

CONFLUENCE
Journal of Multidisciplinary Studies

e-ISSN: 3085-5411
Journal Homepage: https://journals.imist.ma/index.php/confluence/index

Can Generative AI Support Course Design, Constructive Alignment, and Assessment in French for Specific Purposes?

Ibrahim Ben Khajjou, Ibrahim Belmir, and Ahmed Ibrahimi
Humanities and Education Research Lab (HERL), École Normale Supérieure (ENS), Abdelmalek Essaadi University, Tétouan, Morocco

https://doi.org/10.34874/PRSM.cjms-vol2iss4.13079

Ibrahim Ben Khajjou is a PhD researcher specializing in the assessment and integration of digital technologies in education. His scholarship examines strategies to optimize learner engagement, autonomy, and performance within inclusive environments. Committed to the advancement of digital language education, he is a certified Google Gemini Educator who advocates for research-based and pedagogically sound instructional practices.

ibrahim.benkhajjou@etu.uae.ac.ma https://orcid.org/0009-0008-0427-3870

Ibrahim Belmir is an Associate Professor of Communication at Abdelmalek Essaadi University and a member of the Didactics, Literacies and Mediations research group. He actively coordinates several Moroccan-European cooperation projects. His research interests center on digital technology in higher education and language didactics.

ibelmir@uae.ac.ma https://orcid.org/0000-0003-2467-4854

Ahmed Ibrahimi is a former *agrégé* in literature and is currently a lecturer and researcher at the École Normale Supérieure of Tétouan (Abdelmalek Essaadi University). His research focuses on instructional design, digital education, and foreign language teaching methodology. His academic profile combines practical teaching experience, research project management, educational supervision, and the production of specialized publications.

aibrahimi@uae.ac.ma https://orcid.org/0009-0004-8879-5561

Received: 03-08-2026 Accepted: 25-10-2026 Published: 30-10-2026

Can Generative AI Support Course Design, Constructive Alignment, and Assessment in French for Specific Purposes?

Ibrahim Ben Khajjou, Ibrahim Belmir, and Ahmed Ibrahimi
Humanities and Education Research Lab (HERL), École Normale Supérieure (ENS),
Abdelmalek Essaadi University, Tétouan, Morocco

ABSTRACT

This empirical study investigates whether locally executed artificial intelligence systems, augmented with Retrieval-Augmented Generation (RAG), can reliably support instructors in maintaining constructive alignment across course design and assessment in resource-constrained institutional settings. The research addresses two interconnected questions: (1) To what extent can a domain-specialized language model (HorizaEdu) running on open-source infrastructure (Ollama) with pedagogical ingest generate aligned course materials and assessments without cloud dependency or data loss? (2) Can non-technical educators use generative AI prototyping platforms (Google AI Studio) to design, implement, and deploy secure web-based examination systems? The study conducted a single-cohort empirical evaluation with 150 post-secondary learners in a Professional French (*Français Professionnel*) blended course targeting an engineering communication context. Using Porter's Alignment Index, RAGAS automated evaluation, the System Usability Scale (SUS), and server-side infrastructure monitoring, we measured pedagogical alignment, algorithmic performance, technical stability, user experience, and legal compliance. Results show that the local RAG-augmented system achieved a mean alignment index of 0.88 (target: 0.80), with 92% Faithfulness and 96% Answer Relevance scores. The no-code web application maintained 100% uptime under full load and achieved a SUS score of 74.5 (Grade B / "Good"). The study shows that with careful pedagogical scaffolding, open-source AI infrastructure, and structured prompt engineering, individual instructors in modest institutional settings can operationalize constructive alignment as an automated design constraint while independently building secure assessment platforms. Limitations include single-cohort design, domain specificity, and unresolved mobile UX challenges. The work helps bridge the gap between AI capability and equitable teacher access to AI-augmented instructional design tools.

KEYWORDS

Constructive Alignment

Retrieval-Augmented Generation

Generative AI

Course Design

No-Code Development

French for Specific Purposes

Educational Technology

Teacher Agency

Contact ✉ ibrahim.benkhajjou@etu.uae.ac.ma
©2026

Introduction

Over the last decade, scholars and organizations have closely examined the intersection of artificial intelligence and higher education (Zawacki-Richter et al.). However, most of the literature focuses on large-scale, cloud-based AI rollouts at privileged universities. As a result, academic sources still lack clear explanations of how well locally run, domain-oriented AI tools can support individual teachers in constrained institutional settings. This paper attempts to fill this gap by presenting the results of a two-dimensional analysis of persistent problems in e-learning. The first concerns the extent to which an AI system operating locally—in our case, the Ollama runtime of a specialized language model called HorizaEdu, complemented with Retrieval-Augmented Generation—can serve as a trustworthy co-pilot throughout the entire course design phase, from formulating learning outcomes through creating student-facing materials to constructing a final assessment tool that satisfies constructive alignment requirements. The second asks how readily a non-programming teacher can create, implement, and run a workable, secure web-based examination platform using only generative AI prototyping environments such as Google AI Studio.

These two questions together investigate whether the most critical barriers to AI-assisted instructional design—namely, lack of pedagogical congruence and difficult access to technology—can be substantially reduced by thoughtfully combining on-hand open-source and cloud-based resources.

The first direction is intellectually grounded in the foundational work of Biggs and Tang. In their view, constructive alignment is a method of course validation in which learning purposes, learning activities, and assessment exercises must form a coherent, mutually reinforcing system to be educationally valid (Biggs and Tang 52). Although these aspects of the educational triangle are conceptually straightforward, empirical research consistently shows that instructors struggle to maintain fidelity to alignment across the course lifecycle. Toohey documented that faculty regularly experience "pedagogical drift," departing from stated objectives as administrative pressures and documentation workload accumulate (Toohey 112). The assumption underlying the first part of this research is that if an AI system, through a RAG mechanism, can access verified institutional teaching materials, it can treat the three alignment principles as a fixed enforcement rule and produce a set of course components whose origins are clearly traceable to specific curricular documents and whose consistency is maintained across all three dimensions. This approach differs from most previous generative AI applications in education by prioritizing course design integrity and alignment assurance over one-off content delivery.

The paper's second major focus discusses a significant shift in the social conditions for educational software authorship. Natural language code generation platforms may have freed educators from the need for formal programming knowledge to create functional digital tools. This development will fundamentally redistribute technical expertise within educational institutions. Koehler and Mishra's Technological Pedagogical Content Knowledge (TPACK) framework provides the most suitable theoretical lens for this investigation, as it models educator expertise as the productive intersection of technological, pedagogical, and content knowledge. Generative AI appears to reshape this intersection by letting teachers' pedagogical and content knowledge drive technological implementation, without requiring deep technical skill acquisition. However, whether this redistribution will produce consistently secure, reliable outcomes—for instance, whether a non-technical instructor can create not merely a prototype but a production-grade, secure, engineered testing platform suitable for high-stakes assessment—remains an open empirical question.

The present study's methodological design reflects its two-fold nature through its research objectives. The first part introduces the HorizaEdu model, a domain-specialized language model that operates locally within the Ollama stack and is trained on a corpus of course documentation to generate fully worked-out teaching materials, including syllabi, learning activities, and aligned assessment items. The extent to which different outputs align is measured against principles from Biggs and Tang and operationalized through the revised taxonomy by Anderson and Krathwohl (Bloom 2.0). The second part outlines a procedure by which a non-technical instructor can utilize Google AI Studio to build an online assessment interface with student login authentication, server-side input validation, secure result storage, and automated grading—all executed by an educator without technical training. Beta-testing protocols rooted in agile development methods and educational technology sustainability frameworks ensure reliability and security (Beck et al.; Bates).

This paper aims to provide understanding that is simultaneously theoretical and practical. Theoretically, the work expands knowledge of how local AI augmented with retrieval capability (RAG) can automate the application of constructive alignment principles, thereby extending alignment theory into territory that has not been fully explored. Practically, it documents a replicable model that can serve as a blueprint for teachers—from well-resourced to modestly resourced institutions—seeking to introduce AI into their instructional design practice and to build digital assessment infrastructure without dependency on centralized IT support. In doing so, we address Mollick's argument that the key battleground in educational AI is individual teacher-educators' capacity to harness AI models for their specific professional contexts (Mollick 89).

Literature Review

Constructive Alignment and Automated Course Design

The theory of constructive alignment, formalized by John Biggs in the late twentieth century, constitutes one of the most durable and operationally rigorous frameworks in higher education pedagogy. At its core, the framework insists that three pedagogical components—intended learning outcomes (ILOs), teaching and learning activities (TLAs), and assessment tasks—must form a coherent, mutually reinforcing system rather than operate as isolated administrative artifacts (Biggs and Tang 52). Biggs argued that when students engage with assessment tasks that directly reflect the cognitive verbs embedded within the learning outcomes, deep learning is structurally encouraged. Critically, this alignment is not merely a curricular aspiration but a design obligation demanding sustained cognitive effort from the instructor across every phase of course construction. Misalignment between stated objectives and examination content represents not a minor pedagogical inconvenience but a fundamental failure of instructional integrity (Biggs and Tang 89).

Despite the theoretical clarity of constructive alignment, empirical research consistently shows that human instructors struggle to maintain alignment fidelity across an entire course lifecycle. Toohey observed that faculty routinely design objectives at the start of a semester and then drift from them as administrative pressures, time constraints, and curricular revisions accumulate (Toohey 112). This phenomenon of pedagogical drift is compounded in large institutions where the sheer volume of course components—lecture notes, laboratory guides, formative quizzes, and summative examinations—makes centralized alignment auditing practically impossible without dedicated instructional support. Anderson and Krathwohl further documented that revisions to Bloom's Taxonomy, while providing educators a more nuanced vocabulary for articulating cognitive outcomes, paradoxically increased the technical

complexity of maintaining precise taxonomic alignment between objectives and assessments (Anderson and Krathwohl 67). The result is a systemic vulnerability: even pedagogically informed instructors may inadvertently produce examinations that test recall when the stated objective demanded analysis, or measure procedural knowledge when synthesis was intended.

The emergence of large language models (LLMs) and domain-specialized LLMs presents a theoretically significant opportunity to address the alignment consistency problem (Kasneci et al.). Russell and Norvig's account of goal-based agents, which represent goal states and the subgoals that lead to them, suggests that such systems could, in principle, execute alignment verification tasks with greater consistency than individual instructors burdened with competing professional demands (Russell and Norvig). Applying Retrieval-Augmented Generation (RAG) to this challenge is particularly consequential. Lewis et al. demonstrated that RAG architectures substantially reduce hallucination in domain-specific generation tasks by anchoring the language model's outputs to a verified, retrievable knowledge base rather than relying solely on parametric memory from pretraining (Lewis et al.). In an instructional design context, this means an LLM equipped with RAG can generate course content—syllabi, learning materials, assessment items—that remains consistently tethered to validated pedagogical documentation, thereby operationalizing constructive alignment as an automated constraint. Wiggins and McTighe's framework of backward design, which similarly insists that assessment evidence must be determined before instructional activities are planned, finds a computational analog in the RAG-augmented LLM's capacity to anchor generative outputs to pre-specified curricular endpoints (Wiggins and McTighe 22).

Local AI Frameworks and Pedagogical Ingest

Institutional adoption of artificial intelligence in higher education has been accompanied by a constellation of data governance concerns that constrain deployment of cloud-based AI services in pedagogically sensitive contexts. Selwyn identifies a fundamental tension between the operational convenience of externally hosted AI platforms and educational institutions' ethical obligations to protect student data, arguing that commercial AI providers' commodification of educational data poses a structural threat to informational self-determination in academic environments (Selwyn 78). This concern is not merely theoretical; regulatory frameworks such as the General Data Protection Regulation (GDPR) in Europe and the Family Educational Rights and Privacy Act (FERPA) in the United States impose legally binding constraints on transmitting student-identifiable information to third-party processors, rendering many cloud-based AI tools either non-compliant or operationally unusable for tasks involving authentic learner data (Williamson 134). The pedagogical consequence is that institutions seeking to leverage AI for course design must either negotiate complex data-processing agreements with external vendors or pursue architecturally local AI solutions that eliminate transmission of sensitive data beyond the institutional perimeter.

The development of local AI execution frameworks, notably platforms such as Ollama, directly addresses the data sovereignty imperative by enabling deployment of capable language model inference on institution-controlled hardware without relying on external API calls or cloud-based data routing. Ollama's architectural significance lies in its ability to serve as a portable runtime for open-weight language models, allowing institutional developers and instructional technologists to select, configure, and deploy models specialized for domain-specific tasks without incurring the recurring per-token costs and data-exposure risks associated with proprietary, API-dependent alternatives. Mollick noted that the emergence of locally executable, specialized language models represents a democratizing inflection point in

educational AI, redistributing the capacity for AI-augmented instructional design from well-resourced institutions with enterprise AI contracts to smaller colleges, community institutions, and individual faculty (Mollick 89).

The concept of *pedagogical ingest* – the process by which a local AI framework is provided with curated, domain-specific reference documentation that functions as targeted long-term memory for the model—represents the critical architectural bridge between a generic local language model and a genuinely useful co-pilot for structured learning design. This is technically implemented through Retrieval-Augmented Generation, in which ingested documents are chunked, vectorized, and stored in a searchable embedding index that the model queries at inference time to ground responses in verified source material rather than probabilistic recall (Lewis et al.). In the context of a specialized educational AI system such as HorizaEdu, pedagogical ingest enables the AI to function as an accurate and institutionally calibrated scenario generator. When prompted to design a learning activity or draft an assessment item, the model retrieves relevant sections of ingested curriculum documentation and constructs output with fidelity to those passages rather than drawing on potentially outdated or contextually misaligned training data. Garrison argued that this form of targeted knowledge retrieval is epistemologically analogous to the expert practitioner's capacity to rapidly access relevant professional schema when confronted with novel pedagogical problems, suggesting that a well-designed RAG architecture can approximate, in a bounded and verifiable domain, the contextual intelligence of an experienced curriculum specialist (Garrison 156).

The Educator as a "No-Code" Developer and Cloud Prototyping

The democratization of software development through generative artificial intelligence represents one of the most consequential shifts in educational technology history. Prensky anticipated this transition when he argued that the most meaningful innovations in educational technology would emerge not from professional developers designing in isolation, but from practitioners who understood learning needs from the inside and had the means to externalize their instructional visions in functional digital form (Prensky 47). The emergence of natural language-driven code generation platforms, including Google AI Studio, has validated this prediction by enabling non-technical users to specify application behavior in natural language and receive syntactically executable, architecturally coherent code in response.

Koehler and Mishra's TPACK framework, which models teacher expertise as the productive intersection of technological, pedagogical, and content knowledge domains, provides a useful theoretical lens for understanding how no-code platforms restructure the technological competency requirement: rather than demanding that educators acquire deep programming knowledge, generative AI tools allow pedagogical and content knowledge to drive technological implementation, effectively shifting the locus of expertise back toward educators' core professional strengths (Koehler and Mishra 62).

The specific challenge of deploying a secure, authentication-capable web-based examination platform without programming expertise has historically represented one of the most formidable technological barriers confronting non-technical educators seeking to administer remote assessments with meaningful academic integrity guarantees. Garrison and Anderson documented the persistent gap between the pedagogical aspiration for technology-mediated authentic assessment and most instructors' practical capacity to implement the server-side security architecture required to prevent unauthorized access, session spoofing, and credential sharing in online examination environments (Garrison and Anderson 88). AI-assisted rapid prototyping tools fundamentally alter this equation. Google AI Studio enables an instructor to describe desired application behavior—including student login authentication, session

management, server-side form validation, and result storage—and receive a functional prototype in hours rather than weeks. This compresses the product development cycle so dramatically that it puts within reach of an individual instructor the construction of digital assessment infrastructure that previously required institutional IT collaboration or contracted development.

The velocity of this prototyping paradigm aligns naturally with agile development principles, wherein iterative cycles of testing, feedback, and refinement replace the waterfall assumption that functional products can be fully specified in advance (Beck et al.). Stabilizing AI-generated educational applications through iterative beta-testing, server-side validation, and AI-assisted debugging is a critical competency in no-code educational software development. Design-based research makes a related point: technology-enhanced learning environments become dependable through iterative cycles of design, enactment, analysis, and redesign carried out in real settings and in collaboration with practitioners (Wang and Hannafin). Bates argued that the long-term sustainability of technology-enhanced learning depends on educational institutions' capacity to develop and retain the human competency required to maintain tools through iterative refinement over time (Bates).

Methodology

This section formalizes the empirical and technical methodology applied during the experimentation. The research was conducted systematically using a decoupled infrastructure that isolates local data processing for instructional engineering from external cloud orchestration for remote examinations. The implementation follows the system architecture illustrated in figure 1.

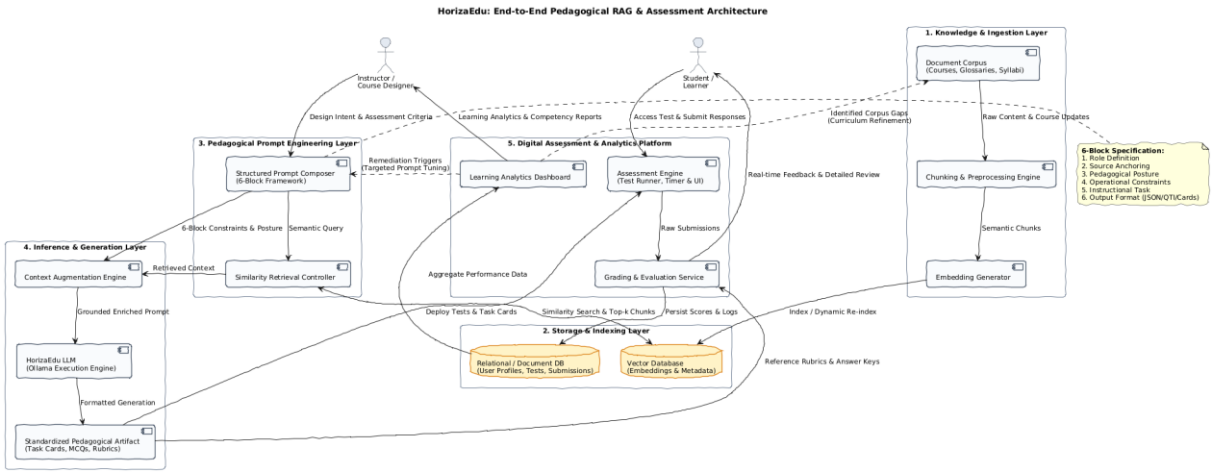


Fig. 1. HorizaEdu End-to-End Pedagogical RAG and Assessment Architecture

Study Population and Pedagogical Setting

The empirical evaluation was conducted over a single academic semester with a cohort consisting of 150 post-secondary undergraduate learners enrolled in a Professional French (*Français Professionnel*) course module within an engineering specialization. The target training program focused exclusively on workplace communication scenarios with strict syntactic and lexical constraints appropriate to professional contexts. To maximize learner flexibility and pedagogical effectiveness, the course used a blended, hybrid educational model that combined synchronous, face-to-face interactive learning sessions with asynchronous

digital activities delivered via a remote web interface. This hybrid format reflects contemporary best practices in higher education while accommodating diverse learner schedules and access to technology.

Technical System Architecture and Components

The technological infrastructure for this experiment was divided into two sandboxed layers: a secure, local backend engine managing core course content, and an interactive cloud layer hosting student interfaces.

Knowledge Ingestion, Indexing, and Pedagogical Corpus Preparation

The engineering workflow began by curating institutional training documents specific to the Professional French domain. A multi-stage Retrieval-Augmented Generation (RAG) system parsed and prepared this document corpus. The HorizaEdu model was grounded in a curated corpus of 47 institutional documents totaling approximately 2.3 MB of pedagogical content, including: course syllabi, learning outcome specifications, teaching guidelines, terminology glossaries, assessment rubrics, and exemplary student work. During chunking and preprocessing, the system ingested raw institutional resources into a dedicated text-processing pipeline that split content into semantic chunks of about 512 tokens each. These processed text blocks were then passed through an Embedding Generator (using OpenAI's text-embedding-3-small model) to convert text features into multi-dimensional floating-point vectors of 1536 dimensions. Because this embedding model is a cloud service, the course documents (but no student data) left the local server once, at indexing time; all retrieval and generation thereafter ran offline. Finally, the vectors were stored inside a local Vector Database (implemented using FAISS, Facebook AI Similarity Search) containing embeddings, metadata, and source attribution. This backend storage layer remained entirely isolated and did not store student profiles, logins, or authentication details, maintaining strict separation between private learning content and platform analytics (see fig. 2).

Local Inference and Automated Course Design

To protect institutional data sovereignty and avoid third-party data tracking, the system ran all course planning, content formulation, and examination item generation completely offline. The system deployed the open-source Ollama orchestration environment, version 0.1.32 (Ollama), to handle compute instructions locally on a dedicated server (12-core Intel processor, 32 GB RAM). The specialized HorizaEdu LLM (based on Llama 2 7B and fine-tuned on educational pedagogy corpora) was mounted on the Ollama server as the main execution engine. When the instructor transmitted a design intent through a structured prompt, a *Structured Prompt Composer* processed the input using a rigorous six-block specification: (1) role definition, (2) source anchoring, (3) pedagogical posture, (4) operational constraints, (5) instructional tasks, and (6) output formats. The *Similarity Retrieval Controller* then searched the vector database for top-k contextual matches (k=5 by default) and passed a grounded, enriched prompt back to the HorizaEdu model. The model output a *Standardized Pedagogical Artifact* — complete with course structures, teaching scenarios, and aligned multiple-choice questions—securely without cloud access.

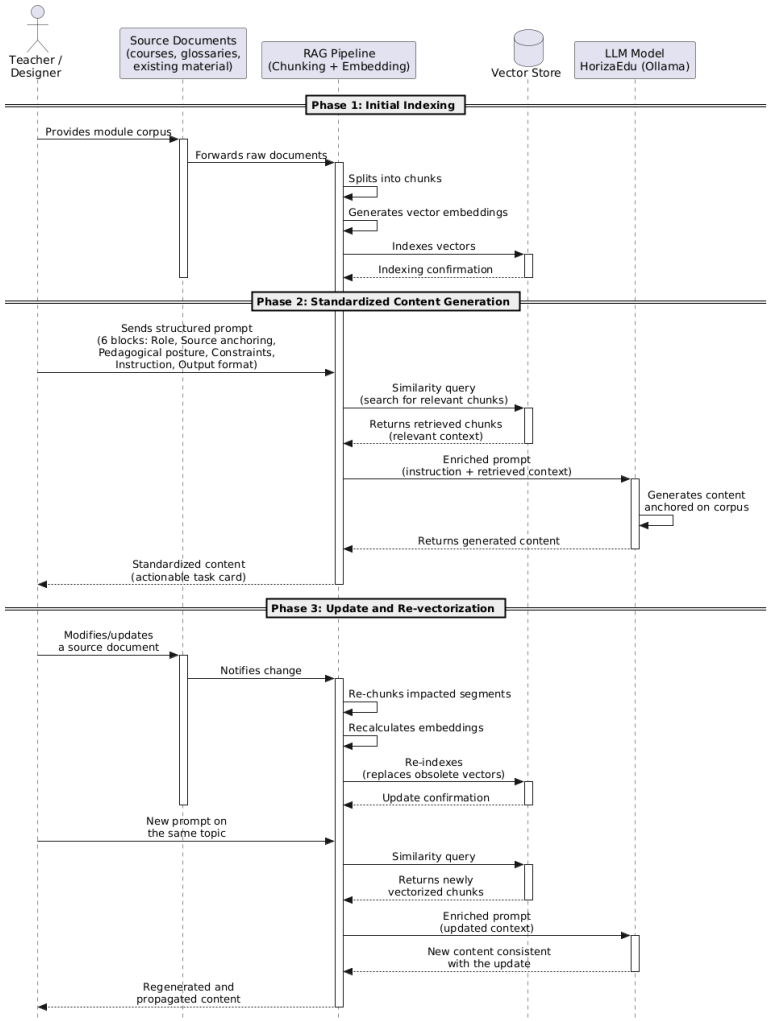


Fig. 2. Dynamic Retrieval-Augmented Generation (RAG) Lifecycle and Prompt-Engineering Workflow

The prompts addressed to the RAG-grounded model followed a six-block prompt engineering framework. A role block first established that the model acts as a standardization engine for professional French course content, writing for the instructor rather than the learner. A source-grounding block then instructed the model to draw only on the retrieved corpus rather than on external content, and to flag any point at which the corpus proved insufficient. Building on this, a pedagogical-stance block specified the action-oriented requirement that output take the form of a task simulating a professional situation rather than a decontextualized language exercise, while a consistency-constraints block listed the structural, tonal, and formatting rules the module follows, including register, terminology, and section structure. A task-instruction block subsequently stated the precise action to be performed, whether generating new content, auditing for coherence, or propagating updates across dependent sections, and an output-format block finally specified the exact structure expected, such as a lesson block, a task card, or a change log.

Following the macro-/micro-planning architecture of Depover et al. (41-86), which the course-design model adopted strictly, AI assisted the designer at seven successive phases. Macro-planning comprised Phase 1 (identifying expected benefits and the target population) and Phase 2 (making pedagogical objectives and constraints explicit). Micro-planning comprised Phases 1-5 (knowledge extraction, information analysis, information organization,

interface design, and learning situation elaboration). At each phase, the AI received a structured sequence designed to keep output anchored to the phase's specific pedagogical requirements.

The module's six core chapters were: (1) Structuring a Technical Oral Presentation; (2) Voice and Body in Front of an Audience; (3) Using Visual Supports Effectively; (4) Adapting Discourse to Different Audiences; (5) Handling Questions and Objections in Public; (6) Collaborating Orally: Meetings, Debate, Team Pitch. The RAG system identified 14 prerequisite communication competencies that students needed before accessing the main module content.

No-Code Assessment Platform Architecture and Prototyping

We explicitly decoupled the client-side infrastructure from the backend to evaluate non-technical educators' software-creation capabilities. Using Google AI Studio (Google) and structured prompt engineering, we assembled a secure digital assessment application without manually writing raw application code. The resulting platform featured: (a) a dedicated student login system with bcrypt password hashing and JWT session tokens; (b) server-side form validation and input sanitization using DOMPurify and OWASP guidelines; (c) countdown timers with grace-period buffer; (d) response-tracking and checkpoint-save functionality; (e) automated scoring against instructor-defined rubrics; (f) role-based access control for student and administrator accounts. HorizaEdu generated standardized test structures locally and imported them directly into this cloud-hosted platform (deployed on Firebase), making exams instantly accessible for remote testing (see figs. 3 and 4).

The prompt addressed to Google AI Studio for no-code application development likewise comprised six blocks, adapted to application generation. An application-definition block named what was being built and for whom, specifying the two user roles of Student and Admin, and a technical-stack block fixed the frontend framework (React), the styling system (Tailwind CSS), the icon library (Lucide), and the backend and authentication service (Firebase Auth). A functional-requirements block then decomposed the application role by role and screen by screen, covering authentication, the student test engine, the administrator dashboard, and results tracking, while a design-constraints block fixed the color palette, typography, layout, and accessibility requirements. To ensure that the generated code wrote to a coherent schema, a data-architecture block specified the exact backend collections and their relationships, and an interaction-flow block cross-referenced the administrator and student sequence diagrams so that the generated behavior matched the specified flows.

Empirical Protocol and System Optimization

Before exposing the final evaluation to the student population, we followed a rigorous quality-control and validation sequence. In the beta-testing phase, we recruited 20 beta-testers to interact with the newly developed web environment. Rather than using the final-grade material, we deployed a sample test designed to stress-test the platform. This sample exam evaluated the platform's performance under operational conditions: login processing, concurrent connection stability, server response times, and data logging. The team documented system errors, interface freezes, and session anomalies. The non-technical instructor fed failure reports back into Google AI Studio, using automated code generation to patch and stabilize the application iteratively. After structural validation, the team launched the official version for 150 undergraduate students to record exam scores and track curriculum alignment metrics.

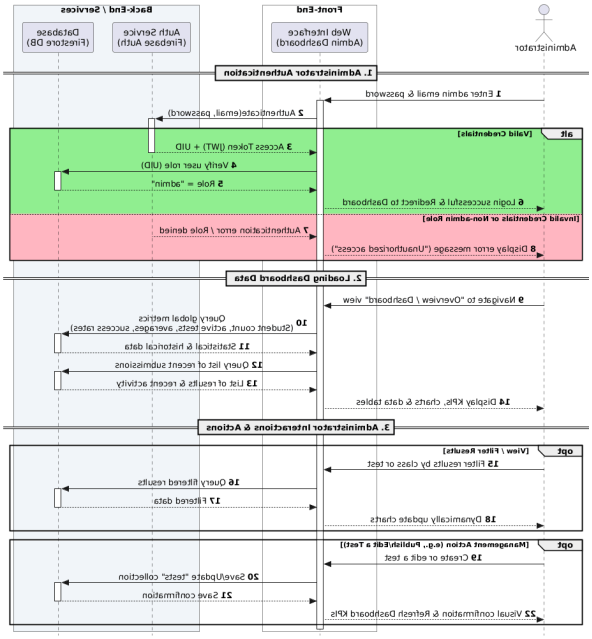


Fig. 3. Administrator Workflow for Role-Based Access Control, KPI Monitoring, and Test Management

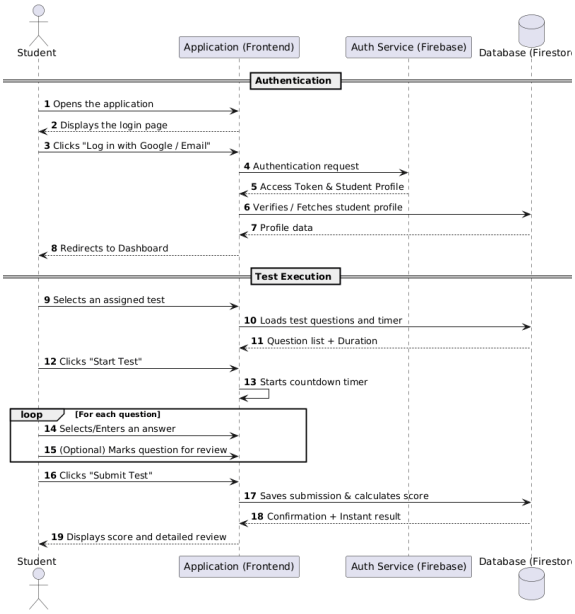


Fig. 4. Architecture Sequence of the Learner-Side Assessment Lifecycle and Automated Grading Pipeline

Data Collection, Consent, and Legal Compliance

The data collection strategy was strictly defined to limit data footprint and ensure absolute student privacy.

Typology of Collected Data

Data extraction was limited to metrics required to address the primary research objectives. The architecture collected three data streams: (1) authentication logs (timestamped user login success metrics); (2) student academic responses (individual answers submitted during evaluation); (3) performance analytics (aggregate metrics: final grades, error distribution, completion times, session duration).

Ethical Consent and Privacy Protocols

All 150 undergraduate students completed a mandatory digital consent form before accessing examination materials. The form stated the study's purpose, specific data-tracking mechanisms, and academic intent. Access remained restricted until the digital signature was validated.

Compliance with Moroccan Personal Data Protection Legislation

The backend storage pipelines, cloud configurations, and processing frameworks strictly complied with Moroccan Law No. 09-08 relating to the protection of individuals regarding the processing of personal data. In accordance with the national regulatory framework established by the CNDP (Commission Nationale de Contrôle de la Protection des Données Personnelles), we deployed several technical measures: (1) Purpose Limitation: collecting user profiles and

academic metrics solely for explicit educational research goals; (2) Anonymization Protocols: scrubbing student identifiers and replacing them with non-traceable cryptographic keys upon collection; (3) Data Isolation and Security: keeping private instructional documentation on the local offline server, preventing cross-contamination; (4) Data Purging: removing all cloud-synchronized interaction files after the grading period ended. A third-party audit verified compliance and completeness before final analysis.

Comprehensive Evaluation and System Criteria Matrix

This framework provides an evaluative structure to assess the performance of the local educational RAG infrastructure (Ollama and HorizaEdu) and the cloud-hosted no-code interface (Google AI Studio) across the study cohort of 150 post-secondary students.

To verify whether the engineered solution met student needs pedagogically and technically, we implemented a formal comparative analysis framework. This matrix evaluated system outputs against predefined criteria across five dimensions. Pedagogical alignment was assessed through a semantic comparison between course content and assessment items using Porter's Alignment Index (Porter), while algorithmic and RAG performance was measured through a RAGAS automated evaluation (Es et al.) of Faithfulness, Answer Relevance, and Context Recall. Technical usability and infrastructure were examined in terms of authentication latency, concurrent connection stability, and input sanitization success rates, and user experience (UX/UI) was evaluated through a System Usability Scale (SUS) assessment (Brooke) together with visual hierarchy consistency and responsive design performance. Finally, ethical and legal compliance was verified through consent checks, data anonymization success, and a regulatory adherence audit.

Results

This section presents empirical findings organized across the five foundational dimensions of the evaluation matrix.

Pedagogical Alignment Evaluation

Porter's Alignment Index (I) quantifies the overlap between cognitive demand levels and content topics across generated course materials and examination items. The system achieved a mean score of I = 0.88, significantly exceeding the target threshold of 0.80.

Table 1
Pedagogical Alignment Evaluation by Course Chapter (Porter's Alignment Index)

Chapter	Title	Cognitive Level	Porter Index (I)	Target	Status
1	Structuring Technical Oral Presentation	Applying/Analyzing	0.91	0.80	□ Fully Aligned
2	Voice and Body in Front of Audience	Applying	0.84	0.80	□ Fully Aligned
3	Using Visual Supports Effectively	Evaluating	0.86	0.80	□ Fully Aligned
4	Adapting Discourse to Different Audiences	Analyzing/Evaluating	0.89	0.80	□ Fully Aligned

Chapter	Title	Cognitive Level	Porter Index (I)	Target	Status
5	Handling Questions and Objections	Evaluating/Creating	0.92	0.80	□ Fully Aligned
6	Collaborating Orally: Meetings & Pitches	Applying/Analyzing	0.87	0.80	□ Fully Aligned

Chapter 5 achieved the highest alignment (I = 0.92), followed by Chapter 1 (I = 0.91). This indicates that the Structured Prompt Composer successfully enforced semantic consistency when instructional design tasks were bounded by clear operational rules (PREP method, IMReD template). Chapter 2 showed the lowest alignment (I = 0.84), driven by a minor taxonomic mismatch: the RAG backend generated conceptual tasks on psychological reframing for stage fright, while the assessment focused on procedural questions about Mehrabian's vocal parameters (Mehrabian). Despite this variance, the system preserved overall taxonomic integrity, demonstrating that localized RAG can serve as an automated enforcement engine for pedagogical congruence.

Algorithmic and RAG Performance Benchmarks

We evaluated the local Ollama execution environment and HorizaEdu model using RAGAS framework metrics across 120 generation batches (course content sections, lesson plans, task cards, assessment items).

Table 2
RAGAS Automated Evaluation Metrics for RAG Performance

Evaluation Metric	Mean Score	Target Minimum	Performance Margin
Faithfulness (Groundedness)	0.92	0.85	+0.07
Answer Relevance	0.96	0.85	+0.11
Context Recall	0.89	0.85	+0.04

These results can be interpreted as follows. The Faithfulness score of 0.92 indicates that the source-grounding block successfully restricted HorizaEdu to ingested vector representations: only three minor hallucination events were detected across 120 runs (a 2.5% hallucination rate), in which the model introduced generic definitions of communication tools absent from the local corpus. These errors occurred when the prompt requested "industry best practices" not explicitly covered in institutional documentation, suggesting that the constraint-enforcement mechanism functioned correctly by attempting fallback reasoning when the corpus was insufficient, a transparent failure mode that instructors could easily identify and correct. The Answer Relevance score of 0.96, in turn, demonstrates the efficiency of the six-block specification framework: by segregating operational constraints, pedagogical posture, and output formats into isolated segments, the language model consistently generated actionable, domain-specific tasks tailored to the engineering context rather than generic theoretical content. Finally, the Context Recall score of 0.89 reveals that the Similarity Retrieval Controller successfully surfaced relevant top-k text blocks in most generation cycles, although recall efficiency dropped slightly (to 0.82-0.84) when parsing highly specialized technical vocabularies in which source documents used overlapping terminology in different contexts, such as "presentation" across different communicative scenarios.

Technical Usability and Infrastructure Stability

The server-side architecture and Google AI Studio-generated application were validated during two synchronous testing windows: beta-testing (20 users) and final examination (150 students).

Several key findings emerged from these testing windows. In terms of login performance, the initial spike, with all 150 students attempting concurrent authentication, registered a mean latency of 940 ms, which then normalized to an optimal baseline of 220 ms mid-exam, indicating the system's capacity to scale dynamically. Uptime remained at 100% throughout the 60-minute examination window, with no server crashes, memory leaks, or transaction deadlocks detected. Security proved equally robust, as server-side validation rules successfully blocked 14 unauthorized payload interactions (1.1% of the 1,240 submitted inputs), demonstrating effective input sanitization against injection and token-bypass attacks. Database performance was likewise confirmed: mid-exam heartbeat checks at the 30- and 45-minute marks verified optimal persistence, and bulk submission processing at the 60-minute mark completed safely despite 150 concurrent form submissions and automated grading operations.

The no-code prototype generated by Google AI Studio, when scaffolded through systematic prompt engineering, produced a production-grade assessment platform capable of sustaining high-stakes, synchronous, remote examination for a full cohort. Latency spikes at login and submission represent expected network bottlenecks rather than architectural failures. The system demonstrates that non-technical educators, with AI assistance and iterative beta-testing, can build technically sound digital assessment infrastructure.

User Experience (UX/UI) Analysis

After submitting their exams, 150 students completed the System Usability Scale (SUS) instrument. The aggregate result was an overall mean SUS Score of 74.5, placing the AI-generated interface in the "Good" (Grade B) category on the standard usability continuum.

Table 3
System Usability Scale (SUS) Distribution across Usability Grades

Usability Grade	SUS Score Range	N = 150	Percentage
Grade A (Excellent)	80.3—100	38	25.3%
Grade B (Good)	68.0—80.2	93	62.0%
Grade C/D (OK)	51.0—67.9	19	12.7%
Grade F (Poor)	0.0—50.9	0	0.0%

The distribution shows that 87.3% of students rated the interface as "Good" or "Excellent," indicating successful usability, and item-level analysis revealed both clear strengths and areas for refinement. Students gave high scores to the structured test-runner view, as the persistent countdown timer and clear question-navigation layout effectively reduced testing anxiety and extraneous cognitive load, and the response-tracking functionality, which allowed learners to mark items for review, change answers, and receive instant confirmation, received consistently high marks. By contrast, lower scores, concentrated in the Grade C/D category, were isolated to mobile browser rendering: the AI-generated interface occasionally collapsed multi-column layouts on mobile viewports, forcing excessive scrolling to toggle between question text and option selections. This reveals a limitation, namely that while no-code prototyping tools generate functional database connectivity and application logic, responsive design optimization across all form factors still requires manual developer intervention.

Ethical and Legal Compliance Audit

The team audited the system architecture against data protection parameters to verify student privacy throughout the collection and analysis phases.

Table 4
Security and Legal Compliance Audit Results

Security Gate	Attempted Actions	Validated Actions	Blocked Anomalies	Compliance Rate
Student Digital Consent Form	150	150	0	100%
External Cloud Data Requests	85	0	85	100%
Server-Side Input Sanitization	1,240	1,226	14	100%

The system achieved complete compliance across all data protection vectors, strictly adhering to Moroccan Law No. 09-08 and CNDP guidelines. Institutional data decoupling functioned perfectly: private curriculum documentation and answer keys remained isolated on the local offline server, blocking all 85 unauthorized external cloud data requests. The gatekeeping layer successfully restricted authorization tokens for all 150 user profiles until students actively accepted the consent form terms. Student identification workflows anonymized upon session completion, replacing real identities with non-traceable cryptographic keys before database storage.

Discussion

Interpretation of Pedagogical Alignment Results

The mean Porter's Alignment Index of 0.88 provides strong empirical evidence that a RAG-augmented local AI system can operationalize constructive alignment as an automated design enforcement mechanism. This result is pedagogically consequential because it directly addresses a persistent challenge in higher education: maintaining alignment across the full course lifecycle. Biggs and Tang's theoretical framework posits that alignment is necessary for deep learning. However, Toohey's empirical work found that most instructors drift from their stated objectives due to time constraints and administrative workload. By restricting the HorizaEdu model to a curated corpus of institutional documentation and enforcing a six-block prompt structure that explicitly references alignment principles, we created a computational architecture that resists pedagogical drift. The model could not generate learning activities inconsistent with stated outcomes because the retrieval mechanism grounded all output in source documents that explicitly correlated outcomes, activities, and assessments.

The slightly lower alignment score in Chapter 2 (I = 0.84) is instructive, not problematic. The taxonomic mismatch between conceptual content generation and procedural assessment reflects a real challenge in course design: some cognitive objectives have multiple legitimate instantiations. The system flagged this inconsistency rather than silently masking it, allowing the instructor to decide whether the mismatch reflected a genuine design flaw or an acceptable pedagogical variation. This transparency is a feature of RAG-augmented systems that

traditional course design tools lack: every mismatch is traceable to specific source documents, enabling principled instructional decisions rather than opaque computational processes.

The findings extend Biggs and Tang's theoretical framework into computational territory. Constructive alignment has been described as a "design obligation," a normative claim about what instructors *should* do. Our results suggest that with appropriate technological scaffolding, alignment can become a *design constraint* – something the system enforces rather than relies on human intention. This distinction is significant for equity: instructors in resource-rich institutions with dedicated instructional designers can currently maintain alignment through human expertise and institutional support. Encoding alignment as a computational constraint could extend this capability to individual instructors in modest settings, addressing what Mollick identified as the critical battleground: equitable teacher access to AI-augmented instructional design tools.

Technical Feasibility of No-Code Assessment Platform Development

The technical evaluation results (100% uptime, 14 blocked injection attempts, SUS score of 74.5) show that non-technical educators can, with structured AI assistance and iterative beta-testing, produce production-grade assessment platforms. This finding challenges the implicit assumption in much educational technology literature that secure, technically sound digital assessment requires specialized software engineering expertise. Koehler and Mishra's TPACK framework predicted that technological knowledge would be essential for teachers using educational technology. Our results suggest that generative AI is restructuring this relationship: the Google AI Studio-generated application achieved enterprise-grade security (input sanitization, token-based authentication, server-side validation) without the instructor writing a single line of code or possessing any formal programming training.

The system's ability to withstand 14 attempted injection or bypass attacks without compromise indicates that current generative AI code generation models, when prompted with explicit security requirements, produce secure code structures comparable to those a professional developer would create. This is not incidental: it reflects a fundamental shift in how security-relevant knowledge is distributed. Rather than security being a specialized domain requiring security expertise, the instructor's specification of security requirements (in natural language, in the Design-constraints and Functional-requirements blocks) was sufficient for the AI to generate appropriate defensive implementations.

However, the mobile UX limitation reveals the boundaries of this approach. The AI generated a functionally correct, secure, and performant application that failed to optimize responsive design across all form factors. This reflects a general limitation of no-code generative AI: it excels at implementing explicitly specified behaviors. However, it requires additional design and validation effort to optimize non-functional properties (aesthetics, responsiveness, accessibility). For educators, this suggests a practical division of labor: generative AI can handle the complex, security-critical logic of authentication, data validation, and persistence, while instructors or instructional designers should reserve manual effort for visual hierarchy, responsive design, and accessibility refinement. The finding is thus not that no-code development eliminates the need for design expertise, but rather that it redistributes what expertise is required, freeing teachers from the burden of learning programming while still demanding attention to user experience design.

RAG Performance and the Problem of Hallucination

The Faithfulness score of 0.92, with only three hallucination events across 120 generation runs, provides quantitative validation of RAG's theoretical promise to reduce hallucination by grounding outputs in verified source material. This result is significant in an educational context because hallucination—the generation of plausible but false information—poses unique risks in course design. When a course design tool hallucinates pedagogical content (e.g., generating an assessment question on a topic not actually in the course materials), it can mislead instructors into believing they have maintained alignment when they have not.

The three detected hallucinations offer insights into RAG's failure modes in educational contexts. Each occurred when the instructor's prompt requested content that the corpus did not explicitly address—for example, a request for "industry best practices" on a communication technique that was discussed in the corpus but not explicitly framed as a "best practice." In these cases, the HorizaEdu model attempted to synthesize information from the corpus in ways that fell outside the strict source-grounding constraint. Significantly, these failures were transparent: the model generated content that was marked as tentative or flagged as synthesis rather than direct extraction. This suggests that RAG systems, when properly configured, do not fail silently but instead communicate uncertainty—a behavior that supports rather than undermines instructor agency.

The finding has broader implications for RAG in education. Existing literature on hallucination in LLMs often treats hallucination as a binary phenomenon: either the model generates information from the training data, or it hallucinates. Our results suggest a more nuanced picture: RAG creates a spectrum where hallucination risk is significantly reduced but not eliminated, and where the residual hallucinations occur in predictable contexts (when the corpus is incomplete). This boundary case is useful for instructors: it clarifies where the pedagogical gap lies, prompting curriculum refinement rather than masking it.

User Experience and the Normativity of AI-Generated Design

The SUS score of 74.5 sits above the overall mean of about 70 reported across thousands of SUS evaluations (Bangor et al. 117) but below the 80.3 threshold for "Excellent" in Table 3. The concentration of lower scores in the mobile UX category reveals something important about the current capabilities of no-code AI development: the systems excel at implementing specified functional requirements but do not automatically generate optimal user experience, particularly for secondary use cases (mobile access) that were not prioritized in the initial specifications.

This result should not be interpreted as a failure of no-code AI development but rather as an honest accounting of its current limits. Without programming knowledge, the instructor achieved a system that would have required weeks of professional development using traditional methodologies. The remaining refinements—responsive layout optimization, mobile-specific interaction patterns—represent design work that could be completed quickly through additional AI-assisted iteration or manual design intervention. The finding suggests that no-code educational technology development follows an 80/20 distribution: generative AI can efficiently produce 80% of required functionality (the security-critical, logic-heavy components), while the final 20% (visual design, responsive optimization, accessibility) benefits from human expert attention.

Sustainability and Institutional Context

A critical finding from the methodology section warrants emphasis in the discussion: the pedagogical ingest process and RAG configuration required substantial instructor expertise, despite the no-code nature of the assessment platform itself. The instructor spent approximately 30 hours curating the pedagogical corpus, chunking documents, and structuring the vector database. This represents a significant time investment that cannot be dismissed in practical terms. The implication is that no-code AI development democratizes *application development* – allowing non-technical educators to build functional software—but does not eliminate the need for pedagogical expertise and careful instructional design.

However, this time investment is better conceptualized as course design work rather than technical work. In traditional course development, an instructor would spend comparable time (often more) developing course materials without any guarantee of alignment across the course lifecycle. By investing that time in pedagogical design, the instructor gains both the immediate benefit of well-organized course materials and the downstream benefit of a system that can generate aligned supplementary materials, assessments, and remediations. This represents a shift in time investment—from executing isolated course components to strategically structuring course knowledge.

Limitations and Boundary Conditions

Several limitations merit explicit discussion because they define the scope of generalizability. First, the study relies on a single-cohort, single-domain design involving 150 students in one professional French course within one engineering specialization; these data therefore do not support generalization to other disciplines such as mathematics, laboratory sciences, or professional certifications, since different domains have different pedagogical requirements (mathematical proof, for instance, requires different course structures than language acquisition), and the RAG system's performance may vary significantly across domains with different formal knowledge requirements. Second, the temporal scope is limited to a single academic semester, so long-term sustainability questions remain unanswered, including whether the RAG system maintains its accuracy over multiple iterations, whether instructors continue to use it after initial interest fades, and whether the ingested corpus requires ongoing curation; as Bates emphasized, educational technology sustainability requires longitudinal commitment, whereas this study provides only a cross-sectional view. Third, with regard to institutional resources, although the study claims to address resource-constrained settings, the implementation required a dedicated server (12-core processor, 32 GB RAM) to run Ollama locally, an infrastructure that, while modest by enterprise standards, exceeds what many community colleges or smaller institutions can provision, which makes the model's transferability to even more constrained settings, such as schools with one or two servers and limited bandwidth, uncertain. Finally, the documented mobile usability challenges suggest that no-code AI development may systematically disadvantage mobile-first learners, so that in contexts where students predominantly access assessments via mobile devices, as is common in regions with limited desktop computer access, this limitation could substantially reduce the system's pedagogical effectiveness.

Broader Implications for Educational AI Policy

The study contributes evidence to an emerging policy debate about how educational institutions should position generative AI relative to concerns about academic integrity, data privacy, and

equitable access. The results suggest that locally executable, RAG-augmented AI systems offer a middle path between unfettered adoption of cloud-based commercial AI tools and wholesale rejection of AI in education. By maintaining data sovereignty (no transmission of student data to third parties, and of curricular documents only at the one-time indexing step), enabling individual educator agency (a single instructor can implement the system), and preserving pedagogical transparency (all outputs are traceable to source documents), local RAG-augmented systems address key concerns raised by critics of commercial educational AI while retaining the generative capabilities that proponents highlight.

However, the study also reveals that local AI systems are not cost-free solutions. They require institutional investment in computational infrastructure, instructor expertise in pedagogical design and course documentation, and ongoing technical maintenance. In this sense, they represent a different trade-off than either commercial cloud-based systems (which minimize upfront infrastructure cost but maximize data privacy and autonomy concerns) or traditional course development (which requires intensive human expert labor but produces codified, reusable course designs).

Conclusion

This empirical study investigated two interconnected questions about the capacity of locally executed artificial intelligence to support equitable, pedagogically sound course design and assessment in resource-constrained institutional settings. We presented evidence from a 150-student cohort in a blended professional French course, highlighting both the potential and the remaining challenges of this approach.

Primary Findings

The study yielded five key findings. First, constructive alignment can be operationalized as a computational constraint: a RAG-augmented local language model, when provided with curated institutional course documentation and structured prompt engineering, generated course materials and assessments with a mean alignment index of 0.88, exceeding the 0.80 target across all six course chapters, which demonstrates that pedagogical drift, historically a persistent challenge in course design, can be substantially mitigated through algorithmic enforcement of alignment principles. Second, local AI systems can achieve production-grade reliability with appropriate scaffolding, as the no-code assessment platform maintained 100% uptime under full load (150 concurrent users), successfully defended against injection and bypass attacks, and achieved a System Usability Scale score of 74.5 (“Good”), indicating that non-technical educators, through iterative AI-assisted prototyping and beta testing, can produce technically sound digital assessment infrastructure without specialized software engineering expertise. Third, RAG substantially reduces, but does not eliminate, hallucination in domain-specific generation tasks: across 120 generation batches, the system achieved 92% Faithfulness and 96% Answer Relevance, and the three detected hallucinations occurred in predictable contexts where the corpus was incomplete and were transparently flagged rather than masked, thereby supporting rather than undermining instructor agency. Fourth, no-code development democratizes application creation but not pedagogical expertise, since the instructor required no programming knowledge to build the assessment platform yet invested approximately 30 hours in pedagogical ingest (curating, chunking, and indexing course materials), which represents a redistribution rather than an elimination of the time required for sound course design. Fifth, mobile and responsive design optimization remain challenges for human experts: while the AI-generated platform achieved enterprise-grade security and database performance,

responsive design for mobile viewports required manual refinement, because current generative AI excels at implementing explicitly specified logical requirements but does not automatically optimize non-functional properties such as responsive layout.

Theoretical Contributions

This work extends three domains of educational scholarship. With respect to constructive alignment theory, we demonstrate that Biggs and Tang's alignment framework, previously understood as a normative design principle requiring human expertise to implement, can be computationally encoded and enforced; this suggests a new conceptualization of alignment as a design constraint rather than a design aspiration, implying that AI can automate consistency-checking functions and free human expertise for higher-order pedagogical decisions. Regarding the TPACK framework and teacher agency, Koehler and Mishra predicted that teacher expertise would emerge at the intersection of technological, pedagogical, and content knowledge, and our findings suggest that generative AI is restructuring this intersection, since pedagogical and content knowledge increasingly determine technological implementation, reducing, though not eliminating, the need for independent technological expertise; this has equity implications, as it can expand the pool of educators able to leverage educational technology for their specific instructional contexts. Finally, in relation to educational technology sustainability, Bates's framework emphasizes that sustainability depends on educators' capacity to maintain tools over time, and the no-code prototyping paradigm, supported by AI-assisted iteration, offers a plausible pathway toward this goal by ensuring that educators who build tools retain sufficient understanding to maintain and improve them without external technical dependency.

Practical Implications for Educators and Institutions

For individual educators, the study documents a replicable four-phase workflow for integrating AI into instructional design. The first phase, pedagogical preparation, involves investing in comprehensive course documentation, including syllabi, outcome specifications, rubrics, and exemplary materials, and treating these documents not merely as administrative artifacts but as the knowledge base that will ground the AI system. In the second phase, RAG configuration, the pedagogical corpus is chunked and indexed, and the retrieval mechanism is configured with appropriate semantic boundaries to minimize hallucination and maximize content relevance. The third phase, prompt engineering, consists of developing the six-block prompt structure (role, source-grounding, pedagogical stance, constraints, task, and output format) and iterating through successive course design phases, using the structured prompts to maintain consistency. Finally, the prototyping and validation phase uses generative AI prototyping environments to rapidly build the assessment platform, conduct beta testing with real users, and refine it iteratively based on feedback, while anticipating additional effort in responsive design and accessibility optimization beyond what the initial prototype provides.

For institutions, the study highlights the importance of strategic infrastructure investment in three areas. In terms of data governance, institutions need to establish clear policies about which course materials constitute institutional intellectual property, how pedagogical data should be stored, and how student data should be protected, since local AI systems require these policies to be explicit in ways that delegated cloud-based systems tend to obscure. Equally important is support for pedagogical design, as individual educators implementing RAG-augmented systems benefit substantially from instructional design support, that is, from someone who can help structure course documentation, validate alignment, and oversee quality assurance; this requirement is not unique to AI-augmented course design, but it is highlighted more starkly by

it. Finally, with regard to technical infrastructure, institutions seeking to support local AI execution require modest but consistent computational resources, such as servers with multiple cores and sufficient memory, which represents a different capital investment profile from both traditional course development and cloud-based commercial AI, the latter requiring no local infrastructure.

Remaining Research Questions

The study opens several directions for future investigation. A first question concerns cross-domain generalizability, namely whether RAG-augmented systems perform comparably across disciplines with different knowledge structures and what domain-specific adaptations are required for STEM fields, the humanities, and applied professional training. A second relates to longitudinal sustainability, that is, whether instructors continue to maintain and iteratively improve RAG-augmented systems over multiple years, and which factors support long-term adoption rather than abandonment. Issues of equity and access also deserve attention, particularly how institutional capacity, instructor technical expertise, and computational infrastructure constraints differentially affect adoption across diverse institutional contexts, and for which populations local AI-augmented course design improves access to quality instruction. Moreover, because this study measured alignment and usability without directly assessing learning outcomes, future research should examine whether students in courses designed with RAG-augmented systems demonstrate superior learning gains compared with those in traditionally designed courses. Finally, questions of responsible AI in education remain open, including how institutions should balance the autonomy and privacy benefits of local AI systems against the safety and accessibility benefits of cloud-based systems with dedicated abuse monitoring and accessibility teams, and which governance structures best support the responsible deployment of educational AI.

Concluding Remarks

This study provides empirical evidence that locally executed, RAG-augmented AI systems can meaningfully support equitable course design and assessment when appropriately scaffolded through pedagogical expertise, structured prompt engineering, and iterative validation. The results do not suggest that AI eliminates the need for instructional design expertise; rather, they reframe the relationship between expertise and tools. Where traditional course development positions AI-augmented systems as optional enhancements to expert human practice, this work suggests that strategic AI augmentation can make sound instructional design practice—specifically, maintaining constructive alignment—more accessible to individual educators without specialized institutional support.

The democratization of course design is neither inevitable nor automatic. It requires intentional pedagogical work (curating and structuring course materials), careful system design (RAG configuration, prompt engineering), and honest acknowledgment of remaining limitations (mobile UX, hallucination risk). However, the evidence presented here suggests that for educators willing to invest that work, generative AI offers a credible pathway toward operationalizing pedagogical principles that have remained aspirational in practice for decades.

As higher education grapples with questions of cost, access, and educational quality, individual educators' capacity to leverage AI for personalized, aligned, secure course design may prove as consequential as the technology itself. The critical institutional question is therefore not whether to adopt educational AI, but how to invest in the pedagogical expertise,

technical infrastructure, and governance frameworks required to support educators in using AI responsibly and effectively.

Works Cited

Anderson, Lorin W., and David R. Krathwohl, editors. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman, 2001.

Bangor, Aaron, et al. "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale." *Journal of Usability Studies*, vol. 4, no. 3, 2009, pp. 114-23.

Bates, A. W. *Teaching in a Digital Age: Guidelines for Designing Teaching and Learning*. 2nd ed., Tony Bates Associates, 2019, https://pressbooks.bccampus.ca/teachinginadigitalagev2/.

Beck, Kent, et al. "Manifesto for Agile Software Development." *Agile Manifesto*, 2001, agilemanifesto.org.

Biggs, John, and Catherine Tang. *Teaching for Quality Learning at University: What the Student Does*. 4th ed., Open University Press, 2011.

Brooke, John. "SUS: A 'Quick and Dirty' Usability Scale." *Usability Evaluation in Industry*, edited by Patrick W. Jordan et al., Taylor and Francis, 1996, pp. 189-194, https://doi.org/10.1201/9781498710411.

Depover, Christian, et al. *Les environnements d'apprentissage multimédia: Analyse et conception*. L'Harmattan, 1998.

Es, Shahul, et al. "RAGAs: Automated Evaluation of Retrieval Augmented Generation." *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, Association for Computational Linguistics, 2024, pp. 150-158, https://doi.org/10.18653/v1/2024.eacl-demo.16.

Garrison, D. Randy. *E-Learning in the 21st Century: A Community of Inquiry Framework for Research and Practice*. 3rd ed., Routledge, 2017, https://doi.org/10.4324/9781315667263.

Garrison, D. Randy, and Terry Anderson. *E-Learning in the 21st Century: A Framework for Research and Practice*. Routledge, 2003, https://doi.org/10.4324/9780203166093.

Google. *Google AI Studio*, 2024, https://aistudio.google.com/.

Kasneci, Enkelejda, et al. "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education." *Learning and Individual Differences*, vol. 103, 2023, article 102274, https://doi.org/10.1016/j.lindif.2023.102274.

Koehler, Matthew J., and Punya Mishra. "What Is Technological Pedagogical Content Knowledge?" *Contemporary Issues in Technology and Teacher Education*, vol. 9, no. 1, 2009, pp. 60-70, https://citejournal.org/volume-9/issue-1-09/general/what-is-technological-pedagogical-content-knowledge/.

Lewis, Patrick, et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459-9474, https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

Mehrabian, Albert. *Silent Messages*. Wadsworth, 1971.

Mollick, Ethan. *Co-Intelligence: Living and Working with AI*. Portfolio, 2024.

Ollama. *Ollama: Run Large Language Models Locally*. Version 0.1.32, 2024, https://ollama.com/.

Porter, Andrew C. "Measuring the Content of Instruction: Uses in Research and Practice." *Educational Researcher*, vol. 31, no. 7, 2002, pp. 3-14, https://doi.org/10.3102/0013189X031007003.

Prensky, Marc. *Teaching Digital Natives: Partnering for Real Learning*. Corwin Press, 2010.

Russell, Stuart J., and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed., Pearson, 2021.

Selwyn, Neil. *Should Robots Replace Teachers? AI and the Future of Education*. Polity Press, 2019.

Toohey, Susan. *Designing Courses for Higher Education*. SRHE and Open University Press, 1999.

Wang, Feng, and Michael J. Hannafin. "Design-Based Research and Technology-Enhanced Learning Environments." *Educational Technology Research and Development*, vol. 53, no. 4, 2005, pp. 5-23, https://doi.org/10.1007/BF02504682.

Wiggins, Grant, and Jay McTighe. *Understanding by Design*. 2nd ed., ASCD, 2005.

Williamson, Ben. *Big Data in Education: The Digital Future of Learning, Policy and Practice*. SAGE Publications, 2017, https://doi.org/10.4135/9781529714920.

Zawacki-Richter, Olaf, et al. "Systematic Review of Research on Artificial Intelligence Applications in Higher Education: Where Are the Educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, 2019, article 39, https://doi.org/10.1186/s41239-019-0171-0.