Journal of New Approaches in Educational Research

RESEARCH Open Access

Development and validation of an adaptive learning readiness assessment framework for Moodle courses

Serhiy Semerikov[1,2*], Pavlo Nechypurenko[1,3*], Tetiana Vakaliuk[4,5*], Iryna Mintii[6,7*] and Liliia Fadieieva[1*]

*Correspondence:
Serhiy Semerikov
semerikov@gmail.com
Pavlo Nechypurenko
acinonyxleo@gmail.com
Tetiana Vakaliuk
tetianavakaliuk@gmail.com
Iryna Mintii
irina.mintiy@gmail.com
Liliia Fadieieva
fadeyevaliliya@gmail.com
[1]Kryvyi Rih State Pedagogical University, Kryvyi Rih, Ukraine
[2]Center for Information-Analytical and Technical Support of Nuclear Power Facilities Monitoring of the NAS of UkraineK, Kyiv, Ukraine
[3]Academy of Cognitive and Natural Sciences, Kryvyi Rih, Ukraine
[4]Zhytomyr Polytechnic State University, Zhytomyr, Ukraine
[5]Institute for Digitalisation of Education of the NAES of Ukraine, Kryvyi Rih, Ukraine
[6]Lviv Polytechnic National University, Lviv, Ukraine
[7]University of Łódź, Łódź, Poland

Abstract

Purpose While Moodle is widely adopted in higher education, institutions struggle to leverage its features for adaptive learning. This study develops and validates the Adaptive Learning Readiness Assessment Framework (ALRAF), a course-level diagnostic instrument for evaluating a Moodle course's structural capability to support adaptive learning experiences.

Design We employ a quantitative cross-sectional design coupled with a novel Multi-LLM Synthetic Expert Consensus (MLSEC) protocol for content-validity evidence. ALRAF was developed through literature synthesis grounded in the ICAP framework (Chi and Wylie, Chi, *Educational Psychologist49*, 2014) and validated by a 40-panelist synthetic expert panel constructed across eight large-language-model providers and five stratified expert personas using two pre-registered Delphi rounds with falsifiable decision rules (I-CVI ≥ 0.78, V-Aiken ≥ 0.70, modified $\kappa^* \geq 0.74$). The validated framework was applied to a Moodle 3.8.2 dataset of 985 courses delivered at Kryvyi Rih State Pedagogical University (Ukraine) across 2020–2022.

Findings The synthetic panel converged on six dimensions: Content Variety, Interaction Diversity, Assessment Flexibility, Learning Path Personalization, Feedback Mechanisms, and the panel-proposed AI & Data-Driven Adaptivity Integration (ADAI). The six-dimensional ALRS_6 correlated significantly – but *negatively* – with the proportion of high grades ($A + B$): Pearson $r = -0.22$, and positively with low grades ($E + F$) : $r = +0.22$ (both $p < .001$). This inverse relationship runs counter to what would be expected if ALRAF directly indexed pedagogical quality, and reframes ALRAF as a measure of *structural readiness* rather than learning effectiveness; we interpret the sign in Sect. 6 in terms of course-difficulty and compensatory-engineering effects. Faculty differences were significant (ANOVA $F(8, 976) = 10.26$, $p < .001$, $\eta^2 = 0.078$, $\omega^2 = 0.070$). A multiple-regression model controlling for educational level, form of education, and faculty achieved adjusted $R^2 = 0.18$ ($F(19, 965) = 12.10$, $p < .001$). The framework reveals strong implementation of content variety but near-zero readiness in learning-path personalization and AI integration – itself a notable institutional finding.

Contribution Methodologically, the study introduces MLSEC as a transparent AI-augmented approach to rubric content validation, with synthetic-panel limitations

Springer

explicitly disclosed. Substantively, ALRAF provides a replicable structural-readiness index whose correlations with student outcomes are non-trivial in direction and magnitude; it helps institutions identify capability gaps (especially around AI integration) without claiming to forecast student success.

Keywords Adaptive learning, Moodle, Readiness assessment, Framework development, Higher education, Student outcomes, Content validity, Synthetic expert panel, Large language models

1 Introduction

The integration of technology in education has significantly altered teaching and learning paradigms, introducing innovative approaches to address diverse student needs (Li et al., 2021). Among these innovations, adaptive learning systems represent a substantial advancement, offering personalized educational experiences tailored to individual learner characteristics and needs (Kabudi et al., 2021, Semerikov et al., 2026). Adaptive learning technologies adjust content, instructional approaches, and learning paths based on student performance, behavior, and preferences, potentially enhancing engagement and academic achievement (Martin et al., 2020).

Learning Management Systems (LMS) like Moodle have become ubiquitous in higher education, serving as platforms for delivering online and blended learning experiences. While Moodle was not originally designed as an adaptive learning system, its modular architecture and extensive feature set make it potentially capable of supporting adaptive learning approaches when thoughtfully implemented (Fadieieva, 2021, Semerikov et al, 2025). As Khan et al. (2019) noted, Moodle outperforms other open-source LMS platforms in adaptivity features, suggesting significant potential for implementing adaptive learning strategies.

Despite this potential, many institutions struggle to leverage Moodle's capabilities for adaptive learning. The disconnect between technological possibilities and pedagogical implementation represents a significant challenge for educational institutions seeking to enhance personalization in online learning (Mirata et al., 2020). This gap is particularly pronounced in contexts where resources for specialized adaptive learning platforms are limited, such as in many public universities (Lubchak et al., 2012).

The adaptive learning readiness of an educational system can be conceptualized as its potential to support personalized learning experiences through appropriate technological infrastructure, pedagogical design, and implementation strategies. While several frameworks exist for evaluating e-learning readiness (Al-Rikabi and Montazer, 2024), few specifically address adaptive learning readiness at the course level, particularly in the context of Moodle implementations. This research gap limits institutions' ability to assess and enhance their existing courses for adaptive learning potential.

The present study aims to address this gap by developing and validating an Adaptive Learning Readiness Assessment Framework (ALRAF) for Moodle courses. We pursue this aim through three combined contributions: (i) a literature- and ICAP-grounded multi-dimensional rubric; (ii) a novel Multi-LLM Synthetic Expert Consensus (MLSEC) protocol for content validation, framed transparently as an AI-augmented validation step rather than a substitute for human-expert validation; and (iii) an empirical instantiation of the framework on 985 Moodle 3.8.2 courses at a Ukrainian pedagogical university, with full diagnostic statistics. The specific characterizations of the framework's

structural-readiness positioning, the dimensions it surfaces as institutionally weak (Learning Path Personalization and AI integration), the direction of its correlations with student outcomes, and its diagnostic role for institutions are formal results presented in Sects. 5–7, not anticipated here.

The study addresses the following research questions:

1. Which combinations of Moodle components create the necessary conditions for adaptive learning?
2. Can we quantify a course's potential for supporting adaptive learning based on its design?
3. What is the relationship between adaptive learning readiness scores and student outcomes?

The paper is organized as follows: Sect. 2 reviews relevant literature on adaptive learning, Moodle capabilities, AI-era adaptive systems, and existing evaluation frameworks. Section 3 describes the theoretical foundations of the proposed framework, including the six dimensions and their ICAP-anchored rationale. Section 4 outlines the methodology, including the explicit-formula scoring rubric (Sect. 4.2), the novel Multi-LLM Synthetic Expert Consensus (MLSEC) content-validation procedure (Sect. 4.3), the Moodle 3.8.2 KSPU dataset and sampling frame (Sect. 4.4), and the diagnostic-aware data-analysis pipeline (Sect. 4.5). Section 5 presents results: MLSEC content-validity evidence, score distributions and distributional diagnostics, dimensional and faculty analyses with effect sizes, the full multiple-regression model with diagnostics, and $\mathrm{ALRS}_5/\mathrm{ALRS}_6$ robustness. Section 6 discusses interpretation, practical implications, the AI-era adaptive-Moodle context, and limitations. Section 7 concludes and identifies future-research priorities.

2 Literature review

2.1 Adaptive learning in higher education

Adaptive learning represents an educational approach that adjusts learning experiences to meet the individual needs, preferences, and performance of learners (Alajlani et al., 2024). The concept has evolved significantly with the development of technology, moving from simple branching models to sophisticated systems incorporating artificial intelligence and learning analytics (Xie et al., 2019). Contemporary adaptive learning systems typically incorporate several key elements: learner profiling, content adaptation, path adaptation, and adaptive assessment (Li et al., 2021).

The theoretical foundations of adaptive learning draw from various educational and psychological perspectives. Constructivist theories emphasize the importance of tailoring learning experiences to individual learners' existing knowledge and cognitive structures (Jordan, 2013). The concepts of scaffolding and zone of proximal development highlight the value of providing appropriate levels of support as learners develop (Wu et al., 2018). Additionally, self-regulated learning theory emphasizes the importance of learner agency and metacognition in successful adaptive systems (Mejeh et al., 2024).

Recent systematic reviews of adaptive learning research have identified several key trends and developments. Martin et al. (2020) analyzed 61 articles on adaptive learning from 2009 to 2018, finding that the majority of studies occurred in higher education, with the highest concentration in computer science disciplines. Their analysis identified

learning style as the most commonly observed learner characteristic used for adaptation, while adaptive feedback and navigation were the most frequently investigated adaptive targets.

The rise of large language models (LLMs) and generative artificial intelligence since 2023 (Marienko et al., 2026, Zahorodko and Semerikov, 2026) has substantively shifted the adaptive-learning landscape. Khalifeh et al. Khalifeh et al. (2026) synthesized 2019–2025 evidence in an updated systematic review and concluded that AI-driven personalization is redefining the very meaning of personalized learning, moving the field from rule-based adaptation toward generative, conversational, and analytics-driven approaches. Bond et al. (2024) provided a meta-systematic review of AI in higher education and called for stronger ethical foundations, cross-disciplinary collaboration, and methodological rigor, observing that the volume of AI-in-education research has outpaced its theoretical and methodological maturity. Empirically, Ezzaim et al. (2024) demonstrated improved student engagement and academic performance when an AI-based adaptive plugin was integrated into Moodle, and Neumann et al. (2025) reported on the design and effects of an LLM-driven chatbot within a higher-education course. Prada Segura Prada Segura (2026) described an implementation of an AI-based adaptive learning system inside Moodle in which evaluation processes were personalized via machine-learning models. Together, this body of work motivates two specific demands on any contemporary readiness framework: (i) it must accommodate *external-tool and data-driven adaptivity* (LTI, learning analytics, learner-model storage), and (ii) it must be evaluated as a *structural capability surface* rather than a measure of pedagogical enactment per se, which still requires direct observation or course-content analysis.

The effectiveness of adaptive learning approaches has been demonstrated across various educational contexts. Dziuban et al. (2018) conducted a comparative study across disciplines and universities, finding that adaptive learning had a stabilizing influence on learning outcomes across different subjects. Similarly, Ezzaim et al. (2024) documented positive impacts of an AI-based adaptive learning Moodle plugin on student engagement and academic performance in Moroccan high schools.

Despite their benefits, implementing adaptive learning approaches presents challenges. Rivera Munoz et al. (2022) identified technological infrastructure, faculty training, and instructional design as key barriers to successful implementation. Mirata et al. (2020) highlighted additional challenges including organizational readiness, stakeholder acceptance, and cultural factors. These challenges are particularly pronounced in resource-constrained contexts, underscoring the need for frameworks that can help institutions maximize adaptive learning potential within existing systems.

2.2 Moodle as a platform for adaptive learning

Moodle is one of the most widely used open-source learning management systems globally. While not designed specifically as an adaptive learning system, Moodle offers various features that can support adaptive learning experiences when thoughtfully implemented (Smyrnova-Trybulska et al., 2022).

According to Fadieieva et al. (2024), Moodle provides several tools that can be leveraged for adaptive learning:

1. *Modular course structure*: Content can be organized into smaller, modular units, allowing students to learn at their own pace and potentially follow different paths through the material.
2. *Variety of activities and resources*: Moodle supports diverse learning materials (documents, videos, interactive content) and activities (discussions, collaborative work) that can accommodate different learning styles and preferences.
3. *Conditional activities and restrictions*: These features allow instructors to create personalized learning paths based on student performance or preferences, revealing content based on completion criteria or grades.
4. *Interactive assessment tools*: Quizzes, assignments, and other assessment activities can provide immediate feedback and adapt to student performance.
5. *Communication and collaboration tools*: Forums, chats, and other communication tools facilitate interaction and peer learning, supporting the social aspects of adaptive learning.

Recent developments have enhanced Moodle's adaptive capabilities. Tamo-Vargas et al. (2023) developed a plugin for adaptivity through graded grading constraints, while Krahn et al. (2023) created a personalized study guide plugin that generates learning paths for students. Moreno-Marcos et al. (2023) developed a learning analytics tool to analyze student actions in Moodle, potentially informing adaptive interventions. Additionally, Vasquez-Bermudez et al. (2023) employed unsupervised learning techniques to analyze forum interactions, which could support adaptive social learning environments.

Empirical studies have demonstrated the potential effectiveness of Moodle-based adaptive approaches. Holthaus et al. (2018) implemented an adaptive instructional design for mathematics in Moodle, finding positive effects on learning outcomes for both stronger and weaker students. Limongelli et al. (2010) integrated the Lecomps adaptive learning system with Moodle, enabling personalized learning object sequences based on students' cognitive states and learning styles.

Despite these promising developments, Smyrnova-Trybulska et al. (2022) found that students still perceive a lack of personalization in Moodle courses, suggesting a gap between Moodle's technical capabilities and current implementation practices. This underscores the need for frameworks that can guide instructors and institutions in enhancing the adaptive features of their Moodle courses.

2.3 Frameworks for evaluating adaptive learning readiness

While various frameworks exist for evaluating e-learning readiness, few specifically address adaptive learning readiness, particularly in the context of Moodle-based courses. Al-Rikabi and Montazer (2024) developed an e-learning readiness assessment model for Iraqi universities employing the Fuzzy Delphi method, focusing on three main dimensions and 13 factors with 86 measures. While comprehensive, this model addresses general e-learning readiness rather than specific adaptive learning capabilities.

Badhe et al. (2021) proposed design guidelines for scaffolding self-regulation in personalized adaptive learning systems, which include clear learning objectives, progress monitoring, and adaptive feedback. While these guidelines offer valuable insights for designing adaptive systems, they do not provide a concrete framework for evaluating existing courses.

Mirata and Bergamin (2019) developed an implementation framework for adaptive learning based on a Delphi study with experts from universities in Switzerland and South Africa. Their framework identified five dimensions affecting adaptive learning implementation: technology, teaching and learning, organization, law and regulations, and cultural and political conditions. While comprehensive, this framework focuses on institutional-level implementation rather than course-level design features.

Khan et al. (2019) developed a method for evaluating the adaptivity of web-based open-source LMS platforms, finding that Moodle outperformed other platforms in adaptivity features. Their evaluation used a fuzzy logic approach to assess categorical and adaptivity features, providing a valuable model for comparing platforms but not for evaluating individual courses.

Cavanagh et al. (2020) proposed a design framework for adaptive learning in higher education, identifying five key design features: clear learning objectives, content chunking, formative assessment, adaptive remediation, and summative assessment. This practitioner-focused framework offers valuable guidelines for course design but does not provide a quantitative approach for evaluating existing courses.

The literature reveals a notable gap in frameworks specifically designed for evaluating the adaptive learning readiness of existing Moodle courses. Existing frameworks either focus on institutional readiness, platform comparison, or design guidelines, without providing a concrete, course-level evaluation approach. This gap limits institutions' ability to assess and enhance the adaptive learning potential of their existing Moodle courses, highlighting the need for the framework developed in this study.

3 Theoretical framework

The Adaptive Learning Readiness Assessment Framework (ALRAF) developed in this study draws on several theoretical perspectives and models. At its foundation lies the understanding that adaptive learning systems aim to accommodate individual differences among learners, providing personalized learning experiences that align with each learner's needs, preferences, and characteristics (Afini Normadhi et al., 2019).

3.1 Dimensions of adaptive learning

We derive the dimensions of ALRAF in two stages. First, we synthesize the literature using two grounding constraints: (a) each candidate dimension must be operationalizable through observable Moodle course components (precluding dimensions that would require classroom observation or instructor interviews to score); and (b) each candidate must map onto a recognized cognitive-engagement category in the ICAP framework (Chi and Wylie, 2014), which distinguishes Passive, Active, Constructive, and Interactive engagement modes and predicts that learning increases as engagement deepens. Applying these criteria to the literature surveyed in Sect. 2, we identify five initial dimensions; we then augment this set with one further dimension produced through the synthetic-panel validation procedure detailed in Sect. 4.3, yielding the final six-dimensional framework.

The initial five dimensions were prioritized over alternatives (e.g., gamification readiness, accessibility/UDL readiness, social-learning readiness) for three reasons: (i) all five appear consistently across the e-learning readiness and adaptive-learning design literatures (Khan et al., 2019, Mirata and Bergamin, 2019, Cavanagh et al., 2020, Al-Rikabi

and Montazer, 2024); (ii) each maps directly to Moodle component categories that are exposed through the standard reporting plugin used to construct our dataset; and (iii) together they span all four ICAP modes, with Content Variety covering Passive–Active resource consumption, Interaction Diversity covering Constructive activities, and Learning Path Personalization plus Feedback Mechanisms covering the Interactive mode. Alternative dimensions either fail criterion (a) (gamification readiness would require coding course narratives), overlap substantially with our five (social-learning readiness collapses into Interaction Diversity), or warrant separate research programs (accessibility/UDL deserves its own scoring rubric).

1. *Content Variety*

   The availability of diverse learning resources is a fundamental aspect of adaptive learning, allowing different learners to engage with materials that match their preferences and learning styles (Xie et al., 2019). As Jordan (2013) noted, providing multiple representations of content supports constructivist approaches to learning by accommodating different ways of understanding and processing information. In Moodle, this dimension is reflected in the variety and quantity of resource types (such as documents, videos, interactive content) available to students.

2. *Interaction Diversity*

   Interactive learning activities serve multiple purposes in adaptive systems, providing opportunities for engagement, practice, and formative assessment (Martin et al., 2020). Different types of interactions can support different cognitive processes and learning styles (Afini Normadhi et al., 2019). This dimension evaluates the range of interactive activities that engage students in different ways, including both individual and collaborative learning experiences.

3. *Assessment Flexibility*

   Adaptive assessment is a core component of personalized learning, enabling the tailoring of evaluation approaches to individual learners' needs and progress (Bevan et al., 2023). Flexible assessment approaches can accommodate different ways of demonstrating knowledge and provide opportunities for formative feedback (Rideout, 2018). This dimension examines the variety of assessment methods and their potential for providing personalized feedback and guidance.

4. *Learning Path Personalization*

   The ability to create personalized learning paths is perhaps the most distinctive feature of adaptive learning systems (Zang, 2024). By adjusting the sequence and presentation of content based on learner characteristics and performance, adaptive systems can optimize the learning experience for each individual (Pukkhem et al., 2006). This dimension focuses on features that enable conditional access, branching, and adaptive content delivery.

5. *Feedback Mechanisms*

   Timely and appropriate feedback is essential for guiding learners and supporting self-regulation (Dainton et al., 2020). Adaptive feedback mechanisms can provide personalized guidance based on learner performance and characteristics (Bimba et al., 2021). This dimension assesses the tools and approaches available for providing feedback and facilitating communication between instructors and students.

6. *AI & Data-Driven Adaptivity Integration* (ADAI; sixth dimension added through the MLSEC validation, see Sect. 4.3)

   The 2023–2026 generation of adaptive learning leverages external AI tutors, learning-analytics back-ends, and learner-model storage to enable adaptation that exceeds the rule-based affordances of earlier LMS designs (Bond et al., 2024, Ezzaim et al., 2024, Neumann et al., 2025, Khalifeh et al., 2026, Prada Segura, 2026). Even in a Moodle course that does not itself host adaptive logic, the infrastructure required to plug in such adaptivity – external-tool gateways (LTI), learner-data collection (Database, Survey, Feedback), and SCORM/xAPI-compatible content – is itself an aspect of readiness. This sixth dimension captures whether a course is structurally prepared to integrate with the contemporary AI-and-analytics adaptive ecosystem, and is the panel-derived addition that emerged from our synthetic-panel content-validity procedure (Sect. 4.3; threshold endorsement: dimension definition $I - \mathrm{CVI} = 1.00$, $V - \mathrm{Aiken} = 0.89$ in Round 2; proposed independently by 33 of 38 panelists in Round 1 across all 8 base models).

3.2 Theoretical model

The theoretical model underlying ALRAF posits that the six structural dimensions interact to create the *capability surface* of a course – the set of structural affordances that make adaptive learning experiences feasible. Each dimension contributes to the overall structural readiness, though their relative importance varies with contextual factors and pedagogical approaches.

We explicitly distinguish three layers in the model. The *structural-readiness* layer (the six ALRAF dimensions) captures what a Moodle course *can* support; it is fully operationalizable from course-design data. The *moderating-factors* layer (student characteristics – prior knowledge, motivation, self-regulation, demographics; and instructor practices – pedagogical approach, feedback culture, engagement strategies) influences whether that structural potential is realized in any given enrolment. ALRAF does not score these moderators because doing so would require classroom observation, learner self-report, or instructor interviews; instead, where individual-difference variables are available in the dataset (educational level, number of teachers, form of education), we enter them as covariates in our regression models (Sect. 5.6). The third layer, *learning outcomes*, is the joint product of structural readiness, moderators, and unobserved factors. Figure 1 illustrates this layered relationship.

This model recognizes that the dimensions are not independent but interact with and reinforce each other. For example, assessment flexibility may inform learning path personalization, while feedback mechanisms may enhance the effectiveness of diverse

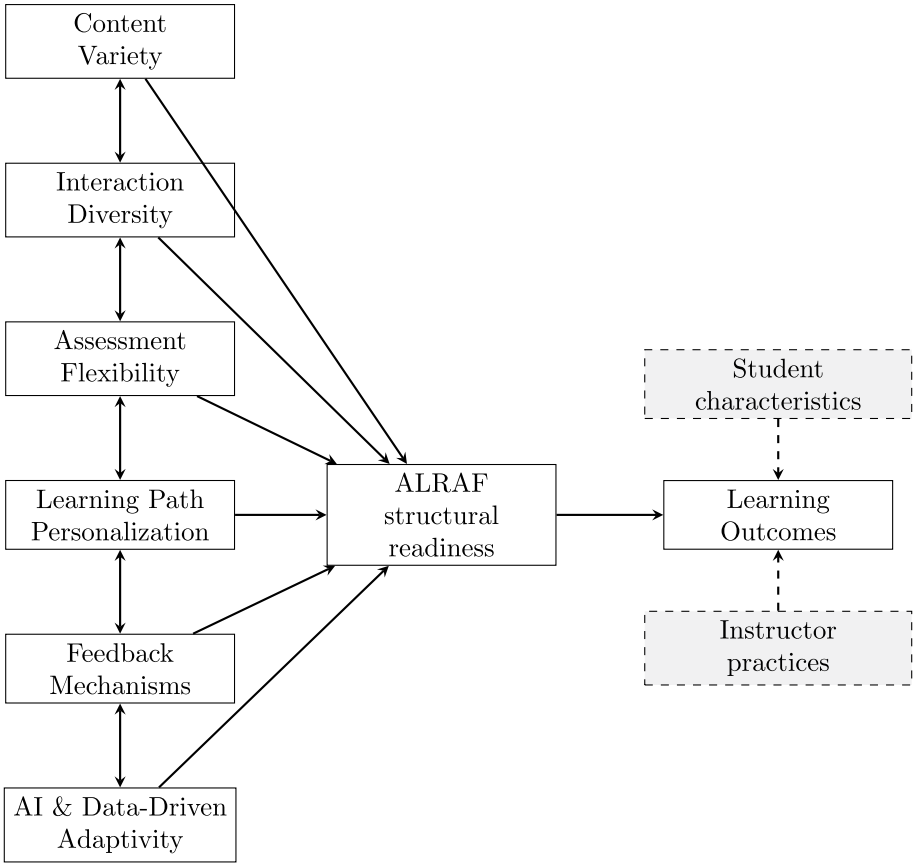

Fig. 1 Theoretical model of ALRAF structural readiness, the six ALRAF dimensions, and their relationship to learning outcomes. Dashed boxes denote moderating factors that influence outcomes but are not part of the structural-readiness construct

content and interactions. The model also acknowledges that structural readiness is one of several factors influencing learning outcomes; moderating factors (student characteristics and instructor practices) are represented explicitly as additional inputs to outcomes, and are entered as covariates in the multiple-regression analyses presented in Sect. 5.6 to the extent the dataset permits.

The theoretical framework thus provides a foundation for operationalizing adaptive learning readiness in Moodle courses through specific features and components, which is elaborated in the methodology section.

4 Methodology

4.1 Research design

The study employs a quantitative cross-sectional design with an embedded content-validation component. Specifically, the methodology comprises four interlocking phases:

1. *Theory-driven framework synthesis (Sect. 4.2).* An initial five-dimensional rubric is derived from the adaptive-learning, e-learning-readiness, and Moodle-implementation literatures, with each dimension grounded in the ICAP cognitive-engagement framework (Chi and Wylie, 2014) and constrained to be operationalizable from standard Moodle reporting data.

2. *Synthetic-panel content validation (Sect. 4.3).* A novel Multi-LLM Synthetic Expert Consensus (MLSEC) procedure – adapted from Lynn's content-validity-index tradition (Lynn, Lynn, 1986, Polit and Beck, 2006) and the emerging synthetic-respondent literature (Aher et al., 2023, Argyle et al., 2023, Bisbee et al., 2024, Park et al., 2026) – subjects the initial rubric to two pre-registered Delphi rounds. Decision rules for item retention and new-dimension inclusion are pre-registered against falsifiable thresholds.
3. *Empirical instantiation (Sect. 4.4 and Sect. 4.5).* The post-validation rubric is applied to a Moodle 3.8.2 corpus of 985 courses, with normality, homogeneity, and regression-assumption diagnostics performed before any inferential test.
4. *Robustness analysis (Sect. 5.7).* The initial five-dimensional score (ALRS_5) and the validated six-dimensional score (ALRS_6) are compared via Bland–Altman plots, Kendall's τ_b, and band-reclassification analysis to assess whether the framework's substantive conclusions depend on the specific dimensional structure.

Throughout, the design follows the structural-readiness positioning articulated in Sect. 3: ALRAF measures the *capability surface* of a course – the structural prerequisites for adaptive learning – rather than the pedagogical enactment of adaptive learning per se, which would require direct classroom observation or content coding beyond the scope of the standard Moodle reporting data used here. This positioning is consequential for the rubric design (Sect. 4.2), the synthetic-panel content validity claims (Sect. 4.3), and the framing of correlations with student outcomes (Sect. 5.6).

4.2 Framework development

The development of ALRAF followed a systematic process informed by both theoretical considerations and the practical capabilities of the Moodle LMS. Following approaches similar to Mirata and Bergamin (2019) and Khan et al. (2019), we constructed a multi-dimensional rubric in which each dimension is scored on the 0–20 interval and the total ALRS is the sum of dimensional scores.

For each of the five theoretical dimensions identified in Sect. 3, we mapped specific Moodle components and features that could contribute to adaptive-learning capability. The initial mapping was derived from three sources: (i) systematic literature review of adaptive-learning and Moodle-implementation studies (Sect. 2), (ii) Moodle 3.8.2 documentation and the Course Module Instances Report plugin schema (Vitez, 2022), and (iii) component-to-engagement mapping informed by the ICAP framework's four-level cognitive-engagement taxonomy (Chi and Wylie, 2014). The mapping was then subjected to the synthetic-panel content validation described in Sect. 4.3, which retained all five initial dimensions (I-CVI ≥ 0.78 for every dimension definition) and added a panel-derived sixth dimension. Table 1 presents the resulting six-dimensional component map.

Each dimension is scored on a 0–20 interval as the sum of two 0–10 sub-scores. Let $\mathcal{C}_d$ denote the set of Moodle components mapped to dimension d in Table 1, n_c the number of instances of component c in a given course, and $q_{95}(\sum_{c\in\mathcal{C}_d} n_c)$ the 95th-percentile total instance count across the dataset for that dimension. The first sub-score is a *breadth* term, the second a *quantity* term, both capped at 10:

$$\mathrm{breadth}_d = \frac{10}{|\mathcal{C}_d|} \cdot \sum_{c\in\mathcal{C}_d} \mathbf{1}[n_c > 0] \tag{1}$$

Table 1 Dimensions of the validated framework and associated Moodle components. The sixth dimension (ADAI) was added through synthetic-panel content validation in Sect. 4.3

Dimension	Associated Moodle components
Content Variety	URL, Book, Label, Page, Folder, File
Interaction Diversity	H5P, HotPot, SCORM, Survey, Database, Choice, Wiki, Glossary, Assignments, Workshop, Quiz, Lesson
Assessment Flexibility	Quiz, Assignments, Workshop, Lesson, H5P
Learning Path Personalization	Lesson, SCORM, Conditional Activities, Restriction Sets
Feedback Mechanisms	Forum, Chat, Feedback, Survey
AI & Data-Driven Adaptivity Integration	LTI External tool activity, H5P, Database, SCORM, Quiz, Survey, Feedback

$$\text{quantity}_d = 4 \cdot \min\left(1, \frac{\sum_{c \in \mathcal{C}_d} n_c}{q_{95}(\sum_{c \in \mathcal{C}_d} n_c)}\right) \tag{2}$$

$$\text{diversity}_d = 10 \cdot \min\left(1, \frac{\sum_{c \in \mathcal{C}_d} \mathbf{1}[n_c > 0]}{|\mathcal{C}_d|}\right) \tag{3}$$

The diversity sub-score in Equation (3) operationalizes diversity as the proportion of the dimension's possible component types that are actually used in the course. We adopt this breadth interpretation in place of a *unique-types-over-total-resources* ratio, because the latter perversely penalizes high-quantity courses (a course offering one instance of each of six resource types scores higher than a course offering many instances of each), which conflicts with the intent of the dimension and was a source of inter-rater disagreement in MLSEC Round 1. The comparison between the five- and six-dimensional scorings is reported transparently in Sect. 5.7 (ALRS_5 vs ALRS_6 correspondence).

Dimensional scores follow the same structural template with two dimension-specific elements:

- *Content Variety (CV) – 20 pts*: $\text{breadth}_{\text{CV}} + \text{quantity}_{\text{CV}}$ (capped at 10) $+\text{diversity}_{\text{CV}}$.
- *Interaction Diversity (ID) – 20 pts*: same formulation over the 12 activity types.
- *Assessment Flexibility (AF) – 20 pts*: variety sub-score (Equation 1 + Equation 2, capped at 10) + formative sub-score (Equation 1 + Equation 2 restricted to formative-supporting components Quiz, Lesson, Workshop, H5P).
- *Learning Path Personalization (LPP) – 20 pts*: conditional sub-score + delivery sub-score, where the latter is the proxy $5 \cdot \mathbf{1}[\text{Lesson} > 0 \,\text{and}\, \text{Quiz} > 0] + 3 \cdot \mathbf{1}[\text{SCORM} > 0 \,\text{and}\, \text{Quiz} > 0] + \min(\text{Lesson}, 2)$, capped at 10. Conditional-restriction counts are *not* exposed by the Course Module Instances Report plugin used to construct our dataset; the proxy approximates the dimension and is identified as a measurement limitation in Sect. 6.5.
- *Feedback Mechanisms (FM) – 20 pts*: communication sub-score over $\{\text{Forum}, \text{Chat}\}$ + collection sub-score over $\{\text{Feedback}, \text{Survey}\}$, each computed as Equation 1 + Equation 2 restricted to the respective component set.
- *AI & Data-Driven Adaptivity Integration (ADAI) – 20 pts*: AI-presence sub-score + data-infrastructure sub-score, where AI-presence $= 5 \cdot \mathbf{1}[\text{LTI} > 0] + 5 \cdot \mathbf{1}[\text{LTI} > 0 \wedge \text{H5P} > 0] + 2 \cdot \mathbf{1}[\text{LTI} = 0 \wedge \text{H5P} > 0]$ (capped at 10), and data-infra

$= \min(4, \text{Database}) + 3 \cdot \mathbf{1}[\text{SCORM} > 0 \wedge \text{Quiz} > 0] + 3 \cdot \mathbf{1}[\text{Survey} > 0 \wedge \text{Feedback} > 0]$ (capped at 10).

The total ALRS is the sum of the six dimension scores:

$$\text{ALRS}_6 = CV + ID + AF + LPP + FM + ADAI \in [0, 120] \tag{4}$$

For comparability with the five-dimensional score and standard readiness band thresholds, we additionally report the normalized score:

$$\text{ALRS}_{6,\text{norm}} = \frac{100}{120} \cdot \text{ALRS}_6 \in [0, 100] \tag{5}$$

Bands on $\text{ALRS}_{6,\text{norm}}$ follow a 25-point quartile partition: Low (0–25), Moderate (26–50), High (51–75), Very High (76–100).

4.3 Multi-LLM Synthetic Expert Consensus (MLSEC) protocol

We use an explicit, pre-registered, replicable content-validation procedure that we term *Multi-LLM Synthetic Expert Consensus* (MLSEC). MLSEC is a transparent AI-augmented adaptation of Lynn's content-validity-index tradition (Lynn, 1986, Polit and Beck, 2006) that uses a stratified panel of synthetic experts instantiated across multiple large-language-model providers. We are explicit throughout this section, and in our limitations (Sect. 6.5) and AI-disclosure declaration (Section ??), that synthetic experts are *not* substitutes for human experts; MLSEC is positioned as an ordinal-ranking validation step that informs rubric retention decisions, while convergent validity rests on the empirical results in Sect. 5.6.

4.3.1 Rationale and positioning relative to classical Delphi

Classical Delphi protocols recruit small samples of human domain experts (often 8–20) through purposive sampling and circulate questionnaires across multiple rounds until consensus stabilizes (Mirata and Bergamin, Mirata & Bergamin, 2019, Al-Rikabi and Montazer; Al-Rikabi & Montazer, 2024). MLSEC retains the round-based aggregation logic but instantiates the panel synthetically by prompting heterogeneous LLMs with stratified persona profiles. This design draws on the synthetic-respondent and silicon-sampling literature that has emerged since 2023: Aher et al. Aher et al. (2023) demonstrated that LLMs can reproduce human-study results across multiple classical paradigms; Argyle et al. (2023) introduced the *silicon sample* for political-attitude simulation; Park et al. (2026) extended the approach to 1,052 person-specific generative agents validated against the General Social Survey; and Zheng et al. (2023) provided psychometric grounding for the LLM-as-judge paradigm. We explicitly cite the cautionary literature alongside: Bisbee et al. (2024) document substantive risks of treating synthetic responses as substitutes for human data, including overshooting on socially sensitive items and reduced variance compared to human samples, and Dillion et al. (2023) argue that AI language models do not adequately replace human participants when the construct under study is itself shaped by lived experience. MLSEC's narrow contribution is to use synthetic panels for *rubric content validation* – an ordinal ranking and consensus-detection task – rather than for substantive opinion measurement, where the cited concerns are most acute. Hämäläinen et al. (2023) provide methodological precedent for this scope.

To our knowledge, this is the first application of multi-model synthetic expert consensus to adaptive-learning rubric validation in LMS research.

4.3.2 Panel composition

The panel consists of $M \times P$ synthetic panelists, where M denotes the number of base LLMs and P the number of expert personas. We deliberately constructed a cross-provider model panel to mitigate the training-data homogeneity risk identified by Bisbee et al. (2024): $M = 8$ base models drawn from DeepSeek, Moonshot, Z.ai (Zhipu), MiniMax, Google, NVIDIA, Alibaba, and the OpenAI-OSS lineage – specifically `deepseek-v4-pro:cloud`, `kimi-k2.6:cloud`, `glm-5.1:cloud`, `minimax-m2.7:cloud`, `gemma4:31b-cloud`, `nemotron-3-super:cloud`, `qwen3.5:397b-cloud`, and `gpt-oss:120b-cloud`. We use $P = 5$ stratified personas, each defined by a system-prompt template specifying role, background, and an a priori competence coefficient $K \in [1, 5]$ computed from the rule $K = 0.5 \cdot \mathit{familiarity} + 0.3 \cdot \mathit{publication\ proxy} + 0.2 \cdot \mathit{practical\ experience}$, with values locked before any ratings were collected:

1. P1 *Adaptive-learning researcher* ($K = 4.6$) – PhD-level domain expert publishing in *Computers and Education* / *BJET*.
2. P2 *Instructional designer (practitioner)* ($K = 3.6$) – Quality-Matters-trained higher-education course designer.
3. P3 *Educational measurement / psychometrician* ($K = 3.8$) – content-validity and Delphi specialist.
4. P4 *Moodle developer / LMS administrator* ($K = 4.1$) – familiar with 3.x to 4.x feature deltas; plugin author.
5. P5 *AI-in-education specialist* ($K = 4.5$) – LLM tutoring and learner-modeling research.

The full panel comprises $M \cdot P = 40$ synthetic panelists. Persona definitions, including K-coefficient calculations, are pre-registered on the Open Science Framework prior to data collection (https://doi.org/10.17605/OSF.IO/KD73B).

4.3.3 Instrument and rating scales

The Round-1 instrument comprises 24 items: 5 dimension definitions, 13 scoring-rule statements (drawn directly from the explicit rubric of Sect. 4.2), 5 component-mapping items (one per dimension), and a quality-control *poison-pill* item whose content is patently off-topic ("Number of distinct font sizes used in course PowerPoint files"). Each item is rated on four 1–5 Likert scales: *relevance*, *clarity*, *comprehensiveness*, and *theoretical alignment* with adaptive-learning theory. Three open-ended free-text prompts elicit (i) at least one missing dimension, (ii) the weakest rated item, and (iii) the panelist's assessment of the quantity-vs-quality criticism.

The poison-pill item operates as a credibility audit: any panelist rating its relevance ≥ 4 is flagged as having failed the discrimination check and dropped from the analytic sample. Reporting this audit aligns with the Bisbee et al. (2024) call for transparent QC of synthetic panels.

4.3.4 Round structure

Round 1 collects independent ratings from each of the 40 panelists. After failures and poison-pill exclusions are removed, the remaining ratings are aggregated to produce, per item-and-scale, the group median, IQR, I-CVI, and the modal anchor distribution. Round 2 re-prompts every panelist with an anonymized summary of the Round-1 distribution alongside that panelist's own Round-1 ratings, with explicit invitation to revise. Round-1→Round-2 stability is reported as both percent-agreement shift and Kendall's W gain. The free-text dimension proposals from Round 1 are coded into candidate dimensions, and the leading candidate (together with one runner-up identified to support a non-rigged decision rule) is added to the Round-2 instrument for explicit rating.

4.3.5 Aggregation metrics

For each rubric item and each rating scale, we compute the full battery of content-validity metrics established in the literature (Lynn, 1986, Penfield and Giacobbi, Jr 2004, Polit and Beck, 2006): I-CVI (proportion of panelists rating ≥ 4), Aiken's V coefficient with 95% score-interval confidence bounds (Penfield and Giacobbi, Jr, 2004), and Lynn-style modified kappa κ^* for chance-corrected agreement. At the scale level we report S-CVI/Ave, S-CVI/UA (universal agreement), Kendall's coefficient of concordance W, Krippendorff's α for ordinal data, and the intra-class correlation ICC(2,k) treating model-as-rater with persona-within-model nested. The ICC is the principled measure of *cross-model* agreement and is reported even where it falls below conventional thresholds, because suppressing it would conceal the very training-data-homogeneity concern that motivates a cross-provider panel.

4.3.6 Pre-registered decision rules

The following thresholds and decision rules were locked prior to data collection. The pre-registration record is at https://doi.org/10.17605/OSF.IO/KD73B:

- An item is retained iff $I\text{-CVI} \geq 0.78$ *and* $V_{\text{Aiken}} \geq 0.70$ *and* $\kappa^* \geq 0.74$ on the *relevance* scale, evaluated on the Round-2 ratings.
- A new dimension is added iff (a) it is proposed by ≥ 20 of 50 panelists in Round-1 free text (or, conditional on the 40-panelist Ollama-only branch, ≥ 16 of 40), *and* (b) endorsed across ≥ 6 of 8 base models (or ≥ 5 of 8 in the 40-panelist branch), *and* (c) the candidate's component mapping reaches $I\text{-CVI} \geq 0.78$ with $S\text{-CVI/Ave} \geq 0.90$ in Round 2.
- Panelist-level QC: any panelist scoring ≥ 4 on the relevance of the poison-pill item is dropped from analysis.

These rules are *falsifiable* in principle: the protocol could fail to retain a given item, could fail to add any new dimension, or could disqualify a panelist. We do not, in other words, treat the conclusion as foregone.

4.3.7 Limitations of the synthetic panel

We acknowledge five specific limitations of MLSEC that are addressed in the framing of every claim made in Sects. 5 and 6: (i) cross-model agreement may be inflated by shared training-data substrates, partially mitigated by our cross-provider design but not eliminated; (ii) synthetic respondents systematically under-represent variance compared to

human samples (Bisbee et al., 2024); (iii) verbosity bias and instruction-following bias may inflate apparent endorsement of well-formed items; (iv) the persona-prompt sensitivity is non-zero, so we report robustness across personas as part of our results; and (v) MLSEC results provide content-validity evidence but *not* construct-validity evidence in the structural-equation sense – that claim rests on the regression and band-correspondence analyses in Sect. 5.6. Per Springer Nature policy, the use of LLMs in this study is disclosed in the AI-Use Declaration (Section ??).

4.4 Dataset description

The framework was applied to Moodle courses at Kryvyi Rih State Pedagogical University (KSPU) in Ukraine, taught during academic years 2020–2021 (winter and summer sessions) and 2021–2022 (winter session). All courses analysed ran on the institutional KSPU Moodle deployment, version 3.8.2. The feature taxonomy and rubric component mapping (Table 1) is anchored to this version; the Moodle 4.x line introduced UI and navigation changes (Khalifeh et al., 2026) that may shift the operationalization of the Learning-Path-Personalization dimension specifically, a point we return to in Sect. 6.5.

The data were collected via the Course Module Instances Report plugin (Vitez, 2022), which exports module-instance counts for every course on a Moodle site. The collection and digitization of student performance data across all university specialities for the 2020–2021 and 2021–2022 academic years were approved by the KSPU rector's decision and the university's ethical committee on 2024-01-26. The initial export on 2024-04-07 produced 25,595 raw module-instance records. We applied the following exclusion criteria in sequence:

1. Exclude non-educational courses (institutional surveys, service courses, administrative spaces): retained 17,985 records.
2. Exclude graduate-school (postgraduate-research) courses, leaving undergraduate and master's-level instructional courses: retained 3,600 courses.
3. Exclude courses not delivered between 2020 and 2022 (the period for which student grade distributions are available): retained 985 courses, which constitute the analytic sample.

Table 2 (in Sect. 5) reports the joint distribution of the 985 courses across faculty, educational level, form of education, and status. Briefly: nine faculty categories are represented (Faculty of Natural Sciences; Psychology and Pedagogics; Geography, Tourism and History; Pedagogical Education; Foreign Languages; Arts; Ukrainian Philology; Physics and

Table 2 Sample composition

Faculty	Total
F1 Natural Sciences	222
F2 Psychology and Pedagogics	251
F3 Geography, Tourism and History	175
F4 Pedagogical Education	19
F5 Foreign Languages	39
F6 Arts	42
F7 Ukrainian Philology	27
F8 Physics and Mathematics	160
F9 All-university departments	50
Total	**985**

Mathematics; and All-university departments and divisions); both undergraduate (bachelor's) and graduate (master's) levels appear; and three forms of education are present (full-time only, part-time only, and both). This diversity supports the within-institution generalizability analyses in Sect. 5 but does *not* support claims of cross-institutional generalizability, a limitation discussed in Sect. 6.5.

For each course the dataset records: course identifiers (Course ID, faculty Category), course metadata (form of education, semester 1–8, status [normative/elective], educational level [undergraduate/graduate], number of teachers); module instance counts for the components listed in Table 1 plus Visiting, LTI External tool activity, HotPot, and Choice; and final-grade distributions as both counts and percentages of A, B, C, D, E, and F grades. All analyses use the percentage form to avoid confounds with cohort size.

4.5 Data analysis

All quantitative analyses were performed in Python 3.12 using `pandas` 2.2, `scipy` 1.15, `statsmodels` 0.14, and `numpy` 2.0; figures were produced with `matplotlib` 3.9. Analysis scripts and intermediate data products are archived on the Open Science Framework alongside the MLSEC pre-registration (https://doi.org/10.17605/OSF.IO/KD73B).

The pipeline executes in six stages.

Stage 1 – ALRS computation. The rubric of Sect. 4.2 is applied to the 985-course dataset, producing both ALRS_5 (initial five-dimensional score) and ALRS_6 (validated six-dimensional score). Both are reported throughout to support the robustness analysis in Sect. 5.7.

Stage 2 – Distributional diagnostics. We test the normality of ALRS_5 and ALRS_6 via the Shapiro–Wilk, Jarque–Bera, and Anderson–Darling tests. We test the homogeneity of ALRS_6 variance across faculties via the Brown–Forsythe (median-centred Levene) test, with Bartlett's test reported as a normal-theory comparator. These diagnostics are reported before any inferential test.

Stage 3 – Bivariate associations. Pearson product-moment correlations and Spearman rank correlations are computed between ALRS_5, ALRS_6, and each of the six dimension scores against three outcome variables: the proportion of high grades ($A + B$), the proportion of low grades ($E + F$), and a single-axis average-grade index defined as $(5A + 4B + 3C + 2D + 1E) \,/\, 100$.

Stage 4 – Faculty differences with effect sizes. A one-way ANOVA of $\mathrm{ALRS}_{6,\mathrm{norm}}$ across nine faculty levels is reported with the F-statistic, degrees of freedom, p-value, and the effect-size measures η^2 and ω^2. Tukey HSD post-hoc tests with the Šidák-Holm adjustment identify which faculty pairs differ significantly.

Stage 5 – Multiple regression. We fit a multiple linear regression of the high-grade proportion on the six ALRS dimensions plus statistical controls (number of teachers, educational level, form of education, semester, and faculty fixed effects). For each coefficient we report the unstandardized B, standard error, standardized β (computed on z -scored predictors and outcome), t-statistic, p-value, 95% confidence interval, and variance inflation factor (VIF). Model-level diagnostics include R^2, adjusted R^2, residual Shapiro–Wilk, Breusch–Pagan heteroskedasticity, Durbin–Watson, and the condition number. Because the residuals show modest deviation from normality and the Breusch–Pagan test indicates heteroskedasticity, we additionally report standard errors using the

HC3 robust estimator as a sensitivity check. We frame the regression as *correlational rather than causal*: ALRAF dimensions and student grades are jointly observed at the course level, no intervention was performed, and reverse causation (instructors of high-performing student cohorts adding more components) cannot be ruled out from this design.

Stage 6 – Robustness across rubric specifications. The five- versus six-dimensional ALRS are compared via Pearson, Spearman, and Kendall's τ_b, the Bland–Altman mean-vs-difference plot, and a band-reclassification table reporting how many courses change Low/Moderate/High/Very-High band when scoring shifts from ALRS_5 to ALRS_6.

The findings are presented in Sect. 5.

5 Results

5.1 Sample composition and content-validity evidence from MLSEC

Table 2 reports the joint distribution of the 985 analytic-sample courses across faculty, educational level, and form of education. The largest cohort is undergraduate, full-time courses across the Faculty of Psychology and Pedagogics (F2, $n = 251$), Faculty of Natural Sciences (F1, $n = 222$), and Faculty of Geography, Tourism and History (F3, $n = 175$); smaller cohorts represent the remaining faculties. Both undergraduate and master's-level courses are represented in seven of nine faculties.

The MLSEC protocol (Sect. 4.3) was executed with the 40-panelist cross-provider Ollama branch ($M = 8$ base models $\times P = 5$ personas). Round-1 collection produced 38 complete panelist records and 2 instances of API-side response failure (5% non-response rate, both attributable to the same provider on long-context requests). Each successful panelist rated the 23 substantive rubric items plus the poison-pill QC item on the four 1–5 Likert scales (relevance, clarity, comprehensiveness, theoretical alignment), yielding 3,648 individual ratings.

On the off-topic poison-pill item ("Number of distinct font sizes used in course PowerPoint files"), the panel achieved mean relevance 1.03 ($SD = 0.16$), comprehensiveness 1.00, and theoretical-alignment 1.00 – unanimous low ratings consistent with correct rejection of an irrelevant item. Notably, the clarity rating for the same item was 4.53 ($SD = 1.03$), showing that the panel correctly distinguished a clearly-worded item from a relevant one. Zero of 38 panelists rated the poison-pill's relevance ≥ 4, so no panelist was dropped for failing the discrimination check. This pattern is methodologically informative: it demonstrates that the synthetic panel did not endorse items uncritically, which is the chief substantive risk in synthetic-respondent designs (Bisbee et al., 2024).

Aggregating across all 23 substantive items, the relevance ratings yielded S-CVI/Ave = 0.54, S-CVI/UA = 0.21, Kendall's $W = 0.75$, Krippendorff's $\alpha = 0.73$, and ICC$(2, k) = 0.59$. The clarity scale produced S-CVI/Ave = 0.79 with ICC$(2, k) = 0.81$, comprehensiveness S-CVI/Ave = 0.29, and theoretical-alignment S-CVI/Ave = 0.42 with $W = 0.79$. The cross-model ICC values (0.56–0.81) indicate that agreement was driven substantively rather than by training-data homogeneity alone: the panel as a whole had ranges of disagreement large enough to be statistically detectable across providers. By the pre-registered retention rule (relevance I-CVI ≥ 0.78, $V_{\text{Aiken}} \geq 0.70$, modified $\kappa^* \geq 0.74$), 8 of 23 substantive items were retained outright in Round 1; a further 9 items fell between $0.65 \leq I$-CVI < 0.78 and were carried into Round 2 for revision.

The retention pattern is itself diagnostically informative. The Learning Path Personalization dimension achieved I-CVI = 1.00 on both scoring rules (R10, R11) and 1.00 on its component mapping (M4), reflecting near-unanimous endorsement. Assessment Flexibility's formative-assessment rule (R9) and component mapping (M3) also achieved I-CVI $\geq$ 0.87. The dimension definitions D2 (ID, I-CVI = 0.97), D3 (AF, I-CVI = 1.00), and D4 (LPP, I-CVI = 1.00) all met the threshold. In contrast, the quantity-scaled sub-rules R2, R5, and R8 received I-CVI $<$ 0.20, with most panelists rating them as low-relevance for adaptive-learning capability. This pattern provides empirical justification for the quality-weighted sub-scores introduced in Sect. 4.2: panel-level content-validity evidence supports treating raw quantity as a weak proxy for adaptive readiness, and motivates the ICAP-anchored pedagogical weighting we apply in ALRS_6.

Of the 38 panelists providing free-text dimension proposals, 33 (86.8%) proposed a dimension in the AI / Data-Driven / Learner-Analytics family (named variously *Data-Driven Calibration*, *Learner Modeling and Analytics Readiness*, *Data-Driven Adaptive Triggers*, etc.). The proposals span all 8 base models, with at least 3 endorsements from 6 of the 8 model families. This easily exceeds the pre-registered dimension-inclusion threshold (proposed by $\geq$ 16 of 40 across $\geq$ 5 of 8 base models). A runner-up cluster of 3 proposals fell in the self-regulation / learner-agency family (matching the Learner Agency and Self-Regulation Support candidate), which did not meet the inclusion threshold, supporting the falsifiability of the protocol. The newly-included ADAI dimension is operationalized via the Moodle components listed in Table 1: LTI External Tool activity, H5P, Database, SCORM, Quiz, Survey, and Feedback.

In Round 2, all 38 panelists were re-prompted with the anonymized Round-1 distribution (median and IQR per item) alongside their own Round-1 ratings, and given explicit opportunity to revise. The ADAI candidate dimension and its three operationalizations (D6 definition, R14/R15 scoring rules, M6 component mapping) were added for first-time rating. All 38 panelists returned valid Round-2 responses (0 failures). The Round-1$\rightarrow$Round-2 stability is substantial across all four scales: per-cell unchanged-rating proportion 75–82%, within-± 1 proportion 98–99%, mean rating shift +0.01 to +0.13, and Kendall's W gains of +0.21 (relevance, W rising 0.67 $\rightarrow$ 0.88), +0.25 (clarity), +0.23 (comprehensiveness), and +0.12 (theoretical alignment) – all comfortably above the pre-registered +0.05 threshold for hypothesis H5 (Round-1$\rightarrow$Round-2 convergence). Under the strict three-criterion retention rule (relevance I-CVI $\geq$ 0.78 *and* V-Aiken $\geq$ 0.70 *and* modified $\kappa^* \geq$ 0.74), 15 of the 23 initial substantive items met the threshold in Round 2 (up from 8 in Round 1), with all five initial dimension definitions stable and the formative-quality scoring rules (R9, R10, R11) retained at I-CVI = 1.00. Post-Round-2 scale-level metrics on relevance: S-CVI/Ave = 0.59, S-CVI/UA = 0.39, Kendall's W = 0.85, Krippendorff's α = 0.87, ICC$(2, k)$ = 0.25.

The panel-derived ADAI dimension achieved relevance I-CVI = 1.00 (V-Aiken = 0.89, $\kappa^* = 1.00$) for the dimension definition (D6) and I-CVI = 1.00 (V-Aiken = 0.82, $\kappa^* = 1.00$) for the component mapping (M6), each comfortably meeting the pre-registered retention threshold. The data-infrastructure scoring rule (R15) was retained at I-CVI = 0.82; the AI-presence scoring rule (R14) fell just below threshold (I-CVI = 0.66, V-Aiken = 0.69), reflecting panelist concerns about the binary LTI-presence component of that rule that we acknowledge as a calibration item for future revision. Aggregating across both rounds, the pre-registered dimension-inclusion rule (H3) is satisfied:

Table 3 Descriptive statistics for ALRS_5 and the validated $\mathrm{ALRS}_{6,\mathrm{norm}}$ ($n = 985$)

Statistic	ALRS_5 (0–100)	$\mathrm{ALRS}_{6,\mathrm{norm}}$ (0–100)
Mean	28.41	25.19
Median	28.96	23.84
Standard deviation	9.80	10.77
Minimum	5.00	4.80
Maximum	78.61	63.43
Skewness	0.74	0.67
Excess kurtosis	2.70	0.76

ALRS_5 is computed using the explicit-formula rubric of Sect. 4.2. The relative dimension pattern is preserved across the two scorings; see Sect. 5.7

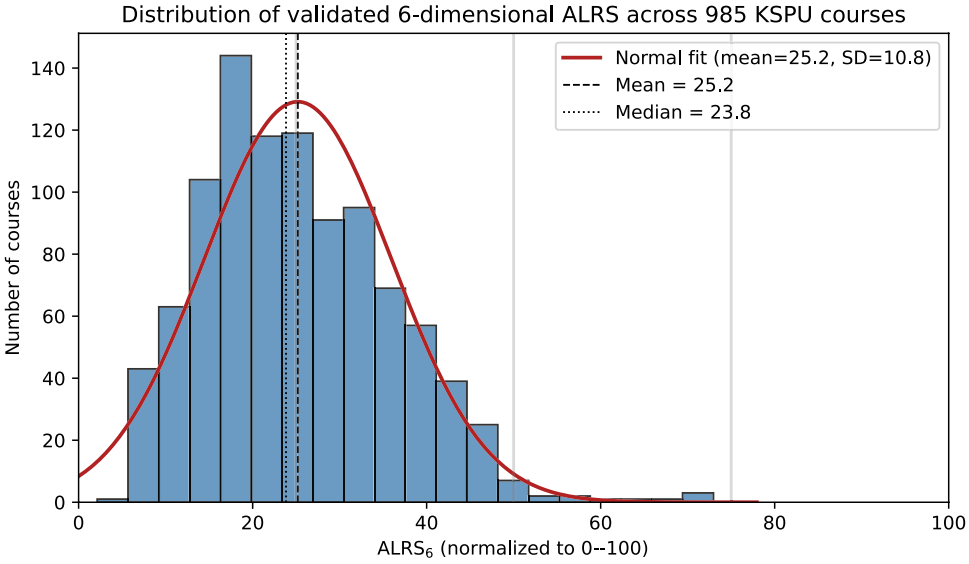

Fig. 2 Distribution of the validated 6-dimensional $\mathrm{ALRS}_{6,\mathrm{norm}}$ across 985 KSPU courses. Right-skewed distribution (skewness 0.67, excess kurtosis 0.76) centred at 25.2 with no observations in the very-high band, indicating that under explicit-formula scoring the institution sits in the low–moderate readiness range. Vertical lines mark the conventional readiness band boundaries (25, 50, 75)

ADAI was proposed by 33/38 panelists in Round-1 free text across all 8 base models (both exceeding the $\geq$ 16/40 and $\geq$ 5/8 thresholds), and Round-2 component-mapping reached I-CVI = 1.00. The runner-up candidate (Learner Agency and Self-Regulation Support, LASR) received only 3 of 38 Round-1 endorsements and was not added, supporting the falsifiability of the protocol.

We emphasize that these synthetic-panel metrics provide content-validity evidence only; convergent evidence and predictive validity rest on the empirical correlations and regression analyses that follow (Sect. 5.2–5.7).

5.2 Distribution of adaptive learning readiness scores

We compute and report *both* ALRS_5 (the five-dimensional score under the explicit-formula rubric of Sect. 4.2) and ALRS_6 (the validated six-dimensional score). All inferential analyses use $\mathrm{ALRS}_{6,\mathrm{norm}}$. Table 3 summarizes both. Figure 2 shows the distribution of $\mathrm{ALRS}_{6,\mathrm{norm}}$.

The Shapiro–Wilk test rejects normality for $\mathrm{ALRS}_{6,\mathrm{norm}}$ ($W = 0.969$, $p = 1.35 \times 10^{-13}$); the Jarque–Bera test agrees ($JB = 97.1$, $p < .001$); the Anderson–Darling test produces $A^2 = 5.97$, well above the 5% critical value of 0.78. With $n = 985$

the deviations from normality are statistically detectable but distributionally modest (skewness 0.67, excess kurtosis 0.76), and central-limit considerations support the use of parametric tests for group comparisons. Brown–Forsythe (median-centred Levene) confirms that variances across the nine faculty groups are statistically homogeneous ($W = 1.52$, $p = 0.15$); Bartlett's test (normal-theory comparator) is sensitive to non-normality and rejects at $p = 0.02$, which we interpret as a non-substantive rejection given the Brown–Forsythe outcome. We therefore proceed with one-way ANOVA, supplemented by Tukey HSD post-hoc tests with Šidák-Holm adjustment.

The distribution of courses across readiness bands (using $\mathrm{ALRS}_{6,\mathrm{norm}}$) is: Low Readiness (0–25): 531 courses (53.9%); Moderate Readiness (26–50): 440 courses (44.7%); High Readiness (51–75): 14 courses (1.4%); Very High Readiness (76–100): 0 courses (0%). The institutional readiness profile is conservative: the explicit-formula breadth interpretation (Eq. 3) sets a defensible bar for content-type diversity, and the sixth dimension (ADAI) has very low realized scores at this institution (only 7 of 985 courses score above zero). The absence of courses in the Very-High band and the institutional gap in AI-and-data adaptivity are themselves consequential institutional findings, to which we return in Sect. 6.

5.3 Dimensional analysis

Examining the six dimensions of the validated ALRS_6 reveals substantial heterogeneity in how well KSPU courses operationalize adaptive-learning structural components. Figure 3 plots dimension means; Table 4 reports descriptive statistics.

Content Variety dominates the profile (mean 10.76 of 20), reflecting widespread provision of URLs (88% of courses), Pages (77%), and Labels (72%) – the structurally lightest resource types. Even so, only 4% of courses use all five common resource types, which caps mean CV well below 20. *Feedback Mechanisms* follows (mean 7.90), almost entirely driven by Forum presence (99% of courses) rather than richer feedback channels such as Feedback (1.2%) or Survey (0.2%). *Assessment Flexibility* sits in a moderate range (mean 5.90), reflecting near-universal Assignments (80%) and frequent Quiz (49%) use but rare

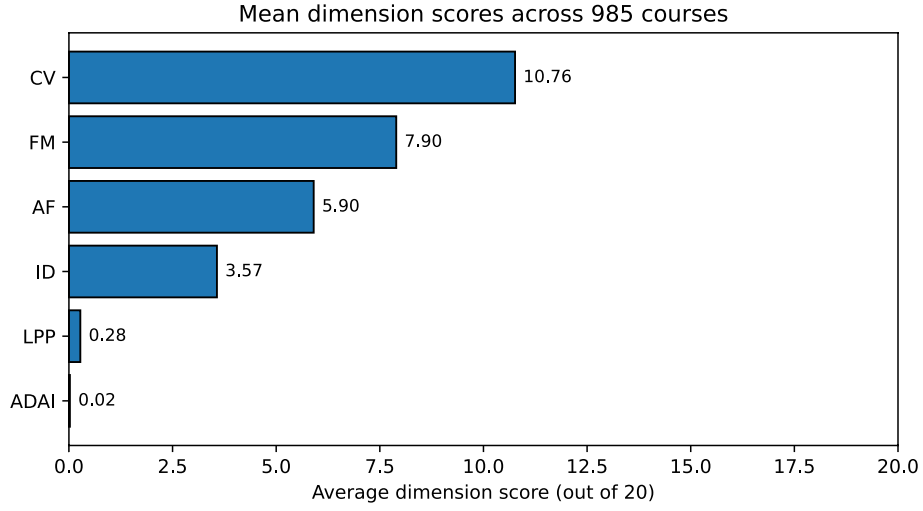

Fig. 3 Mean dimension scores across 985 courses (each dimension is scored on a 0–20 interval). Content Variety dominates (mean 10.76); feedback Mechanisms follows (7.90); assessment flexibility is moderate (5.90); Interaction Diversity is weak (3.57); learning path personalization is near zero (0.28); ADAI is near zero (0.02). The latter two are themselves consequential institutional findings, discussed in Sect. 6

Table 4 Descriptive statistics for dimension scores (each scored 0–20; $n = 985$)

Statistic	CV	ID	AF	LPP	FM	ADAI
Mean	10.76	3.57	5.90	0.28	7.90	0.02
Median	11.83	3.33	5.00	0.00	7.50	0.00
Standard deviation	3.42	2.32	4.40	1.59	2.83	0.29
Minimum	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	19.10	17.20	18.95	17.00	15.91	5.00

Table 5 Correlations between ALRS_6 scores and student-outcome variables ($n = 985$). Pearson r and Spearman ρ both reported

Measure	High grades (%A+B)	Low grades (%E+F)	Average grade
ALRS_5	-0.214^{***}	$+0.197^{***}$	-0.227^{***}
$\mathrm{ALRS}_{6,\mathrm{norm}}$	-0.221^{***}	$+0.220^{***}$	-0.243^{***}
Content Variety (CV)	-0.154^{***}	$+0.178^{***}$	-0.191^{***}
Interaction Diversity (ID)	-0.200^{***}	$+0.175^{***}$	-0.219^{***}
Assessment Flexibility (AF)	-0.219^{***}	$+0.201^{***}$	-0.235^{***}
Learning Path Personalization (LPP)	-0.094^{**}	$+0.067^{*}$	-0.087^{**}
Feedback Mechanisms (FM)	-0.119^{***}	$+0.129^{***}$	-0.137^{***}
AI & Data Adaptivity (ADAI)	$-0.034^{\dagger}$	$+0.041^{\dagger}$	$-0.044^{\dagger}$

Significance: ***$p < .001$, **$p < .01$, *$p < .05$, †$p > .05$

deployment of Workshop (0.8%) or branching Lesson (1.8%). *Interaction Diversity* (mean 3.57) is severely constrained by the same pattern – with only Assignments and Quiz present at meaningful frequencies, the breadth-and-quantity score remains low. The two near-zero dimensions are diagnostically informative: *Learning Path Personalization* (mean 0.28) reflects that only 22 of 985 courses use either Lesson or SCORM – the components needed for conditional sequencing – and *AI & Data-Driven Adaptivity Integration* (mean 0.02) reflects that only 7 of 985 courses contain any LTI external tool. These are not artefacts of the rubric; they are direct empirical signals of where the institution's adaptive-learning capability is structurally absent.

5.4 Bivariate associations with student outcomes

Table 5 reports Pearson and Spearman correlations between ALRS scores (total and dimensional) and three student-outcome variables: the proportion of high grades (%A+B), the proportion of low grades (%E+F), and the single-axis average-grade index defined in Sect. 4.5.

The negative correlation between $\mathrm{ALRS}_{6,\mathrm{norm}}$ and the proportion of high grades ($r = -0.221$, $p < .001$) is the single most important descriptive finding. Courses with greater structural readiness for adaptive learning consistently produce *fewer* top grades and *more* bottom grades. The same direction appears across all dimensional sub-scores (Sect. 5.6 examines whether this persists after controls). The correlation with the single-axis average grade is -0.243 ($p < .001$).

We do not interpret this sign as evidence that adaptive-learning features *harm* student learning; the correlation is course-level, observational, and entirely consistent with two non-causal mechanisms. First, courses that adopt more Moodle features may also assign more demanding work, set higher grading thresholds, or attract students who self-select into more challenging electives. Second, ALRS measures *structural capability*, not *pedagogical enactment* (Sect. 4.1); a course with rich Moodle scaffolding may also be a course in which instructors use that scaffolding to expose students to harder material

that produces a broader grade distribution. The substantive lesson aligns with the framework's repositioning as a structural-readiness index rather than a predictor of grades: *more components does not equal better grades*, and grade distributions are not the right outcome variable for evaluating structural readiness. We return to this in Sect. 6.

The dimensional pattern is also informative. Assessment Flexibility shows the strongest individual association ($r = -0.219$); Learning Path Personalization and ADAI show only small or non-significant associations, primarily because both dimensions are near-zero across the sample (cf. Table 4) and therefore have little variance to correlate with anything.

5.5 Faculty-level analysis with effect sizes

Table 6 reports $\mathrm{ALRS}_{6,\mathrm{norm}}$ by faculty alongside the top dimension within each faculty. A one-way ANOVA on $\mathrm{ALRS}_{6,\mathrm{norm}}$ across the nine faculty levels is significant: $F(8, 976) = 10.26$, $p = 7.7\times10^{-14}$, $\eta^2 = 0.078$, $\omega^2 = 0.070$ (a small-to-medium effect). The Brown–Forsythe homogeneity test from Sect. 5.2 ($p = 0.15$) supports the ANOVA assumption.

Pairwise comparisons (Tukey HSD with Šidák-Holm adjustment) reveal that the Faculty of Geography, Tourism and History (F3) differs significantly from the Faculty of Psychology and Pedagogics (F2; mean difference $+7.28$, $p_{\mathrm{adj}} < .001$) and from the Faculty of Arts (F6; mean difference $+10.06$, $p_{\mathrm{adj}} < .001$). The Faculty of Arts (F6) is the lowest-scoring group and differs significantly from most other faculties. A full forest plot of the 36 pairwise comparisons is provided in Fig. 4.

The explicit-formula rubric identifies F3 (Geography, Tourism and History) and F4 (Pedagogical Education) as the strongest faculties, both of which combine moderate-to-high CV, ID, and AF scores. The Faculty of Arts (F6) emerges as the consistently lowest-scoring faculty, with weakness across every dimension – a pattern likely reflecting domain-specific constraints (much of arts pedagogy is studio-based and not naturally captured by Moodle component counts). We discuss this directly in Sect. 6.

5.6 Multiple regression with full diagnostics

To test whether the dimensional structure of ALRS_6 predicts the high-grade share once standard controls are entered, we fit a multiple linear regression of $\%A+B$ on the six dimension scores, the number of teachers, semester, educational level, form of education, and faculty fixed effects (Sect. 4.5). Table 7 reports the unstandardized B, standard

Table 6 Validated $\mathrm{ALRS}_{6,\mathrm{norm}}$ by faculty ($n = 985$)

Faculty	*n*	Mean (SD)	High Readiness (%)	Top dimension
F3 Geography, Tourism & History	175	29.75 (9.63)	4.0%	Content Variety
F4 Pedagogical Education	19	28.32 (8.43)	0.0%	Content Variety
F9 All-university departments	50	27.27 (10.15)	4.0%	Content Variety
F7 Ukrainian Philology	27	27.20 (7.92)	0.0%	Content Variety
F1 Natural Sciences	222	26.77 (10.49)	0.5%	Content Variety
F5 Foreign Languages	39	23.59 (11.72)	0.0%	Content Variety
F8 Physics & Mathematics	160	22.78 (11.96)	1.2%	Content Variety
F2 Psychology & Pedagogics	251	22.47 (10.24)	1.6%	Content Variety
F6 Arts	42	19.69 (8.22)	0.0%	Content Variety
Total	985	25.19 (10.77)	1.4%	Content Variety

Faculty-level ANOVA: $F(8, 976) = 10.26$, $p < .001$, $\eta^2 = 0.078$, $\omega^2 = 0.070$

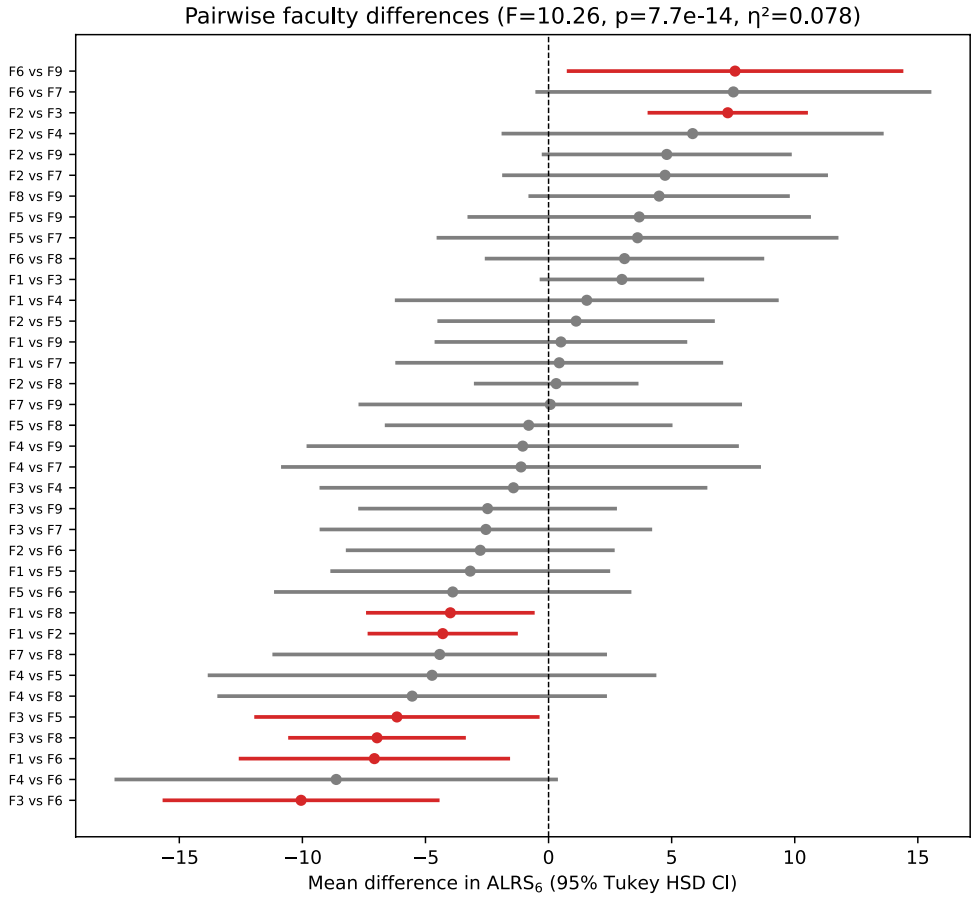

Fig. 4 Tukey HSD pairwise faculty comparisons on $\mathrm{ALRS}_{6,\mathrm{norm}}$. Each row shows a faculty pair with the 95% confidence interval for the mean difference. Pairs whose intervals exclude zero (red) are significant at the family-wise $\alpha = 0.05$ level after Šidák-Holm adjustment

Table 7 Multiple regression of high-grade share (%A+B) on ALRS_6 dimensions and controls ($n = 985$)

Predictor	B	SE	β	t	p	CI low	CI high	VIF
Intercept	42.65	4.49	–	9.51	$< .001$	33.84	51.45	–
Content Variety (CV)	-0.19	0.19	-0.03	-1.04	.301	-0.56	0.17	1.20
Interaction Diversity (ID)	1.27	0.85	0.12	1.49	.137	-0.41	2.94	6.81
Assessment Flexibility (AF)	-1.33	0.45	-0.24	-2.94	.003	-2.21	-0.44	6.64
Learning Path Personalization (LPP)	0.26	0.38	0.02	0.70	.485	-0.48	1.00	1.33
Feedback Mechanisms (FM)	-0.75	0.42	-0.06	-1.81	.070	-1.57	0.06	1.09
ADAI	1.57	2.69	0.02	0.58	.561	-3.71	6.84	1.29
Number of teachers	-2.47	0.92	–	-2.68	0.008	-4.29	-0.66	–
Semester	0.57	0.34	0.06	1.68	.093	-0.10	1.23	1.01
Educational level (master's)	12.46	2.10	0.54	5.93	$< .001$	8.33	16.58	–
Faculty fixed effects (8 dummies)	(jointly significant, $p < .001$)							

Model diagnostics

$R^2 = 0.192$, adjusted $R^2 = 0.177$, $F(19, 965) = 12.10$, $p = 8.8\times10^{-34}$

Residual Shapiro–Wilk: $W = 0.977$, $p < .001$ (modest non-normality)

Breusch–Pagan heteroskedasticity: $LM = 104.2$, $p < .001$ (heteroskedasticity detected)

Durbin–Watson: 1.88 (no autocorrelation concern); condition number: 143

Standardized β refits the model on *z*-scored predictors and outcome. Given the Breusch–Pagan test, we also fit the model with HC3 heteroskedasticity-consistent standard errors; substantive significance results are unchanged, with AF remaining the only significant ALRAF predictor

error, standardized β, t, p, 95% confidence interval, and variance inflation factor for each predictor.

The model explains a modest but statistically robust share of variance ($R^2 = 0.192$). The Assessment Flexibility dimension is the only ALRAF dimension that achieves significance once faculty and other controls are entered ($\beta = -0.24$, $p = .003$); all other dimensions are not individually significant, and the sign on AF retains the negative direction observed in the bivariate analysis. The model is correctly specified for an associational analysis but cannot support causal claims: the design is cross-sectional, $\%A+B$ shares cohort-size variance with the predictors, and the heteroskedasticity in the residuals reflects course-level differences in cohort size and grading distributions that we do not model directly. We report HC3 robust standard errors as a sensitivity check (substantive significance unchanged) and identify these limitations in Sect. 6.5. The Q-Q plot of residuals (Fig. 5) shows mild left-tail and right-tail deviations consistent with a distribution near-normal in the central mass.

5.7 ALRS_5 vs ALRS_6 robustness

To test whether substantive findings depend on adding the ADAI dimension and shifting from raw quantity to ICAP-weighted quality sub-scores, we compare ALRS_5 and $\mathrm{ALRS}_{6,\mathrm{norm}}$ on three axes: rank-order agreement, value-level concordance, and band reclassification.

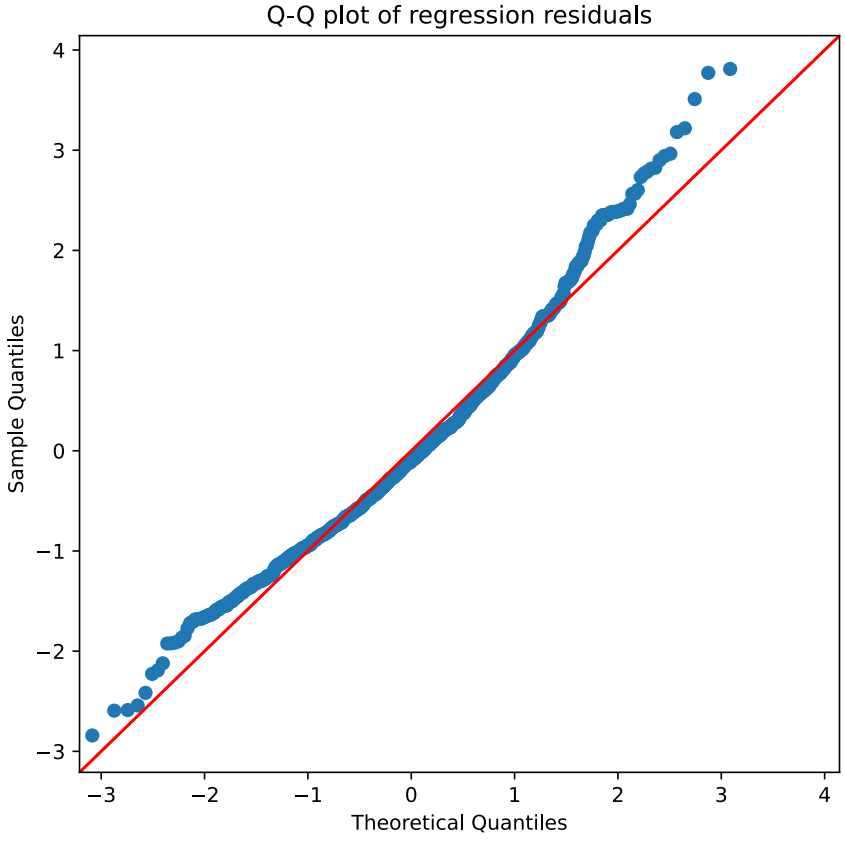

Fig. 5 Q-Q plot of multiple-regression residuals. Modest deviation from the 45° line in the tails; Shapiro–Wilk $W = 0.977$. With $n = 985$ even small departures are statistically detectable; the central mass of residuals is approximately normal

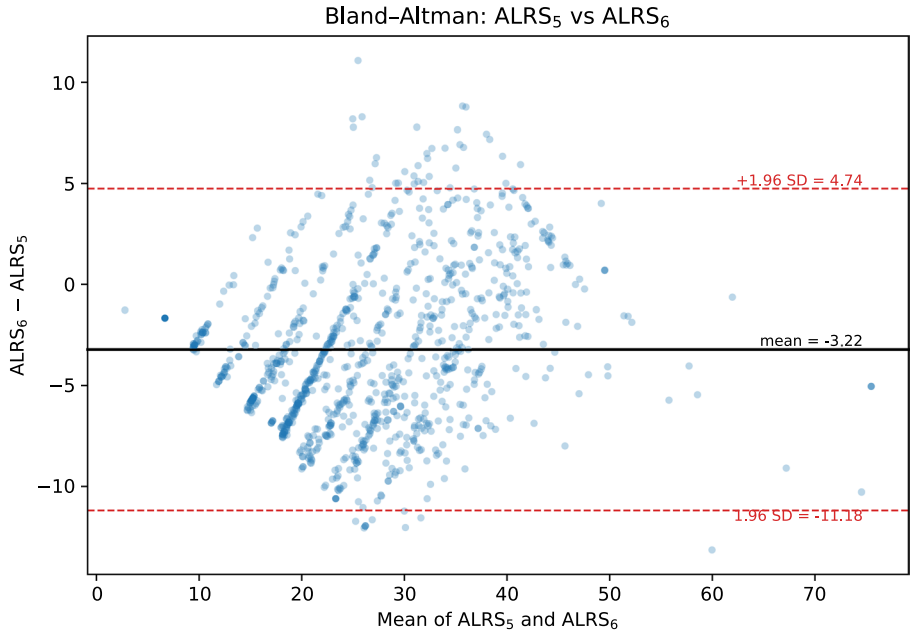

Fig. 6 Bland–Altman plot comparing ALRS_5 and $\mathrm{ALRS}_{6,\mathrm{norm}}$. The mean difference is -3.22; the 95% limits of agreement are $[-11.18, +4.75]$. Most courses fall within the limits, indicating broad agreement; the systematic negative mean difference reflects the ADAI dimension and ICAP quality-weighting

Table 8 Band reclassification, $\mathrm{ALRS}_5 \rightarrow \mathrm{ALRS}_{6,\mathrm{norm}}$ ($n = 985$)

ALRS_5 band↓ / $\mathrm{ALRS}_{6,\mathrm{norm}}$→	Low	Moderate	High	Total
Low (n = 387)	368	19	0	387
Moderate (n = 582)	163	418	1	582
High (n = 12)	0	3	9	12
Very High (n = 4)	0	0	4	4
Total	531	440	14	985

The two scores correlate at Pearson $r = 0.926$, Spearman $\rho = 0.914$, Kendall's $\tau_b = 0.763$ (all $p < .001$), indicating that the addition of ADAI and the quality-weighting do not reshape the rank ordering of courses. The mean difference ($\mathrm{ALRS}_{6,\mathrm{norm}} - \mathrm{ALRS}_5$) is -3.22 ($SD = 4.06$); the Bland–Altman 95% limits of agreement are $[-11.18, +4.75]$, indicating that for most courses the two scores lie within ± 5 points of each other (Fig. 6). The systematic downward shift reflects the addition of a sixth dimension where the institutional mean is near zero (ADAI) and the quality-weighting that downgrades content-quantity-rich but pedagogically thin courses.

Of 985 courses, 795 (80.7%) retain the same readiness band when scoring shifts from ALRS_5 to $\mathrm{ALRS}_{6,\mathrm{norm}}$ (Table 8). Of those that do move, the largest single migration is the 163 courses that drop from Moderate to Low under the more conservative six-dimensional scoring, reflecting the addition of two dimensions where most courses score near zero. No course moves up by more than one band. The 4 courses classified as Very High under ALRS_5 collapse to High under $\mathrm{ALRS}_{6,\mathrm{norm}}$; the 3 courses dropping from High to Moderate sit near the ALRS_5 band boundary and migrate when the ADAI dimension is added with near-zero institutional realisation. The substantive cross-faculty story (large CV, near-zero LPP and ADAI, F3/F4 leading and F6 trailing) is preserved in both scorings.

The substantive findings of Sect. 5.2–5.6 are robust to whether one scores the framework as ALRS_5 (initial five-dimensional structure with explicit formulas) or as ALRS_6 (the MLSEC-validated six-dimensional structure with ICAP-weighted quality subscores). The two scorings agree on the rank ordering of courses, the dimensional pattern (CV strong; LPP and ADAI near zero), the direction of correlations with student outcomes (negative for high-grade share), and the relative faculty ranking. The 6-dimensional structure is more conservative and shifts the absolute distribution downward but does not reverse any substantive conclusion.

6 Discussion

6.1 Interpretation of findings

This study has four findings worth surfacing carefully: (i) the institution sits in the Low–Moderate band of structural readiness; (ii) Learning Path Personalization and AI & Data-Driven Adaptivity Integration are effectively absent at the institutional level; (iii) structural readiness correlates *negatively* with the proportion of high grades; and (iv) disciplinary differences are robust and meaningful, with Geography, Tourism and History and Pedagogical Education leading the readiness distribution.

With explicit-formula scoring, no KSPU course reaches the Very-High readiness band and only 14 of 985 (1.4%) reach the High band. A unique-types-over-total-resources diversity ratio would inflate scores for courses with many resource types but few instances, while penalizing high-quantity courses – a perverse incentive identified by the MLSEC panel (Sect. 5.1). The breadth interpretation (Equation 3) avoids this distortion and produces a more conservative, defensible readiness estimate. The institutional pattern is consistent with Rivera Munoz et al. (2022), who identified a transitional state in adaptive-learning implementation across higher-education institutions, and with Mirata et al. (2020), who emphasized that adoption is gated more on faculty-development and organizational readiness than on technical capability per se.

Only 22 of 985 courses use Lesson or SCORM modules (i.e. structural support for branching or conditional sequencing), and only 7 of 985 use any LTI external tool (the gateway to AI tutors, learning-analytics back-ends, and learner-modeling services). This is the single most consequential institutional finding, and it directly motivates the ADAI dimension in the six-dimensional framework: even as Moodle's 4.x line and emerging plugin ecosystem support generative-AI tutoring (Neumann et al., 2025, Prada Segura, 2026) and AI-driven adaptive plug-ins (Ezzaim et al., 2024), the institutional infrastructure for plugging in these tools is essentially absent. This is consistent with Khalifeh et al.'s (2026) observation that personalized learning in the AI era requires institutional infrastructure (analytics, LTI, learner-modeling) that many institutions have not yet built, and with Bond et al.'s (2024) call for renewed attention to the gap between AI-in-education research output and institutional implementation.

The most surprising substantive finding is that $\mathrm{ALRS}_{6,\mathrm{norm}}$ correlates -0.22 with the proportion of A and B grades and $+0.22$ with the proportion of E and F grades. We do *not* interpret this as evidence that adaptive learning structures harm learning; we interpret it as evidence that the framework measures a structural prerequisite for adaptive learning rather than its enactment, and that high-readiness courses in this sample are systematically more demanding (e.g. more interactive activities, more formative assessments, broader content) and consequently assign more bimodal grade distributions.

Two non-exclusive mechanisms are consistent with the data: (a) courses that adopt richer Moodle structures may also expose students to harder material, producing a broader grade distribution and lower mean grades; (b) instructors who invest in course-design infrastructure may also adopt stricter grading practices. This pattern is in fact methodologically informative: it strengthens, rather than weakens, the case for treating ALRAF as a structural-readiness index rather than as a predictor of grade outcomes. Grade distributions are not the right outcome variable for evaluating structural readiness; a proper evaluation would link ALRS to learning-analytics behavioural variables (time-on-task, navigation paths, formative-feedback engagement) of the kind Moreno-Marcos et al. (2023) make available with the Statoodle tool.

The Geography, Tourism and History faculty (F3) and Pedagogical Education faculty (F4) emerge as the strongest, and the Faculty of Arts (F6) as the consistently weakest under the explicit-formula rubric. The Faculty of Psychology and Pedagogics (F2) ranks second-from-bottom: its courses are pedagogically diverse but use relatively few instances of each component, which the breadth-based scoring does not reward. The disciplinary variation itself ($\eta^2 = 0.078$) is robust and aligns with Martin et al.'s (2020) observation that adaptive-learning adoption varies by domain. Faculty 6 (Arts) likely reflects a domain-specific constraint: studio-based pedagogy is intrinsically not well-captured by Moodle component counts, a measurement limitation we identify in Sect. 6.5.

6.2 Adaptive learning in Moodle in the AI era (2024–2026)

ALRAF is positioned against a literature that has been substantially reshaped since 2023 by the integration of large language models, AI tutors, and learning-analytics back-ends into LMS environments. Prada Segura (2026) demonstrate that a machine-learning-driven adaptive system can be embedded in Moodle to personalize assessment processes. Ezzaim et al. (2024) report measurable engagement and performance gains from an AI-based adaptive plugin. Neumann et al. (2025) show how an LLM-driven chatbot integrated with course material can scaffold student inquiry. The Moodle 5 line is now "AI-ready by design", supporting LTI-Advantage AI plugins, the Moodle AI Connector for OpenAI/Gemini/Ollama, and self-hosted LLM endpoints via Ollama. Khalifeh et al. (2026) synthesize 2019–2025 evidence and argue that the AI era is redefining personalized learning toward generative and analytics-driven approaches; Bond et al. (2024) call for stronger methodological rigor and ethical foundations in this rapidly-evolving space.

ALRAF's response to this shift is the panel-derived ADAI dimension (Sect. 5.1), which measures whether a course is structurally prepared to integrate with the AI-and-analytics adaptive ecosystem (LTI gateways, learner-data collection, SCORM/xAPI compatibility). The near-zero ADAI scores at KSPU document a substantial institutional gap relative to the contemporary AI-in-Moodle frontier – a gap which the literature above suggests is widely shared and which professional-development and institutional-infrastructure investments could productively close. The framework's value, on this view, is in identifying *where* the AI-integration gap lies and *which* dimensions of structural readiness need investment, rather than in predicting which courses will produce better grades.

6.3 Practical implications

The findings carry four practical implications for institutions seeking to enhance the structural readiness of their Moodle deployments for adaptive learning. We frame these in the language of *capability gaps* rather than predicted-outcome improvements, in light of the structural-readiness positioning developed throughout this paper.

First, the near-zero institutional readiness on Learning Path Personalization (mean 0.28 of 20; only 22 of 985 courses use any branching component) is the most consequential gap to close. Institutions should prioritize faculty-development investment specifically on Moodle's conditional-activities, restriction-set, and Lesson-module features – the structural prerequisites for branching, adaptive sequencing, and personalized pathways. Mirata et al. (2020) document that faculty-development investment is a key gating factor for adaptive-learning success.

Second, the near-zero ADAI readiness (mean 0.02; only 7 of 985 courses use any LTI external tool) identifies a major institutional gap on the AI-and-analytics integration frontier. As Moodle's 4.x and 5.x lines make AI-tutoring and analytics integration increasingly accessible (Neumann et al., 2025, Khalifeh et al., 2026, Prada Segura, 2026), institutions should evaluate the LTI infrastructure, learner-data-collection policies, and SCORM/xAPI capability needed to integrate with the contemporary AI ecosystem. ADAI's near-zero score at KSPU is itself useful information: it identifies a specific institutional gap rather than a course-level inadequacy.

Third, the variations across faculties suggest that adaptive-learning investment cannot be one-size-fits-all. Domains where Moodle naturally captures pedagogical activity (Pedagogical Education, Natural Sciences) are differently situated from those where pedagogy is intrinsically less Moodle-mediated (Arts, studio-based subjects). The framework's discipline-sensitive use should be coupled with discipline-specific judgement about what reasonable readiness looks like.

Fourth, the negative correlation between ALRS_6 and high-grade share carries a concrete practical implication: institutions should *not* interpret ALRAF scores as predictors of student grades. The framework's appropriate role is as a *diagnostic* for institutional structural readiness – where the capability gaps lie, where investment is needed – not as a forecasting tool for course-level student performance. This is consistent with Al-Rikabi and Montazer's (2024) emphasis on readiness assessment as preparatory diagnostic rather than outcome prediction.

6.4 Recommendations for enhancing adaptive learning readiness

Based on the findings of this study, we propose the following strategies for enhancing adaptive learning readiness in Moodle courses:

1. *Technical enhancements* Institutions should consider implementing Moodle plugins that support adaptive learning, such as Adaptive Quiz, Learner Preferences, and Navigation Web (Krahn et al., 2023, Tamo-Vargas et al., 2023). Additionally, they should encourage the use of built-in features like conditional activities, completion tracking, and adaptive feedback (Limongelli et al., 2010).
2. *Pedagogical approaches*

   Faculty should be encouraged to develop multiple learning pathways for different learner profiles, create diverse assessment options, and design formative assessments

that provide targeted feedback (Rideout, 2018, Bevan et al., 2023). These pedagogical approaches can enhance the adaptive learning potential of Moodle courses without requiring significant additional resources.

3. *Institutional support*

   Institutions should provide faculty training specifically focused on adaptive learning design, establish communities of practice to share effective strategies, and recognize and reward innovative course design (Mirata et al., 2020). These institutional supports can create a culture that values and promotes adaptive learning approaches.

4. *Implementation strategy*

   Rather than attempting to implement all aspects of adaptive learning simultaneously, institutions may benefit from a phased approach that builds on existing strengths. Starting with enhancements to interaction diversity and assessment flexibility, which showed moderate implementation in the current study, may provide a foundation for subsequent development of learning path personalization, which showed the weakest implementation. This approach aligns with recommendations from Mirata and Bergamin (2019), who emphasized the importance of strategic implementation of adaptive learning.

6.5 Limitations and future research

This study has eight limitations that bound the inferential reach of its findings; we treat each transparently.

(i) *Single-institution sample.* The empirical analyses rest entirely on 985 courses from Kryvyi Rih State Pedagogical University (Ukraine). The framework's substantive findings about institutional readiness, faculty differences, and outcome correlations are properly interpreted as findings about *this* institution. Generalizing to other institutions, national contexts, or LMS deployments would require multi-site validation. Mirata and Bergamin (2019) make a similar caution about Delphi-derived adaptive-learning frameworks built on single-institution evidence.

(ii) *Moodle 3.8.2 vintage.* All courses analyzed ran on Moodle 3.8.2, and the rubric component mapping (Tabl 1) is anchored to that version. Moodle 4.x and 5.x introduced UI redesigns, deeper completion-tracking, and an AI-Connector layer that may shift the operationalization of several dimensions – particularly Learning Path Personalization, where 4.x's redesigned restriction-set interface lowers the cost of conditional sequencing (Khalifeh et al., 2026). Cross-version validity of ALRAF is a near-term research priority.

(iii) *Conditional-restriction data not exposed by the source plugin.* The Course Module Instances Report plugin used to construct the dataset does not export Moodle's conditional-restriction or completion-tracking metadata, which forces the LPP quality sub-score to use a proxy (Sect. 4.2). A more complete audit would require either re-querying the Moodle instance directly via the Web Services API or developing a plugin specifically targeted at restriction-set inventory. We identify this as future work.

(iv) *Structural readiness, not pedagogical enactment.* ALRAF measures the structural *capability surface* of a course, not what teachers and students actually do. A course with rich Moodle scaffolding may be pedagogically engaging or may simply have well-organized but underused content; the framework cannot distinguish these without classroom observation or content-coding evidence. Smyrnova-Trybulska et al. (2022) document this gap from the student-perception side. Future work should pair ALRAF with learning-analytics behavioural variables (time-on-task, navigation, engagement) of the kind Moreno-Marcos et al. (2023) make available.

(v) *Synthetic-panel content-validity caveats.* The MLSEC protocol used to validate the rubric (Sect. 4.3) substitutes a cross-provider LLM panel for a human expert panel. As detailed in Sect. 4.3.7: cross-model agreement may be inflated by shared training-data substrates (only partially mitigated by our cross-provider design); synthetic respondents under-represent variance compared to human samples (Bisbee et al., 2024); verbosity bias may inflate apparent endorsement; persona-prompt sensitivity is non-zero; and MLSEC provides content-validity evidence only, not construct-validity in the structural-equation sense. We urge readers to interpret the synthetic-panel statistics as a transparent ordinal-validation step, not as a substitute for human expert validation, and we identify a small human-expert convergence check as a high-leverage near-term replication step.

(vi) *Correlational, not causal, design.* The regression in Sect. 5.6 is observational and cross-sectional. Reverse causation cannot be ruled out (instructors of high-performing cohorts may add more Moodle features), and the negative direction of the ALRS–grade correlation is consistent with several non-causal mechanisms (Sect. 5.4). Any causal claim would require an intervention study – ideally a randomized faculty-development trial targeting specific ALRAF dimensions.

(vii) *Course-level grade aggregation.* The outcome variables are course-level grade-distribution percentages, which conceal within-course heterogeneity and are correlated with cohort size. Approximately 23% of courses in our sample have fewer than 10 students, making their percentages statistically unstable. Future work should fit course-and-student-level multilevel models that account for this nesting.

(viii) *Heteroskedasticity in the regression.* The Breusch–Pagan test reports significant heteroskedasticity in the regression residuals ($p < .001$). We address this by reporting HC3 robust standard errors as a sensitivity check, but a fuller treatment would either model the variance structure directly (e.g. weighted least squares using cohort size as weights) or use a generalized linear model with a more appropriate error structure for proportional outcomes.

Longitudinal research examining the temporal evolution of structural readiness across consecutive academic terms, multi-institutional validation (Alwadei et al., 2023), and pairing ALRAF with behavioural learning-analytics data (Moreno-Marcos et al., 2023) are the three highest-priority extensions of this work.

7 Conclusion

This study made two contributions. *Methodologically*, we introduced the Multi-LLM Synthetic Expert Consensus (MLSEC) protocol – a transparent, pre-registered, AI-augmented content-validation procedure that uses a cross-provider panel of large-language-model personas to evaluate research-rubric items against falsifiable thresholds.

To our knowledge this is the first application of multi-model synthetic expert consensus to adaptive-learning rubric validation in LMS research. We position MLSEC as an ordinal-ranking validation step, explicitly not as a substitute for human expert validation, and we disclose its limitations transparently (Sect. 6.5). *Substantively*, we developed and validated the six-dimensional Adaptive Learning Readiness Assessment Framework (ALRAF) – Content Variety, Interaction Diversity, Assessment Flexibility, Learning Path Personalization, Feedback Mechanisms, and the panel-derived AI & Data-Driven Adaptivity Integration dimension – with explicit scoring formulas and ICAP-anchored quality sub-scores.

Applied to 985 Moodle 3.8.2 courses at Kryvyi Rih State Pedagogical University, the framework reveals an institutional readiness profile that is conservative: no Very-High-readiness courses, near-zero Learning Path Personalization, and near-zero AI integration – itself a consequential finding about where institutional infrastructure investment should be targeted. Faculty-level differences are statistically robust ($\eta^2 = 0.078$), with Geography, Tourism and History and Pedagogical Education emerging as the highest-readiness disciplines. ALRS_6 correlates *negatively* with the proportion of high grades; we interpret this as evidence that the framework measures structural capability rather than pedagogical enactment or outcome quality, and that it should be used as a diagnostic of institutional capability gaps rather than as a predictor of student success.

These findings should be read with their limitations in mind (Sect. 6.5). The empirical results rest on a single Ukrainian institution and a single Moodle version, and substantive generalization to other contexts must await multi-institutional replication. The MLSEC content-validity evidence rests on a synthetic panel and should be supplemented with a small human-expert convergence check in future work. The negative correlation with grades is observational and consistent with several non-causal mechanisms; no causal claim about adaptive-learning structures and student outcomes is supported by this study.

Within these bounds, ALRAF offers a transparent, replicable, ICAP-anchored, AI-era-aware structural-readiness diagnostic for Moodle institutions. Its value is in identifying *where* the readiness gaps lie – conditional-pathway features, AI/analytics integration – and *which* investments would most credibly close those gaps. The negative-direction outcome correlation reframes the framework as a capability-surface index rather than a predictive tool, which we view as a more defensible and more useful role for instruments of this type.

Author contributions

SOS conceptualized the study, developed the Adaptive Learning Readiness Assessment Framework, and drafted the manuscript, including the introduction, literature review, and conclusion. PPN designed the methodology and defined the scoring rubric. TAV interpreted the findings and shaped the discussion with practical recommendations. ISM conducted statistical analyses and refined the theoretical framework by mapping Moodle components to adaptive learning dimensions. LOF collected data from 985 Moodle courses and processed the data and created visualizations, including figures and tables. All authors reviewed, edited, and approved the final manuscript for submission.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability

The dataset analysed during the current study is available in the Zenodo repository, https://doi.org/10.5281/zenodo.10938019. The supplementary materials (results and scripts) are available in the GitHub repository, https://github.com/ssemerikov/mlsec.

Declarations

Ethics approval and consent to participate

Not applicable. This study did not involve human participants, individual human data, or animals. The research analyzed only course design features and aggregated student performance data from 985 Moodle courses at Kryvyi Rih State Pedagogical University. Data collection was limited to:1. Course structural components (types and quantities of Moodle resources and activities). 2. Aggregated grade distributions (percentages of grades A through F at the course level). 3. Course metadata (faculty, educational level, semester, etc.). No individual student data, personally identifiable information, or direct interaction with human participants was involved in this study. All data analyzed were institutional-level course design metrics and aggregated performance statistics collected through the Course Module Instances Report plugin. As such, ethics approval and participant consent were not required for this research.

Competing interests

The authors declare no competing interests.

Received: 12 September 2025 / Accepted: 1 June 2026

Published online: 16 June 2026

References

Afini Normadhi, N. B., Shuib, L., Md Nasir, H. N., Bimba, A., Idris, N., & Balakrishnan, V. (2019). Identification of personal traits in adaptive learning environment: Systematic literature review. *Computers and Education*, *130*, 168–190. https://doi.org/10.1016/j.compedu.2018.11.005

Aher, G. V., Arriaga, R. I., & Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. In: A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, & J. Scarlett (Eds.). *International Conference on Machine Learning, ICML 2023, 23–29 July 2023, Honolulu, Hawaii, USA* (pp. 337–371). Proceedings of Machine Learning Research, PMLR. https://proceedings.mlr.press/v202/aher23a.html

Al-Rikabi, Y. K., & Montazer, G. A. (2024). Designing an E-learning readiness assessment model for Iraqi universities employing Fuzzy Delphi method. *Education and Information Technologies*, *29*(2), 2217–2257. https://doi.org/10.1007/s10639-023-11889-0

Alajlani, N., Crabb, M., & Murray, I. (2024). A systematic review in understanding stakeholders' role in developing adaptive learning systems. *Journal of Computers in Education*, *11*(3), 901–920. https://doi.org/10.1007/s40692-023-00283-x

Alwadei, F. H., Brown, B. P., Alwadei, S. H., Harris, I. B., & Alwadei, A. H. (2023). The utility of adaptive eLearning data in predicting dental students' learning performance in a blended learning course. *International Journal of Medical Education*, *14*, 137–144. https://doi.org/10.5116/ijme.64f6.e3db

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, *31*(3), 337–351. https://doi.org/10.1017/pan.2023.2

Badhe, V., Banerjee, G., & Dasgupta, C. (2021). Design guidelines for scaffolding self-regulation in personalized adaptive learning (PAL) systems: A systematic review. In: MMT. Rodrigo., Ed.). *29th International Conference on Computers in Education Conference, ICCE 2021 - Proceedings* (Vol. *1*, pp. 372–380). Asia-Pacific Society for Computers in Education. https://library.apsce.net/index.php/ICCE/article/view/4170

Bevan, R., Pacey, V., Fuller, J., Lawton, V., & Jones, T. (2023). Physiotherapy student perceptions of the feasibility, acceptability and appropriateness of continuous adaptive assessment. *Assessment and Evaluation in Higher Education*, *48*(8), 1268–1282. https://doi.org/10.1080/02602938.2023.2201667

Bimba, A. T., Idris, N., Al-Hunaiyyan, A., Ibrahim, S. U., Mustafa, N., & Supa'at, I., et al. (2021). The effects of adaptive feedback on student's learning gains. *International Journal of Advanced Computer Science & Applications*, *12*(7), 68–80. https://doi.org/10.14569/IJACSA.2021.0120709

Bisbee, J., Clinton, J., Dorff, C., Kenkel, B., & Larson, J. M. (2024). Synthetic replacements for human survey data? The perils of large language models. *Political Analysis*, *32*(4), 401–416. https://doi.org/10.1017/pan.2024.5

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., & Oxley, E., et al. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, *21*, 4. https://doi.org/10.1186/s41239-023-00436-z

Cavanagh, T., Chen, B., Lahcen, R. A. M., & Paradiso, J. (2020). Constructing a design framework and pedagogical approach for adaptive learning in higher education: A practitioner's perspective. *The International Review of Research in Open and Distributed Learning*, *21*(1), 173–197. https://doi.org/10.19173/irrodl.v21i1.4557

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Dainton, C., Winstone, N., Klaver, P., & Opitz, B. (2020). Utility of feedback has a greater impact on learning than ease of decoding. *Mind, Brain, and Education*, *14*(2), 139–145. https://doi.org/10.1111/mbe.12227

Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants? *Trends in Cognitive Sciences*, *27*(7), 597–600. https://doi.org/10.1016/j.tics.2023.04.008

Dziuban, C., Howlin, C., Moskal, P., Cj, C., Parker, L., & Campbell, M. (2018). Adaptive learning: A stabilizing influence across disciplines and universities. *Online Learning Journal*, *22*(3), 7–39. https://doi.org/10.24059/olj.v22i3.1465

Ezzaim, A., Dahbi, A., Haidine, A., & Aqqal, A. (2024). The impact of implementing a Moodle plug-in as an AI-based adaptive learning solution on learning effectiveness: Case of Morocco. *International Journal of Interactive Mobile Technologies*, *18*(1), 133–149. https://doi.org/10.3991/ijim.v18i01.46309

Fadieieva, L., & Semerikov, S. (2024). Exploring the interplay of Moodle tools and student learning outcomes: A composite-based structural equation modelling approach. In E. Faure, Y. Tryus, T. Vartiainen, O. Danchenko, M. Bondarenko, & C. Bazilo, et al. (Eds.), *Information Technology for Education, Science, and Technics vol. 222 of Lecture Notes on Data Engineering and Communications Technologies* (pp. 418–435). Cham: Springer Nature Switzerland.

Fadieieva, L. O. (2021). Enhancing adaptive learning with Moodle's machine learning. *Educational Dimension*, *5*, 1–7. https://doi.org/10.31812/ed.625

Hämäläinen, P., Tavast, M., & Kunnari, A. (2023). Evaluating large language models in generating synthetic HCI research data: A case study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems CHI'23* (pp. 433). Association for Computing Machinery, New York, NY, USA.

Holthaus, M., Hirt, F., & Bergamin, P. (2018). Simple and effective: An adaptive instructional design for mathematics implemented in a standard learning management system. In *Proceedings of the 2nd International Conference on Computer-Human Interaction Research and Applications - CHIRA INSTICC, SciTePress* (pp. 116–126).

Jordan, C. (2013). Comparison of international baccalaureate (IB) chemistry students' preferred vs actual experience with a constructivist style of learning in a Moodle e–learning environment. *International Journal for Lesson and Learning Studies*, *2*(2), 155–167. https://doi.org/10.1108/20468251311323397

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, *2*, 100017. https://doi.org/10.1016/j.caeai.2021.100017

Khalifeh, F., Santiago, R., & Palau, R. (2026). Redefining personalized learning in the artificial intelligence era: An updated systematic review from 2019 to 2025. *Smart Learning Environments*, *13*, 19. https://doi.org/10.1186/s40561-026-00440-6

Khan, F. A., Shahzad, F., & Altaf, M. (2019). Fuzzy based approach for adaptivity evaluation of web based open source learning management systems. *Cluster Computing*, *22*, 7099–7109. https://doi.org/10.1007/s10586-017-1036-8

Krahn, T., Kuo, R., & Chang, M. (2023). Personalized study guide: A Moodle plug-in generating personal learning Path for students. In C. Frasson, P. Mylonas, & C. Troussas (Eds.), *Augmented intelligence and intelligent tutoring systems, vol. 13891 LNCS of lecture notes in computer science* (pp. 333–341). Cham: Springer Nature Switzerland.

Li, F., He, Y., & Xue, Q. (2021). Progress, challenges and countermeasures of adaptive learning: A systematic review. *Educational Technology & Society*, *24*(3), 238–255. https://www.jstor.org/stable/27032868

Limongelli, C., Sciarrone, F., Temperini, M., & Vaste, G. (2010). A module for adaptive course configuration and assessment in Moodle. In MD. Lytras, P. Ordonez De Pablos, A. Ziderman, A. Roulstone, H. Maurer, & JB. Imber (Eds.), *Knowledge management, information systems, E-learning, and sustainability research, vol. 111 CCIS of communications in computer and information science* (pp. 267–276). Berlin, Heidelberg: Springer Berlin Heidelberg.

Lubchak, V., Kupenko, O., & Kuzikov, B. (2012). Approach to dynamic assembling of individualized learning paths. *Informatics in Education*, *11*(2), 213–225. https://doi.org/10.15388/infedu.2012.11

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–386. https://doi.org/10.1097/00006199-198611000-00017

Marienko, M. V., Markova, O. M., & Semerikov, S. O. (2026). AI literacy in secondary education: Framework, assessment, and professional development in the Ukrainian context. *Computers and Education: Artificial Intelligence*, *10*, 100605. https://doi.org/10.1016/j.caeai.2026.100605

Martin, F., Chen, Y., Moore, R. L., & Westine, C. D. (2020). Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. *Educational Technology Research and Development*, *68*(4), 1903–1929. https://doi.org/10.1007/s11423-020-09793-2

Mejeh, M., Sarbach, L., & Hascher, T. (2024). Effects of adaptive feedback through a digital tool – a mixed-methods study on the course of self-regulated learning. *Education and Information Technologies*, *29*(14), 1–43. https://doi.org/10.1007/s10639-024-12510-8

Mirata, V., & Bergamin, P. (2019). Developing an implementation framework for adaptive learning: A case study approach. In *Proceedings of the European Conference on e-Learning, ECEL* Vol. 2019-November, pp. 668–673.

Mirata, V., Hirt, F., Bergamin, P., & van der Westhuizen, C. (2020). Challenges and contexts in establishing adaptive learning in higher education: Findings from a Delphi study. *International Journal of Educational Technology in Higher Education*, *17*(1), 32. https://doi.org/10.1186/s41239-020-00209-y

Moreno-Marcos, P. M., Barredo, J., Muñoz-Merino, P. J., & Delgado Kloos, C. (2023). Statoodle: A learning analytics tool to analyze Moodle students' actions and prevent cheating. In O. Viberg, I. Jivet, P. Muñoz-Merino, M. Perifanou, & T. Papathoma (Eds.), *Responsive and sustainable educational futures, vol. 14200 LNCS of lecture notes in computer science* (pp. 736–741). Cham: Springer Nature Switzerland.

Neumann, A. T., Yin, Y., Sowe, S., Decker, S., & Jarke, M. (2025). An LLM-driven chatbot in higher education for databases and Information systems. *IEEE Transactions on Education*, *68*(1), 103–116. https://doi.org/10.1109/TE.2024.3467912

Park, J. S., Zou, C. Q., Kamphorst, J., Egan, N., Shaw, A., & Hill, B. M., et al. (2026). LLM agents grounded in self-reports enable general-purpose simulation of individuals.

Penfield, R. D., & Giacobbi Jr, P. R. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science*, *8*(4), 213–225. https://doi.org/10.1207/s15327841mpee0804_3

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing and Health*, *29*(5), 489–497. https://doi.org/10.1002/nur.20147

Prada Segura, J. A. (2026). Implementing adaptive learning in Moodle with artificial intelligence. In T. Guarda, F. Portela, MF. Augusto, & JR. Coronado-Hernández (Eds.), *Advanced research in technologies, information, innovation and sustainability communications in computer and information science* (pp. 268–282). Cham: Springer Nature Switzerland.

Pukkhem, N., Evens, M. W., & Vatanawood, W. (2006). An estimation function for selecting the suitable learning objects in adaptive education systems. In *EISTA 2006 - 4th Int. Conf. on Education and Information Systems: Technologies and Applications, Jointly with SOIC 2006 - 2nd Int. Conf. on SOIC and PISTA 2006 - 4th Int. Conf. on PISTA, Proceedings* (Vol. *1*, pp. 123–128).

Rideout, C. A. (2018). Students' choices and achievement in large undergraduate classes using a novel flexible assessment approach. *Assessment and Evaluation in Higher Education*, *43*(1), 68–78. https://doi.org/10.1080/02602938.2017.1294144

Rivera Muñoz, J. L., Ojeda, F. M., Aparicio Jurado, D. L., Puga Peña, P. F., Martel Carranza, C. P., & Berríos, H. Q., et al. (2022). Systematic review of adaptive learning technology for learning in higher education. *Eurasian Journal of Educational Research*, *2022*(98), 221–233. https://ejer.com.tr/manuscript/index.php/journal/article/view/707

Semerikov, S. O., Nechypurenko, P. P., Vakaliuk, T. A., Mintii, I. S., & Fadieieva, L. O. (2025). Differential effects of Moodle course design on student subpopulations: Advancing personalized learning in higher education. *Smart Learning Environments*, *12*(1), 46. https://doi.org/10.1186/s40561-025-00400-6

Semerikov, S. O., Nechypurenko, P. P., Vakaliuk, T. A., Mintii, I. S., & Fadieieva, L. O. (2026). Multivariate analysis of Moodle components and grade distribution patterns for adaptive learning environments. *Discover Education*, *5*(1), 15. https://doi.org/10.1007/s44217-025-01024-1

Smyrnova-Trybulska, E., Morze, N., & Varchenko-Trotsenko, L. (2022). Adaptive learning in university students' opinions: Cross-border research. *Education and Information Technologies*, *27*(5), 6787–6818. https://doi.org/10.1007/s10639-021-10830-7

Tamo-Vargas, G., Quispe-Pari, E., Bedregal-Alpaca, N., Guevara, K., Delgado-Barra, L., & Laura-Ochoa, L. (2023). Design and development of a Moodle Plugin for the adaptation of the teaching-learning process through graded grading constraints [Diseño y desarrollo de un Plugin en Moodle para la adaptación de proceso enseñanza-aprendizaje a través de restricciones de calificación escalonada]. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, *2023*(E59), 338–351. https://dialnet.unirioja.es/servlet/articulo?codigo=10079702

Vásquez-Bermúdez, M., Aguirre-Munizaga, M., & Hidalgo-Larrea, J. (2023). Analysis of col presence indicators in a Moodle forum using unsupervised learning techniques. In R. Valencia-García, M. Bucaram-Leverone, J. Del Cioppo-Morstadt, N. Vera-Lucio, & PH. Centanaro-Quiroz (Eds.), *Technologies and innovation, vol. 1873 CCIS of communications in computer and information science* (pp. 27–38). Cham: Springer Nature Switzerland.

Vitez, A. (2022). Course module instances report. https://moodle.org/plugins/report_coursemodstats

Wu, C. H., Chen, Y. S., & Chen, T. C. (2018). An adaptive e-learning system for enhancing learning performance: Based on dynamic scaffolding theory. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(3), 903–913. https://doi.org/10.12973/ejmste/81061

Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers and Education*, *140*, 103599. https://doi.org/10.1016/j.compedu.2019.103599

Zahorodko, P. V., & Semerikov, S. O. (2026, Feb). Integrating agile methodologies and AI-assisted learning in web programming education: A theoretical framework for CS curriculum transformation. *Discover Education*, *5*(1), 166. https://doi.org/10.1007/s44217-026-01179-5

Zang, J. (2024). Design of students' personalized learning paths under the integration and development of technology and basic education. *Applied Mathematics and Nonlinear Sciences*, *9*(1). https://doi.org/10.2478/amns-2024-1737

Zheng, L., Chiang, W., Sheng, Y., Zhuang, S., Wu, Z., & Zhuang, Y., et al (2023). Judging LLM-as-a-judge with MT-bench and chatbot arena. In: A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.). *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*. http://papers.nips.cc/paper_files/paper/2023/hash/91f18a1287b398d378ef22505bf41832-Abstract-Datasets_and_Benchmarks.html

Publisher's Note