal to institutional leaders that AI investment carries reputational risk without accreditation reward. Moreover, Scenario A forecloses the possibility of system-level learning: without structured data collection on AI-augmented assessment practices across institutions, HEEACT cannot build the evidence base needed to inform future policy iterations.

Scenario B – Framework Evolution

Scenario B envisions a deliberate but incremental evolution of the HEEACT accreditation framework to formally recognize AI-generated learning evidence. Under this scenario, HEEACT revises the descriptors for selected Core Indicators – particularly those under Standard 3 (教學與學習) – to incorporate new dimensions reflecting AI-augmented assessment infrastructure. For instance, the descriptor for Core Indicator 3-2, which currently emphasizes "direct and indirect assessment mechanisms for student learning outcomes" (HEEACT, 2023, p. 38), would be expanded to include "continuous learning evidence mechanisms, including AI-mediated formative assessment data where institutionally appropriate."

New indicators might be introduced under Standard 3, such as "AI-augmented assessment infrastructure" and "learning analytics for student support." The Self-Assessment Report (SAR; 自我評鑑報告) template would be modified to include an optional appendix on AI in assessment, requiring institutions that deploy such tools to document their data governance, algorithmic transparency measures, and bias mitigation protocols. Critically, this scenario maintains the existing accreditation cycle structure – periodic self-assessment followed by on-site visits – but enriches the evidentiary palette that evaluators can draw upon.

The strengths of Scenario B lie in its balance between innovation and continuity. Taiwan's accreditation culture has matured over three cycles spanning more than fifteen years; institutions and evaluators have developed shared understandings of quality evidence, visit protocols, and improvement-oriented dialogue (Lin et al., 2021). Scenario B leverages this accumulated QA culture rather than

discarding it. It also aligns with the approach emerging in several peer QA agencies: the European Standards and Guidelines (ESG) revision discussions have emphasized incorporating digital assessment within existing quality frameworks rather than creating parallel systems (ENQA, 2015).

The principal limitation of Scenario B is pace. Incremental revision of accreditation standards typically requires multi-year consultation processes involving MOE, HEEACT staff, institutional representatives, student groups, and industry stakeholders. If AI-augmented assessment tools evolve as rapidly as current trajectories suggest (Gartner, 2025), framework evolution risks perpetual lag – the standards always one generation behind the technology. Additionally, Scenario B demands significant HEEACT staff capacity building; evaluators must develop sufficient technical literacy to assess AI assessment systems meaningfully, not merely check compliance boxes.

Scenario C – Paradigm Replacement

Scenario C represents the most radical departure: an entirely new continuous assurance model that replaces the periodic SAR-plus-on-site-visit structure with real-time quality monitoring. Under this scenario, institutions maintain AI-mediated evidence dashboards that stream learning outcome data continuously to HEEACT. Agentic AI systems generate, curate, and present evidence of student learning in real time – not as static tables in a document submitted every six years, but as living data ecosystems that HEEACT can access on demand. The on-site visit, if retained at all, becomes a targeted "deep dive" triggered by anomalies in the continuous data stream rather than a scheduled ritual.

HEEACT's role transforms fundamentally under Scenario C: from periodic evaluator to continuous quality partner. Rather than assembling evaluation teams every six years, HEEACT would maintain ongoing analytical relationships with institutions, providing formative feedback, benchmarking data, and early warning signals. This vision aligns with what Temper et al. (2025) described in the HEAT-AI framework as "embedded quality intelligence" – AI systems that do not merely measure quality but actively participate in its improvement.

The appeal of Scenario C is its conceptual elegance: it fully realizes the paradigm shift from assessment-of-learning to assessment-as-learning that the theoretical framework in Section 2 articulated. It eliminates the assessment lag problem – the months or years between data collection and accreditation judgment – that has long been recognized as a structural weakness of periodic QA systems (Ewell, 2009).

Yet the obstacles are formidable. The infrastructure costs alone – standardized data architectures, secure API connections between every HEI and HEEACT, AI system maintenance – would be substantial. Equity concerns are paramount: Taiwan's higher education system encompasses more than 140 institutions ranging from research universities with robust IT departments to small colleges of technology with minimal digital infrastructure (MOE, 2024). A continuous assurance model risks excluding precisely those institutions that most need quality support. International recognition presents another challenge; no major QA framework globally has fully operationalized continuous assurance, meaning Taiwan would be pioneering without established benchmarks. Finally, the South Korean experience with AI-driven educational transformation – discussed below – offers a sobering reminder that speed of technological deployment does not guarantee quality of educational outcomes.

Comparative Analysis of Scenarios

To systematically evaluate the three scenarios, Table 4 presents a multi-criteria assessment matrix. The criteria reflect the dimensions most salient to Taiwan's policy context: implementation feasibility, equity impact across diverse institution types, QA integrity maintenance, international recognition compatibility, estimated cost, implementation timeline, and institutional readiness requirements.

Table 4

Comparative Evaluation Matrix for Agentic AI Integration Scenarios

Criterion	Scenario A: Conservative	Scenario B: Evolution	Scenario C: Replacement
Feasibility	High – no regulatory change	Moderate – requires HEEACT handbook revision	Low – requires new legislation, infrastructure
Equity impact	Negative – widens gap between resourced and under-resourced institutions	Moderate – optional indicators reduce pressure on less-ready institutions	High risk – may exclude institutions lacking digital infrastructure
QA integrity	Maintained – existing standards unchanged	Maintained with enhancement – new indicators add rigor	Uncertain – untested model with no international precedent
International recognition	High – aligns with current Washington Accord, ESG	High – consistent with ESG revision trajectory	Low initially – no peer QA agency has adopted this model
Estimated cost	Minimal (institutional-level only)	Moderate (HEEACT capacity building + handbook revision)	Very high (national infrastructure + ongoing maintenance)
Timeline	Immediate	2-4 years for standards revision	5-10+ years for full implementation
Institutional readiness requirement	Low – voluntary adoption	Moderate – institutions must document AI practices	Very high – all institutions must maintain real-time data systems
Risk of unintended consequences	Low but stagnation risk	Moderate – depends on indicator design quality	High – South Korea cautionary parallel

No single scenario dominates across all criteria. Scenario A minimizes risk but maximizes opportunity cost; Scenario C maximizes transformative potential at prohibitive near-term cost; Scenario B occupies the pragmatic middle ground. This convergence on hybrid approaches is consistent with Bardach and Patashnik's (2019) observation that effective policy analysis sequences radical goals through incremental steps.

The South Korean Cautionary Tale

South Korea's 2024 plan to deploy AI-powered digital textbooks across all schools by 2025 provides a cautionary parallel. The rollout provoked severe backlash: parents protested increased screen time, teachers reported poor calibration to diverse needs, and the government was forced into significant rollback by mid-2025 (Rest of World, 2025). Three lessons apply directly: technological readiness does not equal stakeholder readiness; equity of access is a precondition, not an afterthought; and phased piloting with genuine evaluation is essential. These lessons argue against Scenario C as a near-term strategy.

Recommended Phased Approach

Drawing on the scenario analysis and the South Korean cautionary evidence, this paper recommends a three-phase implementation pathway that begins with conservative piloting, transitions to framework evolution based on empirical evidence, and selectively introduces elements of paradigm transformation for institutions with demonstrated readiness.

Phase 1: Structured Piloting (2026-2028). Under Phase 1, MOE and HEEACT jointly select 10-15 higher education institutions representing diverse institutional types – research universities, comprehensive universities, and universities of technology – to participate in a structured pilot of AI-augmented assessment. Participating institutions receive modest funding support and technical guidance to implement agentic AI tools in selected programs. HEEACT develops provisional guidelines for documenting AI-generated learning evidence, and evaluators conduct supplementary reviews (not formal accreditation) of AI assessment practices. Critically, Phase 1 includes a rigorous evaluation component: independent researchers assess the validity, reliability, equity, and

pedagogical impact of AI-augmented assessment across pilot sites. Taiwan's AI Basic Act (人工智慧基本法), enacted in December 2025 with its seven governing principles – including human autonomy, privacy protection, transparency, and fairness (Legislative Yuan, 2025) – provides the legal framework for ensuring that pilot implementations comply with national AI governance standards.

Phase 2: Framework Evolution (2028-2030). Phase 2 commences only after Phase 1 evaluation data demonstrate that AI-augmented assessment can produce valid, equitable, and pedagogically valuable learning evidence. Based on pilot findings, HEEACT initiates formal revision of Core Indicator descriptors under Standards 3 and 4, incorporating new language for AI-mediated assessment evidence. The SAR template is updated to include the AI assessment appendix described under Scenario B. HEEACT launches an evaluator capacity-building program, including international benchmarking visits to peer QA agencies (see Section 5.5) and specialized training modules on AI system audit. Phase 2 aligns with the anticipated launch window for the fourth cycle of institutional accreditation, creating a natural integration point for revised standards.

Phase 3: Selective Transformation (2030+). For self-accrediting institutions (自辦評鑑機構) that have demonstrated mature AI assessment infrastructure through Phases 1 and 2, Phase 3 introduces selective elements of Scenario C: continuous quality monitoring dashboards, real-time evidence streams, and a more fluid relationship between institution and quality assurance body. Critically, Phase 3 maintains the traditional accreditation pathway as a fully legitimate alternative. No institution is compelled to adopt continuous assurance; rather, it becomes an option for institutions whose digital maturity warrants it. This dual-pathway model ensures that the system does not replicate the equity failures observed in the South Korean rollback (Rest of World, 2025).

5.3.1 The Political Economy of Implementation

The phased approach outlined above assumes a degree of coordination among stakeholders that the political economy of Taiwan's higher education governance does not guarantee. The relationship between the key actors — MOE, HEEACT, and universities — involves overlapping but distinct

authorities, incentives, and constraints that must be explicitly addressed.

The political economy of this transition should not be underestimated. HEEACT's position as an independent foundation provides operational flexibility but limits its authority to mandate AI adoption; MOE controls funding but may prioritize other policy objectives; and universities face competing demands for limited resources. Specifically, HEEACT develops accreditation standards and administers the evaluation process, but MOE retains approval authority over the framework's overall design and the regulatory consequences of accreditation outcomes. Universities implement assessment practices but have varying degrees of autonomy: self-accrediting institutions enjoy considerable procedural flexibility, while institutions under standard accreditation operate within tighter regulatory constraints.

Budget allocation presents a particularly thorny coordination challenge. Who funds the AI infrastructure required for even modest piloting? Three potential funding streams exist, each with limitations. First, MOE could allocate resources through the Higher Education Sprout Project, which already includes technology-enhanced teaching quality as a funded dimension; however, Sprout Project funding is competitive and time-limited, creating sustainability risks. Second, institutions could fund AI assessment infrastructure from their own budgets; however, the institutions most in need of improved assessment are precisely those with the most constrained budgets (see Section 5.3.2 below). Third, private sector partnerships with educational technology vendors could provide infrastructure; however, such partnerships raise vendor dependency and data governance concerns that must be carefully managed (see Section 5.3.3 below).

Among these stakeholders, faculty deserve particular attention as the primary implementers of any assessment change. Continuous AI-augmented assessment implies a fundamental shift in faculty workload — from periodic grading to ongoing engagement with AI-generated evidence streams — that may be experienced as an additional burden rather than a liberation, particularly at teaching-intensive institutions where course loads are already high. Faculty AI literacy is a prerequisite, not an afterthought: without understanding how AI assessment systems generate evidence, faculty

cannot meaningfully exercise the "meta-assessor" role described in Section 4.5. There is also a legitimate concern that AI-mediated continuous monitoring may be perceived by faculty as surveillance of their teaching practice rather than support for student learning — a perception that, if unaddressed, could generate resistance analogous to the South Korean backlash. These concerns argue for faculty involvement in the design process from the outset, not merely as recipients of a predetermined system.

Beyond faculty, multiple additional stakeholders have interests that must be coordinated: the Taiwan AI College Alliance (TAICA), which could provide shared infrastructure; university presidents and academic vice presidents, who must champion AI assessment integration; student associations, whose members are the ultimate subjects of the transformation; and technology vendors, whose commercial incentives may not align with educational values. Aligning these diverse stakeholders around a common implementation vision requires the kind of deliberate, inclusive consultation process that HEEACT has successfully managed in previous accreditation cycle transitions — but on a compressed timeline, given the pace of technological change.

5.3.2 The Resource Paradox

A fundamental resource paradox underlies any proposal for AI-augmented assessment: institutions most vulnerable to the demographic crisis are precisely those least capable of investing in AI infrastructure. Without deliberate policy intervention, AI-augmented assessment could deepen rather than narrow the quality divide. Four equity mechanisms are essential: shared AI infrastructure through TAICA's 55-member network; dedicated MOE funding streams prioritizing under-resourced institutions; open-source assessment tools from publicly funded research; and regional AI assessment hubs anchored at resource-rich institutions serving clusters of smaller ones. These are preconditions for equitable implementation, not optional supplements.

5.3.3 AI Vendor Dynamics

Commercial AI providers pose three risks: vendor lock-in limiting institutional autonomy, data extraction for non-educational purposes, and misaligned incentives favoring rapid deployment over

careful implementation. HEEACT's proposed "AI in Assessment Standards" (Section 6.4) should mandate data ownership clarity, interoperability requirements, and open standards (xAPI, cmi5, Open Badges) to ensure student competency records remain portable regardless of vendor platform.

Implications for the Fourth Cycle of Institutional Accreditation

The third cycle of institutional accreditation (第三週期大學校院校務評鑑) concludes in academic year 2025; the design of the fourth cycle represents the most immediate and consequential opportunity to position Taiwan's QA framework for the agentic AI era. This subsection offers specific, actionable recommendations for five dimensions of fourth-cycle design: Core Indicator revision, SAR template updates, on-site visit protocol adjustments, the role of self-accrediting institutions, and evaluator capacity building.

Revised Standard 3 Core Indicators

The third cycle's Standard 3, "Student Learning and Outcomes" (學生學習與成效), contains four Core Indicators (3-1 through 3-4) that address curriculum design, assessment mechanisms, learning support, and teaching quality improvement respectively (HEEACT, 2023). Table 5 proposes specific revisions to three of these indicators to accommodate AI-generated learning evidence while preserving the standard's existing evaluative logic.

Table 5

Proposed Fourth-Cycle Core Indicator Revisions Under Standard 3

Core Indicator	Current Descriptor (3rd Cycle)	Proposed Additional Dimension (4th Cycle)	Rationale
3-1: Curriculum Design and Institutional Goals	Examines alignment between institutional mission, program learning outcomes, and curriculum structure	Add: "Digital assessment infrastructure readiness – the institution demonstrates capacity to collect, manage, and analyze digital learning evidence, including AI-mediated assessment data where deployed"	Ensures institutions developing AI assessment tools have the foundational infrastructure; does not mandate AI adoption but recognizes it as a legitimate dimension of curriculum delivery

Core Indicator	Current Descriptor (3rd Cycle)	Proposed Additional Dimension (4th Cycle)	Rationale
3-2: Assessment of Student Learning Outcomes	Evaluates direct and indirect assessment mechanisms; examines use of assessment results for continuous improvement	Add: "Continuous learning evidence mechanisms – the institution may demonstrate learning outcome achievement through continuous, technology-mediated evidence streams (e.g., learning analytics, AI-formative assessment data, competency tracking systems) alongside or as complement to traditional periodic assessments"	Expands the evidentiary palette without displacing traditional assessment; aligns with UNESCO (2023) guidance on diverse assessment modalities

Core Indicator	Current Descriptor (3rd Cycle)	Proposed Additional Dimension (4th Cycle)	Rationale
3-3: Student Support and Learning Resources	Examines tutoring, counseling, learning resources, and support for diverse learners	Add: "AI-augmented learning analytics for student support and early intervention – where institutions deploy predictive analytics or AI-driven early warning systems, evaluators examine their effectiveness, equity, and integration with human advising"	Recognizes growing use of learning analytics dashboards while insisting on human-in-the-loop support structures; addresses equity by examining whether AI tools serve all student populations

Several design principles undergird these proposals. First, each addition is framed as an "additional dimension" rather than a replacement, preserving backward compatibility with institutions that have not adopted AI tools. Second, the language consistently uses conditional phrasing – "where deployed," "the institution may demonstrate" – to avoid mandating specific technologies. Third, each proposed dimension explicitly references equity and human oversight, reflecting the AI Basic Act's principles of fairness and human autonomy (Legislative Yuan, 2025). Fourth, the proposals are calibrated to be evaluable: evaluators can assess whether an institution's digital infrastructure is adequate, whether continuous evidence mechanisms produce valid data, and whether AI-driven student support reaches all learner populations.

SAR Template Updates

The Self-Assessment Report is the primary documentary vehicle through which institutions present their quality evidence to HEEACT. The fourth-cycle SAR template should incorporate two new elements:

AI in Assessment Appendix. Institutions deploying AI-driven assessment tools would complete an optional appendix documenting: (a) the specific AI tools in use, their purposes, and the scope of their deployment; (b) data governance protocols, including data sources, storage, access controls, and retention policies; (c) algorithmic transparency measures – how the institution ensures that AI-generated judgments (e.g., automated scoring, learning trajectory predictions) are interpretable to faculty and students; (d) bias testing results – evidence that AI tools have been evaluated for differential performance across student subpopulations defined by gender, socioeconomic status, disability, and indigenous identity; and (e) student consent and communication protocols – how students are informed about AI's role in their assessment and what opt-out mechanisms exist. This appendix draws on the framework proposed by Temper et al. (2025) and aligns with the transparency and privacy requirements of Taiwan's AI Basic Act.

Learning Evidence Dashboard Option. In lieu of or in addition to traditional static data tables (e.g., grade distributions, graduation rates, employer survey results), institutions may provide evaluators with access to a real-time learning evidence dashboard. Such dashboards would display longitudinal learning outcome data, including formative assessment trends, competency achievement trajectories, and learning analytics summaries. To ensure comparability, HEEACT would publish minimum display standards specifying required data fields, visualization formats, and data currency requirements. This option recognizes that static tables in a document submitted months before an on-site visit represent an increasingly anachronistic mode of evidence presentation in a data-rich environment (OECD, 2023).

On-Site Visit Protocol Adjustments

The on-site visit (實地訪評) remains the cornerstone of Taiwan's accreditation process, providing evaluators with contextual understanding that documentary evidence alone cannot convey (HEEACT, 2023). Three adjustments would prepare the on-site visit protocol for the agentic AI era.

First, evaluator competency requirements should be expanded to include the ability to assess AI assessment system quality and integrity. This does not mean every evaluator must be a machine learning expert; rather, at least one member of each evaluation team should possess sufficient technical literacy to examine an institution's AI tools with informed skepticism – understanding, for instance, the difference between a validated adaptive assessment engine and a superficially impressive chatbot that lacks psychometric grounding.

Second, HEEACT should develop an AI system audit checklist for use during on-site visits. Drawing on Singapore's AI Verify toolkit (IMDA, 2020) and TEQSA's guidance on learning analytics (TEQSA, 2024), the checklist would cover: data sources and their provenance; model training procedures and validation evidence; algorithmic transparency and explainability provisions; bias testing methodology and results; system reliability and failure mode protocols; student consent documentation; and faculty governance structures for AI assessment oversight. The checklist serves a dual purpose: it provides evaluators with a structured assessment tool and signals to institutions the specific dimensions of AI governance that HEEACT considers essential.

Third, the on-site visit interview protocol should be expanded to include questions about AI governance, faculty AI literacy, and student experience with AI-mediated assessment. Sample interview questions might include: "How does the institution's assessment committee oversee AI-driven assessment tools?" (for administrators); "What training have you received on interpreting AI-generated learning analytics, and how do they inform your teaching?" (for faculty); and "How were you informed about the role of AI in your coursework assessment, and do you feel the process is transparent and fair?" (for students). These questions ensure that the human dimensions of AI

integration – governance, literacy, trust, and experience – receive evaluative attention commensurate with the technical dimensions.

Self-Accrediting Institutions as Early Adopters

Taiwan's QA system includes a category of self-accrediting institutions (自辦評鑑機構) – universities that have earned the privilege of conducting their own accreditation processes subject to HEEACT meta-evaluation. These institutions, typically research-intensive universities with mature QA cultures, possess two characteristics that make them natural early adopters for AI-augmented accreditation practices. First, they have greater procedural flexibility; their accreditation processes need not follow the standard HEEACT template in all particulars. Second, they typically command greater financial and technical resources, reducing the equity barriers that would constrain less-resourced institutions.

HEEACT should develop "AI-Ready Self-Accreditation Guidelines" (AI 就緒自辦評鑑指引) that enable self-accrediting institutions to serve as pilot sites for Scenario B practices during Phase 2 of the recommended implementation pathway. These guidelines would specify: minimum requirements for AI assessment tool documentation; expected evidence of validity and equity testing; governance structures for AI oversight; and reporting protocols that allow HEEACT to aggregate learning across self-accrediting institutions. The meta-evaluation process for self-accrediting institutions would then include assessment of AI integration maturity as a supplementary dimension, generating the institutional evidence base needed to inform broader framework evolution.

Evaluator Training and Capacity Building

Perhaps the most critical – and most frequently underestimated – element of fourth-cycle preparation is evaluator capacity building. Taiwan's accreditation evaluators are predominantly senior academics and administrators who volunteer their expertise; few have backgrounds in artificial intelligence, learning analytics, or educational data science. Without targeted capacity building, even well-designed indicators and protocols will be implemented superficially.

HEEACT should establish a three-component evaluator development program. The first compo-

nent is foundational workshops on AI in assessment, covering core concepts (machine learning, natural language processing, learning analytics), common applications in higher education, and the critical questions evaluators should ask when examining AI systems. These workshops need not produce technical experts; rather, they should cultivate what Selwyn (2019) termed "critical digital literacy" – the ability to ask penetrating questions about technology claims without being intimidated by technical jargon.

The second component is international benchmarking. HEEACT should organize study visits to peer QA agencies that are furthest advanced in AI governance – particularly Singapore's Council for Private Education (CPE), which benefits from the national Model AI Governance Framework and AI Verify testing toolkit (IMDA, 2020); Australia's TEQSA, which has published leading guidance on learning analytics and AI in higher education (TEQSA, 2024); and selected European agencies participating in ENQA's digital quality assurance working group. These visits would expose HEEACT staff and evaluator candidates to international best practices and help Taiwan avoid reinventing solutions that peer agencies have already developed.

The third component is the creation of a new evaluator specialty profile: the "digital assessment specialist" (數位評量專家). Individuals holding this designation would have demonstrated competence in evaluating AI assessment systems and would be assigned to evaluation teams visiting institutions with significant AI deployment. Over time, as AI integration becomes more widespread, the skills associated with this specialty would diffuse into the general evaluator population; in the near term, however, a dedicated specialist role ensures that at least one team member can conduct technically informed assessment of AI systems.

International Comparative Reference

Taiwan does not confront the challenge of AI-augmented quality assurance in isolation. A brief survey of peer QA agencies' responses to AI provides both benchmarks and cautionary reference points.

The European Standards and Guidelines (ESG), last revised in 2015, are currently undergoing review discussions that explicitly consider digital transformation in higher education. ENQA member agencies have published working papers on AI in quality assurance, generally favoring an evolutionary approach consistent with Scenario B – integrating AI considerations within existing standards rather than creating parallel frameworks (ENQA, 2015). The ESG revision process, however, is notoriously slow; any updated standards are unlikely before 2027 at the earliest, suggesting that Taiwan could position itself as a regional leader by moving more decisively.

Australia's Tertiary Education Quality and Standards Agency (TEQSA) has been among the most proactive QA agencies globally. TEQSA published guidance on artificial intelligence in higher education in 2024, addressing both teaching applications and quality assurance implications. Notably, TEQSA has emphasized provider responsibility for ensuring the integrity of AI-mediated assessment – a framing that aligns with the proposed AI in Assessment appendix for Taiwan's SAR template (TEQSA, 2024).

The United Kingdom's Quality Assurance Agency (QAA) has published multiple position papers on AI and quality, including guidance on academic integrity in the age of generative AI. QAA's approach has emphasized adaptability – encouraging institutions to develop local AI policies within a national framework of quality expectations – rather than prescribing specific technological solutions (QAA, 2023).

In the United States, the Council for Higher Education Accreditation (CHEA) has established an AI and accreditation working group, reflecting growing recognition that regional accreditors must develop AI-related competencies. However, the decentralized nature of U.S. accreditation – with multiple regional and specialized accreditors operating independently – has slowed the development of unified guidance.

Singapore merits particular attention as a comparator for Taiwan. The Infocomm Media Development Authority's (IMDA) Model AI Governance Framework, now in its second edition, provides

a national-level structure for responsible AI deployment across sectors, including education. The AI Verify toolkit – an open-source testing framework that allows organizations to assess their AI systems against governance principles – represents precisely the kind of practical tool that HEEACT could adapt for accreditation purposes (IMDA, 2020). Singapore's EdTech Masterplan 2030, which articulates a vision for technology-enhanced education grounded in evidence and equity, offers a strategic planning model that Taiwan's MOE could reference in developing its own AI-in-education roadmap (MOE Singapore, 2023).

Collectively, the international landscape reveals a convergent pattern: no major QA agency has adopted anything resembling Scenario C's paradigm replacement, but most are actively moving beyond Scenario A's conservative stance toward some version of Scenario B's framework evolution. Taiwan's recommended phased approach is thus consistent with international trends while positioning the nation to lead within the Asia-Pacific region – particularly given HEEACT's established reputation as one of Asia's most mature QA agencies and its active roles in INQAAHE and APQN (Lin et al., 2021).

The policy and accreditation recommendations presented in this section are predicated on a fundamental assumption: that the integration of agentic AI into learning outcome measurement can be governed ethically, equitably, and transparently. This assumption is far from trivial. The following section examines the ethical dimensions of AI-augmented assessment in depth, interrogating the tensions between algorithmic efficiency and human dignity, between data richness and privacy, and between innovation and the irreducible value of human judgment in education.

Ethical Considerations and Risk Governance

The deployment of agentic artificial intelligence in student success measurement introduces ethical challenges that are qualitatively distinct from those posed by conventional educational technology. Unlike passive analytics dashboards or rule-based early warning systems, agentic AI operates with

autonomous decision-making authority, persistent memory, and iterative reasoning capabilities that fundamentally alter the power dynamics between institutions, educators, and students. As Taiwan advances toward integrating such systems within its higher education quality assurance infrastructure, a rigorous ethical analysis is not merely advisable—it is a prerequisite for responsible innovation. This section applies the principlist bioethics framework of Beauchamp and Childress (2019) to the educational AI context, constructs a comprehensive risk matrix, identifies risks unique to agentic architectures, and proposes a three-tier governance framework calibrated to Taiwan's legal, cultural, and institutional landscape.

The Four-Principle Ethical Analysis

The principlist framework, originally developed for biomedical ethics, has been increasingly adopted in technology ethics scholarship as a structured approach to evaluating complex sociotechnical systems (Floridi et al., 2018). Its four principles—autonomy, beneficence, non-maleficence, and justice—provide a systematic lens through which the ethical implications of agentic AI in assessment can be examined.

Autonomy

Autonomy concerns manifest along three dimensions: informed consent, the right to opt out, and data sovereignty. Continuous monitoring of student learning behaviors raises fundamental consent questions, as agentic AI deepens the asymmetric power relationship Slade and Prinsloo (2013) identified in learning analytics: students are no longer merely observed but subject to autonomous decisions regarding their competency status and intervention pathways. The right to opt out is particularly vexing — if AI is embedded in core assessment architecture, opting out may mean opting out of the educational experience itself (Zuboff, 2019). Taiwan's Personal Data Protection Act (個人資料保護法) provides a legal foundation, but its application to educational contexts remains untested. Data ownership presents a third concern: when agentic AI constructs longitudinal competency profiles, who owns that profile? Taiwan's AI Basic Act (2025) — a framework law (基本法) that articulates governing principles but requires implementing regulations (施行細則) for enforcement — does not yet resolve data ownership questions specific to educational contexts.

Beneficence

The principle of beneficence requires that interventions produce demonstrable benefits that justify their costs and risks. The evidence base for AI-enhanced learning, while promising, remains unevenly distributed across contexts and time horizons. Kestin et al. (2025), in a randomized controlled trial at Harvard University, found that students using an AI tutoring system achieved learning gains approximately twice those of students in traditional active learning environments—a striking result that has attracted considerable attention. However, this study involved a specific, well-resourced context (introductory physics at an elite research university) with a narrowly defined outcome measure (immediate post-instruction assessment performance). Whether such gains persist over time, transfer to other disciplines, or replicate in the resource-constrained environments typical of many Taiwanese higher education institutions remains an open empirical question.

More broadly, the benefits attributed to agentic AI in assessment—personalized feedback, early intervention for at-risk students, continuous competency tracking, and adaptive learning pathway optimization—are largely theoretical or demonstrated only in short-term pilot studies (Temper et al., 2025). The UNESCO Guidance for Generative AI in Education and Research cautioned against conflating the potential benefits of AI with demonstrated benefits, noting that the evidence base for AI in education is "still emerging and largely inconclusive" for many claimed applications (UNESCO, 2023, p. 18). A responsible application of the beneficence principle thus requires institutions to treat agentic AI as an experimental intervention subject to ongoing evaluation, rather than a proven solution to be deployed at scale.

Non-Maleficence

Four categories of harm merit attention. First, *surveillance normalization*: continuous behavioral monitoring risks eroding the intellectual freedom and exploratory risk-taking essential to genuine learning (Zuboff, 2019). Second, *algorithmic bias*: Baker and Hawn (2022) documented systematic under-prediction of academic performance for minoritized students. In Taiwan's context, analogous risks exist for indigenous students (原住民), new immigrant children (新住民子女), and

socioeconomically disadvantaged students — populations whose learning behaviors may diverge from majority-trained AI models' expectations. Third, *de-professionalization of teaching*: when AI assumes assessment decisions previously requiring faculty judgment, professional expertise may erode through "skill degradation through disuse" (Parasuraman et al., 2000) — a risk with particular cultural resonance in Taiwan's tradition of faculty autonomy. Fourth, *AI hallucination and error propagation*: LLM reliability research shows minor input errors can reduce assessment accuracy by 30% or more (Chinta et al., 2024), and within agentic architectures, such errors compound through autonomous decision chains.

Justice

The justice principle demands equitable distribution of benefits and burdens. In Taiwan's higher education system, structural inequalities between national and private universities, between urban and rural institutions, and between well-resourced research universities and smaller teaching-focused institutions create conditions under which the benefits of agentic AI are likely to accrue disproportionately to already-advantaged institutions.

The resource requirements for deploying agentic AI—computational infrastructure, technical expertise, data science personnel, and ongoing system maintenance—are substantial. Without deliberate policy intervention, the digital divide between institutions may widen rather than narrow, producing a two-tier quality assurance system in which elite institutions leverage sophisticated AI-driven assessment while under-resourced institutions rely on manual processes. This concern is not speculative: Gandara et al. (2024), analyzing predictive analytics systems in United States higher education, found that algorithmic models produced 19% false negatives for Black students and 21% for Latinx students—meaning that nearly one in five students from these groups who would have succeeded were incorrectly flagged as at-risk. The implications for Taiwan's marginalized populations—indigenous peoples, new immigrant children, students with disabilities, and those from economically disadvantaged families—are significant. If agentic AI systems trained predominantly on data from majority student populations are deployed without rigorous equity testing,

they risk systematically mischaracterizing the abilities and potential of precisely those students who most need institutional support.

Risk Matrix

Table 6 synthesizes the ethical analysis into a structured risk matrix that can serve as a practical governance tool for institutional decision-makers.

Table 6

Risk Matrix for Agentic AI Deployment in Student Success Measurement

Risk Category	Likelihood	Severity	Mitigation Strategy
Algorithmic bias in assessment	High	High	Regular bias audits with disaggregated demographic analysis; diverse and representative training data; mandatory equity impact assessments prior to deployment
Student surveillance normalization	High	Medium	Clear data governance policies communicated to students; informed consent frameworks with genuine opt-out mechanisms; purpose limitation principles

Risk Category	Likelihood	Severity	Mitigation Strategy
Faculty de-professionalization	Medium	High	Human-in-the-loop mandates for all consequential assessment decisions; faculty AI literacy programs; professional development integrating AI as augmentation rather than replacement
Data breach or privacy violation	Medium	Critical	Privacy-by-design architecture; full compliance with the Personal Data Protection Act (個資法); data minimization; encryption at rest and in transit; regular penetration testing

Risk Category	Likelihood	Severity	Mitigation Strategy
Digital divide amplification	High	High	MOE infrastructure funding earmarked for under-resourced institutions; shared AI infrastructure through inter-institutional consortia; cross-institutional resource pooling modeled on the TAICA framework
Over-reliance on AI judgment	Medium	High	Mandatory human review thresholds for consequential decisions; student appeal mechanisms with guaranteed human adjudication; periodic calibration between AI and faculty assessments

Risk Category	Likelihood	Severity	Mitigation Strategy
International non-recognition of AI-assessed credentials	Low	Medium	Alignment with ESG (Standards and Guidelines for Quality Assurance in the EHEA) principles; Washington Accord compliance for professional programs; leverage HEEACT's INQAAHE full membership standing
AI-generated assessment gaming	Medium	Medium	Continuous validation against external measures; multi-modal evidence requirements; process-oriented assessment design that resists metric optimization

Risks Unique to Agentic AI

A critical contribution of this analysis is the distinction between ethical risks that are common to all AI applications in education and those that are specific to agentic AI architectures. Much existing scholarship on AI ethics in education addresses general concerns—privacy, bias, transparency—that apply to any computational system processing student data. The risks enumerated below, however, arise specifically from the autonomous, persistent, and multi-agent characteristics that

define agentic AI.

Autonomy amplification. Unlike passive AI tools that present recommendations for human decision-makers, agentic AI operates with delegated authority to make assessment decisions. When an AI agent autonomously determines that a student has not met a competency threshold—triggering consequences such as course failure, remediation requirements, or delayed graduation—the question of accountability becomes genuinely novel. Traditional accountability structures assume a human decision-maker whose reasoning can be interrogated, challenged, and overridden. Agentic AI disrupts this assumption. The institution deploys the system; the developer designs the architecture; the AI agent executes the decision. Responsibility is diffused across a chain of actors, none of whom may have full visibility into the specific reasoning that produced a particular outcome. This "accountability gap" (Danaher, 2016) is not merely a theoretical concern—it has practical implications for student grievance procedures, legal liability, and institutional accreditation.

Unpredictable behavior through iterative reasoning. Agentic AI systems reason iteratively, decomposing complex assessment tasks into sub-tasks, executing them in sequence, and adjusting their approach based on intermediate results. While this iterative capacity enables sophisticated assessment strategies, it also introduces a fundamental unpredictability that distinguishes agentic systems from rule-based alternatives. A rule-based early warning system produces deterministic outputs: given the same inputs, it will always generate the same alert. An agentic AI system, by contrast, may pursue different reasoning pathways on different occasions, producing assessment outcomes that are difficult to predict, reproduce, or explain. This stochasticity challenges the principles of consistency and fairness that are foundational to assessment practice (Mislevy et al., 2003).

Persistent memory and compounding bias. One of the defining features of agentic AI architectures for student success measurement is their capacity to maintain longitudinal student profiles—accumulating data across courses, semesters, and years to construct comprehensive competency trajectories. While this persistence enables valuable longitudinal analysis, it also creates a mechanism by which early assessment errors or biased judgments can compound over time. A student

who is incorrectly assessed as deficient in a foundational competency during the first year may find that this initial error shapes all subsequent AI-driven assessments, creating a self-reinforcing cycle of under-evaluation. This "bias compounding" risk is distinct from the static bias documented in conventional algorithmic systems (Baker & Hawn, 2022) and represents a uniquely pernicious threat to equitable assessment.

Multi-agent coordination failures. Advanced agentic architectures envision multiple specialized AI agents collaborating on assessment tasks—one agent monitoring engagement, another evaluating written work, a third tracking competency progression. When these agents coordinate successfully, the result is a comprehensive assessment ecosystem. When coordination fails, however, responsibility for errors becomes diffused to the point of untraceability. If a student receives an incorrect competency assessment, determining which agent contributed to the error—and how to correct it—requires forensic analysis of multi-agent interaction logs that may not exist in interpretable form.

Goal misalignment and metric gaming. Perhaps the most insidious risk unique to agentic AI is the possibility that systems optimized for "student success metrics" may learn to optimize the metrics rather than the underlying learning they are intended to measure. This phenomenon, sometimes termed Goodhart's Law in computational contexts ("when a measure becomes a target, it ceases to be a good measure"), takes on new dimensions when the optimizer is an autonomous agent capable of strategic behavior. An agentic AI system that is rewarded for improving retention rates, for example, might learn to lower assessment thresholds rather than improve learning support—a strategy that improves the metric while undermining the educational mission it purports to serve.

A Governance Framework for Agentic AI in Assessment

Addressing the risks identified above requires a governance framework that operates simultaneously at national, institutional, and technical levels. Drawing on Taiwan's existing legal infrastructure, the recently enacted AI Basic Act, and international best practices, this section proposes a three-tier governance architecture.

Tier 1: National Governance (MOE and HEEACT)

At the national level, governance should be anchored in Taiwan's AI Basic Act, which articulates seven guiding principles: human autonomy, privacy protection, transparency, fairness and non-discrimination, safety and security, accountability, and sustainability. These principles provide a normative foundation, but they require operationalization in the specific context of educational assessment.

The Ministry of Education (教育部), in collaboration with HEEACT, should develop "AI in Assessment Standards" (AI 評量標準) that translate the Act's principles into concrete requirements for higher education institutions. These standards should mandate (a) registration of all AI systems used in consequential assessment decisions as part of institutional accreditation documentation, (b) annual equity impact assessments that disaggregate AI system performance by demographic characteristics including indigenous status, new immigrant background, socioeconomic status, and disability, and (c) minimum transparency requirements specifying that students must be informed when AI systems contribute to assessment decisions and must have access to explanations of how those decisions were reached. HEEACT's established role as Taiwan's national quality assurance agency, and its full membership in INQAAHE, positions it to develop these standards with both domestic credibility and international alignment.

Tier 2: Institutional Governance

At the institutional level, each university deploying agentic AI in assessment should establish an AI Assessment Ethics Committee (AI 評量倫理委員會), modeled on the institutional review boards (IRBs) that govern human subjects research. This committee should include faculty representatives, student representatives, data science experts, and external ethics advisors. Its mandate should encompass pre-deployment review of AI assessment systems, ongoing monitoring of system performance and equity outcomes, and adjudication of student complaints regarding AI-driven assessment decisions.

Institutions should further adopt a Student Data Bill of Rights (學生資料權利宣言) that guar-

antees five core rights: (a) informed consent—students must affirmatively consent to AI-driven assessment, with genuine alternatives available for those who decline; (b) access—students must be able to view all data collected about them and the competency assessments derived from that data; (c) correction—students must be able to challenge and correct inaccurate data or assessments; (d) deletion—upon graduation or withdrawal, students must be able to request deletion of their behavioral and assessment data; and (e) portability—students must be able to export their competency records in interoperable formats. Taiwan's Personal Data Protection Act provides a legal basis for these rights, but institutional policies must operationalize them in the educational context where the Act's general provisions require domain-specific interpretation.

Faculty AI literacy should be treated not as an optional professional development opportunity but as a prerequisite for any instructor whose courses incorporate AI-driven assessment. Literacy programs should address not only the technical operation of AI systems but also their ethical implications, their limitations, and the critical judgment required to interpret and override AI recommendations when warranted. This approach positions AI as a tool that augments professional expertise rather than one that supplants it, directly addressing the de-professionalization risk identified in the ethical analysis.

Tier 3: Technical Governance

Technical governance requirements should ensure that the systems themselves are designed and operated in accordance with ethical principles. Four requirements are essential.

First, algorithmic transparency. Institutions deploying agentic AI must be able to provide meaningful explanations of how assessment decisions are reached. This does not necessarily require full model interpretability—which may be technically infeasible for complex agentic systems—but it does require what Floridi et al. (2018) termed "explicability": the capacity to provide an understandable account of why a particular decision was made, even if the full computational process cannot be rendered transparent.

Second, bias testing protocols. AI assessment systems must undergo bias testing both prior to deployment and on an ongoing basis. Pre-deployment testing should use synthetic and historical data to evaluate system performance across demographic groups. Post-deployment monitoring should continuously track outcome disparities and trigger automatic review when disparities exceed pre-defined thresholds. The FairAIED framework proposed by Chinta et al. (2024) provides a technically rigorous methodology for such testing that could be adapted to Taiwan's demographic context.

Third, data minimization. Consistent with both the Personal Data Protection Act and the AI Basic Act's privacy principle, agentic AI systems should collect only the minimum data necessary for their assessment functions. The temptation to collect comprehensive behavioral data because it is technically feasible must be resisted in favor of principled data collection guided by clearly articulated assessment purposes.

Fourth, interoperability standards. To protect student data portability and prevent vendor lock-in, AI assessment systems should adhere to established learning data standards, including xAPI and cmi5 for learning records and Open Badges for competency credentials. Interoperability is not merely a technical convenience—it is an ethical requirement that ensures students retain meaningful control over their educational records.

The ethical analysis and governance framework presented in this section reveal that the deployment of agentic AI in student success measurement is not merely a technical undertaking but a fundamentally normative one. The risks are real, the evidence base is incomplete, and the governance infrastructure is nascent. Yet Taiwan possesses distinctive institutional advantages—a mature quality assurance ecosystem in HEEACT, a recently enacted AI Basic Act, a strong data protection legal tradition, and a culture of inter-institutional collaboration—that position it to develop governance practices that could serve as regional models. The question is not whether to engage with agentic

AI in assessment, but how to do so in a manner that honors the autonomy of students, the expertise of faculty, and the equity commitments that define a just higher education system. The following section synthesizes the findings across all research questions and considers their implications for policy, practice, and future inquiry.

Discussion and Future Directions

Integrating the ADAPT Framework: Key Insights

This paper set out to examine whether and how agentic AI could transform student learning outcome measurement in Taiwan's higher education system. Having traversed the conceptual terrain of agentic AI capabilities (Section 2), mapped the structural limitations of the current measurement paradigm (Section 3), proposed the ADAPT framework as an integrative analytical tool (Section 4), evaluated policy scenarios and accreditation recommendations (Section 5), and conducted a rigorous ethical analysis (Section 6), it is now possible to step back and consider what the assembled analysis reveals when its parts are read as a whole.

The most consequential finding is the convergence between technological capability and structural need. The six limitations identified in Section 3 — temporal, granularity, modality, agency, competency capture, and indirect measurement dominance — are not arbitrary deficiencies; they are the predictable consequences of an assessment architecture designed for a world of periodic review, documentary evidence, and aggregate judgment. The six capabilities of agentic AI identified in Section 2 — autonomous planning, dynamic adaptation, tool use, multi-step reasoning, persistent memory, and multi-agent collaboration — map onto these limitations with striking directness. Persistent memory addresses the temporal limitation by enabling continuous rather than periodic evidence collection. Individual learner modeling addresses the granularity limitation by tracking trajectories at the student level rather than the program level. Multimodal evidence integration addresses the modality limitation by capturing learning processes alongside learning products. Autonomous observation addresses the agency limitation by generating evidence that is not exclusively institution-curated. Multi-step reasoning addresses the competency capture limitation by

enabling evaluation of complex competencies — critical thinking, creative problem-solving, ethical reasoning — that resist reduction to countable outputs. And the shift from survey-based to performance-based evidence addresses the indirect measurement dominance that characterizes the current paradigm.

This convergence is what the Kuhnian framework would predict: paradigm shifts occur not because a new theory is abstractly superior, but because it resolves the specific anomalies that the reigning paradigm cannot (Kuhn, 1962/2012). The ADAPT framework, as proposed in Section 4, operationalizes this convergence by mapping the pathway from anomaly identification (the "D" in ADAPT — Diagnostic Mapping) through assessment reconception (the second "A" — Assessment Reconception) to policy implementation (the "P" — Policy Pathways) and ethical safeguarding (the "T" — Trust & Ethics Safeguards). The framework's analytical utility lies precisely in this integration: it does not treat technology, policy, and ethics as separate conversations but as interdependent dimensions of a single transformation.

A second key insight concerns the seven dimensions of paradigm shift identified in Section 4. These dimensions — temporality (periodic to continuous), granularity (aggregate to individual), agency (institution-reported to AI-observed and student-co-created), feedback latency (retrospective to real-time), evidence type (documents and surveys to multimodal learning traces), assessment purpose (accountability to improvement), and assessor identity (human-only to human-AI collaborative) — do not merely describe a technological upgrade. They describe a fundamentally different epistemology of learning measurement. The current paradigm asks: "Did this institution demonstrate, through its own curated documentary evidence, that it has mechanisms in place for assessing learning outcomes?" The emerging paradigm asks: "What is actually happening in student learning, as captured through continuous, multi-source, multi-modal evidence streams that no single stakeholder controls?" The shift is from measuring institutional compliance to measuring educational reality — a distinction whose implications for accreditation practice are profound.

A third insight, developed in Section 4.7, concerns the co-evolutionary dynamic between AI and

human learning. The analysis reveals that learning definitions are not stable targets awaiting better measurement instruments but moving constructs that are continuously reshaped by the technologies embedded in educational practice. Cognitive offloading changes what students remember; cognitive augmentation changes what students can accomplish; cognitive restructuring changes how students think (Clark & Chalmers, 1998; Sparrow et al., 2011; Hayles, 2012; Sterelny, 2010). This co-evolutionary loop — AI changes learners, changed learners change how AI should be adopted, which further changes learners — means that the paradigm shift theorized in this paper should not be conceived as a one-time transition that then stabilizes. The ADAPT framework must therefore incorporate mechanisms for *definitional monitoring* (tracking how learning definitions evolve), *human-AI boundary tracking* (recalibrating what is attributed to the learner versus the tool), and *generational sensitivity* (acknowledging that successive student cohorts have qualitatively different cognitive relationships with AI). Without these adaptive mechanisms, any assessment framework — however well-designed at inception — will gradually drift from the educational reality it purports to measure.

A fourth insight concerns the reusability of the ADAPT framework beyond Taiwan's specific context. While this paper has applied the framework to the intersection of agentic AI and HEEACT's accreditation system, the framework's five-layer analytical structure — define the technology, diagnose the current paradigm's limitations, reconceive the assessment model, evaluate policy pathways, and govern ethically — is applicable to any national quality assurance system confronting the same technological disruption. Quality assurance agencies in Southeast Asia (AQAN members), South Asia (NAAC in India), and the Middle East (CAA in the UAE) face structurally analogous challenges: periodic accreditation cycles, document-based evidence, and limited capacity for continuous quality monitoring. The ADAPT framework offers these agencies a structured analytical pathway for examining how agentic AI intersects with their own assessment architectures, without prescribing a one-size-fits-all solution. Future comparative research could apply the framework across national contexts to identify which dimensions of the paradigm shift are universal and which are context-dependent.

What This Paper Does NOT Claim

Intellectual honesty requires explicit delineation of what this analysis does and does not assert. Four boundaries are particularly important.

First, this paper does not claim that agentic AI will inevitably transform higher education assessment. The paradigm shift theorized here is a possibility, not a prophecy. It depends on contingent factors — institutional readiness, faculty acceptance, regulatory will, infrastructure investment, and the pace of technological maturation — any one of which could stall or redirect the transformation. The South Korean cautionary tale discussed in Section 5 demonstrates that even aggressive governmental commitment to AI deployment can be reversed by stakeholder resistance. Technology determinism — the assumption that technological capability automatically translates into institutional adoption — is precisely the analytical error this paper seeks to avoid.

Second, this paper does not claim that Taiwan's current assessment paradigm is worthless. The HEEACT accreditation framework, developed and refined over three cycles spanning more than fifteen years, represents a sophisticated and internationally recognized quality assurance system. It has cultivated a culture of self-evaluation, continuous improvement, and evidence-based decision-making across Taiwan's higher education sector (Lin et al., 2021). The argument is not that this framework should be discarded but that it should be evolved — that its foundational commitments to quality, equity, and accountability can be more fully realized through assessment architectures that leverage technological capabilities its original designers could not have anticipated.

Third, this paper does not claim that AI should replace human judgment in educational assessment. The ethical analysis in Section 6 and the governance framework proposed therein are predicated on the opposite conviction: that human judgment — the expertise of faculty, the deliberation of evaluation panels, the contextual understanding that comes from sustained engagement with educational communities — is irreducible and irreplaceable. Agentic AI can augment this judgment by providing richer, more timely, more granular evidence; it cannot and should not supplant the

normative and relational dimensions of educational evaluation that only human beings can provide.

Fourth, this paper does not provide empirical evidence that agentic AI improves learning outcomes. As a theoretical and policy-analytical paper, it constructs conceptual frameworks, evaluates policy scenarios, and proposes governance structures. The empirical validation of these frameworks — through longitudinal implementation studies, randomized controlled trials, and quasi-experimental designs — is the work of subsequent research, not of this paper.

Limitations of the Analysis

Several limitations constrain the analysis and should be acknowledged transparently.

The most fundamental limitation is the paper's theoretical character. The claims advanced here — that agentic AI can address the structural limitations of the current paradigm, that the ADAPT framework accurately maps the dimensions of the paradigm shift, that the recommended policy pathway balances innovation with caution — are arguments from theoretical reasoning and analogical evidence, not from empirical observation. Until agentic AI assessment systems are implemented in Taiwanese higher education institutions and their effects are rigorously evaluated, the claims remain provisional. The evidence base for AI in education, as UNESCO (2023) has cautioned, is "still emerging and largely inconclusive" for many applications; this caution applies with even greater force to agentic AI, which is newer and less studied than the generative AI tools that dominate the current literature.

Second, the analysis is Taiwan-specific. The ADAPT framework was developed with reference to HEEACT's accreditation structure, MOE's policy instruments, Taiwan's AI Basic Act, and the particular demographic, economic, and institutional conditions of Taiwan's higher education system. While Section 7.1 argued for the framework's broader applicability, this applicability is conjectural until tested in other national contexts. Quality assurance systems differ significantly in their regulatory authority, cultural context, institutional diversity, and technological infrastructure; what works in Taiwan may not transfer directly to other settings.

Third, the technological landscape is evolving with a rapidity that challenges any analysis published in a traditional academic format. Agentic AI capabilities that are speculative today may be routine by the time this paper is read; conversely, capabilities that appear promising may encounter technical barriers that are not yet apparent. The four-level taxonomy proposed in Section 2, while grounded in the best available evidence as of early 2026, may require revision as the technology matures.

Fourth, this paper lacks primary stakeholder data. The perspectives of faculty, students, institutional administrators, employers, and accreditation evaluators — the individuals who would be most directly affected by the paradigm shift theorized here — are represented only through secondary sources. A comprehensive policy analysis would ideally incorporate structured consultations, surveys, or interviews with these stakeholders. Their absence here is acknowledged as a significant gap that future research must address.

Fifth, the co-evolutionary thesis advanced in Section 4.7 — that AI and human cognition are mutually reshaping each other — is, as a theoretical framework, difficult to falsify: virtually any observed change in learning behavior could be interpreted as evidence of co-evolution. While this is a common feature of broad theoretical frameworks (and is acknowledged rather than denied), it means the framework's value lies in its generative and analytical utility — the design principles it produces — rather than in its predictive specificity.

Sixth, the analysis does not engage deeply with the technical feasibility of implementing agentic AI assessment systems at scale. Questions of computational infrastructure, data architecture, system interoperability, and maintenance costs are acknowledged as critical but are treated at a level of generality that does not do justice to their complexity. The phased implementation pathway proposed in Section 5 assumes that these technical challenges are tractable; that assumption must be validated through pilot implementations.

Future Research Agenda

The limitations identified above define, in inverse, a research agenda of considerable scope and urgency. Six priorities merit particular attention.

Priority 1: Longitudinal Empirical Studies. The Phase 1 structured piloting recommended in Section 5 should be accompanied by rigorous mixed-methods evaluation research spanning multiple semesters — quantitative analysis of learning outcome validity and reliability, qualitative investigation of faculty and student experience, and quasi-experimental comparison with traditional assessment. The shift from snapshots to trajectories cannot be validated through short-term pilots.

Priority 2: Equity Impact Studies. Dedicated research should examine whether agentic AI assessment produces differential outcomes for indigenous students (原住民), new immigrant children (新住民子女), students with disabilities, and socioeconomically disadvantaged populations, adapting the methodological models of Baker and Hawn (2022) and Gandara et al. (2024) to Taiwan's context.

Priority 3: Faculty Experience and Acceptance. Research drawing on technology acceptance models (TAM), I-TPACK frameworks, and qualitative studies of faculty professional identity would provide essential evidence for designing implementation strategies that secure faculty engagement rather than resistance.

Priority 4: Cross-National Comparative Studies. Systematic comparison of approaches by HEEACT (Taiwan), TEQSA (Australia), QAA (United Kingdom), CHEA (United States), and emerging frameworks in Singapore and South Korea would identify which challenges are universal and which are context-specific.

Priority 5: ADAPT Framework Validation. Case study research applying the framework to specific institutional implementations would test its analytical utility and identify dimensions requiring refinement.

Priority 6: Technical Interoperability Standards. Research on applying existing learning data standards (xAPI, cmi5, Open Badges) to agentic AI assessment contexts is essential for ensuring portable, auditable, and institutionally interoperable learning evidence.

Conclusion

This paper began with a paradox: a higher education system that aspires to prepare students for a rapidly changing world, assessed through mechanisms that change only incrementally. It ends with a proposition: that the convergence of agentic AI capabilities and accumulated structural limitations in Taiwan's assessment architecture creates the conditions for a paradigm shift — not a gradual improvement in existing methods, but a fundamental reconception of what it means to measure student learning.

The question facing Taiwan's quality assurance community is how the intersection of AI and assessment will be managed: through deliberate governance or ad hoc accommodation, through equity-centered policy or market-driven diffusion. Three key findings warrant emphasis.

First, agentic AI creates the technical preconditions for qualitatively new measurement modalities — continuous competency tracking, personalized assessment, multimodal evidence integration, and longitudinal developmental trajectories — that address precisely the structural limitations the current paradigm cannot resolve from within its own logic.

Second, the shift requires governance, not just technology. The ethical analysis in Section 6 demonstrated that the risks of agentic AI in assessment — algorithmic bias, surveillance normalization, faculty de-professionalization, accountability diffusion, and persistent memory bias compounding — are not incidental side effects but intrinsic features of the technology's autonomous, adaptive, and persistent character. The three-tier governance framework proposed in this paper — national standards anchored in the AI Basic Act, institutional ethics committees modeled on research ethics boards, and technical governance mandating transparency, bias testing, data minimization, and in-

teroperability — is not an optional supplement to AI deployment but a precondition for responsible deployment. Technology without governance is not innovation; it is negligence.

Third, Taiwan possesses institutional advantages that position it to lead. HEEACT's international standing — full INQAAHE compliance, CHEA recognition, active APQN membership — provides a platform from which Taiwan's approach to AI-augmented quality assurance can influence regional and global practice. The AI Basic Act provides a legal foundation that many peer nations lack. The mature quality assurance culture cultivated over three accreditation cycles, encompassing shared understandings of evidence, evaluation, and improvement across more than 140 institutions, provides the institutional substrate on which AI-augmented assessment can be built. And Taiwan's demographic crisis, while deeply challenging, creates a clarity of purpose: in a system where institutional survival depends on demonstrated educational value, the incentive to develop more valid, more timely, and more actionable learning outcome measurement is not theoretical but existential.

The fourth cycle of institutional accreditation represents the decisive moment. The concrete recommendations advanced in Section 5 — revised Core Indicator descriptors under Standard 3, an AI in Assessment appendix for the Self-Assessment Report, on-site visit protocol adjustments including an AI system audit checklist, self-accrediting institutions as early adopters, and a three-component evaluator capacity-building program — are calibrated to be implementable within the fourth-cycle design window without requiring legislative action or prohibitive infrastructure investment. They represent Scenario B, the framework evolution pathway: ambitious enough to position Taiwan at the forefront of AI-augmented quality assurance in the Asia-Pacific, cautious enough to avoid the stakeholder backlash and equity failures that derailed South Korea's more aggressive approach.

The call to action is specific: begin with structured pilots in the 2026–2028 window, build evaluator capacity through international benchmarking, and embed the results into a fourth-cycle framework that formally recognizes AI-generated learning evidence while preserving human judgment and institutional autonomy. The risks of premature deployment without governance are too consequential to ignore; equally, the risks of inaction while peer systems advance are too significant to justify in-

definite delay.

The shift from snapshots to trajectories is, ultimately, a shift from measuring institutions to understanding learners. But this paper has argued for an additional recognition: that the learners we seek to understand are themselves being transformed by the technologies we deploy. Any assessment system designed for a fixed definition of learning will eventually be overtaken by the co-evolutionary dynamic between AI and human cognition. The deepest implication is therefore not merely that we need better measurement but that we need *adaptive* measurement: assessment architectures that can track the ongoing evolution of what learning means. The paradigm worth pursuing is not a destination but a capacity — the capacity to continuously reimagine what it means to learn, to know, and to grow, as the technological and human elements of education co-evolve in ways we cannot fully predict. That capacity must be pursued with rigor, with humility, and with unwavering attention to the learners it serves.

References

Agent4EDU. (2024). Advancing AI for education with agentic workflows. In *Proceedings of the ACM International Conference on AI in Education (ICAIE '24)*. ACM.

Arunkumar, V., Gangadharan, G. R., & Buyya, R. (2026). Agentic AI: Architectures, taxonomies, and evaluation of LLM agents. *arXiv preprint arXiv:2601.12560*.

Association of American Colleges and Universities. (2018). *Fulfilling the American dream: Liberal education and the future of work*. AAC&U.

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, *32*(4), 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Bandi, A., Kongari, B., Naguru, R., Pasnoor, S., & Vilipala, S. V. (2025). The rise of agentic AI. *Future Internet (MDPI)*, *17*(9), 404.

Banihashem, S. K., Gasevic, D., Noroozi, O., Jarodzka, H., Joosten-ten Brinke, D., & Drachsler, H. (2025). Optimizing formative assessment with learning analytics: A systematic review. *Review of Educational Research*, *95*(2), 215–258.

Bardach, E., & Patashnik, E. M. (2019). *A practical guide for policy analysis: The eightfold path to more effective problem solving* (6th ed.). CQ Press.

Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, *54*(5), 1160–1173. https://doi.org/10.1111/bjet.1

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics* (8th ed.). Oxford University Press.

Biggs, J. (1999). *Teaching for quality learning at university*. Society for Research into Higher Education & Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Chinta, S. V., Wang, Z., Yin, Z., Hoang, N., Gonzalez, M., Le Quy, T., & Zhang, W. (2024). FairAIED: Navigating fairness, bias, and ethics in educational AI applications. *arXiv preprint arXiv:2407.18745*.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19. https://doi.org/10.1093/analys/58.1.

Coates, H., & Zlatkin-Troitschanskaia, O. (2019). The governance, policy and strategy of learning outcomes assessment in higher education. *Higher Education Policy*, *32*, 507–512. https://doi.org/10.1057/s41307-019-00161-1

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3–12. https://doi.org/10.3102/0013189X032006003

Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. *Philosophy & Technology*, *29*(3), 245–268. https://doi.org/10.1007/s13347-015-0211-1

Eckstein, H. (1992). *Regarding politics: Essays on political theory, stability, and change*. University of California Press.

ENQA. (2015). *Standards and guidelines for quality assurance in the European Higher Education Area (ESG)*. European Association for Quality Assurance in Higher Education.

Ewell, P. T. (2009). Assessment, accountability, and improvement: Revisiting the tension (NILOA Occasional Paper No. 1). National Institute for Learning Outcomes Assessment.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Gandara, D., Anahideh, H., Ison, M. P., & Picchiarini, L. (2024). Inside the black box: Detecting and mitigating algorithmic bias across racialized groups in college student-success prediction. *AERA Open*, *10*(1), 1–15. https://doi.org/10.1177/23328584241258741

Gartner. (2025). *Top strategic technology trends for 2025: Agentic AI*. Gartner, Inc.

Goffman, E. (1959). *The presentation of self in everyday life*. Anchor Books.

HEEACT. (2023a). *Third cycle of institutional accreditation handbook (2023–2025)*. Higher Education Evaluation and Accreditation Council of Taiwan.

HEEACT. (2024). *Program accreditation handbook (2024 edition)*. Higher Education Evaluation and Accreditation Council of Taiwan.

Hayles, N. K. (2012). *How we think: Digital media and contemporary technogenesis*. University

of Chicago Press.

Hou, A. Y. C., Morse, R., & Chiang, C. L. (2012). An analysis of mobility in global rankings: Making institutional strategic plans and positioning for building world-class universities. *Higher Education Research & Development*, *31*(6), 841–857.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

IMDA. (2020). *Model AI governance framework* (2nd ed.). Infocomm Media Development Authority, Singapore.

Inside Higher Ed. (2026, January). AI agent "Einstein" passes university courses autonomously: What it means for assessment design. *Inside Higher Ed*.

Kasparov, G. (2017). *Deep thinking: Where machine intelligence ends and human creativity begins*. PublicAffairs.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms active learning. *Scientific Reports*, *15*, 17458. https://doi.org/10.1038/s41598-025-97652-6

Kuhn, T. S. (2012). *The structure of scientific revolutions* (4th ed.). University of Chicago Press. (Original work published 1962)

Legislative Yuan. (2025). *Artificial Intelligence Basic Act* (人工智慧基本法). Republic of China (Taiwan).

Lin, A. S. R., Hou, A. Y. C., Chan, S. J., & Chiang, T. L. (2021). Quality assurance in Taiwan higher education: Regulation, model shift, and future prospect. In A. Y. C. Hou, T. L. Chiang, &

S. J. Chan (Eds.), *Higher Education in Taiwan: Global, political and social challenges and future trends* (pp. 65–81). Springer. https://doi.org/10.1007/978-981-15-4554-2_4

Luckin, R. (2018). *Machine learning and human intelligence: The future of education for the 21st century*. UCL IOE Press.

Masterman, M. (1970). The nature of a paradigm. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 59-89). Cambridge University Press.

Masterman, T., Besen, S., Sawtell, M., & Chao, A. (2024). The landscape of emerging AI agent architectures for reasoning, planning, and tool calling: A survey. *arXiv preprint arXiv:2404.11584*.

Ministry of Education. (2024). *Education statistical indicators at a glance 2024*. Ministry of Education, Republic of China (Taiwan).

Ministry of Education. (2025). *2025 Education White Paper*. Ministry of Education, Republic of China (Taiwan).

Ministry of Education Singapore. (2023). *EdTech Masterplan 2030: Technology-transformed learning to prepare students for the future*. Ministry of Education, Singapore.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design (ETS Research Report No. RR-03-16). Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2003.tb01908.x

Ng, A. (2024, March 13). Agentic design patterns. *The Batch* (DeepLearning.AI Newsletter).

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology*, *58*(1), 504–509.

OECD. (2023). *OECD digital education outlook 2023: Towards an effective digital education*

ecosystem. OECD Publishing. https://doi.org/10.1787/c74f03de-en

OpenAI. (2025, July). New tools for understanding AI and learning outcomes. https://openai.com/index/understanding-ai-and-learning-outcomes/

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, *2*, Article 100020. https://doi.org/10.1016/j.caeai.2021.100020

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Academies Press.

QAA. (2023). *Artificial intelligence: Guidance for UK higher education providers*. Quality Assurance Agency for Higher Education.

Rest of World. (2025, June). South Korea scales back AI textbook rollout after parental backlash. *Rest of World*.

Ritzer, G. (1975). Sociology: A multiple paradigm science. *The American Sociologist*, *10*(3), 156-167.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Sharma, Y. (2024, September). Taiwan faces wave of university closures. *University World News*.

Shavelson, R. J. (2010). Measuring college learning responsibly: Accountability in a new era. *Stanford University Press*.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14. https://doi.org/10.3102/0013189X029007004

Shute, V. J., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. MIT Press.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1509–1528. https://doi.org/10.1177/0002764213479366

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, *333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Spencer, L. M., & Spencer, S. M. (1993). *Competence at work: Models for superior performance*. John Wiley & Sons.

Sterelny, K. (2010). Minds: Extended or scaffolded? *Phenomenology and the Cognitive Sciences*, *9*(4), 465–481. https://doi.org/10.1007/s11097-010-9174-y

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., … & Gasevic, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *3*, Article 100075. https://doi.org/10.1016/j.caeai.2022.100075

Taiwan News. (2024, August). Seven Taiwan universities shut down amid enrollment crisis. *Taiwan News*.

Tam, M. (2001). Measuring quality and performance in higher education. *Quality in Higher Education*, *7*(1), 47–54. https://doi.org/10.1080/13538320120045076

Temper, M., Tjoa, A. M., & David, K. (2025). Higher Education Act for AI (HEAT-AI): A frame-

work to regulate the usage of AI in higher education institutions. *Frontiers in Education*, *10*, Article 1505370. https://doi.org/10.3389/feduc.2025.1505370

TEQSA. (2024). *Artificial intelligence in higher education: Guidance note for providers*. Tertiary Education Quality and Standards Agency, Australian Government.

UNESCO. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization.

UNESCO. (2025, October 27). What's worth measuring? The future of assessment in the AI age. https://www.unesco.org/en/articles/whats-worth-measuring-future-assessment-ai-age

van der Linden, W. J., & Glas, C. A. W. (Eds.). (2010). *Elements of adaptive testing*. Springer. https://doi.org/10.1007/978-0-387-85461-8

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Yan, L. (2025). From passive tool to socio-cognitive teammate: Reconceptualizing AI's role in learning through the lens of agentic cognition. *arXiv preprint arXiv:2508.14825*.

Zhong, L., & Zhao, X. (2025). Education paradigm shifts in the age of AI: A spatiotemporal analysis of learning. *ECNU Review of Education*, *8*(2), 319–342. https://doi.org/10.1177/20965311251315204

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.