

INTEGRATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN THE FORMATION OF STUDENTS' COGNITIVE INDEPENDENCE

ALBINA VOLKOTRUBOVA^{1*}, JIE SUN², OLHA PROKOPENKO³, OKSANA NIKOLAIENKO⁴, KHRYSTYNA KONDAUROVA⁵

¹Department of Linguistics, International University of Kyrgyzstan, Bishkek, Kyrgyzstan.

²Department of Higher Education Pedagogy, Kyrgyz National University named after Zh. Balasagyn, Bishkek, Kyrgyzstan.

³Armenian State University of Economics, Yerevan, Armenia.

⁴Department of Language Education, Bogomolets National Medical University, Kyiv, Ukraine.

⁵Department of Ukrainian Studies, Bogomolets National Medical University, Kyiv, Ukraine.

E-mail: ¹albinavolkotrubova@gmail.com, ²sun_j@gmail.com, ³prokopenko.olga@gmail.com, ⁴oksana_nikolayenko@ukr.net, ⁵k.kondaurova@gmail.com

ABSTRACT

Generative artificial intelligence is increasingly embedded in higher education, yet its contribution to students' cognitive independence remains uncertain because prior research has often examined self-regulation, critical thinking, learning outcomes, and AI acceptance as separate effects. The problem addressed in this study was the absence of an integrated, empirically validated framework that simultaneously measures cognitive independence and structures AI-supported learning without replacing students' own cognitive effort. The study therefore developed and experimentally validated an AI-driven adaptive model combining learning analytics, generative support, adaptive scaffolding, reflective revision, and teacher pedagogical control. A multi-site quasi-experimental pre-test/post-test design involved 240 students from four higher education institutions. The experimental groups used an integrated environment comprising ChatGPT, Gemini, Claude, and Microsoft Copilot, whereas the control groups followed comparable learning activities without systematic generative AI support. Effectiveness was assessed using the Cognitive Independence Index (CII) and the supplementary SRL-S, TPQ, and DSS indicators. In the experimental groups, CII increased from 0.50 to 0.71, compared with 0.49 to 0.56 in the control groups; SRL-S increased from 0.52 to 0.73 versus 0.51 to 0.58; TPQ from 0.55 to 0.75 versus 0.54 to 0.61; and DSS from 0.51 to 0.72 versus 0.50 to 0.57. Analytical processing and self-correction showed the greatest sensitivity to the intervention. The main contribution is a validated measurement-and-support architecture linking cognitive outcomes to behavioral patterns of editing, critical evaluation, and reflective revision. The findings indicate that pedagogically structured human-AI interaction, rather than the frequency of AI use alone, is associated with stronger cognitive independence.

Keywords: *Learning, Pedagogy, Knowledge, Artificial Intelligence, Generative Artificial Intelligence, Cognitive Independence, Adaptive Learning Model, Learning Analytics, Self-Regulated Learning.*

1. INTRODUCTION

The integration of artificial intelligence (AI) into higher education has intensified debate about its pedagogical value, transparency, and measurable effects on learning [1–3]. Within this debate, cognitive independence is particularly important because generative systems can support planning, information processing, and feedback while also reducing students' own cognitive effort when

outputs are accepted without verification [4–6]. In this study, cognitive independence was defined as the capacity to formulate learning goals, identify and evaluate information, process it analytically, revise errors, and reflect on decisions. Operationally, the construct was represented by five interrelated dimensions: goal formulation, information search and source evaluation, analytical processing, self-correction, and reflective regulation. Self-regulated

learning provides an important theoretical basis for these processes because it emphasizes planning, monitoring, adaptation, and reflection. However, available evidence indicates that AI-related benefits depend on coherent instructional design rather than on access to individual tools alone [7–9].

A related body of research focuses on AI literacy and competence. Meaningful use of generative AI requires an understanding of system capabilities, reliability limits, hallucination risks, verification requirements, and the possibility of algorithmic dependence [4]. This literature clarifies the competencies required for responsible AI use, but two methodological gaps remain. First, validated measures commonly assess separate constructs such as self-regulation, critical thinking, AI literacy, or academic performance rather than cognitive independence as an integrated construct. Second, many studies evaluate individual tools or isolated interventions without combining learning analytics, generative feedback, self-correction, reflective regulation, and pedagogically grounded personalization within one adaptive system.

Empirical findings also show that the educational effects of AI vary across contexts. Studies in Armenia have reported generally positive student experiences and expectations regarding the modernization of higher education [10, 11]. Research in Kyrgyzstan has emphasized the development of critical thinking and analytical skills [12, 13], whereas studies in Ukraine have identified both opportunities for critical thinking and risks associated with the mechanical acceptance of ready-made answers [14, 15]. These findings are relevant but remain difficult to compare because they rely on different outcome variables. Consequently, critical thinking, analytical skills, self-regulated learning, engagement, and academic performance cannot by themselves represent cognitive independence as a coordinated process of problem formulation, information search, analytical synthesis, self-correction, and responsible use of AI tools.

The research problem therefore concerns more than the effectiveness of individual generative AI tools. It concerns the absence of a coherent approach that connects a valid measurement of cognitive independence with a pedagogical mechanism for developing it. Existing approaches provide evidence about separate cognitive or behavioral outcomes, but they do not sufficiently explain how learning analytics, adaptive support, reflective revision, and teacher control can be coordinated to preserve students' independent decision-making while generative AI is used during learning.

The purpose of the study was to develop and experimentally validate an AI-driven adaptive model for fostering students' cognitive independence in higher education by integrating generative artificial intelligence, learning analytics, and personalized pedagogical support.

Three research tasks were defined. The first was to formalize a system of indicators for assessing cognitive independence across cognitive, behavioral, and performance-related characteristics of learning activity. The second was to design an adaptive architecture integrating learning-data collection and analysis, generative feedback, adaptive scaffolding, reflective revision, and teacher pedagogical control. The third was to evaluate the model experimentally in terms of cognitive independence, self-regulation, task performance quality, decision stability, and behavioral patterns of interaction with generative AI.

The study hypothesis assumed that generative AI would be associated with stronger cognitive independence when it functioned as an adaptive support system rather than as a mechanism for automatically producing ready-made answers. Under this condition, learning analytics and personalized feedback were expected to support self-monitoring, analytical processing, adaptive scaffolding, and reflective revision while preserving students' responsibility for final decisions.

Contemporary research on AI-supported learning has developed along several interconnected directions. One major perspective treats generative AI as an environment for self-regulated learning, with potential benefits for planning, monitoring, self-correction, and reflection. Xia et al. [16] synthesized this field and identified substantial variation in intervention design, indicating that the mechanisms linking generative AI to self-regulation remain insufficiently standardized. Xu et al. [17] showed that metacognitive support is critical to the quality of learning in generative AI environments, while Weng et al. [18] demonstrated that learner characteristics also condition the effectiveness of ChatGPT-supported self-regulation. Anders and Dux Speltz [19] connected productive AI use with generative AI literacy, and He [20] emphasized the socio-cognitive foundations of learner autonomy. Taken together, these studies support the importance of regulation and metacognition but do not fully operationalize cognitive independence as a construct that also includes independent information search, analytical synthesis, revision of decisions, and responsible use of intellectual tools.

A second research direction links generative AI to critical thinking, argumentation, and other higher-order cognitive processes. De la Puente et al. [21] found that ChatGPT can facilitate argumentation and critical evaluation in structured student debates, while Avsheniuk et al. [14] identified both the potential for critical-thinking development and the risk of superficial acceptance of generated answers. Oates and Johnson [22] emphasized evaluative and reflective learning processes, and Suriano et al. [23] and Wang et al. [24] associated positive cognitive effects with the structure of human-AI interaction and constructivist instructional design. Romero-Rodríguez et al. [25], by contrast, focused primarily on students' perceived usefulness of ChatGPT. This distinction is important because perceived usefulness or improved performance does not necessarily demonstrate cognitive independence. Critical thinking concerns the evaluation of claims, evidence, and arguments; cognitive independence is broader and additionally includes goal setting, independent information search, analytical synthesis, self-correction, and reflective regulation of decisions.

A third research direction consists of systematic reviews, meta-analyses, and technology-acceptance studies. Deng et al. [26], Wang and Fan [27], and Mo et al. [28] reported generally positive but heterogeneous effects of ChatGPT on learning outcomes, with effect sizes varying by task, instructional design, and implementation conditions. These syntheses strengthen the evidence that generative AI can support learning, but they do not establish whether higher performance reflects durable cognitive independence or more efficient completion of particular tasks. Strzelecki [29] further showed that technology acceptance depends on motivational and contextual factors, confirming that access alone is insufficient for pedagogically meaningful use. Oliinyk et al. [30] examined AI within a methodology for developing doctoral students' research competencies, but the focus remained on research skills rather than an integrated model combining learning analytics, generative feedback, behavioral indicators, self-correction, and cognitive outcomes.

Recent 2026 literature further clarifies the unresolved problem. Nguyen et al. [31] concluded that AI-supported self-regulated learning still concentrates disproportionately on task enactment and evaluation, with less systematic support for goal definition and adaptation. Vendrell and Johnston [32] argued that unstructured large-language-model use can promote cognitive offloading and metacognitive disengagement and therefore requires

deliberate pedagogical scaffolding. Alubthane [33] similarly found that the effects of generative AI on higher-order cognition are conditional on instructional strategy rather than uniformly positive. Jou and Kao [34] showed that the quality of human-AI collaboration is linked to critical-thinking and self-regulatory pathways. These findings strengthen the need for a model that measures and supports an integrated cycle of independent cognition rather than isolated outcomes. The specific gap addressed by the present study was therefore the lack of a validated architecture combining cognitive-independence diagnostics, learning analytics, generative support, adaptive revision, reflective regulation, and teacher control within one empirically testable system.

Accordingly, the formal research problem was stated as follows: how can generative AI be integrated into higher education so that it strengthens measurable cognitive independence rather than merely accelerating task completion? Three research questions were examined. RQ1: Which cognitive, behavioral, and performance dimensions can provide a valid integrated operationalization of students' cognitive independence? RQ2: How can learning analytics, generative feedback, adaptive scaffolding, reflective revision, and teacher control be combined within one AI-driven adaptive model? RQ3: Does participation in this model produce stronger changes in CII, self-regulation, task performance quality, decision stability, and critical interaction with AI outputs than a comparable conventional digital-learning condition?

2. METHODOLOGY

2.1. Research Design

The study employed a multi-site quasi-experimental pre-test/post-test design. Generative AI was conceptualized as part of an adaptive learning-support system rather than as a collection of isolated tools. The methodological framework integrated pedagogical, analytical, and technological dimensions. The pedagogical dimension focused on students' ability to formulate learning goals, locate and evaluate information, revise their work, and make reflective decisions. The analytical dimension addressed behavioral patterns and learning outcomes, whereas the technological dimension concerned the operation of an AI-supported environment that delivered personalized feedback and adaptive assistance. The empirical component involved four higher education institutions representing economic, linguistic, pedagogical, and professionally oriented educational contexts. Students were assigned to control or experimental

groups at the level of existing academic subgroups. The control groups completed comparable learning activities in a conventional digital environment without systematic generative AI support. The experimental groups worked within the adaptive model developed for the study, which combined learning analytics, generative AI, and pedagogically guided personalization. The generative environment incorporated ChatGPT [35], Gemini [36], Claude [37], and Microsoft Copilot [38]. These systems functioned as interconnected components of the broader learning environment rather than as stand-alone answer generators. The study comprised three stages: baseline assessment, an intervention stage involving problem-solving, analytical, interpretive, and reflective activities, and post-intervention assessment of cognitive, behavioral, and performance-related changes. A quasi-experimental design was used because implementation occurred under authentic university conditions without individual random assignment.

2.2. Participants and Research Procedure

The study population consisted of bachelor's and master's students who regularly completed analytical, interpretive, or problem-oriented assignments in digital learning environments. Initially, 252 students from the Armenian State University of Economics, International University of Kyrgyzstan, Kyrgyz National University named after Zh. Balasagyn, and Bogomolets National Medical University were recruited. After the exclusion criteria were applied, the final sample comprised 240 participants, with 60 students from each institution; the attrition rate was 4.8%. The institutions were selected because they represented different educational profiles, provided digital platforms suitable for standardized learning-analytics collection, and allowed comparable learning tasks and a common pre-test/post-test procedure to be implemented. Because purposive sampling was used, generalization was limited to educational contexts with similar characteristics. Participants were 18–26 years old ($M = 21.3$, $SD = 2.1$). The sample included 154 female students (64.2%) and 86 male students (35.8%); 192 participants (80.0%) were undergraduates and 48 (20.0%) were master's students. Four academic areas were represented equally: economics and management, linguistics, higher education pedagogy, and language education at a medical university.

Eligibility required enrollment in the second to fourth year of a bachelor's program or the first to second year of a master's program, active

participation in the educational process, access to the institutional digital platform, basic digital competence, and willingness to follow the study protocol. Participants also could not have prior experience with the systematic use of ChatGPT, Gemini, Claude, or Microsoft Copilot within a formally organized adaptive-support model. Exclusion criteria were absence from more than 20% of learning sessions, incomplete pre-test or post-test data, missing learning-analytics records, violation of experimental instructions, or unauthorized use of external AI services by control-group participants. Of the 252 students initially recruited, 12 were excluded: five did not complete the post-test, four had incomplete digital-activity records, and three missed more than 20% of the formative sessions.

The study followed a multi-site quasi-experimental design with pre-test and post-test measurements. Full randomization was not used; instead, group assignment occurred at the level of existing academic subgroups within each institution. At each university, one experimental group and one control group were formed, each with 30 students. Assignment prioritized comparability in field of study, academic level, workload, and timetable rather than random selection. The absence of randomization was treated as a methodological limitation. Baseline equivalence was examined before the formative stage. Mean CII scores were 0.49 and 0.50 ($p = 0.417$), SRL-S scores were 0.51 and 0.52 ($p = 0.389$), TPQ scores were 0.54 and 0.55 ($p = 0.402$), and DSS scores were 0.50 and 0.51 ($p = 0.436$) in the control and experimental groups, respectively. No statistically significant baseline differences were identified, supporting comparability before the intervention.

The research procedure lasted 16 weeks. Baseline assessment occupied the first three weeks and measured cognitive independence, self-regulation, performance on analytical and problem-oriented tasks, decision-making behavior, and baseline digital activity. The formative intervention then continued for 11 weeks. Students in the experimental groups completed learning tasks in an integrated generative AI environment comprising ChatGPT, Gemini, Claude, and Microsoft Copilot. The tools were embedded in a structured pedagogical scenario that supported problem clarification, generation of alternative decisions, organization of analytical information, identification of contradictions, and self-correction. Control-group students completed tasks equivalent in topic, complexity, duration, and cognitive structure without systematic generative AI support. The final two weeks comprised post-testing,

repeated completion of the task set, and within-group and between-group comparison of changes observed during the intervention.

Several measures were applied to strengthen internal validity. Experimental conditions were standardized across the four institutions, and both control and experimental groups worked with the same educational content; the planned difference was the mode of instructional support. Within each institution, the same instructor or a pair of instructors from the same department worked with both groups according to a unified methodological protocol. Eight instructors participated in total, two from each institution, and used identical assessment rubrics, task templates, and result-recording procedures.

The activity schedule, session duration, number of assignments, submission deadlines, digital logging procedures, and technical conditions were also standardized. All participants completed 22 formative sessions of 90 minutes each. Data were collected on task performance, number of AI interactions, frequency of prompt reformulation, extent of response editing, patterns of self-correction, consistency of selected strategies, and the ability to justify final decisions. These indicators formed the empirical basis for the comparative and statistical analyses.

2.3. Assessment Tools and Indicators

The assessment framework combined analytical and problem-oriented tasks, structured expert assessment, learning-analytics records, and short student reflection protocols because cognitive independence was expected to appear simultaneously at cognitive, behavioral, and performance levels. The selection of the five core CII dimensions was theory-driven rather than arbitrary. Goal formulation and monitoring were grounded in self-regulated-learning research [7–9, 16, 31]; independent information search and source evaluation reflected AI-literacy and responsible-use requirements [4, 19]; analytical processing represented the evaluative and higher-order-thinking dimension emphasized in critical-thinking research [21–24, 32, 33]; and self-correction together with reflective regulation reflected metacognitive-support mechanisms [17, 31, 34]. The assessment system therefore registered task quality, independent information search, self-correction, decision stability, and reflective control as complementary manifestations of the same broader construct. The principal measure was the Cognitive Independence Index (CII), which operationalized the student's ability to organize and manage cognitive activity

independently in an AI-supported environment without substituting personal intellectual effort with generated content. The CII comprised goal formulation and problem structuring, independent information search and source evaluation, analytical processing and synthesis, self-correction and revision, and reflective regulation and responsible AI use.

Each component was represented by three standardized sub-indicators, producing a total of 15 primary indicators. Every sub-indicator was coded on a scale ranging from 0 to 4, where 0 indicated the absence of the characteristic, 1 represented weak manifestation, 2 indicated partial manifestation, 3 reflected sufficient manifestation and 4 meant complete and stable manifestation. For each component, the arithmetic mean of the three sub-indicators was calculated and then normalized to a range between 0 and 1 according to the following formula:

$$X_{norm} = X_{mean} / 4 \quad (1)$$

where X_{mean} represents the average score for the dimension and 4 is the maximum possible value on the scale.

The integral CII score was calculated as a weighted sum of the five normalized components:

$$CII = 0.20 \times GF + 0.20 \times IS + 0.25 \times AP + 0.20 \times SC + 0.15 \times RR \quad (2)$$

Analytical processing and synthesis received the highest weight because it represented the central transformation of information into reasoned conclusions. The GF, IS, and SC dimensions received equal weights because they represented key stages of autonomous learning, whereas RR received a slightly lower weight because it served primarily as a regulatory dimension. The weighting scheme was study-specific and was not intended to establish a universal hierarchy among components. Coefficients were derived from theoretical modeling and expert agreement. Six specialists in higher education pedagogy, digital learning, and educational assessment participated in the validation procedure, and Kendall's $W = 0.82$ indicated a high level of agreement.

In addition to the CII, three supplementary indicators were employed. The Self-Regulated Learning Support Score (SRL-S) measured the extent to which students planned, monitored, and adjusted their learning activities. The scale consisted of eight items scored from 0 to 4 and subsequently normalized using the following formula:

$$SRL - S = (\Sigma srli / 32) \quad (3)$$

where $\Sigma srli$ denotes the total score across the eight items and 32 represents the maximum possible score.

Task Performance Quality (TPQ) assessed the quality of final task outputs according to four criteria: accuracy, argumentation, completeness and structural coherence. Each criterion was evaluated on a scale from 0 to 4. The indicator was calculated as follows:

$$TPQ = (Acc + Arg + Comp + Coh) / 16 \quad (4)$$

where *Acc* refers to accuracy, *Arg* to argumentation, *Comp* to completeness and *Coh* to structural coherence.

The Decision Stability Score (DSS) was used to evaluate the consistency of students' final decisions after several rounds of feedback and revision. It included four components: stability of argumentation, consistency of logic, preservation of the core decision after clarification and justification of the final choice. Each component was scored on a scale from 0 to 4. The indicator was computed using the following formula:

$$DSS = (Stab + Cons + Core + Just) / 16 \quad (5)$$

Behavioral interaction with generative AI was analyzed separately from the CII. Indicators included the number of interactions with each AI tool, frequency of prompt reformulation, proportions of directly accepted, modified, and rejected generated responses, and time between iterations. These variables were not incorporated into the CII because they described the process of AI use rather than the cognitive construct itself. Instead, they were used as explanatory indicators to determine whether changes in cognitive independence were accompanied by more active, critical, and reflective engagement with generated content.

Content validity was examined through a two-stage expert-validation procedure. First, six independent experts evaluated the correspondence between each sub-indicator and the theoretical construct of cognitive independence. The overall S-CVI/Ave was 0.91, and I-CVI values ranged from 0.83 to 1.00, indicating strong content validity. Second, exploratory and confirmatory factor analyses were used to examine construct validity. The proposed five-factor structure showed acceptable fit: the Kaiser–Meyer–Olkin value was 0.87, Bartlett's test of sphericity was statistically significant ($p < 0.001$), and all factor loadings exceeded 0.60. Confirmatory factor analysis yielded CFI = 0.94, TLI = 0.92, RMSEA = 0.052, and SRMR

= 0.047. These results supported interpretation of the CII as a multidimensional measure of cognitive independence.

Convergent validity was assessed through relationships between the CII and the supplementary indicators. The overall CII score correlated positively with SRL-S ($r = 0.71$, $p < 0.001$), TPQ ($r = 0.68$, $p < 0.001$), and DSS ($r = 0.64$, $p < 0.001$). The pattern indicated that cognitive independence was closely related to self-regulation, task quality, and decision stability while remaining distinguishable from each construct.

Reliability was assessed using Cronbach's alpha. The overall CII showed high internal consistency ($\alpha = 0.89$). Reliability coefficients for the individual dimensions were GF = 0.81, IS = 0.84, AP = 0.87, SC = 0.85, and RR = 0.79. The supplementary scales also demonstrated acceptable reliability: SRL-S = 0.86, TPQ = 0.83, and DSS = 0.81.

Inter-rater assessment was conducted by two independent experts who did not teach the study groups. They evaluated written responses and reflection protocols using the common coding rubric. Pilot calibration was completed on 24 papers, representing 10% of the analytical dataset. After discrepancies were discussed, the experts proceeded to the main assessment. Inter-rater consistency was estimated using Cohen's kappa for categorical decisions and the intraclass correlation coefficient for quantitative ratings. Cohen's was 0.82 and ICC(2,1) was 0.88, indicating high agreement. Ratings differing by more than one point on an individual indicator were jointly reviewed and resolved by consensus.

2.4. Proposed AI-driven Adaptive Model for the Formation of Students' Cognitive Independence

The proposed model represented an AI-driven adaptive system designed to support cognitive independence in a digital learning environment. Its components were selected to connect mechanisms that prior research often examined separately: learning analytics for identifying learning patterns [2, 8], cognitive diagnostics for targeting support, generative AI for producing alternatives, adaptive scaffolding for guided revision [17, 31, 32], reflective revision for metacognitive control, and teacher pedagogical control for preserving student agency [32–34]. The model therefore linked the learning task and the student's initial response with activity tracking, cognitive-independence diagnostics, generative support, adaptive revision, reflection, and final evaluation. Generative AI functioned as an intermediary for problem

clarification, critical information processing, self-correction, and reflective revision rather than as a provider of final answers. Model effectiveness was evaluated primarily through the CII and was supplemented by SRL-S, TPQ, and DSS. This

architecture allowed both learning outcomes and the process through which students interacted with AI-supported feedback to be monitored within a single framework.

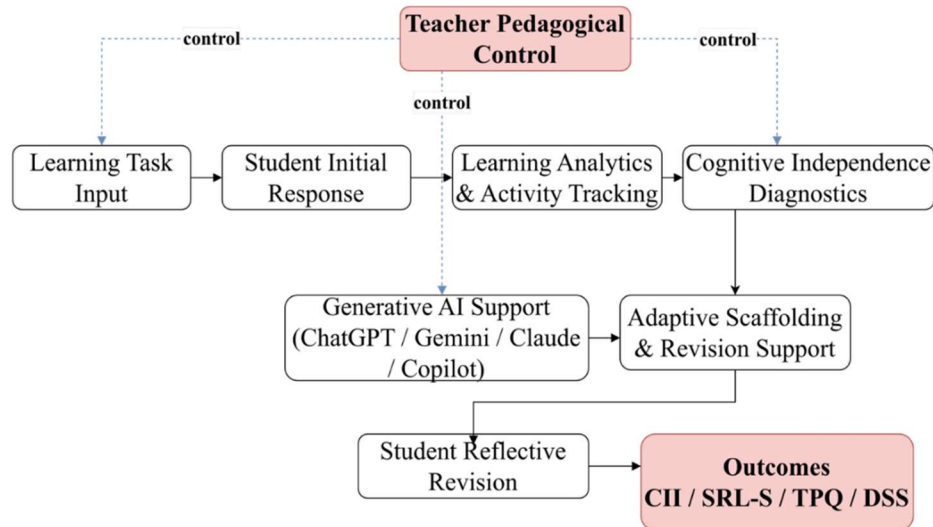


Figure 1: AI-driven adaptive model for developing students' cognitive independence

Source: consolidated by the author

Figure 1 presents the structural cycle of the AI-driven adaptive model. The sequence moved from the learning task and the student's initial response to learning-activity analysis, diagnosis of cognitive independence, generative AI support, adaptive scaffolding, reflective revision, and outcome assessment. A separate teacher-control loop regulated the use of AI tools and maintained the pedagogical boundaries of the model. Feedback from outcome assessment to the analytical unit represented the iterative character of the system and its orientation toward progressively stronger independent cognitive activity.

2.5. Data Processing, Statistical Analysis and Technical Environment

Data processing and statistical analysis were conducted using Python and IBM SPSS Statistics. Python was used to organize educational records, integrate data from multiple sources, normalize indicators, and generate consolidated analytical datasets. IBM SPSS Statistics was used for descriptive analysis, reliability testing, within-group and between-group comparisons, and analysis of relationships between CII and behavioral characteristics of generative AI use. The empirical dataset comprised learning-task outcomes, expert evaluations, reflection protocols, and digital-activity logs. After data collection, indicators were integrated

into a unified observation matrix and normalized to the 0–1 range before the weighted CII was calculated. Descriptive statistics included means, standard deviations, and ranges. Internal consistency was evaluated using Cronbach's alpha. Depending on distributional assumptions, pre-test/post-test changes within groups were assessed with either the paired-samples t-test or the Wilcoxon signed-rank test, while between-group comparisons used either the independent-samples t-test or the Mann–Whitney U test. Effect sizes were calculated for group differences. Spearman's rank correlation coefficient was used to examine relationships between CII growth and behavioral patterns of generative AI use.

3. RESULTS

3.1. General Dynamics of Results in the AI-Supported Learning Environment

Following the 11-week formative intervention, students in the experimental groups showed greater improvements than those in the control groups across all principal indicators used to evaluate the model. The largest changes were observed in the Cognitive Independence Index, Self-Regulated Learning Support Score, Task Performance Quality, and Decision Stability Score. The control groups also improved during regular educational activity, but the magnitude of change was smaller. Table 1

summarizes the pre-test and post-test values in both conditions.

Table 1: Comparative values of CII, SRL-S, TPQ and DSS indicators in the control and experimental groups at the pre-test and post-test stages

Indicator	Pre-test control group	Post-test control group	Pre-test experimental group	Post-test experimental group
CII	0.49	0.56	0.50	0.71
SRL-S	0.51	0.58	0.52	0.73
TPQ	0.54	0.61	0.55	0.75
DSS	0.50	0.57	0.51	0.72

Source: consolidated by the author

Table 1 showed that the control and experimental groups were comparable at pre-test, with no statistically significant differences in CII, SRL-S, TPQ, or DSS. After the intervention, both groups improved, but gains were substantially larger in the experimental condition. CII increased by 0.21 in the experimental groups compared with 0.07 in the control groups. The same pattern was observed for SRL-S (0.21 versus 0.07), TPQ (0.20 versus 0.07), and DSS (0.21 versus 0.07). Post-test differences

were statistically significant for all four indicators: CII, $t(238) = 5.94$, $p < 0.001$, $d = 0.77$; SRL-S, $t(238) = 6.02$, $p < 0.001$, $d = 0.78$; TPQ, $t(238) = 5.61$, $p < 0.001$, $d = 0.72$; and DSS, $t(238) = 5.88$, $p < 0.001$, $d = 0.76$. These results indicated that the observed difference extended beyond task performance to self-regulation and decision stability, with medium-to-large standardized effects.

3.2. Functional Results of the AI-Driven Adaptive Model

The proposed AI-driven adaptive model operated as an integrated instructional system rather than as a set of separate digital tools. Learning analytics, cognitive-independence diagnostics, generative support, adaptive scaffolding, reflective revision, and pedagogical supervision functioned together throughout the intervention. The highest functional effects were associated with response revision, analytical refinement, and pedagogically guided generative AI use. Activity-tracking components mainly provided the information required for adaptation and monitoring rather than directly producing cognitive change. Table 2 summarizes the functional results of the model components.

Table 2: Functional results of the AI-driven adaptive model units in the formative stage of the study

Model function unit	Normalized score	Dominant effect in the educational process	Interpretation of the result
Learning analytics and activity tracking	0.74	Recording behavioral and temporal parameters	Basis for diagnostics and personalization
Cognitive independence diagnostics	0.77	Identification of weaker components of cognitive independence	Basis for targeted support
Generative AI support environment	0.81	Problem clarification and generation of alternative decisions	Support for analytical processing
Adaptive scaffolding and revision support	0.84	Self-correction and content editing	Support for self-review of decisions
Student reflective revision	0.86	Reflection and stabilization of decisions	Key mechanism of cognitive control
Teacher pedagogical control	0.88	Didactic validity and limitation of AI dependence	Guided use of AI

Source: consolidated by the author

Table 2 showed the highest normalized scores for teacher pedagogical control (0.88), student reflective revision (0.86), and adaptive scaffolding and revision support (0.84). This pattern indicated that effectiveness was associated less with access to AI tools than with structured guidance and substantive revision of generated outputs. Learning analytics and cognitive-independence diagnostics performed a system-forming function by supplying information

for targeted adaptation. The generative AI environment supported problem clarification and response organization, but its contribution depended on the diagnostic, reflective, and pedagogical mechanisms surrounding it. The model therefore functioned as an interconnected system in which generative AI served as a cognitive intermediary rather than an autonomous learning agent.

3.3. Results by Components of the Cognitive Independence Index

Growth in the overall Cognitive Independence Index was distributed across all five components rather than being driven by a single dimension. The largest gains occurred in analytical processing, self-correction, and revision following intermediate

feedback. Figure 2 presents the dynamics of goal formulation and problem structuring (GF), independent information search and source evaluation (IS), analytical processing and synthesis (AP), self-correction and revision (SC), and reflective regulation and responsible AI use (RR) in the control and experimental groups.

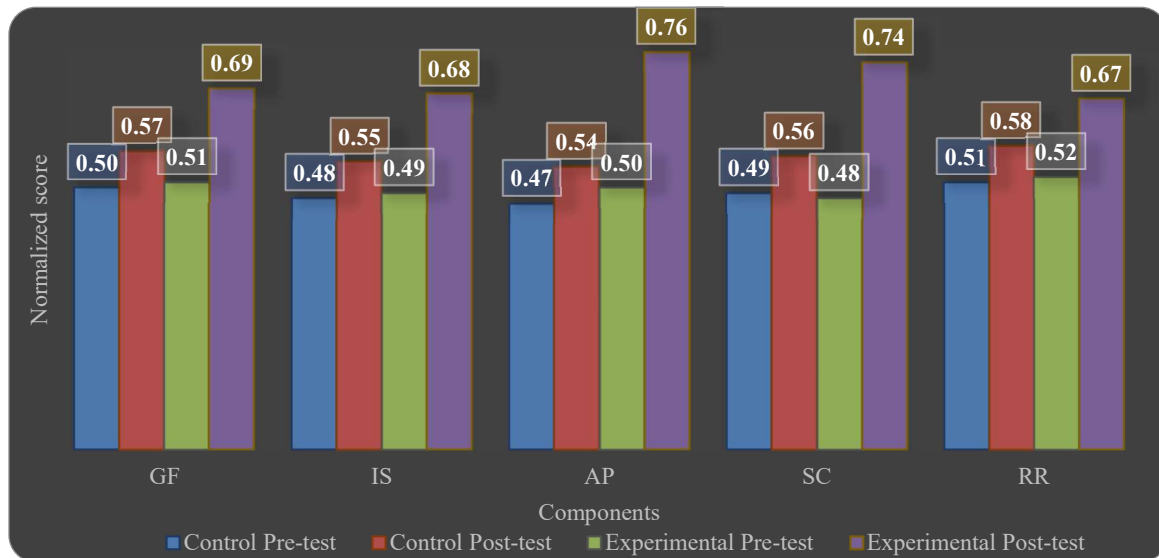


Figure 2: Dynamics of Cognitive Independence Index Components in Control and Experimental Groups
Source: consolidated by the author

Figure 2 showed moderate changes in all CII components in the control groups: GF increased from 0.50 to 0.57, IS from 0.48 to 0.55, AP from 0.47 to 0.54, SC from 0.49 to 0.56, and RR from 0.51 to 0.58. In the experimental groups, GF increased from 0.51 to 0.69, IS from 0.49 to 0.68, AP from 0.50 to 0.76, SC from 0.48 to 0.74, and RR from 0.52 to 0.67. The largest post-test effects were recorded for AP, $t(238) = 6.21$, $p < 0.001$, $d = 0.80$, and SC, $t(238) = 6.08$, $p < 0.001$, $d = 0.79$. Significant effects were also found for GF, $t(238) = 5.14$, $p < 0.001$, $d = 0.66$, and IS, $t(238) = 5.32$, $p < 0.001$, $d = 0.69$. RR showed the smallest but still statistically significant effect, $t(238) = 4.27$, $p < 0.001$, $d = 0.55$. Thus, analytical

processing and self-correction were the components most responsive to the intervention.

3.4. Additional Indicators of the Model's Efficiency: SRL-S, TPQ and DSS

The effect of the AI-supported intervention was also reflected in the supplementary indicators of self-regulation, task performance quality, and decision stability. All three indicators changed more strongly in the experimental groups than in the control groups, providing outcome evidence complementary to the CII. Figure 3 presents the comparative dynamics of SRL-S, TPQ, and DSS.

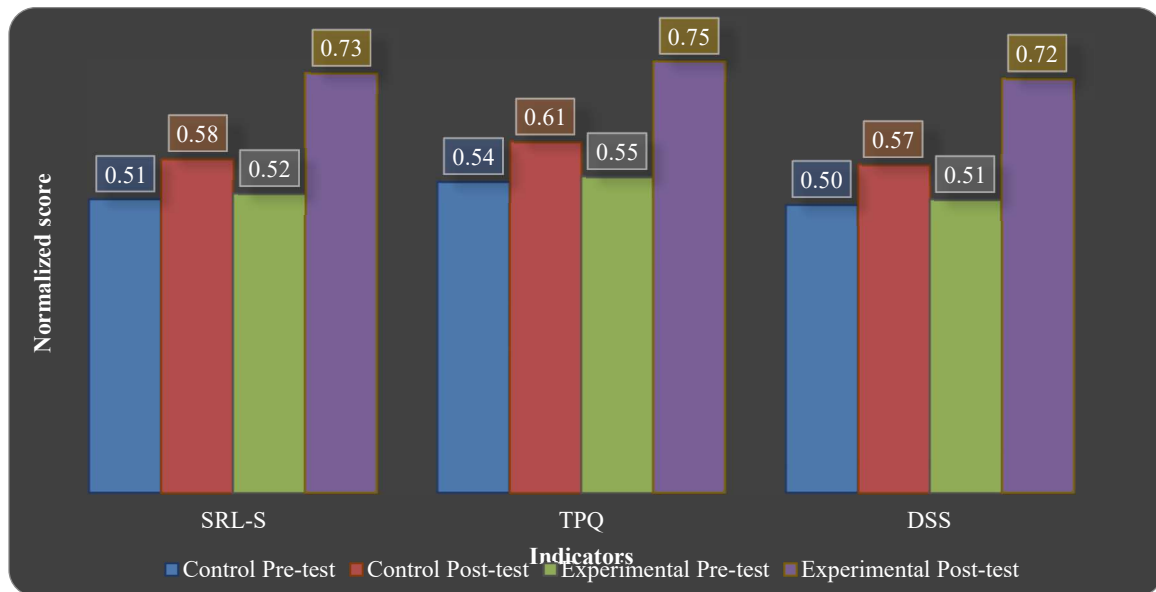


Figure 3: Comparative dynamics of SRL-S, TPQ and DSS indicators in control and experimental groups
Source: developed by the author

Figure 3 showed moderate increases in the control groups: SRL-S rose from 0.51 to 0.58, TPQ from 0.54 to 0.61, and DSS from 0.50 to 0.57. In the experimental groups, SRL-S increased from 0.52 to 0.73, TPQ from 0.55 to 0.75, and DSS from 0.51 to 0.72. The increases were 0.21 for SRL-S and DSS and 0.20 for TPQ. Post-test differences between groups were statistically significant for SRL-S, $t(238) = 6.02$, $p < 0.001$, $d = 0.78$; TPQ, $t(238) = 5.61$, $p < 0.001$, $d = 0.72$; and DSS, $t(238) = 5.88$, $p < 0.001$, $d = 0.76$. The convergent pattern across the three measures indicated that the model was associated with changes in both process-related and performance-related outcomes.

3.5. Behavioral Patterns of Interaction with Generative AI Tools

Digital-trace analysis showed that interaction with generative AI was functionally differentiated across the four tools. ChatGPT, Gemini, Claude, and Microsoft Copilot were used to different degrees and for different purposes, including problem clarification, exploration of alternative explanations, refinement of arguments, response structuring, and revision. The most informative behavioral characteristics were not the number of interactions alone but the frequency of prompt reformulation and the proportions of responses that were modified or rejected. These indicators represented active evaluation of generated content more directly than simple usage frequency. Table 3 summarizes the principal behavioral characteristics.

Table 3: Behavioral characteristics of students' interaction with generative AI tools in experimental groups

AI-tool	Average number of interactions per student	Average number of prompt reformulations	Proportion of directly accepted responses	Proportion of modified responses	Proportion of rejected responses	Dominant function in the educational process
ChatGPT	13.8	1.9	0.19	0.56	0.25	Problem clarification and development of the initial response logic
Gemini	11.2	1.7	0.17	0.58	0.25	Search for alternative explanations and interpretation options
Claude	8.6	1.5	0.21	0.54	0.25	Refinement of argumentation, logic, and wording

Microsoft Copilot	7.9	1.4	0.24	0.49	0.27	Response structuring, formatting, and revision support
-------------------	-----	-----	------	------	------	--

Source: consolidated by the author

Table 3 showed that ChatGPT and Gemini were used most frequently, whereas direct acceptance of generated responses remained low at 0.19 and 0.17, respectively. Modified use predominated across tools, with proportions ranging from 0.49 to 0.58, indicating that generated content was commonly selected, edited, and reworked rather than accepted passively. Mean prompt reformulations were highest for ChatGPT (1.9) and Gemini (1.7), followed by Claude (1.5) and Microsoft Copilot (1.4). Rejection rates remained between 0.25 and 0.27. CII growth was weakly associated with the overall number of AI interactions ($r = 0.24$, $p = 0.011$) but showed stronger relationships with prompt reformulation ($r = 0.48$, $p < 0.001$), modified responses ($r = 0.54$, $p < 0.001$), and rejected outputs ($r = 0.41$, $p < 0.001$). Thus,

cognitive independence was more closely associated with critical and reflective interaction patterns than with usage intensity.

3.6. Integral Indicator of CII and Generalization of Model Efficiency

The integrated CII results showed a substantially greater increase in cognitive independence in the experimental groups than in the control groups. Unlike individual indicators, the CII represented coordinated development across problem formulation, information search, analytical synthesis, self-correction, and responsible AI use. Figure 4 presents the pre-test and post-test CII values for both conditions.

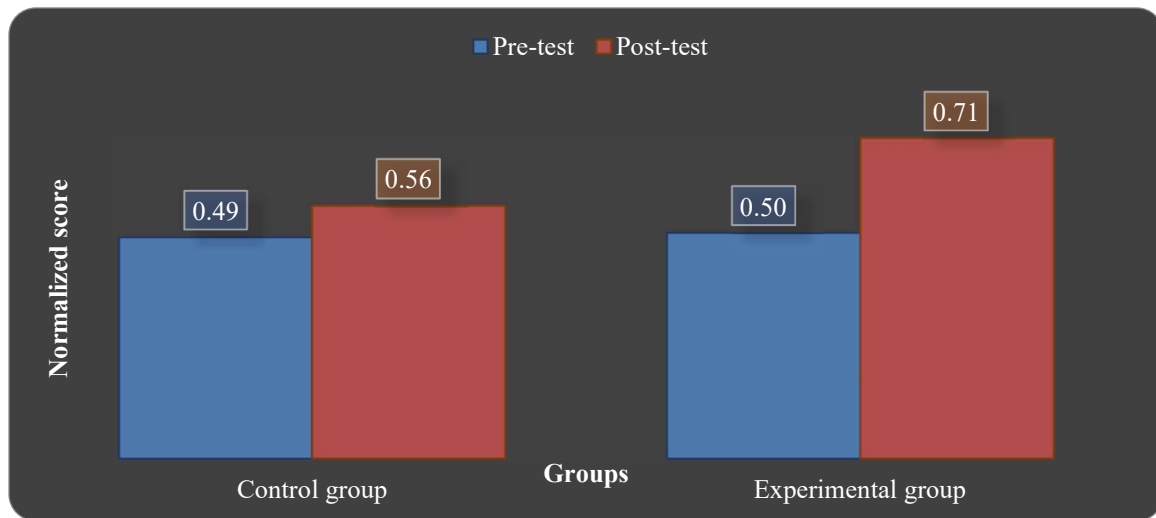


Figure 4: Integral indicator of the Cognitive Independence Index in control and experimental groups

Source: consolidated by the author

Figure 4 showed that CII increased from 0.49 to 0.56 in the control groups and from 0.50 to 0.71 in the experimental groups, corresponding to gains of 0.07 and 0.21. The within-group pre-test/post-test effect was large in the experimental condition, $t(119) = 11.42$, $p < 0.001$, $d = 1.04$, and smaller in the control condition, $t(119) = 4.36$, $p < 0.001$, $d = 0.40$. The pattern was consistent with the between-group results for CII, SRL-S, TPQ, and DSS and with the behavioral evidence showing that editing, critical evaluation, and meaningful modification of generated responses predominated. AP and SC were the most responsive CII components, while GF, IS,

and RR also improved. Within the limits of the quasi-experimental design, the findings supported the study hypothesis that cognitive independence was associated more strongly with the quality of pedagogically structured AI use than with the frequency of interaction itself.

4. DISCUSSION

The findings clarified both the magnitude and the mechanism of the observed association between generative AI support and cognitive independence. The difference between groups could not be reduced to access to AI tools because the experimental

condition required a structured cycle of clarification, comparison, verification, revision, and reflection. This interpretation was consistent with meta-analytic evidence showing that generative AI effects depend on instructional design and task conditions [26–28] and with recent reviews emphasizing that structured scaffolding is necessary to avoid cognitive offloading [31–33]. The behavioral results provided additional mechanism-level evidence: students who reformulated prompts, modified outputs, and rejected inadequate responses showed stronger CII growth than students characterized mainly by higher interaction frequency. This pattern was consistent with Jou and Kao [34], who linked human–AI collaboration to self-regulatory and critical-thinking pathways. The strongest component effects for AP and SC also supported the role of metacognitive regulation identified by Xu et al. [17]. In the present model, however, metacognition was embedded within a larger architecture that connected diagnostics, learning analytics, adaptive scaffolding, reflective revision, and teacher control rather than being treated as an isolated learner characteristic.

The model components contributed differently to this mechanism. Learning analytics and cognitive-independence diagnostics supplied evidence for targeted support; the generative environment provided alternative explanations and response options; adaptive scaffolding transformed those alternatives into revision activity; reflective revision required students to reconsider their own decisions; and teacher pedagogical control maintained the instructional boundaries of AI use. This configuration addressed a limitation of studies centered on perceived usefulness or final performance. Romero-Rodríguez et al. [25] primarily examined perceived usefulness, while Wang and Fan [27] synthesized learning outcomes without treating cognitive independence as a distinct integrated construct. Wang et al. [24] demonstrated the value of constructivist design, and Avsheniuk et al. [14] identified critical-thinking benefits in a language-learning context, but neither approach provided an integrated measurement architecture comparable to the CII. The present study therefore extended prior work in three ways: it operationalized cognitive independence as a multidimensional measurable construct, linked change in that construct to behavioral traces of human–AI interaction, and tested a model in which generative support was coordinated with diagnostics, adaptive revision, reflection, and teacher control. This interpretation also accorded with the 2026 design principles proposed by Vendrell and Johnston [32], while

adding empirical multi-site evidence and a formal outcome measure.

4.1. Limitations

Several limitations constrain interpretation and generalization. First, the quasi-experimental design did not provide individual randomization, so institutional context, disciplinary differences, and implementation variation could not be eliminated completely. Second, the formative intervention lasted 11 weeks, which allowed short- and medium-term change to be examined but did not establish long-term stability. Third, some indicators depended on expert interpretation of behavioral and textual data, although content validity, factor analysis, internal consistency, and inter-rater agreement were used to reduce this source of uncertainty. Fourth, the CII represented a formalized operational definition of cognitive independence and could not capture every manifestation of the construct. Finally, ChatGPT, Gemini, Claude, and Microsoft Copilot were used within one integrated environment; consequently, the findings represented the combined intervention rather than the isolated effect of any single platform. These limitations support future longitudinal, randomized, tool-specific, and cross-disciplinary validation.

4.2. Difference from Prior Work: Strengths and Trade-offs

Relative to earlier studies, the principal difference of the proposed approach was the integration of measurement and intervention within the same architecture. Reviews of AI-supported self-regulated learning have typically organized evidence around planning, monitoring, feedback, and adaptation [16, 31], while critical-thinking research has emphasized evaluation, metacognition, and preservation of epistemic agency [21–24, 32, 33]. The present study connected these processes through a validated five-component CII and linked the outcome measure to learning analytics and behavioral traces. A further strength was the multi-site sample of 240 students and the simultaneous assessment of cognitive, process, and performance outcomes. The use of four generative systems also reduced dependence on the behavior of one platform and allowed the model to be evaluated as a pedagogical architecture rather than a product-specific intervention.

These strengths were accompanied by trade-offs. A broad integrated architecture provided greater ecological relevance but made it impossible to isolate the contribution of ChatGPT, Gemini, Claude, or Microsoft Copilot. The purposive multi-site design increased disciplinary diversity but did

not provide random assignment, and the 11-week intervention supported only short- and medium-term inference. In addition, the CII offered a structured measure of cognitive independence but necessarily reduced a complex construct to predefined dimensions and weights. Compared with more narrowly focused studies, the model therefore gained integrative explanatory value at the cost of tool-specific causal precision. Future studies should retain the integrated measurement framework while using randomized, longitudinal, and factorial designs to separate platform effects, pedagogical components, and discipline-specific responses.

5. CONCLUSION

The study addressed the problem that prior generative-AI research has frequently evaluated separate outcomes without integrating their measurement into a pedagogically controlled model of cognitive independence. The first research question concerned which dimensions could operationalize cognitive independence in an AI-supported environment. The validated CII represented five dimensions – goal formulation, information search and source evaluation, analytical processing and synthesis, self-correction and revision, and reflective regulation and responsible AI use – with overall reliability of $\alpha = 0.89$ and acceptable construct-validity indices. The second question concerned how these dimensions could be supported within an adaptive architecture. The resulting model combined learning analytics, diagnostics, generative support, adaptive scaffolding, reflective revision, and teacher pedagogical control, with the highest functional scores recorded for teacher control (0.88), reflective revision (0.86), and adaptive scaffolding (0.84). The third question concerned whether the model was associated with stronger outcomes than a conventional digital-learning condition. CII increased by 0.21 in the experimental groups compared with 0.07 in the control groups, and the post-test difference reached $t(238) = 5.94$, $p < 0.001$, $d = 0.77$; comparable effects were observed for SRL-S, TPQ, and DSS. Behavioral evidence further showed that CII growth was more strongly related to response modification and prompt reformulation than to interaction frequency. In relation to prior research [31–34], these findings support a conditional rather than tool-deterministic interpretation of generative AI: cognitive benefits were associated with structured human-AI collaboration and pedagogical control. The study's contribution is therefore the empirical linkage of an integrated cognitive-independence measure with a

specific adaptive support architecture and process-level behavioral evidence. Because the design was quasi-experimental, the intervention lasted 11 weeks, and four AI systems were combined, the conclusions remain bounded to comparable contexts. Future research should test the architecture longitudinally, with randomized assignment, discipline-specific replication, and separate experimental arms for individual AI tools.

AUTHOR CONTRIBUTIONS

Albina Volkotrubova, Jie Sun, Olha Prokopenko, Oksana Nikolaenko, and Khrystyna Kondaurava jointly contributed to the conceptualization, methodology, investigation, data interpretation, manuscript preparation, and critical revision of the study. All authors reviewed and approved the final version of the manuscript.

REFERENCES:

- [1] T. K. Chiu F., Q. Xia, X. Zhou, C. S. Chai and M. Cheng, "Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education", *Computers and Education: Artificial Intelligence*, Vol. 4, 2023, Art. No. 100118. <https://doi.org/10.1016/j.caeai.2022.100118>
- [2] R. Alfredo, V. Echeverria, Y. Jin, L. Yan, Z. Swiecki, D. Gašević and R. Martinez-Maldonado, "Human-centred learning analytics and AI in education: A systematic literature review", *Computers and Education: Artificial Intelligence*, Vol. 6, 2024, Art. No. 100215. <https://doi.org/10.1016/j.caeai.2024.100215>
- [3] W. K. Monib, A. Qazi and M. M. Mahmud, "Exploring learners' experiences and perceptions of ChatGPT as a learning tool in higher education", *Education and Information Technologies*, Vol. 30, 2025, pp. 917–939. <https://doi.org/10.1007/s10639-024-13065-4>
- [4] T. K. Chiu F., Z. Ahmad, M. Ismailov and I. T. Sanusi, "What are artificial intelligence literacy and competency? A comprehensive framework to support them", *Computers and Education Open*, Vol. 6, 2024, Art. No. 100171. <https://doi.org/10.1016/j.caeo.2024.100171>
- [5] M. Lan and X. Zhou, "A qualitative systematic review on AI empowered self-regulated learning in higher education", *npj Science of Learning*, Vol. 10, 2025, Art. No. 21. <https://doi.org/10.1038/s41539-025-00319-0>
- [6] J. Blahopoulou and S. Ortiz-Bonnin, "Student perceptions of ChatGPT: Benefits, costs, and attitudinal differences between users and non-

- users toward AI integration in higher education”, *Education and Information Technologies*, Vol. 30, 2025, pp. 19741–19764. <https://doi.org/10.1007/s10639-025-13575-9>
- [7] S.-H. Jin, K. Im, M. Yoo, I. Roll and K. Seo, “Supporting students’ self-regulated learning in online learning using artificial intelligence applications”, *International Journal of Educational Technology in Higher Education*, Vol. 20, 2023, Art. No. 37. <https://doi.org/10.1186/s41239-023-00406-5>
- [8] M. Khalil, J. Wong, B. Wasson and F. Paas, “Adaptive support for self-regulated learning in digital learning environments”, *British Journal of Educational Technology*, Vol. 55, No. 4, 2024, pp. 1281–1289. <https://doi.org/10.1111/bjet.13479>
- [9] D. T. Ng K., C. W. Tan and J. K. L. Leung, “Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study”, *British Journal of Educational Technology*, Vol. 55, No. 4, 2024, pp. 1328–1353. <https://doi.org/10.1111/bjet.13454>
- [10] G. Tovmasyan, “Higher education in Armenia adopting AI and digital technologies: Students’ experiences and perspectives”, *Issues in Educational Research*, Vol. 35, No. 2, 2025, pp. 798–817.
- [11] M. Faraon, K. Rönkkö, M. Milrad and E. Tsui, “International perspectives on artificial intelligence in higher education: An explorative study of students’ intention to use ChatGPT across the Nordic countries and the USA”, *Education and Information Technologies*, vol 30, 2025, pp. 17835–17880. <https://doi.org/10.1007/s10639-025-13492-x>
- [12] D. Zhylchieva and A. Egamberdieva, “Developing critical thinking and analytical skills of students using artificial intelligence”, *Bulletin of the Jusup Balasagyn Kyrgyz National University*, Vol. 2024, No. 3, 2024, pp. 451–455.
- [13] C. K. Chan Y. and W. Hu, “Students’ voices on generative AI: Perceptions, benefits, and challenges in higher education”, *International Journal of Educational Technology in Higher Education*, Vol. 20, 2023, Art. No. 43. <https://doi.org/10.1186/s41239-023-00411-8>
- [14] N. Avsheniuk, O. Lutsenko, T. Svyrydiuk and N. Seminikhyna, “Empowering language learners’ critical thinking: Evaluating ChatGPT’s role in English course implementation”, *Arab World English Journal (AWEJ) Special Issue on ChatGPT*, 2024, pp. 210–224. <https://doi.org/10.24093/awej/ChatGPT.14>
- [15] A. Nguyen, Y. Hong, B. Dang and X. Huang, “Human-AI collaboration patterns in AI-assisted academic writing”, *Studies in Higher Education*, Vol. 49, No. 5, 2024, pp. 847–864. <https://doi.org/10.1080/03075079.2024.2323593>
- [16] Q. Xia, Q. Liu, A. Tlili and T. K. F. Chiu, “A systematic literature review on designing self-regulated learning using generative artificial intelligence and its future research directions”, *Computers & Education*, Vol. 240, 2026, Art. No. 105465. <https://doi.org/10.1016/j.compedu.2025.105465>
- [17] X. Xu, L. Qiao, N. Cheng, H. Liu and W. Zhao, “Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support”, *British Journal of Educational Technology*, Vol. 56, No. 5, 2025, pp. 1842–1863. <https://doi.org/10.1111/bjet.13599>
- [18] X. Weng, Q. Xia, Z. Ahmad and T. K. F. Chiu, “Personality traits for self-regulated learning with generative artificial intelligence: The case of ChatGPT”, *Computers and Education: Artificial Intelligence*, Vol. 7, 2024, Art. No. 100315. <https://doi.org/10.1016/j.caeai.2024.100315>
- [19] A. D. Anders and D. E. Speltz, “Developing generative AI literacies through self-regulated learning: A human-centered approach”, *Computers and Education: Artificial Intelligence*, Vol. 9, 2025, Art. No. 100482. <https://doi.org/10.1016/j.caeai.2025.100482>
- [20] G. He, “Predicting learner autonomy through AI-supported self-regulated learning: A social cognitive theory approach”, *Learning and Motivation*, Vol. 92, 2025, Art. No. 102195. <https://doi.org/10.1016/j.lmot.2025.102195>
- [21] M. de la Puente, J. Torres, A. L. Blanco Troncoso, Y. Y. Hernández Meza and J. X. Marquez Carrascal, “Investigating the use of ChatGPT as a tool for enhancing critical thinking and argumentation skills in international relations debates among undergraduate students”, *Smart Learning Environments*, Vol. 11, No. 1, 2024, Art. No. 55. <https://doi.org/10.1186/s40561-024-00347-0>
- [22] A. Oates and D. Johnson, “ChatGPT in the classroom: Evaluating its role in fostering critical evaluation skills”, *International Journal of Artificial Intelligence in Education*, Vol. 35,

- 2025, pp. 1793–1824.
<https://doi.org/10.1007/s40593-024-00452-8>
- [23] R. Suriano, A. Plebe, A. Acciai and R. A. Fabio, “Student interaction with ChatGPT can promote complex critical thinking skills”, *Learning and Instruction*, Vol. 95, 2025, Art. No. 102011.
<https://doi.org/10.1016/j.learninstruc.2024.102011>
- [24] P. Wang, K. Yin, M. Zhang, Y. Zheng, T. Zhang, Y. Kang and X. Feng, “The effect of incorporating large language models into teaching on critical thinking disposition: An “AI + constructivism learning theory” attempt”, *Education and Information Technologies*, Vol. 30, 2025, pp. 11625–11647.
<https://doi.org/10.1007/s10639-024-13244-3>
- [25] J.-M. Romero-Rodríguez, M.-S. Ramírez-Montoya, M. Buenestado-Fernández and F. Lara-Lara, “Use of ChatGPT at university as a tool for complex thinking: Students’ perceived usefulness”, *Journal of New Approaches in Educational Research*, Vol. 12, No. 2, 2023, pp. 323–339.
<https://doi.org/10.7821/naer.2023.7.1458>
- [26] R. Deng, M. Jiang, X. Yu, Y. Lu and S. Liu, “Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies”, *Computers & Education*, Vol. 227, 2025, Art. No. 105224.
<https://doi.org/10.1016/j.compedu.2024.105224>
- [27] J. Wang and W. Fan, “The effect of ChatGPT on students’ learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis”, *Humanities and Social Sciences Communications*, Vol. 12, 2025, Art. No. 621. <https://doi.org/10.1057/s41599-025-04787-y>
- [28] F. Mo, J. Huang, Y. Yang, Z. Özen, Y. Maeda and F. R. Olenchak, “Undergraduate students’ learning outcomes with ChatGPT: A meta-analytic study”, *Computers and Education: Artificial Intelligence*, Vol. 10, 2025, Art. No. 100536.
<https://doi.org/10.1016/j.caeai.2025.100536>
- [29] A. Strzelecki, “Students’ acceptance of ChatGPT in higher education: An extended unified theory of acceptance and use of technology”, *Innovative Higher Education*, Vol. 49, 2024, pp. 223–245.
<https://doi.org/10.1007/s10755-023-09686-1>
- [30] Oliinyk, I., O. Bulavina, T. Romanenko, A. Tatarnikova and A. Smirnov, “Artificial intelligence in developing doctoral students’ research competencies”, *Revista Eduweb*, Vol. 18, No. 3, 2024, pp. 294–305.
<https://doi.org/10.46502/issn.1856-7576/2024.18.03.22>
- [31] H. Nguyen, H.-L. Chen and K. Asfani, “Supporting self-regulated learning in the age of artificial intelligence: A comparative systematic review of earlier AI and generative AI”, *Journal of Educational Computing Research*, 2026, advance online publication.
<https://doi.org/10.1177/07356331261450647>
- [32] M. Vendrell and S.-K. Johnston, “Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education”, *Computers and Education: Artificial Intelligence*, Vol. 10, 2026, Art. No. 100572.
<https://doi.org/10.1016/j.caeai.2026.100572>
- [33] F. O. Alubthane, “Amplifier or substitute? A systematic review of generative AI’s impact on higher-order cognitive skills among university students”, *Frontiers in Psychology*, Vol. 17, 2026, Art. No. 1863931.
<https://doi.org/10.3389/fpsyg.2026.1863931>
- [34] M. Jou and C.-C. Kao, “Reframing human–AI collaboration in higher education: Behavioral pathways from generative AI to critical thinking and ethical creativity”, *Journal of Computing in Higher Education*, 2026, advance online publication. <https://doi.org/10.1007/s12528-026-09501-8>
- [35] OpenAI. *ChatGPT* [Large language model], n.d. <https://openai.com/chatgpt/>
- [36] Google. *Gemini* [Large language model], n.d. <https://gemini.google.com/>
- [37] Anthropic. *Claude* [Large language model], n.d. <https://www.anthropic.com/claude>
- [38] Microsoft. *Microsoft Copilot* [Artificial intelligence assistant], n.d. <https://copilot.microsoft.com/>