

The Assessment Vacuum: How the AI-First Campus Stopped Grading for Anything It Can Name

Weekly Analysis — <https://ainews.social>

There is a particular kind of confidence that arrives when an institution decides it is done deliberating. This week's higher-education discourse is saturated with it. Universities are not asking whether to embed generative AI; they are announcing that they already have, and the announcements arrive dressed in the language of the inevitable — transformation, readiness, the future of learning. What is strange, once you start reading closely, is how little of this confidence rests on any evidence that students are learning more, or better, or anything at all. The confidence rests instead on usage: on how many people have logged in. And usage, it turns out, is the one thing the modern campus can now measure with precision — because it has quietly stopped measuring the thing it was built to certify.

Consider the figures that anchor the week. Adoption is described as near-total: surveys report something on the order of 88 percent of students and 77 percent of teachers already using these tools, and the structured reviews of the field treat those numbers as the headline finding, the proof that the transformation is underway [11]. The Iberoamerican mapping of AI's arrival in higher education tells the same story, cataloging penetration across institutions as though penetration were the outcome of interest [16]. Notice what these numbers are and what they are not. They are self-reports of behavior. They are not measurements of learning. Nobody in this week's corpus has published a study showing that a cohort taught with these tools knows more, reasons better, or retains longer than a cohort taught without them. The adoption statistic has been promoted to stand in for the outcome statistic — and almost no one seems to notice the substitution.

This essay is about that substitution and its consequences. My claim is not the familiar one that universities are moving too fast, or that they need more teacher training and better infrastructure — the standard complaint, and a real one, but not the interesting one this week. My claim is that the instruments through which a university knows what its own students have learned — homework, examination,

[11] Generative AI in Higher Education: A Structured Integrative Review

[16] La llegada de la IA a la educación superior en Iberoamérica: Un mapa

criterion-referenced grading — are being dismantled at precisely the moment the tools that stress them are being embraced, and that the institution is filling the resulting void with numbers that describe adoption rather than achievement. The result is a legitimacy transfer hiding in plain sight. An institution that once certified graduates is learning to certify only that its graduates used the software. What it can no longer tell you — what it can no longer tell *itself* — is what they learned.

The Numbers That Aren't Outcomes

Begin with the epistemics of the week, because everything else follows from them. When a field wants to justify a costly and irreversible change, it reaches for the strongest available evidence. What is remarkable about the current higher-education discourse is that the strongest available evidence points the wrong way, and so the discourse reaches instead for the weakest: surveys of what people are doing and how they feel about it.

The structured reviews are candid about this if you read them carefully. The integrative literature on generative AI in higher education is largely a literature of perceptions, attitudes, and reported use — an aggregation of what students say they do and what faculty say they fear or hope [11]. The Spanish-language state-of-the-field report for 2025 follows the same contour, describing transformations in terms of institutional adoption, policy posture, and stakeholder sentiment rather than measured learning gains [14]. France's national report on AI in higher education is titled, tellingly, around *formation, structuration, appropriation* — training, structuring, appropriation — a vocabulary of institutional absorption, not of demonstrated pedagogical effect [10]. These are serious documents produced by serious people. But none of them can hand you the sentence a sculpture professor would actually want: *students who used this tool demonstrably learned X better than students who did not*.

Into this evidentiary gap flows a different kind of proof entirely — the proof of business optimism. The week's discourse leans on employer surveys and industry projections of skills demand, the Microsoft-style forecasts of a workforce remade, to supply the sense of urgency that the learning data cannot. This is worth naming precisely because it is so easy to miss: a claim about what employers *want* is being used to settle a question about what universities should *teach and test*. The gap between those two claims is the whole ballgame, and it is being papered over with confidence. Even the aspirational

[11] Generative AI in Higher Education: A Structured Integrative Review

[14] Inteligencia Artificial en la Educación Superior 2025: Transformaciones

[10] Formation, Structuration Et Appropriation Par La Société

literature admits, when pressed, that the promised gains remain projections. The genre of the AI-readiness report — vertically integrated design projects, interdisciplinary teams tackling industry-relevant problems, the whole vocabulary of the reinvented curriculum — reads as a description of a race to be won rather than a result to be verified [3]. The horizon is always 2030; the evidence is always forthcoming.

[3] After Shock

Here is the move to watch. When adoption is high and outcomes are unmeasured, the institution has two honest options: measure the outcomes, or slow the adoption until it can. It is choosing a third. It is declaring victory on the strength of the adoption figure and treating the absence of outcome data not as an alarm but as a convenience. The UNESCO competency framework for teachers is unusual in the corpus precisely because it refuses this convenience, insisting that most AI applications in education are properly considered high-risk and therefore require independent institutional mechanisms for validation before deployment [3]. That is the standard the discourse should be meeting. It is not meeting it. It is measuring logins.

[3] AI competency framework for teachers - UNESCO

The Flagship's Empty Center

If the aggregate discourse hides the assessment vacuum inside averages, the flagship cases expose it in detail. The most instructive material this week comes from the campuses that have gone furthest — the ones proudly rebuilding curricula around generative AI — because they show what happens to assessment when the tool it was designed to defend against becomes the medium of instruction itself.

The pattern is consistent and it is quietly astonishing. At the most visible AI-forward institutions, faculty describe abandoning the very instruments through which grading used to mean something. One professor reports grading no longer on correctness or demonstrated command of material but on "originality and maturity of ideas" — criteria that sound humane and are, on inspection, unfalsifiable. A colleague has stopped assigning homework altogether, on the reasonable but unexamined logic that homework is now indistinguishable from the tool's output. And in the same institutions, referrals for AI-related academic misconduct are climbing. Set those three facts beside one another and the shape of the problem appears: the rules for what counts as cheating are undefined, the criteria for what counts as good work have gone soft and subjective, and the enforcement apparatus is nonetheless spinning up. Undefined rules meeting unmeasured effects — that is the assessment vacuum in operation.

It helps to be precise about why "originality and maturity of ideas"

is not a rescue but a symptom. Broussard’s account of the epistemology of grading is useful here. She describes standing over a worksheet — a cow, an umbrella — and confronting the gap between the many defensible answers the world permits and the single answer that earns full credit [3]. Grading, at its most rigorous, is the institution committing to a criterion and defending it. When a professor retreats to “maturity of ideas,” the institution has not found a more sophisticated criterion; it has abandoned the discipline of having one it can name and check. This matters because a criterion you cannot name is a criterion you cannot audit, and a criterion you cannot audit cannot certify anything to anyone. The scholarship on assessment validity in the generative-AI era makes exactly this point: validity is not a vibe, it is the demonstrable relationship between what a task measures and what a grade claims, and unsupervised, tool-assisted work severs that relationship unless the assessment is deliberately rebuilt [6].

The homework decision deserves the same scrutiny. Abandoning homework because it can be automated is a genuinely defensible pedagogical response — but only if it is *replaced* by something that recovers the diagnostic function homework performed. In the cases the week documents, it is not replaced; it is simply subtracted, and the subtraction is described as innovation. The instrument that told a teacher whether a student could work through a problem alone is gone, and nothing has taken its diagnostic place. The teacher now knows less about the student than before, and calls this progress.

None of this is an argument for nostalgia about the blue-book examination, which had its own well-documented pathologies. It is an argument that when you retire an instrument of measurement you incur a debt: you owe the institution a replacement that measures something. The flagship campuses are retiring instruments and not paying the debt. Marc Watkins, who has watched this closely, warns from the opposite direction about the temptation to let the tool grade — to close the loop by having AI both produce and assess the work, which would complete the evacuation of human judgment from the center of the enterprise [22]. Between the professor who grades only for “maturity” and the fantasy of the machine that grades everything, the same thing disappears: a defensible, nameable account of what a grade certifies.

What the Effect Research Actually Says

The most damning feature of the week’s discourse is not that outcome data is absent. It is that effect data is present — and ignored. There

[3] Artificial Unintelligence - How Computers Misunderstand

[6] Assessment Validity in the Age of Generative AI

[22] The Dangers of using AI to Grade

exists a growing body of peer-reviewed work on what generative AI does to the cognition of the people who use it passively, and that work is being cited in the very reports that recommend adoption, and it is being operationalized in the design of exactly nothing.

The finding has a name: metacognitive laziness. Studies of cognitive offloading in the generative-AI era document a consistent pattern in which learners, offered a tool that will plan, structure, and evaluate for them, stop doing those things themselves — they cease to plan their approach, monitor their progress, and self-assess their output, delegating the executive functions of thinking to the system [18]. This is not a claim that AI makes people stupid. It is a more specific and more troubling claim: that the *mechanism* of learning — the effortful, self-directed regulation that turns exposure into competence — is precisely the process that a fluent assistant is best at short-circuiting. The American Psychological Association’s synthesis of how these tools reshape human skills points the same direction, toward a redistribution of cognitive labor in which the parts a student most needs to practice are the parts most readily surrendered [12].

Sit with the relationship between this finding and the flagship’s practices. The effect research says: passive use erodes the self-regulatory capacities that assessment is supposed to develop and verify. The flagship response says: let us reduce assessment, remove homework, and grade for “maturity.” These are not compatible. One is a warning that the tool attacks the process of learning; the other is a decision to dismantle the instruments that could detect whether learning survived. The stochastic, unreliable nature of generative output — the same input yielding different outputs, the content least trustworthy for exactly the factual and conceptual knowledge foundational courses are meant to build — is precisely why UNESCO flags educational deployment as high-risk [3]. The risk is legible in the research. The institution has simply declined to convert legibility into policy.

Why the decline? Part of the answer is that operationalizing the effect research would be expensive and disruptive in ways the adoption narrative cannot absorb. To take metacognitive laziness seriously, an institution would have to redesign assessment to *require* the very planning and self-evaluation the tool offers to remove — supervised reasoning, oral defense, process portfolios, in-class production under observation. That is difficult, labor-intensive work, and it cuts directly against the efficiency promise that justified adoption in the first place. It is far cheaper to cite the research in a footnote as evidence of institutional awareness and then proceed as though it did not exist. The AI-ethics literature’s aspiration — that doing this work well would require a genuinely interdisciplinary, pluralistic reformation of the

[18] Pereza metacognitiva y descarga cognitiva en la era de la IA generativa

[12] How AI is reshaping human skills and thinking

[3] AI competency framework for teachers - UNESCO

curriculum, an ethics woven into method rather than bolted on afterward — describes a road not taken [3]. The road actually taken runs through the survey dashboard.

The Detector Arms Race as Displacement Activity

Here is where the governance-versus-pedagogy imbalance becomes impossible to miss. Having lost the ability to measure learning, the institution has poured its energy not into rebuilding assessment but into policing the tool — and the policing is failing on every front the week documents, expensively and often unjustly.

Start with the detectors, because the money is real. Colleges have paid millions for AI-detection software whose accuracy does not survive contact with reality, producing false positives that fall hardest on the students least able to contest them [8]. The reliability problem is now serious enough that major institutions have formally repudiated the evidentiary value of these tools — Yale and Johns Hopkins among them, stripping detector output of any probative standing in misconduct proceedings [25]. The courts are arriving at the same place: the *Newby v. Adelphi* ruling marks the entry of the judiciary into a domain the universities failed to govern themselves, testing whether an accusation founded on detector output can survive due-process scrutiny [17]. When the courts have to define your academic-integrity standard because you could not, governance has already failed.

And the enforcement mechanism produces genuine casualties. In Toulouse, a student excluded for five years for alleged AI use during an exam won suspension of the sanction — a reminder that these determinations are frequently overturned when actually examined, which is to say frequently wrong [24]. Meanwhile the surveillance apparatus expands in the other direction, with AI monitoring of student devices generating false alarms serious enough that students have been called to offices and, in the school-age tier, even arrested over jokes flagged by algorithms [21]. The mechanics of that surveillance — the third-party monitoring embedded in institutional devices, the privacy it quietly liquidates — are now well documented [13].

The futility is total, and this is the part that should end the arms race but won't. Proctoring cannot hold the line either: AI-enabled glasses are already passing supervised, proctored examinations, and the essay documenting this is blunt that bans will not fix the problem because the problem is structural, not tactical [4]. Think about what this combination means. The institution is spending millions on detectors that don't work, generating due-process disasters that courts

[3] AI Ethics - The MIT Press Essential Knowledge series

[8] Colleges pay millions for AI detectors that are flawed

[25] Yale et Johns Hopkins retirent toute valeur de preuve aux détecteurs d'IA

[17] Newby v. Adelphi: First AI-Detection Court Ruling

[24] Triche avec l'IA lors de son partiel : une étudiante toulousaine exclue cinq ans

[21] Students have been called to the office — and even arrested — for AI surveillance false alarms

[13] How AI monitors school Chrome-books and what it means for privacy

[4] AI Glasses Are Quietly Passing Proctored Exams, and Bans Won't Fix It

have to clean up, surveilling students in ways that produce false arrests, and *still* cannot secure the exam. Every dollar and every hour spent on this displacement activity is a dollar and an hour not spent on the only thing that would actually help: redesigning assessment so that it measures something the tool cannot counterfeit. The clearest-eyed voices in the corpus say exactly this — that the answer runs through unambiguous, pedagogically grounded rules about what is and is not permitted, defined in advance and integrated into the design of the task, rather than through an ever-escalating enforcement war [2]. Some institutions are groping toward this — Lyon 2’s posture of permitting AI use conditionally, on the requirement that it be declared, at least replaces a phantom prohibition with a nameable rule [20]. But the dominant institutional investment remains in the machinery of catching rather than the labor of measuring. Governance has crowded pedagogy out.

[2] Academic Integrity in the Age of AI: Why Clear Proctoring Rules Matter

[20] Plagiat et IA à l’Université Lyon 2 : l’IA est permise si vous la déclarez

Two Traditions Running the Discourse

Step back far enough and you can see that the week’s discourse is powered by two old ideas wearing new clothes, and that both of them make the assessment vacuum worse rather than better. Naming them is the fastest way to understand why the smartest people in the field keep making the same move.

The first is what an older tradition of educational criticism called banking pedagogy — the conception of education as the transmission of deposits from a knowing source into a receiving student, with the student’s mind as the account and knowledge as the deposit. The generative-AI discourse is banking pedagogy’s dream and its reductio at once. When a university’s implementation literature describes AI as the engine of “personalized learning,” delivering the right content to the right student at the right moment, it is treating the tool as the transmitter and the student as the recipient — the banking model with a faster teller [19]. The trouble is that the effect research is a direct refutation of banking pedagogy’s core assumption. If learning were the reception of deposits, a more efficient depositor would be an unambiguous good. But learning is the effortful self-regulation that the metacognitive-laziness studies show the tool eroding [18]. A pedagogy that mistakes transmission for learning will happily adopt a transmitter that damages learning, because within its own frame the damage is invisible. That is the intellectual engine behind the flagship’s cheerful subtraction of homework.

[19] Personnaliser l’apprentissage pour les étudiants handicapés à l’aide de l’IA

[18] Pereza metacognitiva y descarga cognitiva en la era de la IA generativa

The second tradition is the logic by which the vendor becomes the

architect of the campus’s infrastructure — the arrangement in which the institution’s core knowledge processes come to run on platforms it does not own, cannot inspect, and does not govern. This is where the surveillance-capitalist reading earns its place, not as decoration but as description. The AAUP’s analysis of AI and academic freedom sees the stakes clearly: when instruction, assessment, and even faculty labor are mediated by proprietary systems, the terms of the academic enterprise migrate into contracts written by companies whose interests are not the university’s [5]. The dependency is not hypothetical. When a single upstream failure at a model provider can ripple through the institutions that built their workflows on it — the oversight gap that incidents at the platform layer expose — the university discovers that it has outsourced part of its operational nervous system to a vendor it cannot hold accountable [23].

Put the two traditions together and the assessment vacuum stops looking like an accident. Banking pedagogy supplies the justification for adopting a transmitter that erodes learning; vendor-as-architect supplies the infrastructure and the product pages that circulate as de facto curricula. The institution ends up embedding tools whose cognitive effects it refuses to measure, on platforms whose governance it does not control, and calls the whole arrangement a strategy. The vendor’s product page becomes the syllabus; the adoption metric becomes the outcome; and the question of what students learned falls through the gap between them.

The Partnership That Never Gets Framed

There is a conspicuous absence running through all of this, and it is the absence that would actually resolve the vacuum if anyone framed it. Nowhere in the week’s dominant discourse is the relationship between institution, faculty, student, and tool framed as a *partnership* — a shared, negotiated practice with reciprocal obligations and a common account of what counts as learning. The framing is always hierarchical and always adversarial: the institution deploys, the faculty comply or resist, the student is surveilled, the vendor supplies. Everyone is a party to be managed. No one is a partner in a shared inquiry.

The cost of that missing frame is visible in the global division of labor the technology is reshaping. For years, an offshore essay-writing economy — Kenyan contractors completing assignments for students in wealthier countries — supplied a hidden service to a system that pretended not to need it; then the generative tools arrived and col-

[5] Artificial Intelligence, Academic Freedom, and the Evolving Debate over Institutional Governance

[23] The Oversight Gap: What the OpenAI/Hugging Face incident means for universities

lapsed that market, absorbing into software what had been outsourced human labor [9]. That story is usually told as a curiosity. It is better understood as a diagnosis: the assignment that could be completed, invisibly, by a stranger on another continent was never measuring what the institution claimed it measured. The generative tool did not create the assessment vacuum. It revealed a vacuum that contract cheating had already proven was there, and it scaled the revelation to everyone at once. A partnership frame would treat this as the occasion to ask what an assessment is *for* — and to rebuild it around the things a partner in learning actually owes: demonstrated reasoning, defended judgment, work whose provenance the student can stand behind.

The materials that gesture toward partnership are precisely the ones the dominant discourse marginalizes. The equity literature argues for building an everyday, deliberate practice around these tools — a shared classroom culture negotiated with students rather than imposed on them — as the alternative to both naive adoption and futile prohibition [7]. Quebec’s ethics work on generative AI in higher education frames the questions in terms of collective deliberation over values and stakes rather than compliance with a deployment mandate [15]. And the emerging policy-ecosystem frameworks at least insist that adoption without a governance architecture that includes pedagogy — not just procurement and risk — is not a strategy but an abdication [1]. These voices exist. They are simply not the ones setting the adoption metrics, and the adoption metrics are what the flagship campuses report to their boards.

What partnership would require is uncomfortable, which is why it is avoided. It would require the institution to concede that it does not yet know what these tools do to learning, to say so publicly, and to run the assessment redesign as a genuine collaboration with the faculty who teach and the students who learn — with real measurement, published outcomes, and the willingness to slow adoption where the evidence turns bad. That is slower and less impressive than announcing an AI-first campus. It is also the only path that ends with the institution able to certify learning again rather than merely certifying use.

What the Institution Can No Longer Say

Return, at the end, to the substitution we started with, because its meaning has sharpened. When a university reports that 88 percent of its students and 77 percent of its teachers use generative AI and

[9] Durante años, los kenianos hicieron las tareas de estudiantes universitarios. Luego llegó la IA

[7] Building an everyday equitable practice for AI in schools

[15] Intelligence artificielle générative en enseignement supérieur : enjeux éthiques

[1] 2025 AI Education Policy & Practice Ecosystem Framework

offers that as evidence of successful transformation, it is telling you the one thing it can still measure and hoping you will not notice that it has stopped measuring the other. The adoption figure is real. The outcome figure does not exist. And the gap between them is not neutral — it is a transfer of legitimacy from the institution to the vendor whose product generated the adoption.

This is the quiet event of the week. The university’s historic function was to certify learning; to stand behind a credential as a warrant that a person had, in fact, come to know and to be able to do certain things. Every instrument the flagship campuses are now abandoning — the graded homework, the supervised examination, the criterion the professor could name and defend — existed to make that warrant true rather than merely asserted. As those instruments fall away and are replaced by “originality,” by “maturity of ideas,” by the soft language of unfalsifiable judgment, the warrant hollows out. The effect research tells the institution, in peer-reviewed detail, that passive use of these tools erodes the very self-regulation its credential is supposed to certify [12]. The high-risk designation tells it that educational deployment specifically requires independent validation before it can be trusted [3]. The institution has both documents on its desk and is measuring logins anyway.

An institution that replaces its examinations with adjectives has not modernized its assessment; it has abolished it and kept the ceremony. It can still hold graduation. It can still print the credential. What it can no longer do — for the employer, for the public, for the student, and most damningly for itself — is say what the credential certifies. It knows that its students used the software. Whether they learned anything the software did not do for them is a question it has, with a confidence that ought to frighten us, decided it would rather not be able to answer.

[12] How AI is reshaping human skills and thinking

[3] AI competency framework for teachers - UNESCO

References

1. 2025 AI Education Policy & Practice Ecosystem Framework
2. Academic Integrity in the Age of AI: Why Clear Proctoring Rules Matter
3. AI competency framework for teachers - UNESCO
4. AI Glasses Are Quietly Passing Proctored Exams, and Bans Won’t Fix It
5. Artificial Intelligence, Academic Freedom, and the Evolving Debate over Institutional Governance

6. Assessment Validity in the Age of Generative AI
7. Building an everyday equitable practice for AI in schools
8. Colleges pay millions for AI detectors that are flawed
9. Durante años, los kenianos hicieron las tareas de estudiantes universitarios. Luego llegó la IA
10. Formation, Structuration Et Appropriation Par La Société
11. Generative AI in Higher Education: A Structured Integrative Review
12. How AI is reshaping human skills and thinking
13. How AI monitors school Chromebooks and what it means for privacy
14. Inteligencia Artificial en la Educación Superior 2025: Transformaciones
15. Intelligence artificielle générative en enseignement supérieur : enjeux éthiques
16. La llegada de la IA a la educación superior en Iberoamérica: Un mapa
17. Newby v. Adelphi: First AI-Detection Court Ruling
18. Pereza metacognitiva y descarga cognitiva en la era de la IA generativa
19. Personnaliser l'apprentissage pour les étudiants handicapés à l'aide de l'IA
20. Plagiat et IA à l'Université Lyon 2 : l'IA est permise si vous la déclarez
21. Students have been called to the office — and even arrested — for AI surveillance false alarms
22. The Dangers of using AI to Grade
23. The Oversight Gap: What the OpenAI/Hugging Face incident means for universities
24. Triche avec l'IA lors de son partiel : une étudiante toulousaine exclue cinq ans
25. Yale et Johns Hopkins retirent toute valeur de preuve aux détecteurs d'IA