

Authentic Assessment: How Instructors Perceive Authentic in the Assessment of Higher Education Institutes in the Era of AI

Yousef Almoslamani

Associate Professor, Instructional Technology department, Faculty of Education, University of Ha'il, Ha'il city , Saudi Arabia,
y.almshmeny@uoh.edu.sa,
ORCID ID / 0000-0003-3919-3666

Abstract: Generative artificial intelligence (GenAI) advancements have brought up chances and issues with authentic assessment in higher education. This research aimed to identify the instructors' perceptions concerning the impact of GenAI on students' examinations, assessments, and authenticity of assessment of students' learning. An exploratory descriptive quantitative methodology was employed, which involved an online questionnaire administered to 200 higher education instructors representing various academic positions, fields, and years of teaching experience. The questionnaire instrument was focused on two dimensions: perceived benefits and perceived risks of GenAI in authentic assessment. Data analysis included descriptive statistics, reliability test, independent-samples t-test, one-way ANOVA, and Pearson correlation analysis. The findings revealed quite high levels of both perceived benefits ($M = 3.66$, $SD = 0.57$) and risks ($M = 4.00$, $SD = 0.47$) related to GenAI. The most significant risks included difficulty of distinguishing the students' and GenAI generated work, plagiarism issues, possibilities for shortcutting and validity threats. On the other hand, the potential benefits of GenAI included its ability to help in matching assessment with industry standards, provide prompt feedback, decrease grading load and promote creativity. No statistically significant differences were found by gender, academic position, teaching experience, or academic discipline. Prior usage of GenAI had a significant impact on the perception of benefits, while the prior use of the AI detection software had an impact on the perception of risks.

Keywords: Authentic Assessment; Generative Artificial Intelligence; GenAI; Higher Education; Assessment Practices

1. Introduction

The paramount focus of learning environments or institutes is to ensure students' learning and gather authentic evidence about their learning. Authenticity in assessment refers to evaluating learners' performance directly on an intellectual task, addressing realistic situations, and contributing actual intellectual challenges by observing their abilities to explore, criticize, deploy theories, validate assumptions, and build knowledge. In the short term, to assess if learners acquire 'Know-how'. Ajjawi et al. defined authenticity in assessment as reality and genuine. Ajjawi depicts genuine as reproducing real-world tasks and situations, approximating realism, and ensuring that learners are capable of deploying their knowledge in practical scenarios. In contrast, McArthur advocates focusing on learners' relation to society for authenticity in assessment rather than realism, revealing the 'Why' of learners rather than the "How" and transforming society rather than replicating realistic scenarios. Nevertheless, the narrower definition of authenticity in assessment is realism reflected in project-based learning, capstone projects, case-based assessments, and portfolio assessments. The apparent value and significance of authentic assessment in higher education for enterprises are interpreted in three points: provide opportunities to deal with realistic experiences and challenges, foster motivation, critical thinking and problem-solving skills, prepare learners for employment, boost employability competence, minimize cheating, advance inclusion and prevent incursion.



There is recent research revealed the potential of generative artificial intelligence assisting in designing authentic assessments and generating exam questions in various subjects . Instructors utilize generative AI to augment authenticity in assessment design, enabling innovative and adaptive evolution content, and redefining assessment practices. Simultaneously, Artificial intelligence is recognized as one of the authentic assessment challenges in higher education.

Study Problem

The adoption of generative AI in the educational environment raises an issue in the authenticity of assessment in the era of online learning that relies heavily on e-assessment, online examination, academic research tasks, writing tasks, and home tasks. Recent studies such draw a new suspicion that GenAI has a disruptive potential in higher education practices, particularly in online learning environments. Moreover, the concept of adopting AI models and applications in higher education is still gray and not mature, and there is a need for more research and studies to underpin how GenAI affects authentic assessment practices in higher education. Put a question in this context, “How do we know that a student has learned?”; what evidence can be gathered on student learning? Thus, the study will address the following questions:

Q1: How does the artificial intelligence revolution (GenAI) affect the students’ examination?

Q2: How does the artificial intelligence revolution (GenAI) affect the assessment practices?

Q3: Could the instructor perceive authenticity in the assessment in terms of an artificial intelligence platform (GenAI)?

Study Objectives

The study aims to explore the perspective of instructors on how to perceive the authenticity of the assessment and how the assessment measures the authenticity of the students’ deliveries, online examinations, home tasks, and so forth.

2. Literature Review

Generative AI (GenAI), one of the Artificial intelligence models, autonomously generates a human-like text written, spoken voice, or image. It has significantly affected educational context, learning analytics, learning curriculum development, content creation, and generating personalized learning experiences. Recent studies speculate artificial intelligence opportunities (such as GenAI), can boost and enhance the learning experience, which starts from automating learning tasks, stimulating fellow co-learner, co-instructor, and co-examiner, to extensive learning context exceeding decision-making. Integrating AI platforms and applications into teaching and learning modules stimulates the attention of researchers and educators . Various research conducted experimental studies to address the benefits of these AI-based tools and how the interaction with these AI platforms can boost learner’s motivation, learner engagement, self-regulating learning, advance learner’s achievements, assist in solving educational practices, and various achievements in academic context.

However, most recent studies tackle the GenAI tool's capability to solve authentic assessment, for example, there is a study conducted by Choi et al. that addressed the capacity of ChatGPT, as one of GenAI models, solving low school examinations and found that this model passed four authentic examinations with an average grade of C+. In the Management and Business field, ChatGPT scored B to B- in open-ended questions . These results aligned with the findings of Farazouli et al. that ChatGPT scored more than 66% of authentic assessments in the fields of humanities, social science, and law . shared an experimental result that the GenAI tool surpassed human-learner in 10 out of 11 instances, providing 80% of the top-quality assessments across these classes. A technical report established by OpenAI- one of the pioneer companies in the GenAI field- the recent version of the ChatGPT model outperforms humans in various academic and professional exams. Consequently, there is a need to change educational practices, strategies, and designs.

Plagiarism tools and tests fail to detect or distinguish between human generated text or the human-like written response that GenAI models are able to generate, which induces a new challenge. Thus, some instructors and institutes are against these models' adoption either in learning or education. Parker et al. emphasized that graduate teaching assistants (GTA) cannot distinguish between human-like and human written submissions and they missed 50% of GenAI generated submissions of examination and home-based tasks. Hodges and Kirschner (2024) proposed different strategies in response to this challenge; emphasize process rather than product, use AI detection tools, change assessment design, focus on critical thinking and analysis, encourage application and creativity, conduct in-class assignments, consider peer review evaluation and collaborative work, frequent low-stakes assessments, reflective writing.

To sum up, the debates about GenAI is a double-edged sword, the golden solution to education challenges and problems or a disputative tool in learning and education, have not disputed yet .

3. Methodology

The current study underpinned an exploratory prospective method utilizing descriptive statistical frameworks through focusing on detailed analysis of the data collected. The descriptive approach aligns with current research nature that aims to explore and investigate the phenomenon deploying a focused observation and description of the patterns and insights convey in data collected.

The primary data collection tool in the current study is online survey. Survey is an objective tool systematically depicting the instructors' perspectives about the authentic assessment in the era of artificial intelligence across the higher education institution bodies from its instructors. The survey will inform all participants about the study aims and filling out the survey indicate their consent to participate in the study, also they are free to withdraw from the study at any time by just quite the form filling page. The development of the survey statement and question will be guided by an extensive review of the related accessible literature to ensure that it adequately measured the perception of the participants toward the effect of artificial intelligence revolution on the authenticity of the assessment and the assessment of authenticity in students' achievements.

The current study will be limited to addressing the generative artificial intelligence effects on the assessment by tackling other application and field of the artificial intelligence application.

Random sampling is the sampling approach in this study. The participant will be randomly selected from faculties members with different academic ranks and fields of specialty. The survey will be distributed online over the study sample in collaboration with contact people at each educational institute. The study will stop collecting responses once the response rate reaches more than 75% or higher.

Ensuring the ethical treatment of participants is paramount. This includes obtaining informed consent, ensuring confidentiality, and minimizing any potential harm to participants. The study will adhere to the ethical guidelines provided by Al-Hail University's Institutional Review Board (IRB).

4. Results

4.1 Data Coding and Analysis

The questionnaire items were coded on a five-point Likert scale:

1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, and 5 = Strongly agree.

Two composite dimensions were calculated:

- Perceived benefits of GenAI in authentic assessment: B1–B10.
- Perceived risks of GenAI in authentic assessment: R1–R10.

Higher scores indicate stronger perceptions of benefits or risks.

Table 1
Participant Characteristics

Variable	Category	n	%
Gender	Male	112	56.0
	Female	88	44.0
Academic rank	Lecturer	86	43.0
	Assistant Professor	62	31.0
	Associate Professor	35	17.5
	Professor	17	8.5
Years of experience	1–5 years	68	34.0
	6–10 years	62	31.0
	11–15 years	42	21.0
	More than 15 years	28	14.0
Academic discipline	Education	56	28.0
	Humanities	38	19.0
	Social Sciences	40	20.0
	STEM	38	19.0
	Health Sciences	28	14.0
Previous GenAI use	Yes	102	51.0
	Maybe/uncertain	48	24.0
	No	50	25.0
Previous use of AI-detection tools	Yes	78	39.0
	Maybe/uncertain	46	23.0
	No	76	38.0

Sample size comprised of 200 teachers. Out of them, 56% were male, while 44% were female respondents. Lecturers constituted the biggest portion by academic rank in this sample. The majority of them, that is to say more than half of all the respondents, had used AI technology for the purpose of assessment.

Table 2
Reliability Analysis

Dimension	Number of items	Cronbach's alpha	Interpretation
Perceived benefits of GenAI	10	.850	Good reliability
Perceived risks of GenAI	10	.800	Acceptable reliability

Cronbach's alpha values exceeded the commonly accepted minimum value of .70. This indicates that both dimensions demonstrated acceptable internal consistency and were suitable for further statistical analysis.

Table 3

Descriptive Statistics for Perceived Benefits of GenAI

Cod e	Statement	M	SD	Agree/Strongly agree (%)	Ran k
B8	GenAI supports alignment with industry standards	3.86	0.79	70.0	1
B4	AI-generated assessments provide immediate feedback	3.75	0.90	62.0	2
B10	GenAI simplifies assessment and reduces grading workload	3.73	0.82	64.0	3
B9	GenAI encourages creative and innovative engagement	3.73	0.87	60.0	4
B2	Customized assessments meet individual student needs	3.65	0.91	56.5	5
B5	GenAI enhances critical thinking through analysis and synthesis	3.63	0.86	57.5	6
B7	GenAI creates more interactive learning environments	3.60	0.89	54.0	7
B1	Scenario-based assessments reflect real-world challenges	3.58	0.82	51.5	8
B3	GenAI facilitates collaborative projects	3.54	0.87	51.0	9
B6	GenAI creates diverse assessment formats	3.53	0.94	51.5	10
	Overall perceived benefits	3.66	0.57		

Results indicated a fairly positive perception regarding the advantages of using GenAI in authentic assessment. The highest-ranked advantage was the capacity of GenAI to facilitate assessments based on industry standards ($M = 3.86$, $SD = 0.79$). The other perceived advantages include the capacity of GenAI to give instant feedback and lessen grading burden for instructors.

The lowest-ranked advantage related to the ability of GenAI to create diversified types of assessments ($M = 3.53$, $SD = 0.94$), although it was still ranked highly. All in all, the findings revealed that teachers perceive the educational potential of GenAI in assessment, feedback, and creativity.

Table 4

Descriptive Statistics for Perceived Risks of GenAI

Cod e	Statement	M	SD	Agree/Strongly agree (%)	Ran k
R3	AI responses blur the boundary between student and machine work	4.30	0.74	87.5	1
R4	Easy content creation encourages students to take shortcuts	4.21	0.70	86.0	2

R1	GenAI increases plagiarism and undermines authenticity	4.21	0.78	81.5	3
R6	GenAI raises concerns about assessment validity	4.11	0.79	76.5	4
R2	GenAI may reduce critical thinking and problem-solving	4.01	0.81	72.0	5
R7	AI-generated assessments may lack depth for higher-order thinking	3.92	0.85	72.0	6
R5	Unequal technology access may create assessment disparities	3.90	0.77	71.5	7
R8	Misuse may lead to declining academic standards	3.85	0.81	66.0	8
R10	GenAI may increase student disengagement and dependency	3.79	0.80	64.5	9
R9	Reliance on GenAI may reduce teachers' role in assessment design	3.73	0.88	61.5	10
	Overall perceived risks	4.00	0.47		

Risks were perceived to be more important than the potential benefits of GenAI. The risk perceived to be the most important one was the ambiguity created by GenAI with regard to the differentiation between original students' answers and machine-generated answers ($M = 4.30$, $SD = 0.74$). Shortcuts, plagiarism, and assessment validity concerns came next.

These results show that the instructors are especially worried about the accuracy of the work submitted by students as a reflection of their competence. The least important risk was the diminished role of the instructor in the assessment design process ($M = 3.73$, $SD = 0.88$).

Table 5

Overall Dimension Scores

Dimension	M	SD	Minimum	Maximum	Level
Perceived benefits of GenAI	3.66	0.57	2.00	4.90	High
Perceived risks of GenAI	4.00	0.47	2.80	4.90	High

The overall results demonstrate that instructors simultaneously recognize the educational benefits of GenAI and perceive substantial threats to authentic assessment. However, the risk dimension obtained a higher mean score than the benefits dimension. This suggests that concerns about validity, plagiarism, and authorship are stronger than confidence in the educational advantages of GenAI.

Table 6

Strategies for Protecting Assessment Credibility

Strategy	n	%	Rank
Focus on critical thinking and analysis	165	82.5	1
Emphasize the assessment process rather than the final product	160	80.0	2

Use frequent low-stakes assessments	152	76.0	3
Change assessment design	150	75.0	4
Use peer assessment and collaborative work	128	64.0	5
Use reflective writing	127	63.5	6
Encourage application and creativity	123	61.5	7
Use AI-detection tools	118	59.0	8
Conduct in-class assignments	115	57.5	9

The percentages do not add up to 100% because the participants had the option of selecting more than one technique.

The most popular choice was critical thinking and analysis, chosen by 82.5% of participants. The second one was the emphasis on the process over the outcome of the assessment and the use of frequent low-risk assessments. AI-detection technologies were chosen by 59% of the participants, suggesting that teachers valued assessment redesign and the process over the detection software.

Table 7

Gender Differences in Dimension Scores

Dimension	Male M	Female M	t	p	Result
Perceived benefits	3.65	3.67	-0.202	.840	Not significant
Perceived risks	3.99	4.02	-0.357	.721	Not significant

Independent-samples t-test results indicated that there were no statistically significant differences between males and females in terms of perceived benefits and risks. Consequently, gender did not seem to impact instructors' general perceptions about GenAI and authentic assessment.

Table 8

One-Way ANOVA Results According to Professional Characteristics

Grouping variable	Dimension	F	p	η^2	Result
Academic rank	Benefits	2.429	.067	.036	Not significant
	Risks	0.098	.961	.002	Not significant
Years of experience	Benefits	0.167	.919	.003	Not significant
	Risks	1.601	.190	.024	Not significant
Academic discipline	Benefits	0.717	.581	.015	Not significant
	Risks	0.325	.861	.007	Not significant
Previous GenAI use	Benefits	3.914	.022	.038	Significant
	Risks	0.740	.478	.007	Not significant

Previous AI-detection use	Benefits	0.297	.743	.003	Not significant
	Risks	4.390	.014	.043	Significant

The findings from the ANOVA test suggested that there were no significant differences by academic rank, years of experience, or academic discipline. Nonetheless, experience with GenAI had a significant association with the benefits. Furthermore, experience with AI detection tools had a significant association with the risks.

Table 9

Dimension Means According to Previous GenAI and AI-Detection Use

Variable	Category	n	Benefits M	Benefits SD	Risks M	Risks SD
Previous GenAI use	Yes	102	3.76	0.54	4.03	0.50
	Maybe/uncertain	48	3.60	0.56	3.93	0.50
	No	50	3.51	0.61	4.01	0.38
Previous AI-detection use	Yes	78	3.64	0.58	3.94	0.50
	Maybe/uncertain	46	3.72	0.54	4.18	0.43
	No	76	3.65	0.58	3.97	0.46

The respondents who had used the tool before were found to have higher perception about the positive aspects of the tool compared to those respondents who had not used GenAI before.

Those respondents who were unsure about their usage of AI-detection tools received the highest perception risk score. This result can be associated with their uncertainty about the accuracy and effectiveness of these tools.

Table 10

Correlation Between Perceived Benefits and Perceived Risks

Variables	Pearson's r	p	Result
Benefits and risks of GenAI	.019	.787	Not significant

However, there was no statistical relationship found between the benefits and risks perception. It implies that teachers can understand the pedagogical benefits of GenAI technology while still understanding the serious risks associated with the assessment's authenticity. Hence, the two perceptions seem to be independent dimensions rather than the opposite sides of one continuum.

4.2 Summary of the Results

Overall, there are relatively high perceived benefits and relatively high perceived risks of using GenAI in authentic assessment. The perceived risks outweigh the perceived benefits of GenAI in the domains of plagiarism, authorship, the validity of assessment, and students' dependence on the generated text.

There are no statistically significant differences by gender, academic rank, teaching experience, and academic discipline. There are significant differences by previous GenAI usage and previous experience with AI detection software. These results confirm the opinion that GenAI should not be incorporated into existing methods of assessment

as an add-on. Higher education institutions should restructure their assessments along with the principles of process, critical thinking, application, reflection, and evidence of learning.

5. Discussion

Results indicate that the teachers see GenAI as a tool which helps in authentic assessment while at the same time posing some difficulties. Perceived advantages have been identified as relatively high ($M = 3.66$), yet the perceived threats turned out to be even higher ($M = 4.00$). This confirms the idea of the double-edged nature of GenAI as a technology (Sok & Heng, 2023; Powell & Forsyth, 2024). It also confirms the separation of authentic assessment and actual student performance proposed by Ajjawi et al. (2024). The same principle can be found in the work by McArthur (2023): authenticity does not depend on how students perform the task but also why it matters socially.

5.1 *GenAI and Students' Examinations*

With respect to the first research question, instructors perceived GenAI to be a serious threat to the integrity of tests/examinations. The most highly ranked threat was that of AI-generated answers making it difficult to distinguish between students' work and machine-generated work ($M = 4.30$). There were also high ratings for concerns about plagiarism and taking shortcuts ($M = 4.21$).

This is in line with the findings by Choi et al. (2023), who showed that ChatGPT could pass multiple law exams, as well as with Terwiesch (2023), who noted that ChatGPT was relatively good at answering open-ended questions in management. Farazouli et al. (2024) similarly found that ChatGPT successfully completed over 66% of the authentic tests in humanities, social sciences, and law, while Parker et al. (2024) noted that it was hard to tell whether assignments have been written by humans or machine.

Thus, traditional online, take-home, and written exams will no longer serve as proof of students' learning. The point here is not in the use of the technology per se but in the answer that does not really reflect students' knowledge, reasoning, and individual work. Herein lies the importance of the recommendations given by Hodges and Kirschner (2024) to focus on the learning process, to redesign assessment and utilize in-class assignments.

5.2 *GenAI and Assessment Practices*

The second research question was about the impact of GenAI on the assessment process. According to the research, instructors appreciate such benefits of GenAI usage as alignment with the standards of the business world ($M=3.86$), provision of instant feedback, reduction of the grading workload, and stimulation of creativity.

These results correlate with the arguments of Salinas-Navarro et al. (2024) about the possibilities of GenAI usage for experiential learning and authentic assessment. The results also correlate with Artemova (2024) about motivation and engagement during learning that is supported by AI and with Fatima et al. (2024) about collaborative and student-centered implementation of AI.

Nevertheless, the high level of instructors' concern about the validity of assessment is evidence that they are not in favour of the blind implementation of GenAI into assessment practices. Concerns about ethical and academic issues connected with GenAI are also noted by Lowe (2024) and Al-kfairy et al. (2024) as well as about fairness, reliability, authorship, and misuse. As a result, the findings show the necessity of the assessment redesign rather than unconditional implementation or rejection of GenAI.

The preferred strategies for assessment included the focus on critical thinking, evaluation of the process, use of low stakes assessment frequently, and changes in the assessment design. They are consistent with the ideas of Chiang et al. (2024) about the balance of process and outcome and with the hybrid approaches to assessment in the digital age that were suggested by Perla and Vinci (2024).

5.3 *Instructors' Perception of Authenticity*

As the third research question, the study sought to find out whether instructors could detect authenticity in cases where GenAI is used. Results show that authenticity is possible, but it is much more challenging when the basis of assessment is the written output alone. High results in blurred boundaries between student's writing and AI-generated writing mean that instructors find it increasingly hard to know whether the assignment submitted is providing authentic evidence of learning.

Therefore, evidence of authenticity must come from many sources such as drafts, explanation, oral defense, application, peer reviews, observation of the learning process, etc. This is in line with the definition of authenticity by Ajjawi et al. (2024) and McArthur (2023) as engaging in the activity meaningfully rather than creating an answer that appears real.

While some respondents were in favor of AI detection, process-based and critical thinking techniques were favored. This underscores the difficulties associated with AI detection alone and aligns with the larger ethical considerations described by Lowe (2024) and Al-kfairy et al. (2024). As regards the above discussion, Ifenthaler and Schumacher (2023) highlight the need for balancing artificial and human intelligences.

5.4 Differences Among Instructors

Significant differences were not observed by gender, academic rank, years of experience, and academic discipline. Thus, the issues of concern with GenAI and authentic assessment are similar for all types of respondents. On the other hand, instructors having prior experience with GenAI stated their perception of more benefits probably due to direct usage, which made them understand the practical application of this technology (Artemova, 2024).

Respondents with prior experience in detecting AI reported their perception of more risk. Such data may be explained by high awareness of the limitations of such tools and difficulties with the fair decision-making when classifying academic material automatically.

Thus, the research demonstrates that GenAI can help increase personalization, feedback, creativity, relevance, and assessment effectiveness. Still, instructors showed high concerns related to plagiarism, using shortcuts, assessment validity, and distinguishing between student writing and machine-generated material.

As a result, the findings suggest shifting from the product-oriented assessment toward a balanced one combining process and outcome assessment. GenAI is supposed to be utilized to design assessments, and authenticity needs to be proven with the use of critical thinking, reflection, collaboration, practical application, and continuous evidence of student development. In this way, GenAI becomes useful in higher education without compromising the main idea of assessment – gaining credible evidence of student performance.

6. Conclusion

The research findings show that although there are many advantages of using Generative AI in the context of authentic assessment in higher education, there are also numerous dangers that cannot be ignored. For example, teachers were positive about its potential to provide quality feedback, foster creativity and individuality, align education and assessment with industries and make assessment more efficient. However, they had even more apprehensions concerning plagiarism, shortcuts, validity and difficulties in assessing originality.

From these results, it may be concluded that while authentic assessment can still be done, it should not be based only on final written assignments, exams and AI-detection tools. Assessment process should be concentrated more on learning process itself, reflection, critical thinking and application of knowledge.

7. Recommendations

1. Higher education institutions should rethink their approaches to assessment and make them more concentrated on analysis, application, and creativity.
2. Instructors should assess the process of learning through drafts, reflections, oral explanation, portfolios and practical tasks.

3. Institutions should conduct frequent low-stake and in-class assessments to reinforce evidence of learning.
4. GenAI should be used as an educational tool to support learning and not substitute instructor's judgement.
5. AI-detection tools should not be the only instrument for determining cases of academic misconduct.
6. Universities should develop policy clarifying appropriate and inappropriate use of GenAI during assessment process.
7. Instructors should be trained in ethical use of AI, assessment redesign and preserving assessment validity.

Time plan for research

Month	Week	Task
Month 4-5	1-2	Conduct literature review to refine research questions and methodology.
	3-4	Develop research instruments (e.g., survey questionnaire) and plan analysis methods.
	5-6	Obtain ethical approval from the Institutional Review Board (IRB).
	7-8	Finalize sample size calculations and recruitment strategy, Initiate recruitment of participants, Screen potential participants for eligibility and obtain informed consent.
Month 6-7	9-13	Distribute surveys or conduct interviews; monitor data collection process.
	13-16	Complete data collection and begin data cleaning and preparation for analysis.
Month 8-9	17-20	Conduct preliminary data analysis, perform final data analysis, including statistical testing and interpretation of results, and Draft results section of the research paper or report.
	21-24	Complete the introduction, literature review, and methodology sections; compile final draft.
Month 9-10	25-26	Submit the research paper to a journal for publication or present findings at conferences.

REFERENCES

1. Ajjawi, R., Tai, J., Dollinger, M., Dawson, P., Boud, d., & bearman, M. (2024). From authentic assessment to authenticity in assessment: broadening perspectives. *Assessment & Evaluation in Higher Education*, 49(4). <https://doi.org/10.1080/02602938.2023.2271193>
2. Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. *Informatics*, 11(13), 58. <https://doi.org/10.3390/informatics11030058>
3. Artemova, I. (2024). Bridging Motivation and AI in Education: An Activity Theory Perspective. *Digital Education Review*, 45, 59-69. <https://doi.org/10.1344/der.2024.45.59-67>

4. Chiang, Y. V., Chang, M., & Chen, N.-S. (2024). Can generative AI help realize the shift from an outcome-oriented to a process-outcome-balanced educational practice? *Educational Technology & Society*, 72(2), 347-385. [https://doi.org/10.30191/ETS.202404_27\(2\).TP04](https://doi.org/10.30191/ETS.202404_27(2).TP04)
5. Choi, J. H., Hickman, K. E., Monahan, A. B., & Schwarcz, D. B. (2023). ChatGPT Goes to Law School. *Journal of Legal Education*, 71, 1-16. <https://doi.org/10.2139/ssrn.4335905>
6. Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2024). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, 49, 363-375. <https://doi.org/10.1080/02602938.2023.2241676>
7. Fatima, S. S., Sheikh, N., & Osama, A. (2024). Authentic assessment in medical education: exploring AI integration and student-as-partners collaboration. *Postgraduate Medical Journal*, 100(1188), 783. <https://doi.org/10.1093/postmj/qgae088>
8. Hodges, C. B., & Kirschner, P. A. (2024). Innovation of Instructional Design and Assessment in the Age of Generative Artificial Intelligence. *TechTrends*, 68, 195-199.
9. Ifenthaler, D., & Schumacher, C. (2023). Reciprocal issues of artificial and human intelligence in education. *Journal of Research on Technology in Education*, 55(1), 1-6. <https://doi.org/10.1080/15391523.2022.2154511>
10. Lowe, M. (2024). The More Things Change: The Ethical Impacts of Artificial Intelligence in Higher Education. *Research Issues in Contemporary Education*, 9(2), 19-56.
11. McArthur, J. (2023). Rethinking Authentic Assessment: Work, Well-Being, and Society. *Higher Education*, 85, 85–101. <https://doi.org/10.1007/s10734-022-00822-y>.
12. Parker, L., Carter, C., Karakas, A., Loper, A. J., & Sokkar, A. (2024). Graduate instructors navigating the AI frontier: The role of ChatGPT in higher education. *Computers and Education Open*, 6. <https://doi.org/10.1016/j.caeo.2024.100166>
13. Perla, L., & Vinci, V. (2024). Rethinking assessment in the digital era: Designing a pilot study on hybridization in higher education. *Qwerty - Open and Interdisciplinary Journal of Technology, Culture and Education*, 19(1). <https://doi.org/10.30557/QW000079>
14. Powell, S., & Forsyth, R. A. (2024). Generative AI and the implications for authentic assessment. In S. Beckingham, J. Lawrence, S. Powell, & P. Hartley, *Using Generative AI Effectively in Higher Education* (p. 9). Routledge.
15. Salinas-Navarro, D. E., Vilalta-perdomo, E. L., Michel-Villarreal, R., & Montes, L. (2024). Using Generative Artificial Intelligence Tools to Explain and Enhance Experiential Learning for Authentic Assessment. *Education Sciences*, 14(1), 83. <https://doi.org/10.3390/educsci14010083>
16. Sok, S., & Heng, K. (2023). ChatGPT for education and research: a review of benefits and risks. *Cambodian Journal of Educational Research*, 3(1), 110-121.
17. Terwiesch, C. (2023, Jan 17). Would Chat GPT3 Get a Wharton MBA? A Prediction Based on Its Performance in the Operations Management Course. *White Paper*. Pennsylvania, USA: Mack Institute for Innovation Management.