Human TECHNOLOGY

ISSN: 1795-6889

https://ht.csr-pub.eu Volume 21(3), December 2025, 620–639

THE IMPACT OF ARTIFICIAL INTELLIGENCE ON TASK PERFORMANCE AND DECISION-MAKING: EMPIRICAL EVIDENCE ON GENERATION Z

Adam P. Balcerzak
University of Warmia and Mazury in Olsztyn, Poland;
Brno University of Technology
Czechia
ORCID 0000-0003-0352-1373

Marek Zinecker
Brno University of Technology
Czechia;
Nicolaus Copernicus University in Torun
Poland
ORCID 0000-0003-1764-0904

Jiří Mičánek
Brno University of Technology
Czechia
ORCID 0009-0001-7237-9369

Abstract: *This study examines how generative artificial intelligence (AI) reshapes task performance, decision-making, and evaluative judgement in higher education assessments, with a focus on emerging human-AI assemblages among Generation Z university students. A controlled three-stage scenario-based experiment was conducted with the same cohort of students of business and economics, comparing a baseline session (no AI), independent reasoning (no AI), and identical AI-assisted conditions. Participants completed tasks involving situational judgment, quantitative reasoning, and short written responses. Results reveal that AI access increased average performance but markedly compressed score variance and reduced internal reliability, undermining the assessment's diagnostic capacity to differentiate independent abilities. Qualitative findings indicate that students perceived non-AI conditions as more cognitively effortful and educationally valuable, with AI shifting agency toward tool management and oversight. Together, these results highlight how AI redistributes agency in assessment, raising questions about responsibility and validity in sociotechnical contexts. Based on these insights, the study recommends hybrid assessment designs that separately evaluate independent reasoning and AI-augmented performance, incorporating reflective components to render distributed agency visible and preserve evaluative judgement.*

Keywords: *Artificial Intelligence; AI; Higher Education; Task Performance; Decision-Making; Experimental Study*.

DOI: https://doi.org/10.14254/1795-6889.2025.21-3.7

INTRODUCTION

Generative artificial intelligence (AI) is becoming an integral part of modern problem solving processes in academic and professional contexts (Kliestik et al., 2024; Andronie et al., 2024; Balcerzak & Valaskova, 2024; Lazaroiu & Rogalska, 2024; Moravec et al., 2024; Durica et al., 2023). In higher education, it challenges traditional assessment models that emphasise memorisation, recall, and individual performance (Barlybayev et al., 2025; Melisa et al., 2025; Ibieta et al., 2024; Kozova et al., 2024). Generative AI tools form human-AI assemblages in which cognitive agency is distributed between student and technology raising fundamental questions about responsibility, authorship, and what constitutes achievement in assessment (O'Dea, 2024; Southworth et al., 2023; Rensfeldt & Rahm, 2023; Durica et al., 2023; Zafari et al., 2022; Huang, 2021).

A growing body of research examines AI's transformative potential in education, highlighting benefits such as enhanced efficiency and personalisation alongside risks including overreliance, reduced critical engagement, and homogenisation of outputs (Gavira Durón & Jiménez-Preciado, 2025; Mishchuk et al., 2025; Wang et al., 2024; Chiu, 2024; Adiguzel et al., 2023; Wojciechowski et al., 2023; Chen et al., 2020). Particular attention has been paid to evaluative judgement – the learner's capacity to appraise the quality of work (Bearman et al., 2024) – which must now adapt to AI-rich environments where students evaluate AI outputs, reflect on AI-use processes, and risk undermining independent appraisal through uncritical dependence. Despite these conceptual advances, empirical evidence remains limited on AI's direct effects on assessment validity.

This study addresses this gap by quantifying a core tension: AI augments performance but simultaneously degrades the diagnostic power of assessments, shifting the focus of agency from individual cognition to sociotechnical assemblages. More specifically, we attempt to address these challenges and investigate how AI-assisted tools influence task performance and decision-making of a sample of educted generation Z representatives, namly speaking university students in economics-related problem-solving. We examine whether AI access improves their time efficiency and accuracy and how it affects cognitive engagement. The guiding intent is not to assess specific knowledge, but in particular to observe how participants reason, decide, and adapt to changing cognitive conditions.

A controlled three-stage experimental design was employed: a baseline diagnostic session (no AI), an independent reasoning session (no AI) and an identical AI-assisted session. This parallel comparison isolates the impact on task outcomes, decision-making quality, and perceptions of effort and learning.

The study is guided by the following research questions:

RQ1: *To what extent does access to generative AI affect task performance (accuracy and efficiency) in scenario-based economics assessments?*

The RQ1 draws on cognitive offloading and distributed cognition theories (Kirsh, 2013; Salomon et al., 1991). Research shows that external cognitive tools can reallocate mental resources, boosting efficiency but not necessarily deep understanding (Barr et al., 2015). In education, technology-enhanced problem-solving often improves surface performance while obscuring true competence (Baker & Siemens, 2014). Scenario-based assessments, grounded in situated cognition, highlight that performance depends on contextual supports rather than isolated skills (Mislevy et al., 2003).

RQ2: *How does AI access influence the psychometric properties (variance and internal reliability) of the assessment?*

This question draws on classical test theory and construct representation theory, which posit that score variance reflects differences in the latent construct rather than shared access to external resources (Embretson, 1983; Cronbach, 1951). Reliability and variance thus indicate an assessment's discriminatory power. Empirical studies on calculator use, open-book testing, and computer-assisted assessments show that external aids can compress score distributions, reduce item discrimination, and alter internal consistency by introducing construct-irrelevant or suppressing construct-relevant variance (Wainer & Thissen, 1993). These findings question whether AI access fundamentally changes score meaning.

RQ3: *What are students' perceptions of cognitive effort, learning value, and agency in AI-assisted versus independent conditions?*

This research question draws on Cognitive Load Theory (CLT), which distinguishes intrinsic, extraneous, and germane load and predicts that instructional supports may either aid learning or replace learning-relevant processes (Paas & van Merrienboer, 2020). Perceptions of learning value and agency are informed by self-regulated learning and self-determination theories, emphasizing autonomy and ownership (Deci & Ryan, 2000). Studies on automation bias and human-AI interaction suggest that reliance on intelligent systems can diminish perceived agency and critical engagement, especially when AI outputs are treated as authoritative (Dzindolet et al., 2003).

RQ4: *What implications arise for designing hybrid human–AI assessments that preserve evaluative judgement and diagnostic validity?*

This question draws on assessment validity theory, particularly Messick's unified framework, which views validity as an argument about score interpretation, use, and consequences (Messick, 1995), and argument-based approaches emphasizing evidentiary reasoning (Kane, 2013). Evidence-Centered Design (ECD) underscores modeling how tasks, tools, and scoring generate interpretable evidence of competence (Mislevy et al., 2003). Emerging AI-era scholarship advocates shifting focus from unaided performance to evaluative judgment, justification, and epistemic responsibility in AI-rich contexts (Luckin & Holmes, 2016).

The rest of this paper is structured as follows. The next section reviews the relevant literature on AI in education and assessment. This is followed by the research methodology, including the experimental design, data col-lection, and analytical approach. The subsequent sections present the re-sults, discuss their implications in relation to existing research, and con-clude with recommendations for integrating AI into higher education as-sessment.

LITERATURE REVIEW

The question of how best to integrate artificial intelligence into higher education has attracted increasing scholarly attention in recent years. Scholars have explored its applications in learning environments, the po-tential to personalise education, and the associated risks of over-reliance on technology (Gavira Durón & Jiménez-Preciado, 2025; Chiu, 2024; Adiguzel et al., 2023; Chen et al., 2020; Temerbulatova et al., 2025). AI is recognised not only as a technological innovation but also as a driver of pedagogical change, challenging traditional models based on memorisation and recall (Barlybayev et al., 2025; Melisa et al., 2025; Ibieta et al., 2024; Lazaroiu et al., 2024; Sarker, 2022). From a distributed cognition perspective, such technologies function as external cognitive resources that reshape how thinking and problem-solving are enacted, rather than simply supporting existing cognitive processes (Salomon et al., 1991; Kirsh, 2013).

Technological and social shifts pose new challenges for higher education, requiring students to master decision-making and problem-solving, as disciplinary knowledge alone is no longer sufficient (van Berkum et al., 2023). Traditional assessments often measure knowledge recall, yet research suggests that scenario-based and project-based approaches can better evaluate applied reasoning and strategic thinking (Gratchev, 2023; Ghosh et al., 2020; Koh, 2017; Clark, 2009; Haynes et al., 2009; Iverson & Colky, 2004). However, research on technology-enhanced problem solving cautions that improvements in observable performance may reflect surface-level efficiency gains rather than deeper conceptual understanding, particularly when external tools assume cognitively demanding processes (Barr et al., 2015; Baker & Siemens, 2014).

A growing body of scholarly literature underscores the transformative potential of education, examining its practical applications, emerging opportunities, and associated risks. Wang et al. (2024) conducted a bibliometric analysis of 2,223 research articles, complemented by a content review of 125 selected papers, which revealed a comprehensive conceptual framework within the existing literature. The findings highlight that, while AI tools in higher education offer notable benefits such as enhanced personalisation, efficiency, and feedback mechanisms, they also raise concerns regarding ethical implications, data privacy, algorithmic bias, and an overreliance on automation. Earlier scholarly contributions, including those by Bond et al. (2024), Castillo-Martínez et al. (2024), and Batista et al. (2024), also examine how artificial intelligence is reshaping higher education, with particular attention to its advantages and associated challenges. A key concern is the potential homogenisation of outputs and erosion of human judgement when AI is permitted, with scholars advocating blended approaches that preserve critical engagement while leveraging AI for efficiency (Lee et al., 2025; Frankford et al., 2024; Xia et al., 2024; Rensfeldt & Rahm, 2023).

A growing strand of scholarly inquiry advocates for increased emphasis on cultivating university students' evaluative judgement – that is, their capacity to appraise the quality of their own work and that of others – in response to the evolving educational landscape. Bearman et al. (2024) present a conceptual analysis of how evaluative judgement, a learner's ability to assess the quality of work, must evolve in response to generative AI. The authors identify three key intersections: evaluating AI-generated out-puts, assessing the processes of using AI, and using AI to appraise students' evaluative judgement. The paper highlights that while generative AI can support learning, it also risks undermining critical appraisal if students rely uncritically on its

outputs. To address this, the authors propose adapting established assessment strategies – such as self-assessment, peer review, feedback, rubrics, and exemplars – to foster deliberate and contextualised evaluative judgement. These arguments align with assessment validity frameworks that emphasise the interpretation, use, and consequences of scores, rather than performance alone (Messick, 1995; Kane, 2013).

Although the body of literature is expanding, there remains a paucity of empirical studies that directly compare student performance and per-ceptions under AI-assisted versus non-AI assessment conditions. Han et al. (2025) calls for further empirical research to gain insights in areas such as surveillance during data collection, algorithmic bias, standardisation, and uniform design, and excessive reliance on AI for support.

Elshall & Badir (2025) addresses the pedagogical challenge of integrating AI tools into university-level education while preserving rigorous academic standards. The empirical data were gained under involvement of upper-level undergraduate and graduate students from disciplines of civil and environmental engineering and Earth sciences. The FACT framework comprising four components was applied: (F) Fundamental skills assessed through non-AI coding assignments; (A) Applied projects using AI tools to simulate real-world tasks; (C) Conceptual understanding evaluated via traditional paper-based exams; and (T) Critical thinking measured through complex, multi-step case studies involving AI. Research findings indicates that both AI tools and guided AI support enhanced performance and ena-bled learners to engage with complex tasks and authentic applications more effectively than AI tools used in isolation. Survey data revealed that while many students regarded AI tools as advantageous for problem-solving, some voiced apprehensions about potential over-reliance. By incorporating both AI-supported and traditional assessment formats, the FACT framework fosters AI-assisted learning whilst upholding rigorous academic standards, thereby equipping students for professional practice in an increasingly AI-driven landscape. Furthermore, using a student-centred approach, Han et al. (2025) employed the Story Completion Method to explore perceptions of AI-driven educational tools. Qualitative analysis of 71 university students revealed that AI may influence learner autonomy, pedagogical roles, and educational relationships, with impacts often interlinked. The study introduces a novel speculative design ap-proach and offers design recommendations to support ethically responsive AI development. Lee et al. (2025) conducted a large-scale study involving 319 knowledge workers to examine how generative AI influences critical thinking in professional contexts. Drawing on 936 real-world task exam-ples, the authors found that higher confidence in AI correlates with re-duced critical engagement, while greater self-confidence promotes more critical thinking, albeit with increased perceived effort. The study revealed that AI shifts cognitive effort from execution to oversight, particularly in tasks involving analysis, synthesis, and evaluation. The survey design in-corporated quantitative regression models and qualitative coding to ex-plore motivators and inhibitors of critical thinking, highlighting the need for AI tools to support user awareness, motivation, and the ability in fos-tering reflective practices.

Despite progress, empirical research quantifying AI's impact on psychometric properties – such as variance and internal reliability – in assessment remains scarce. This study fills this gap with controlled evidence of AI's dual effects: augmenting performance while degrading diagnostic validity and redistributing agency in human-AI assemblages.

METHODS

This study employed a controlled, sequential experimental design within an International Economics course at Brno University of Technology (Spring 2025), involving the same cohort of students across three thematically aligned assessment sessions. Each of the three sessions introduced a distinct context developed on the basis of a text book *International Economics* (Krugman et al., 2015).

Session 1 (diagnostic baseline, no AI) used situational judgement tests (SJTs) requiring selection of best/worst responses in economic scenarios, establishing independent judgement reference (Bardach et al., 2021; Whetzel & McDaniel, 2009). In session 2, expanded assessment was conducted that included SJT tasks, evaluative reasoning, numerical problem-solving, fill-in-the-blank and a short essay. Again, the use of AI or other digital tools was not per-mitted. This session foregrounded independent cognitive effort, sustained attention, and unaided problem-solving. Session 3 was identical to Session 2 in structure, order of questions, and scoring. However, an AI-assisted framework was introduced, where students were permitted to use tools such as GPT models, AI-driven cal-culators, or language models to support their responses. While time con-straints remained, students were encouraged to critically think about when and how to rely on AI, reflecting the kind of decision-making in-creasingly expected in digital policy and consulting environments. For details, see Table 1.

Data was collected across the three sessions, including (a) quantitative performance metrics, (b) post-session qualitative reflections, and (c) comparative survey responses that evaluated students' perceptions of AI and non-AI versions. Scenario-based methodology was applied (Ghosh & Francia, 2021; Clark, 2009; Haynes et al., 2009; Iverson & Colky, 2004). Each session was time-limited (40 minutes for Session 1, 20 minutes for Sessions 2 and 3). Assessments were administered via the Moodle e-learning platform (Athaya et al., 2021), which also facilitated data collec-tion.

Participants were 31 students in Session 1, decreasing slightly to 28 in Session 3 (attrition typical for longitudinal designs). Participation was voluntary; data were anonymised and handled in compliance with GDPR and ethical protocols.

The parallel design of Sessions 2 and 3 (differing only in AI permission) enabled rigorous isolation of AI's effects on performance, reliability, and agency.

After each session, the students completed a post-assessment ques-tionnaire designed to capture (a) perceived clarity of instructions, (b) cognitive difficulty and time pressure, and (c) overall satisfaction with the quiz format. Furthermore, following Session 3, students completed an additional comparative reflection survey where they evaluated their experiences across both non-AI (Session 2) and AI-assisted (Session 3) conditions. These included: categorical responses (e.g., "Which version was more challenging?"), Likert-scale items (e.g., satisfaction, perceived fairness, learning value) and open-ended questions (e.g., "What advice would you give a future student using AI?").

Quantitative analysis included descriptive statistics and Cronbach's α for internal reliability; qualitative data underwent inductive thematic analysis (themes: cognitive load, learning value, AI strategies, agency).

Researchers used generative AI ethically for data organisation, summarisation, and writing support only - mirroring student AI access but not influencing participant responses.

Table 1. Structure of the Assessment Sessions [Source: own processing]

Session	Objective	Format	Conditions	Duration	No of Questions	Scoring
Session 1: Diagnostic assessment	Establish a baseline for students' judgment, decision-making, and economic knowledge	Situational judgment test (SJT), using a dual-response structure (choose the "Best" and "Worst" option)	No AI tools or digital assistance permitted	40 minutes	20	Each question scored up to 1 point (0.5 for each correct BEST/WORST pair), maximum 20 points
Session 2: Independent reasoning (No-AI support)	Assess performance under increased complexity and diverse question types, without external assistance	Included SJT-style items, "Best + Neither Best Nor Worst" formats, numerical calculations, fill-in-the-blank, and a short written essay	No AI, calculators, or digital tools permitted	20 minutes	15	Mixed scoring (0.5–1 point per item), maximum 15 points
Session 3: AI-Assisted assessment	Evaluate how students interact with complex economic problems when supported by AI tools, reflecting digitally augmented decision-making contexts	Identical in structure to Session 2—same task types, sequence, and scoring—allowing direct comparison	AI tools were permitted (e.g., GPT models, online calculators), but students were advised to manage time and think critically	20 minutes	15	Same as Session 2, maximum 15 points

RESULTS

Session 1: Diagnostic Assessment

Session 1 served as a diagnostic baseline to establish students' independent applied economic reasoning and decision-making performance under closed-book conditions, without AI or digital support. The assessment used a Situational Judgment Test (SJT) dual-response format (BEST and WORST option), intended to elicit evaluative judgement and prioritisation in realistic economic scenarios.

The diagnostic quiz was a closed-book, independent assessment with no AI or digital tools, consisting of 20 items completed within 40 minutes to simulate high-pressure conditions.

A total of 31 students participated in the Session 1 diagnostic assess-ment. For descriptive statistics, see Table 2.

Table 2. Diagnostic Assessment – Descriptive Statistics [Source: own processing]

Metric	Average score	Median score	Standard deviation	Skewness	Kurtosis	Internal consistency (α)	Standard error
Value	72.42%	72.50%	10.83%	-0.7866	1.1617	68.80%	6.05%

In terms of performance overview, 31 students participated in Session 1, achieving a mean score of 72.42% (median = 72.50%, SD = 10.83%). Scores clustered in the upper-middle range, with mild negative skewness (−0.79) and moderate kurtosis (1.16), indicating that most students performed well with relatively few low-scoring outliers. The majority of participants scored between 14 and 19 out of 20, suggesting strong engagement with the assessment logic and accessibility of the task despite the unfamiliar dual-response format.

Regarding the psychometric quality of the baseline instrument, internal consistency was α = 0.688, reflecting moderate reliability appropriate for a formative diagnostic assessment. This level of reliability supports the interpretability of students' independent judgement in the situational judgement test (SJT) format and provides a stable baseline for interpreting performance and measurement changes in later AI-assisted sessions.

With respect to student experience and baseline perceptions, post-assessment feedback was largely positive. Instructional clarity was rated as "very clear" by 77% of students, and most reported low time pressure, indicating that the 40-minute allocation was sufficient. More than 70% of students expressed satisfaction with the assessment. Qualitative comments described the task as realistic, cognitively engaging, and distinct from traditional quizzes, with students highlighting the value of the dual-response format in encouraging judgement from multiple perspectives.

In terms of alignment with the research questions, Session 1 establishes essential baseline evidence for RQ2 by documenting psychometric properties under no-AI conditions and contributes to RQ3 by demonstrating active evaluative judgement and learner agency in the absence of AI. While it does not directly address RQ1, it provides the necessary reference point for interpreting AI-related performance effects. Indirectly, the findings also inform RQ4 by supporting the use of scenario-based judgement tasks as a human-only component within hybrid assessment designs.

Session 2: Independent Reasoning (Non-AI Condition)

Session 2 extended the diagnostic baseline by increasing task diversity, cognitive demands, and time pressure under strictly non-AI conditions, thereby functioning as the human-only reference point for later AI comparisons. The assessment comprised 15 items across five task formats (Table 3), including scenario-based judgement, numerical reasoning, verbal logic, and a short written response, completed within a 20-minute time limit without access to AI tools, calculators, or digital resources. A total of 29 students completed the session.

Table 3. Assessment Format and Structure [Source: own processing]

Task type	Description	Quantity
Situational judgment (SJT)	Dual response: Best and worst option from four	5
Evaluative reasoning	Choose “Best” and “Neither Best nor Worst”	5
Quantitative Reasoning (ABCD)	Choose a correct numerical answer	2
Fill-in-the-Blank (Verbal Logic)	One question with three missing words	1
Short Essay	Write a response explaining a core economic concept	1
Conditions: No access to AI, calculators, search engines, or external support.		
Duration: 20 min.		
Scoring: Max score = 15 points		

Table 4. Session 2 – Descriptive Statistics [Source: own processing]

Metric	Average score	Median score	Standard deviation	Skewness	Kurtosis	Internal consistency (α)	Standard error
Value	56.44%	60.00%	22.64%	-0.1717	-0.7205	86.59%	8.29%

In terms of performance overview, the mean score was 56.44% (median = 60.00%), representing a substantial decline relative to Session 1 (72.42%). Score dispersion increased markedly (SD = 22.64%), indicating wide variation in students’ ability to manage the mixed-format, time-constrained assessment. Most students scored between 7 and 12 out of 15, fewer reached the upper performance range (13–15), and no participant achieved a perfect score, suggesting that the assessment was calibrated to stretch cognitive limits rather than reward surface familiarity. This pattern indicates that while students were able to engage with individual task types, maintaining accuracy and pace across heterogeneous demands proved challenging.

Regarding the psychometric quality of the baseline instrument, internal consistency was high (Cronbach’s $\alpha = 0.866$), despite the diversity of task formats. The combination of strong reliability and increased score variance suggests that the assessment coherently measured a single underlying construct - applied economic reasoning - while providing greater discriminatory power than Session 1. From a classical test theory perspective, the broader score distribution reflects meaningful differences in the latent construct rather than random error, establishing Session 2 as a psychometrically robust human-only benchmark.

With respect to student experience and baseline perceptions, post-session reflections indicated that Session 2 was perceived as substantially more demanding than the earlier diagnostic session. Students reported heightened time pressure, sustained concentration requirements, and difficulty switching rapidly between judgement-based, numerical, and verbal reasoning modes. Many described engaging in strategic triage, prioritising certain questions while responding heuristically to others. The short written response emerged as a key source of differentiation, with some students producing concise, well-structured arguments under pressure, while others struggled to articulate reasoning or left responses incomplete.

Numerical items also posed challenges, primarily due to calculation fluency and confidence under time constraints rather than conceptual misunderstanding. Despite these difficulties, many students characterised the experience as realistic and educationally valuable, noting increased awareness of their own reasoning strategies, time management, and cognitive limits – evidence of strong learner agency under independent conditions.

In terms of alignment with the research questions, the results demonstrate that, under non-AI conditions, the assessment exhibits both high internal reliability and substantial variance, supporting its diagnostic validity (RQ2). It also contributes to RQ3 by documenting high perceived cognitive effort, strategic self-regulation, and ownership of decision-making in the absence of external cognitive support, consistent with predictions from Cognitive Load Theory and self-regulated learning frameworks. While Session 2 does not address RQ1 directly, it establishes the essential control condition against which changes in performance, efficiency, and strategic behaviour observed under AI access can be meaningfully interpreted. Indirectly, the coherence between task demands, performance patterns, and student perceptions also informs RQ4 by illustrating how scenario-based, judgement-oriented assessments can preserve evaluative reasoning and diagnostic value in human-only conditions, providing a principled foundation for hybrid human–AI assessment design.

Session 3: AI-Assisted Assessment

Session 3 introduced AI access into the assessment environment while maintaining identical task structure, sequencing, scoring, and time constraints as in Session 2. Students were permitted to use AI-based tools, including GPT models, calculators, and other digital assistants, to support their responses. The purpose of this session was to observe how access to AI alters performance outcomes, score distributions, and student perceptions when solving the same scenario-based economics tasks under digitally augmented conditions. This session provides direct empirical evidence for RQ1, RQ2, and RQ3. The assessment format and conditions are summarised in Table 5. A total of 28 students participated. Self-reports indicate that most students used AI selectively – most frequently for the short essay, conceptual clarification, and numerical verification – while patterns of use varied from confirmatory checking to partial or full substitution of reasoning.

Table 5. Assessment Format and Conditions [Source: own processing]

Task type	Identical to Session 2 (SJT, Best+Neither, numerical, fill-in-the-blank, essay)	
Number of Questions	15	
Maximum Score	15	
Duration	20 minutes	
AI Policy	Students were allowed to use AI to support their responses, but were reminded to think critically and manage time effectively	
Key Control Factors	**Session 2**	**Session 3**
Task types	✔	✔
Question order	✔	✔
Scoring method	✔	✔
AI permitted	✘	✔

Table 6 shows descriptive statistics.

Table 6. Session 3 – Descriptive Statistics [Source: own processing]

Metric	Average score	Median score	Standard deviation	Skewness	Kurtosis	Internal consistency (α)	Standard error
Value	86.90%	86.67%	7.14%	-0.2945	-0.7329	31.02%	5.93%

In terms of performance overview, descriptive statistics (Table 6) show a substantial increase in performance relative to the non-AI condition. The mean score rose to 86.90% (median = 86.67%), representing a gain of more than 30 percentage points compared to Session 2. At the same time, score dispersion decreased markedly (SD = 7.14%), indicating strong convergence toward the upper performance range. The distribution exhibited slight negative skewness and reduced kurtosis, reflecting a flattened, top-heavy score pattern under AI-assisted conditions. Although task completion time was not directly measured, efficiency gains can be inferred from the combination of higher accuracy under identical time constraints and student reports of faster response formulation and verification.

Regarding the psychometric quality of the baseline instrument, Session 3 showed pronounced changes relative to the human-only control condition. Internal consistency declined sharply, with Cronbach's α dropping to 0.31 compared to 0.87 in Session 2, alongside a substantial compression of score variance. These shifts indicate a loss of discriminatory power under AI access, suggesting that observed scores no longer consistently reflected differences in applied economic reasoning. Instead, performance appears to have been influenced by heterogeneous AI-use strategies, including tool management, verification behaviour, and output editing, introducing construct-irrelevant variance into the measurement process.

With respect to student experience and baseline perceptions, post-session reflections indicated significantly lower perceived cognitive effort in Session 3 than in the non-AI condition. When asked which version was more challenging, 20 of 25 respondents identified Session 2 as more difficult, indicating that AI access reduced perceived difficulty despite higher scores. Qualitative comments revealed a shift in perceived agency: while many students appreciated the increased confidence and support provided by AI, several described reduced engagement in independent reasoning and greater focus on supervising or managing the tool, particularly for the essay and numerical items. At the same time, students reported selective trust in AI outputs, expressing reluctance to rely on AI for judgement-based SJT items involving ethical or evaluative considerations.

In terms of alignment with the research questions, Session 3 provides direct evidence for RQ1 by demonstrating that AI access substantially increases observable task performance and perceived efficiency under identical assessment conditions. It addresses RQ2 by showing that AI access compresses score distributions and sharply reduces internal reliability, thereby diminishing diagnostic validity. It also informs RQ3 by documenting reduced perceived cognitive effort and a shift in learner agency from independent reasoning toward tool-mediated oversight. Overall, Session 3 illustrates how AI access transforms both the product and process of assessment, improving output while weakening the assessment's capacity to function as a reliable diagnostic measure of independent applied reasoning.

Comparative Reflections – AI vs. Non-AI

Following the completion of both Session 2 (independent reasoning) and Session 3 (AI-assisted problem solving), students participated in a comparative reflection exercise designed to capture perceived differences in cognitive effort, learning value, and agency across the two assessment conditions. This phase provides direct descriptive evidence addressing RQ3, complementing performance-based findings with student-reported experiences that are not directly observable through score data.

The reflection instrument combined multiple-choice comparisons, short semantic descriptors, and open-ended questions. A total of 25 students completed the post-Session 3 survey. Responses to the question "Which version of the quiz was more challenging?" are reported in Table 7

Table 7. Research results when asked "Which version of the quiz was more challenging?" [Source: own processing]

Perceived challenge	Number of students	In per cent
Non-AI version	20	80
AI-assisted version	2	8
Both equally	1	4
Not sure	2	8

The results indicate a strong consensus that the non-AI version (Session 2) was more challenging, with 80% of respondents identifying it as such. Only 8% perceived the AI-assisted version as more challenging. Students frequently attributed the greater difficulty of the non-AI condition to higher mental workload, sustained concentration, time pressure, and the need to generate answers without external scaffolding ("I had to rely entirely on myself"; "The non-AI one felt like problem-solving").

Inductive thematic analysis of the open-ended student reflections revealed five recurring themes relevant to perceived cognitive effort, learning value, and agency (RQ3). Theme 1 (effort versus outcome) captured a commonly reported trade-off between ease of performance and depth of engagement. Students described the AI-assisted condition as more comfortable and confidence-enhancing, yet also as requiring less effort and offering weaker cognitive challenge, with several noting diminished ownership of their answers despite higher scores. Theme 2 (strategic adaptation and AI use) reflected heterogeneous patterns of tool deployment: some students used AI selectively to restructure essays, verify calculations, or explore alternative formulations, while others relied more extensively on AI in response to time pressure, indicating meaningful variation in how AI was integrated into individual reasoning processes. Theme 3 (selective trust in AI outputs) highlighted task-dependent reliance, with students expressing greater willingness to use AI for calculations and factual clarification than for judgement-based or ethically sensitive SJT items, suggesting the use of task-specific heuristics rather than uniform dependence on automation. Theme 4 (perceived learning value) revealed that many students reported learning more during the non-AI assessment despite lower performance, associating independent reasoning with deeper engagement, improved

knowledge retrieval, and prioritisation under pressure. Finally, Theme 5 (metacognitive awareness) captured increased self-reflection on cognitive habits, including tendencies to overthink, underthink, or allocate time inefficiently, with several students explicitly contrasting their reasoning strategies across AI-assisted and non-AI conditions. A summary of student perceptions across both assessment conditions is presented in Table 8. Taken together, these themes indicate that AI assistance reduced perceived cognitive effort and altered students' sense of agency, whereas independent conditions promoted deeper engagement and metacognitive reflection, providing qualitative evidence directly addressing RQ3.

Table 8. Student Perceptions Across Sessions [Source: own processing]

Dimension	Non-AI (Session 2)	AI-Assisted (Session 3)
Cognitive effort	High	Low to moderate
Score performance	Lower average, higher range	Higher average, compressed
Authenticity	Strong (realism emphasized)	Mixed
Learning experience	Deep, effortful	Fast, surface-level
Tool use	N/A	Mixed: thoughtful to passive
Emotional impact	Stressful but rewarding	Comfortable but detached
Self-reflection	High	Moderate

DISCUSSION

This study examined how access to generative artificial intelligence reshapes task performance, assessment validity, and students' perceived cognitive effort and agency in scenario-based economics assessments. Using a controlled three-stage design comprising a diagnostic baseline, an independent reasoning condition, and an identical AI-assisted condition, the study provides a robust empirical basis for addressing RQ1-RQ4 and situating the findings within established theoretical frameworks.

With respect to RQ1, the results demonstrate that access to generative AI substantially improves observable task performance. Under identical task structures and time constraints, students achieved markedly higher scores when AI was permitted, indicating increased accuracy and inferred efficiency. These findings align with theories of cognitive offloading and distributed cognition, which posit that external tools reduce execution-related cognitive demands and enable faster problem resolution (Angammana & Jayawardena, 2022; Kirsh, 2013; Salomon et al., 1991). However, consistent with prior research, these performance gains should not be interpreted as evidence of deeper understanding (Barr et al., 2015; Baker & Siemens, 2014). Instead, they indicate that AI alters the cognitive conditions under which performance is produced. In scenario-based assessments grounded in situated cognition, performance is inherently dependent on contextual supports, and AI therefore transforms performance into an outcome of a human-AI assemblage rather than a measure of individual reasoning alone (Mislevy et al., 2003).

Findings related to RQ2 reveal a pronounced contrast between AI-assisted and non-AI conditions in terms of psychometric behavior. Under independent reasoning conditions, including both the diagnostic baseline and Session 2, the assessment exhibited meaningful score variance and strong internal reliability, indicating coherent measurement of applied economic reasoning. When AI access was introduced, however, score distributions compressed sharply and internal consistency declined substantially. From the perspective of classical test theory and construct representation theory, this pattern reflects a loss of discriminatory power and a disruption in construct measurement, such that the assessment no longer differentiated reliably between students with differing levels of independent reasoning ability (Embretson, 1983; Cronbach, 1951). These results are consistent with prior research on open-book and calculator-assisted assessments, which has shown that external aids can suppress construct-relevant variance and introduce construct-irrelevant influences (Wainer & Thissen, 1993). The present findings extend this literature by demonstrating that generative AI can fundamentally alter score meaning, thereby challenging the validity of traditional interpretations of assessment outcomes in AI-rich contexts.

With respect to RQ3, the results indicate substantial differences in students' subjective experience across assessment conditions. Independent, non-AI assessments were consistently perceived as more cognitively demanding and more educationally valuable, despite lower average scores, a pattern that was evident in the diagnostic baseline and intensified under the increased time pressure and task diversity of Session 2. This aligns with Cognitive Load Theory, as AI appears to reduce extraneous and execution-related cognitive load, enabling faster and more accurate responses (Paas & van Merrienboer, 2020). However, student reflections suggest that this reduction may also displace germane cognitive processing, particularly when AI outputs are accepted without critical evaluation. Perceptions of agency further reflect insights from self-determination and self-regulated learning theories: students reported stronger ownership over decisions and outcomes in non-AI conditions, whereas AI-assisted performance often shifted agency toward tool management and oversight (Deci & Ryan, 2000). At the same time, selective skepticism toward AI – especially in judgment-based or ethically sensitive tasks – indicates that students are not uniformly dependent on automation, but instead develop task-specific heuristics for trust, consistent with research on automation bias and human–automation interaction (Dzindolet et al., 2003).

Taken together, these findings provide a clear response to RQ4 concerning assessment design. Allowing AI into assessment environments without redesigning assessment logic risks undermining both evaluative judgment and diagnostic validity. From the perspective of Messick's unified validity framework and argument-based validity theory, validity depends on the interpretability and consequences of scores, and when AI mediates cognitive processes, correctness alone is insufficient evidence of competence (Messick, 1995; Kane, 2013). Drawing on Evidence-Centered Design principles, the results suggest that future assessments must explicitly model the role of tools in performance (Mislevy et al., 2003). Hybrid designs that distinguish between independent reasoning and AI-augmented performance, and that require students to justify, critique, and reflect on AI use, offer a viable pathway for preserving evaluative judgment in AI-rich contexts and align with emerging calls to foreground epistemic responsibility over unaided output (Luckin & Holmes, 2016).

CONCLUSIONS

This study examined the impact of access to AI-based models on university students' performance, decision-making quality, and learning perceptions within a scenario-based economic evaluation context. Students completed three sequential assessment sessions – baseline (no AI), independent reasoning (no AI), and AI-assisted – allowing for controlled comparison across conditions. The findings indicate that AI access substantially increased average performance (by 30.46 percentage points) and reduced variability in outcomes. While this performance amplification demonstrates AI's potential to enhance accuracy and efficiency, it was accompanied by a marked reduction in diagnostic value, evidenced by compressed score distributions and weakened internal reliability.

Student reflections reinforced this central trade-off. Most participants perceived the non-AI assessment as more challenging yet more educationally beneficial, citing higher cognitive effort, sustained concentration, time management demands, and deeper critical engagement. In contrast, AI assistance was valued for speed, convenience, and confidence support but was frequently described as reducing the need for deep reasoning. Notably, patterns of selective AI use – particularly reluctance to rely on AI for value-based or judgment-oriented SJT tasks – suggest emerging AI literacy among students, reflecting an ability to judge when AI support is appropriate and when human reasoning should take precedence.

Several limitations should be acknowledged. The study was conducted with a modest sample size within a single institutional context, and novelty effects may have influenced how students engaged with AI tools. Future research should replicate this design with larger and more diverse cohorts, incorporate inferential statistical analyses, and examine longitudinal effects of AI integration on reasoning skill development. Comparative studies across disciplines would also help determine whether the observed tension between performance gains and diagnostic validity is domain-specific or generalizable across educational contexts.

Overall, the results underscore the dual nature of AI-based models in higher education assessment: they function as powerful accelerators of performance while simultaneously posing challenges to assessment validity and interpretability. The key challenge for educators and policymakers is not whether AI should be used, but how assessment systems can be designed to harness AI's strengths while safeguarding the irreplaceable role of human reasoning, evaluative judgment, and accountability in academic and professional practice.

IMPLICATIONS FOR RESEARCH AND APPLICATION

From a research perspective, this study opens several important avenues for future inquiry. Replication studies with larger and more diverse samples are needed to assess the generalisability of the observed effects across disciplines, institutions, and cultural contexts. In particular, comparative research across fields with different epistemic traditions – such as economics, engineering, humanities, and professional education – could reveal whether the trade-off between performance amplification and diagnostic loss is domain-specific or universal.

Longitudinal research designs are also essential. While the present study captures immediate performance and perception effects, future research should examine how sustained

exposure to AI-assisted assessment influences the development of reasoning skills, metacognitive awareness, and evaluative judgement over time. Such studies could clarify whether reliance on AI stabilises, intensifies, or diminishes as students gain experience, and whether hybrid assessment models can mitigate potential long-term erosion of independent reasoning capacity.

Methodologically, further work is needed to develop and test alternative psychometric and validity frameworks capable of capturing competence in human-AI assemblages, rather than in isolated individuals. Traditional indicators such as variance and internal consistency may be insufficient in AI-mediated contexts. Future research could explore process-oriented measures, trace data, or justification-based scoring models that better reflect distributed agency and epistemic responsibility.

From a pedagogical perspective, the findings suggest that AI should not be treated as a simple add-on to existing assessment formats. Instead, educators are encouraged to adopt hybrid assessment models that intentionally combine human-only tasks, AI-permitted tasks, and reflective components. Human-only components remain essential for diagnosing independent reasoning, while AI-assisted components can reflect authentic professional practice. Reflective tasks – such as requiring students to explain when, why, and how AI was used, or to critique AI-generated outputs – can help make cognitive processes visible and preserve evaluative judgement.

In terms of practical application, higher education institutions should develop transparent and pedagogically grounded AI policies that go beyond prohibition or unrestricted allowance. Integrating critical AI literacy into curricula is essential. Students must be supported not only in learning how to use AI tools effectively, but also in understanding their cognitive trade-offs, limitations, and ethical implications. Explicitly acknowledging the role of AI in assessment design can help align evaluation practices with real-world professional environments while safeguarding the central role of human judgement, reflection, and responsibility in learning and evaluation.

REFERENCES

Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary educational technology*, *15*(3), ep429. https://doi.org/10.30935/cedtech/13152

Andronie, M., Blažek, R., Iatagan, M., Skypalova, R., Uță, C., Dijmărescu, A., Kovacova, M., Grecu, G., Pârvu, I., Strakova, J., Guni, C., Zabojnik, S., Chiru, C., Sedláčková, A. N., Novák, A., & Dijmărescu, I. (2024). Generative artificial intelligence algorithms in Internet of Things blockchain-based fintech management. *Oeconomia Copernicana*, *15*(4), 1349-1381. https://doi.org/10.24136/oc.3283

Angammana, J. S. K., & Jayawardena, M. (2022). Influence of artificial intelligence on warehouse performance: The case study of the Colombo area, Sri Lanka. *Journal of Sustainable Development of Transport and Logistics*, *7*(2), 80–110. https://doi.org/10.14254/jsdtl.2022.7-2.6

Athaya, H., Nadir, R. D. A., Indra Sensuse, D., Kautsarina, K., & Suryono, R. R. (2021, September). Moodle implementation for e-learning: A systematic review. In *Proceedings of the 6th International Conference on Sustainable Information Engineering and Technology* (pp. 106–112).

Baker, R. S. J. d., & Siemens, G. (2014). *Educational data mining and learning analytics*. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 253–274). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.015

Balcerzak, A. P., & Valaskova, K. (2024). Artificial intelligence: Financial management under pressure of transformative technology. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *19*(4), 1127-1137. https://doi.org/10.24136/eq.3394

Bardach, L., Rushby, J. V., Kim, L. E., & Klassen, R. M. (2021). Using video-and text-based situational judgement tests for teacher selection: a quasi-experiment exploring the relations between test format, subgroup differences, and applicant reactions. *European Journal of Work and Organizational Psychology*, *30*(2), 251–264. https://doi.org/10.1080/1359432X.2020.1736619

Barlybayev, A., Razakhova, B., Sharipbay, A., Nazyrova, A., Tursynova, N., Zulkhazhav, A., & Yelibayeva, G. (2025). Comparative analysis of grading models using fuzzy logic to enhance fairness and consistency in student performance evaluation. *Cogent education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2481008.

Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015). The brain in your pocket: Evidence that smartphones are used to supplant thinking. *Computers in Human Behavior*, *48*, 473–480. https://doi.org/10.1016/j.chb.2015.02.029

Batista, J., Mesquita, A., & Carnaz, G. (2024). Generative AI and Higher Education: Trends, Challenges, and Future Directions from a Systematic Literature Review. Information (Basel), *15*(11), 676. https://doi.org/10.3390/info15110676

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment and evaluation in higher education*, *49*(6), 893-905. https://doi.org/10.1080/02602938.2024.2335321

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (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*(1), 4-41. https://doi.org/10.1186/s41239-023-00436-z

Bonfield, C. A., Salter, M., Longmuir, A., Benson, M., & Adachi, C. (2020). Transformation or evolution?: Education 4.0, teaching and learning in the digital age. *Higher education pedagogies*, *5*(1), 223–246. https://doi.org/10.1080/23752696.2020.1816847

Castillo-Martínez, I. M., Flores-Bueno, D., Gómez-Puente, S. M., & Vite-León, V. O. (2024). AI in higher education: A systematic literature review. In Frontiers in Education (Vol. 9, p. 1391485). Frontiers Media SA. https://doi.org/10.3389/feduc.2024.1391485

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE access*, 8, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chiu, T. K. F. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and education. Artificial intelligence*, *6*, 100197. https://doi.org/10.1016/j.caeai.2023.100197

Clark, R. (2009). *Accelerating expertise with scenario-based learning. Learning Blueprint*. Merrifield, VA: American Society for Teaching and Development, 10.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), 297-334. https://doi.org/10.1007/BF02310555

Davis, K. A., Grote, D., Mahmoudi, H., Perry, L., Ghaffarzadegan, N., Grohs, J., Hosseinichimeh, N., Knight, D. B., & Triantis, K. (2023). Comparing Self-Report Assessments and Scenario-Based Assessments of Systems Thinking Competence. *Journal of science education and technology*, *32*(6), 793-813. https://doi.org/10.1007/s10956-023-10027-2

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, *11*(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01

Durica, M., Frnda, J., & Svabova, L. (2023). Artificial neural network and decision tree-based modelling of non-prosperity of companies. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *18*(4), 1105-1131. https://doi.org/10.24136/eq.2023.035

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International journal of human-computer studies*, *58*(6), 697-718. https://doi.org/10.1016/S1071-5819(03)00038-7

Elshall, A. S., & Badir, A. (2025, June). Balancing AI-assisted learning and traditional assessment: the FACT assessment in environmental data science education. In *Frontiers in Education* (Vol. 10, p. 1596462). Frontiers Media SA.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological bulletin*, *93*(1), 179-197.

Frankford, E., Sauerwein, C., Bassner, P., Krusche, S., & Breu, R. (2024, April). AI-tutoring in software engineering education. In *Proceedings of the 46th International Conference on Software Engineering: Software Engineering Education and Training* (pp. 309-319).

Gavira Durón, N., & Jiménez-Preciado, A. L. (2025). Exploring the role of AI in higher education: a natural language processing analysis of emerging trends and discourses. *TQM journal*. https://doi.org/10.1108/TQM-10-2024-0376.

Ghosh, S., Brooks, B., Ranmuthugala, D., & Bowles, M. (2020). Authentic Versus Traditional Assessment: An Empirical Study Investigating the Difference in Seafarer Students' Academic Achievement. *Journal of navigation*, *73*(4), 797-812. https://doi.org/10.1017/S0373463319000894

Ghosh, T., & Francia, G. (2021). Assessing Competencies Using Scenario-Based Learning in Cybersecurity. Journal of cybersecurity and privacy, 1(4), 539-552. https://doi.org/10.3390/jcp1040027

Gratchev, I. (2023). Replacing Exams with Project-Based Assessment: Analysis of Students' Performance and Experience. *Education sciences*, *13*(4), 408. https://doi.org/10.3390/educsci13040408

Han, B., Nawaz, S., Buchanan, G., & McKay, D. (2025). Students' perceptions: exploring the interplay of ethical and pedagogical impacts for adopting AI in higher education. *International Journal of Artificial Intelligence in Education*, 1-26. https://doi.org/10.1007/s40593-024-00456-4

Haynes, S. R., Spence, L., & Lenze, L. (2009, October). Scenario-based assessment of learning experiences. In *2009 39th IEEE Frontiers in Education Conference* (pp. 1–8). IEEE. https://doi.org/10.1109/FIE.2009.5350730

Huang, X. (2021). Aims for cultivating students' key competencies based on artificial intelligence education in China. *Education and information technologies*, *26*(5), 5127–5147. https://doi.org/10.1007/s10639-021-10530-2

Ibieta, A., Aguilera, S., Constanza-Medina, M., Ignacia-Sepúlveda, M., & Castellanos-Alvarenga, L. M. (2024, November). Academic Use of Artificial Intelligence Tools by University Students. In International Conference on New Media Pedagogy (pp. 343-353). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-95627-0_23

Iverson, K., & Colky, D. (2004). Scenario-based E-learning design. *Performance Improvement*, *43*(1), 16-22. https://doi.org/10.1002/pfi.4140430105

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of educational measurement*, *50*(1), 1-73. https://doi.org/10.1111/jedm.12000

Kirsh, D. (1995). The intelligent use of space. *Artificial intelligence*, *73*(1-2), 31-68. https://doi.org/10.1016/0004-3702(94)00017-U

Kliestik, T., Kral, P., Bugaj, M., & Durana , P. (2024). Generative artificial intelligence of things systems, multisensory immersive extended reality technologies, and algorithmic big data simulation and modelling tools in digital twin industrial metaverse. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *19*(2), 429-461. https://doi.org/10.24136/eq.3108

Kliestik, T., Nica, E., Durana, P., & Popescu, G. H. (2023). Artificial intelligence-based predictive maintenance, time-sensitive networking, and big data-driven algorithmic decision-making in the economics of Industrial Internet of Things. *Oeconomia Copernicana*, *14*(4), 1097-1138. https://doi.org/10.24136/oc.2023.033

Krugman, P. R., Obstfeld, M., & Melitz, M. J. (2015). *International economics: theory and policy* (Tenth edition, global edition). Pearson Education.

Koh, K. (2017). Authentic Assessment. Oxford Research Encyclopedia of Education. Retrieved 11 Aug. 2025, from https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-22

Kozová, K., Grenčíková, A., & Habánik, J. (2024). Building a sustainable future: Gender, education & workforce needs of Gen Z. *Economics and Sociology*, 17(2), 209-223. doi:10.14254/2071-789X.2024/17-2/10

Lazaroiu, G., Gedeon, T., Valaskova, K., Vrbka, J., Šuleř, P., Zvarikova, K., Kramarova, K., Rowland, Z., Stehel, V., Gajanova, L., Horák, J., Grupac, M., Caha, Z., Blazek, R., Kovalova, E., & Nagy, M. (2024). Cognitive digital twin-based Internet of Robotic Things, multi-sensory extended reality and simulation modeling technologies, and generative artificial intelligence and cyber–physical manufacturing systems in the immersive industrial metaverse. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *19*(3), 719-748. https://doi.org/10.24136/eq.3131

Lazaroiu, G., & Rogalska, E. (2024). Generative artificial intelligence marketing, algorithmic predictive modeling, and customer behavior analytics in the multisensory extended reality metaverse. *Oeconomia Copernicana*, *15*(3), 825-835. https://doi.org/10.24136/oc.3190

Lee, H. P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025, April). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1-22)

Luckin, R., & Holmes, W. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.

Melisa, R., Ashadi, A., Triastuti, A., Hidayati, S., Salido, A., Luansi Ero, P. E., Marlini, C., Zefrin, Z., & Al Fuad, Z. (2025). Critical Thinking in the Age of AI: A Systematic Review of AI's Effects on Higher Education. Education process: international journal, *14*(1). https://doi.org/10.22521/edupij.2025.14.31

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, *50*(9), 741-749. https://doi.org/10.1037/0003-066X.50.9.741

Mishchuk, H., Samoliuk, N., Krol, V., & Rącka, I. (2025). Digital competences of university graduates: Implications for career success. *Human Technology*, *21*(2), 274–292. https://doi.org/10.14254/1795-6889.2025.21-2.2

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, *2003*(1), i-29. https://doi.org/10.1002/j.2333-8504.2003.tb01908.x

Moravec, V., Hynek, N., Gavurova, B., & Kubak, M. (2024). Everyday artificial intelligence unveiled: Societal awareness of technological transformation. *Oeconomia Copernicana*, *15*(2), 367-406. https://doi.org/10.24136/oc.2961

O'Dea, X. (2024). Generative AI: is it a paradigm shift for higher education? Studies in higher education (Dorchester-on-Thames), *49*(5), 811-816. https://doi.org/10.1080/03075079.2024.2332944

Paas, F., & Van Merrienboer, J. J. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science*, *29*(4), 394-398. https://doi.org/10.1177/0963721420922183

Perkins, D. N. (1993). Person-plus: A distributed view of thinking and learning. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 88-110). Cambridge University Press.

Rensfeldt, A. B., & Rahm, L. (2023). Automating teacher work? A history of the politics of automation and artificial intelligence in education. *Postdigital Science and Education*, *5*(1), 25–43. https://doi.org/10.1007/s42438-022-00344-x

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational researcher*, *20*(3), 2-9. https://doi.org/10.3102/0013189X020003002

Sarker, P. K. (2022). Macroeconomic effects of artificial intelligence on emerging economies: Insights from Bangladesh. *Economics, Management and Sustainability*, *7*(1), 59–69. https://doi.org/10.14254/jems.2022.7-1.5

Southworth, J., Migliaccio, K., Glover, J., Glover, J. 'net, Reed, D., Mccarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. Computers and education. Artificial intelligence, 4, 100127. https://doi.org/10.1016/j.caeai.2023.100127

Temerbulatova, Zh., Zhidebekkyzy, A., Sagiyeva, R., & Ludwiczak, A. (2025). Artificial intelligence as a driver of innovation and patent activity: An empirical analysis of cross-country data. *Economics and Sociology*, 18(3), 184-201. doi:10.14254/2071-789X.2025/18-3/11

van Berkum, M., Diederen, J., Buijsse, C. A. P., Boom, R. M., & den Brok, P. J. (2023). Competencies in higher education: identifying and selecting important competencies based on graduates & professionals in food technology. European Journal of Engineering Education, *49*(3), 434–453. https://doi.org/10.1080/03043797.2023.2245768

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert systems with applications*, *252*, 124167. https://doi.org/10.1016/j.eswa.2024.124167

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, *19*(3), 188-202. https://doi.org/10.1016/j.hrmr.2009.03.007

Wojciechowski, A., & Korjonen-Kuusipuro, K. (2023). How artificial intelligence affects education?. *Human Technology*, *19*(3), 302–306. https://doi.org/10.14254/1795-6889.2023.19-3.0

Xia, Q., Weng, X., Ouyang, F. *et al.* A scoping review on how generative artificial intelligence transforms assessment in higher education. *Int J Educ Technol High Educ* 21, 40 (2024). https://doi.org/10.1186/s41239-024-00468-z

Zafari, M., Bazargani, J. S., Sadeghi-niaraki, A., & Choi, S. -Mi. (2022). Artificial Intelligence Applications in K-12 Education: A Systematic Literature Review. *IEEE access*, 10, 61905–61921. https://doi.org/10.1109/ACCESS.2022.3179356.

Authors' Note

All correspondence should be addressed to

Adam P. Balcerzak
Faculty of Economic Sciences, University of Warmia and Mazury in Olsztyn, Poland
Brno University of Technology, Faculty of Business and Management, Czechia
pl. Cieszyński 1/327
10-720 Olsztyn
Email address:
adam.p.balcerzak@gmail.com

Human Technology
ISSN 1795-6889
https://ht.csr-pub.eu