

How Artificial Intelligence Is Changing Knowledge Assessment, and Why University Institutional Autonomy Matters?

Author: Tinatin Nikoleishvili¹

Editor: Irina Gurgenchashvili²

The debate over which method is best for assessing knowledge is older than universities themselves. In Plato's dialogue *Phaedrus*,³ Socrates recounts an Egyptian myth in which the god Theuth presented his inventions, among them writing, to King Thamus, claiming that it would make the Egyptians wiser and improve their memory. Thamus replied that writing would, in fact, be a remedy not for memory but for forgetting. In his view, without active learning, people would merely receive information and master knowledge only superficially.

According to Socrates, a written text cannot by itself prove that its author actually possesses the knowledge it conveys, since a text cannot answer questions, defend its position, or develop an argument. For centuries, this limitation was rarely seen as a practical problem, because producing a high-quality text independently, without the relevant knowledge and intellectual effort, was difficult. The development of generative artificial intelligence has changed that premise. When a quality text can be produced without real knowledge, written work can no longer serve the function it once did. Universities have thus faced the need to rethink how they assess student knowledge and to introduce formats that measure it for real.

From the Oral Exam to the Assessment of Written Work

In the early university tradition, the oral examination was considered an effective form of assessment, yet over time the oral component gradually shrank in syllabi and, in some cases, disappeared altogether. The problem with this form of assessment was not its effectiveness but its efficiency, as oral exams carried significant costs in time and organization. In medieval European universities, teaching and assessment were almost indistinguishable: students defended theses publicly, while

¹ Researcher, Research Institute Gnomon Wise; e-mail: t.nikoleishvili@ug.edu.ge

² Executive Director, Research Institute Gnomon Wise; e-mail: i.gurgenchashvili@ug.edu.ge

³ Plato, *Phaedrus*, trans. Benjamin Jowett (Project Gutenberg, 2008).
<https://www.gutenberg.org/files/1636/1636-h/1636-h.htm>.

professors used questions to probe both factual knowledge and the quality of reasoning. That tradition survives to this day in the name given to the thesis defense - viva voce ("with living voice").⁴

Things changed from the nineteenth and twentieth centuries onward, as the existing model could no longer meet the challenges posed by the Industrial Revolution and the expansion of access to higher education. Where it had once been possible to hold a lengthy discussion with each student, with hundreds and thousands of students the time and resources needed for individual dialogue no longer existed. It was in that period that the written examination took hold as an efficient form of assessment. This model made it possible to assess large numbers of students against a single standard, to compare results, and to run the process at optimal cost. Thus, owing to the mass character of education, oral assessment retreated.⁵

After November 30, 2022, when OpenAI released the first model of ChatGPT, producing written text with the help of language models became widely accessible. This has complicated both summative and formative assessment, since a final written product no longer provides sufficient information about its author's knowledge, and it is becoming difficult to determine how much of the intellectual contribution is actually the student's own. Naturally, this has pushed universities to reconsider their forms of assessment.⁶

What Do We Test When We Assess?

The search for new forms of assessment is preceded by one fundamental question: what should assessment measure - the knowledge a student has accumulated, or the ability to use that knowledge? In educational debates, the prevailing answer leans toward skills. In the digital age, the main task of higher education is increasingly seen as developing skills rather than transmitting knowledge. This vision is reflected both in educational frameworks built around "21st-century skills" and in university curricula, which attach ever less importance to memorizing factual knowledge.

⁴ Christopher Stray, "The Shift from Oral to Written Examination: Cambridge and Oxford 1700–1900," *Assessment in Education* 8, no. 1 (2001), <https://www.tandfonline.com/doi/abs/10.1080/09695940120033243>.

⁵ Ku-ming (Kevin) Chang, "From Oral Disputation to Written Text: The Transformation of the Dissertation in Early Modern Europe," *History of Universities* 19, no. 2 (2004), <https://academic.oup.com/book/48250/chapter/422693430> (open access: https://www.academia.edu/92626748/From_Oral_Disputation_to_Written_Text_The_Transformation_of_the_Dissertation_in_Early_Modern_Europe).

⁶ Tinatin Nikoleishvili, "Generative Artificial Intelligence in Teaching: Policies of the World's Leading Universities," Research Institute Gnomon Wise, 17.06.2026. Available at: <https://gnomonwise.org/en/publications/reports/325>

The rationale behind this approach is understandable. When any piece of information can be found in a matter of seconds, the question naturally arises: how justified is it to spend time and effort memorizing it? Generative artificial intelligence has sharpened the question further: when technology can retrieve, summarize, and process information, does accumulating knowledge matter less, and the ability to use it matter more? This approach often rests on the assumption that critical thinking is a transferable skill - one that can be developed in one context and then applied to any other content, much like the rules of formal logic. Cognitive science tells us that thinking does not work that way. Critical thinking is intertwined with knowledge of the domain in which it is exercised.⁷

To evaluate an argument, a person must notice what it lacks, what in it is questionable, and what it contradicts. Noticing this is possible only when there is prior knowledge to compare it against. Someone unfamiliar with history will not spot the weakness of a historical claim, and someone without medical knowledge will not catch an error in a medical record. In other words, knowledge is the material on which thinking works, and without the material, the skill itself cannot fully manifest.

This dependence has a simple cognitive explanation. Reasoning takes place in working memory, which can process only a few elements at a time. The main mechanism for overcoming this limitation is knowledge stored in memory. It lets us recognize large chunks of information as single, familiar patterns, freeing up working memory for analysis.⁸ This is the ability that separates an experienced reader from a child still sounding out letters: once decoding is automatic, attention is devoted entirely to the meaning of the text. The same logic applies to knowledge and thinking. A person who has to look up every fact in an external source spends cognitive resources on searching and understanding rather than on evaluating and drawing conclusions. Moreover, a person cannot look up what one does not suspect exists. Asking a question, raising a doubt, seeking a counterargument - each of these rests on knowledge already in place. Search rewards those who know what to ask.

Accordingly, the fact that information has become cheap and accessible does not mean that knowledge has become cheap as well. Information is stored in external sources, whereas organized, interconnected, internalized knowledge takes shape in a particular mind, at the cost of intellectual effort. This distinction essentially follows the same line Thamus drew between relying on external marks and possessing knowledge within. If judgment rests on internalized knowledge, then the task of assessment is precisely to verify what actually exists in the student's mind. This was one of the

⁷ Daniel T. Willingham, "Critical Thinking: Why Is It So Hard to Teach?," *American Educator* 31, no. 2 (2007): 8–19, https://www.aft.org/sites/default/files/periodicals/Crit_Thinking.pdf.

⁸ Paul A. Kirschner, John Sweller, and Richard E. Clark, "Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching," *Educational Psychologist* 41, no. 2 (2006): 75–86, https://doi.org/10.1207/s15326985ep4102_1.

purposes of written work - and it is exactly what the written product, by itself, can no longer demonstrate.

The Return of Oral Exams

As a result, if a written paper was once regarded as reliable proof of a student's competence, attention has now shifted to the student's ability to explain their reasoning, defend their arguments, and develop them in response to unexpected questions. Only internalized knowledge makes this possible, and it cannot be delegated to another source in real time. This reasoning has naturally brought oral examinations back onto the agenda.⁹ The London School of Economics (LSE), for example, is turning ever more often to oral assessment and in-class written exams in response to generative artificial intelligence, while Princeton University, which had administered exams without supervision for 133 years, has made proctoring mandatory again amid AI-related academic integrity violations.¹⁰

The active return of oral exams to syllabi is still hindered by the same circumstance that gradually drove it out from the nineteenth century onward - scale. One attempt to answer this problem is an experiment conducted at New York University (NYU),¹¹ in which a voice-based artificial intelligence gave each participating student an individual oral exam, with the answers then graded by large language models.¹² By the authors' calculations, the system's additional cost did not exceed 50 cents per student.¹³ The experiment came with its share of challenges: inconsistent questions, for instance, and very tight time limits for answering, which the researchers are working to resolve. It shows that artificial intelligence may become not only a challenge to assessment but also one of the instruments of its transformation, although the reliability of such systems is still open to question and requires further research.

⁹ Times Higher Education, "Are Universities Returning to In-Person Exams to Combat AI Cheating?," 2026. <https://www.timeshighereducation.com/depth/are-universities-returning-person-exams-combat-ai-cheating>.

¹⁰ The Daily Princetonian, "ADPOL: Proctoring In-Person Examinations Passed by Faculty After 133 Years' Precedent," May 2026, <https://www.dailyprincetonian.com/article/2026/05/princeton-news-adpol-proctoring-in-person-examinations-passed-faculty-133-years-precedent>.

¹¹ Panagiotis Ipeirotis and Konstantinos Rizakos, "Scalable and Personalized Oral Assessments Using Voice AI," (2026), <https://arxiv.org/abs/2603.18221>.

¹² Panagiotis Ipeirotis, "Fighting Fire with Fire: Scalable Oral Exams," Behind the Enemy Lines (blog), December 2025, <https://www.behind-the-enemy-lines.com/2025/12/fighting-fire-with-fire-scalable-oral.html>.

¹³ "NYU Professor Fights AI Cheating with AI-Powered Oral Exams That Cost 42 Cents per Student," The Decoder, January 4, 2026, <https://the-decoder.com/nyu-professor-fights-ai-cheating-with-ai-powered-oral-exams-that-cost-42-cents-per-student/>.

The Institutional Precondition: Autonomy as a Mechanism of Adaptation

The international experience discussed above reveals one common feature: in none of these cases was the adaptation of the assessment system carried out by central directive. The NYU experiment is the product of the research freedom of academic staff, while at LSE and Princeton assessment policy was revised independently, on the basis of deliberation within the academic community. At all three levels - individual, programmatic, and institutional - a key precondition for adaptation was academic freedom and institutional autonomy: the university's ability to decide for itself what it teaches, how it teaches, and how it assesses.

Georgia's higher education reform, launched in 2025,¹⁴ envisages changes in the opposite direction. The state itself will define universities' academic profiles, staffing policy, and student quotas,¹⁵ while the principle of "one city, one faculty" weakens competition between institutions - one of the main incentives for adaptation and experimentation. These changes substantially restrict the constitutionally guaranteed autonomy of universities: the very mechanism through which the institutions discussed above are responding to the challenge of artificial intelligence. The question of how knowledge should be assessed in the age of artificial intelligence confronts Georgian universities with the same urgency as it does Western institutions. The difference is that in Georgia's case, the institutional space for seeking the answer independently is narrowing considerably.

The attempts to bring oral assessment back to Western universities are not merely an echo of medieval practice; they are an element of a new educational environment in which historical experience, artificial intelligence, and modern principles of assessment converge. The central question universities are trying to answer today is how to build an assessment system that adapts to technological change yet still reliably evaluates a student's knowledge and independent reasoning, without sacrificing academic integrity and fairness.

International experience suggests that this problem is unlikely to be solved by one-off decisions. What is needed is a continuously updated policy grounded in research, in the evaluation of technological developments, and in deliberation within the academic community. Such a policy is hard to imagine

¹⁴ Tinatin Nikoleishvili, "The Higher Education Reform Concept: A (Counter)Reform in Education?," Research Institute Gnomon Wise, 31.10.2025. Available at: <https://gnomonwise.org/en/publications/analytics/292>

¹⁵ Tinatin Nikoleishvili, "What Guided Quota Allocation in Public Universities? - Evaluating Labor Market Research," Research Institute Gnomon Wise, 07.04.2026. Available at: <https://gnomonwise.org/en/publications/analytics/311>

without academic freedom and institutional autonomy - the very guarantees that Georgia's ongoing higher education reform puts at risk. Whether these preconditions are in place will largely determine whether the university can remain relevant in the age of artificial intelligence.