When the Shortcut Is Rational: Designing Incentive Structures for Deep Learning in the Age of AI

Shilin Zhang*, Yuman Zhang
School of Electrical Engineering, North China University of Technology, Beijing, China
**Corresponding Author*

Abstract: A professor at Brown University assigned a midterm homework and found the class average score near 96%. However, closed-book exam gave another story: 48.6% failed in the exam. The incentive mechanism that once students used to get academic score has failed, and LLM is the cause. When giving away cognitive stage costs almost minimal effort, working hard and getting best score fall into two different jobs, and the best student takes the short path. Based on Self-Determination Theory, Cognitive Load Theory, and Productive Failure, this paper puts forward a framework rebuild the incentive architecture of education. The first rule goes through AI-Off points, making hard work necessary. The second one resorts to process-oriented assessment to make hard work visible. The third makes struggle through open communication, and the fourth, multi-AI agents debate with each other, changes remarks from getting answers to judging them. What AI faces higher education with is a problem of motivation rather than a technological issue, and the task is straightforward: make the difficult path the one students must to face.

Keywords: Incentive Structure; Generative AI; Assessment Design; Productive Struggle; Self-Determination Theory; Cognitive Load.

1. Introduction

In the first term of 2026, a professor of economics and social science at Brown University in US, let his students to take the exam off campus. The average score of the students is at 96 almost. The professor compare the student's exam with ChatGPT and found evidence closely to match what his students had produced. He gave students a second chance to take exam, and held the final exam in person with no LLM allowed. The result is 48.6 on average, the lowest score in the course's all history. Eighteen students then quit the course, and 11 ended up coming without formally withdrawing [1].

The midterm and final exam compared the two score distributions with different story and made the event for education an alarming tale about honest. The two score distributions differ from each other for reasons that go out of the observation. They reveals what happens in 2026 when an incentive structure in education built up over the last decades fails when a new technology that paves a shortcut to get best performance.

Higher education is based on a formal logic: best effort leads to good grades, and those last scorers embodying a sign for hard word. AI broke the logic in both part. To get a grade no longer requires efforts for learning in the process. A good essay, a sound math proof, an elegant data analysis, all of which are now available at minor cognitive burden. The output looks appealing even without paying much attention to the learning process. And because the students records still direct people into graduate programs and job markets, the same process make sense whoever did less effort. The imminent learning stages of hard work is lost. The idea presentation, the knowledge searching practice, the meta-cognitive showcasing are all the indispensable learning processes.

At classes of the University of Pittsburgh, 82% students believed that LLM can hurt their own study, but they had to used it [2]. When asked what made them cross academic code lines, students did not ascribe to laziness but good marks is the first priority and attraction [3]. One student stated it directly: "If it's either I use LLM versus I fail exam, I'd rather use LLM" [4].

The result is clear and AI's challenge to higher education is about student motivation rather than technical application. What students aim to unchanged but the cost of achieving it without working hard is changed. AI Detection cannot remove that cost, and moral standards cannot either. Only a incentive structure is redesigned, and students walk through the difficult path, the productive struggle path, the retrieval practice, and deep processing is the correct path students should focus on.

The paper is structured as follows. Section 2 diagnoses the incentive paradox LLM has caused. Section 3 applies three theoretical lenses, Self-Determination Theory, Cognitive Load Theory, and Productive Failure, to the question of why students do not strive even when they know it is important. Section 4 sets out four learning principles for restructuring incentives, each with its boundary conditions. Section 5 focus on implementation at the institutional, classroom, and systemic levels, Section 6 draws the main limitations, and Section 7 concludes.

2. The Incentive Paradox

2.1 How Grades Became the Point

Class scores behave as a proxy of every day effort, a standard routine for measuring a student what they had really mastered. The arrangement hold true because their learning process required hard work. The learning endeavour earned a decent grade which is lined up meaningfully with the put into it is applied. However, LLM has broken that correlation and the learning outcome no longer calls for everyday effort which leads to the corresponding outcomes.

The problem is not simple as seems. Grades have become positional rewards which is related to entrance to graduated school, eligibility for research scholar, placement in job markets. Everything depends on where a student gets in the school learning results. In this system, where you stand is more important than what your knowledge is. This logic becomes a competing game. Once enough students adopt LLM, the one who refuse to use it falls behind. Students have to use it as the individually rational choice even for the ones who would prefer nobody used it. Cognitive science has a term for this phenomena. Risko and Gilbert [5] named it cognitive offloading, where people delegate tasks to external tools even while knowing the process weakens their own learning, as scores as present wins learning later.

2.2 When Grades Crowd Out Learning

Self-Determination Theory, proposed by Deci and Ryan [6,7], explains what reward systems do that once external incentives become the main reason for an activity, intrinsic motivation fails. Curious students shift toward grade maximization once the system explains what it actually pays for. LLM aggravate the degradation by making the learning-versus-performing trade visible. Before LLM is adopted, the two goals usually aligned, which means the same activities that working hard also produced good scores. LLM separates the above in two. A student can now achieve a high score with no measurable learning process, and the reward evaluation cannot distinguish the two objectives.

3. Three Theoretical Lenses on Effort and Its Avoidance

3.1 Self-Determination Theory: Competence, Autonomy, and the LLM Shortcut

SDT locates three basic psychological needs that sustain intrinsic motivation: autonomy, competence, and relatedness [6]. Hurt any of them and motivation changes from autonomous to controlled, depending on external pressure alone. Among the three, competence suffers most in LLM learning. In SDT, competence is not the sense of achievement but the experience that your own behave brought about the outcome. When LLM is used to produce an assignment, the outcome lands but the merit belongs to the machine. The process is competent but student does not feel competent. Repeat the pattern more and the sense of achievement and engagement alive becomes less.

Independent learning faces the same situation when students think that the system wants submission rather than deep understanding. Motivation changes from autonomous to be controlled by external forces, which depends on external contingencies and fails when those becomes inconsistent. The final exam case demonstrated precisely what this phenomena looks like. Students who had excelled on the off-campus tasks struggled in person where the contingency was the knowledge itself.

3.2 Cognitive Load Theory: When Efficiency Blocks Learning

Cognitive Load Theory, proposed by Sweller [8], separates the cognitive effort which builds schema from extraneous load, the effort that prohibit from learning. LLM's direct-answer fashion clears out cognitive effort by delivering final solutions. What looks like efficient work is just in fact the absence of direct learning. The student gets the answer without constructing the cognitive process that should repeat it

independently when facing the same questions.
The most relevant research for incentive design is present by Fraundorf, Kirk-Johnson, and Galla [9]. Their experiments found that students dislike effort learning even when those strategies can get better results. Effort itself is deemed as failure. That is why students may shy away from productive struggle. They just cannot distinguish the struggle that yields learning from the incompetence. LLM takes advantage on exactly this ambiguity. It just puts away the discomfort of struggle and the learning together.

3.3 Productive Failure: Why Struggle Is the Mechanism, Not the Obstacle

Kapur's Productive Failure theory [10,11] gives the reverse standpoint. Research shows that students who solve problems before receiving instruction end up ahead of others who get instruction first, especially on conceptual understanding. Sinha and Kapur's meta-analysis [12] confirmed that the effect persists across different contexts.
The issue is present as Vygotsky [13] located development in the space between what a learner can do alone and what the same learner can do with support. The space is precisely where productive struggle matters. What students can do with help today leads to they can do alone tomorrow. Inside the gap struggle learning persists.
Loibl and Rummel [14] name the mechanism failure awareness. Struggle with a problem first, and the questions you bring to instruction are specific and targeted. The instruction then addresses them directly. Ignore the struggle and the instruction stays generic. The instruction fits exactly on the gaps. The implication for incentive design is implicit. If struggle is where learning happens, assessment must make it a necessary step. Brown's take-home midterm made struggle skipped and the in-person final made it necessary. The distance between the two type exams is the distance between struggle learning and its absence.

4. Four Design Principles for Restructuring Incentives

Figure 1 gives the overview of the whole three theory. The left panel gives the broken structure, where generative LLM severs the shortcut between student effort and graded output, leaving students with unconfirmed knowledge states. The right panel shows the redesigned structure, in which three theoretical foundations (SDT, CLT, and PF) feed four design principles that realign effort, cognition, and assessment in the same logic.
None of the above principles stands alone. Each reinforces the other and comes with its own boundary conditions. They mark a starting point for institutional experimentation.

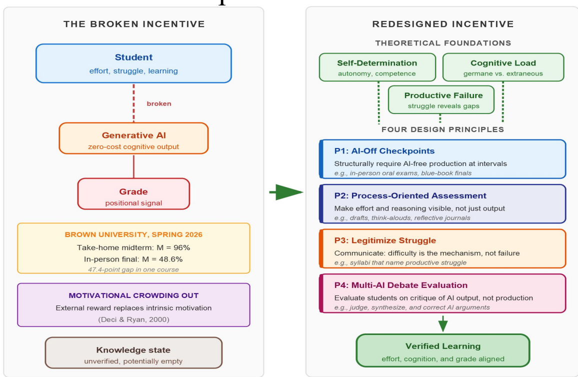

Figure 1. Overview of the Incentive Redesign Framework

4.1 Principle 1: Make Struggle Structurally Necessary Through AI-Off Checkpoints

The simplest approach is to require, at set points in the term, that all submitted task comes entirely without LLM involvement. The AI-Off checkpoint is not surveillance but a design feature, where a space students confront their own knowledge state directly.
Several law schools have already took steps. The University of Chicago Law School has banned laptops, tablets, and phones from first-year classrooms and added an in-person oral examination requirement for graduation [15]. UC Berkeley Law has barred LLM from conceptualizing, drafting, revising, translating, or editing any work submitted for assessment [15]. The dean of the University of Texas at Austin Law School urged faculty to make full use of classroom time to make students in sustained and rigorous communications that computers cannot mediate [15]. None of this is anti-technology. Chicago Law still permits AI for research, revision, and oral argument preparation. The principle is a structural separation as some learning activities need an LLM-free environment the way like that medical procedures need a sterile one.
One caution applies as students accustomed to LLM may experience AI-Off checkpoints and the design and evaluation cost depends on faculty. Schools adopting the requirement should make it with a clear rule, stated openly from the first day

of class.

4.2 Principle 2: Make Effort Visible through Process-Oriented Assessment

If the final outcome deserve a grade, the learning process lost and the youth have no incentive to put more effort to upgrade the learning. Process-oriented assessment can make the effort back into visibility. Drafts with tracked changes with different versions, think-aloud strategies and any intermediate submissions can make the process visible to lead to final scores.

The assessment must distinguish an exam result with struggle from one related to outsourcing. AIGC detection is unreliable at present [3]. The stable solution is to assess what detection cannot see, such as the reasoning trail, the homework iteration history, and the student's own oral explanation of how their answer was concluded.

Chambliss [16] demonstrated an oral examination format that captures precisely what the above strategy presents. In a five-minute structured oral exam, students processed 3 questions, a basic theory concept, and a critical thinking analysis. Outsourcing cannot survive under this situation, and the method allows it to scale to classes with more students involved. Pittsburgh students have required exactly this strategy, making teachers to bring back common exam styles because the temptation to use LLM just can not accept the others use AI [4]. They should need attention for teachers. Some students experience LLM adoption not as self-liberation tool but as a collective dilemma.

4.3 Principle 3: Legitimize Productive Struggle through Transparent Communication

The cheapest measurement redesign is also the most ignored method. Tell students the truth about learning effort behind the final exam score. Students who misunderstand productive struggle for a self incompetence [9] tend to avoid effort because they misread their own disability as proof that they are not fit to the subject. State to them struggle is the true learning mechanism, and students' misunderstanding of their own process can begin to dissolve.

That means changing the assignment instructions and feedback. A syllabus that says that, this course is designed to make you struggle, and the struggle is the mechanism, which sends a different message from one that simply lists office hours and late policies [4]. List competence as growth through effort rather than born ability, and students stay engaged when struggle arrives [6]. Students expecting struggle are also likely to read difficulty as a cue to grow.

4.4 Principle 4: Reorient Evaluation through Multi-AI Debate Tasks

Principle 4 is the hypothesis of the four, aimed at a structural flaw in current assessment. A single student answer, whether LLM-assisted or not, is evaluated in isolation and the evaluator never knows the thinking that produced it. Multi-AI debate tasks alter the situations. Several agent instances approach a problem from different perspectives, and the student's task is to evaluate, synthesize, and critique their arguments.

Mollick and Mollick [17] have shown that LLM can serve as debate partner, interviewer, or the student with a weak standpoint that needs improving. These cases change the usual relationship, that the student evaluates LLM output rather than consuming it. Judgment, comparison, synthesis, these are precisely the skills higher education must to teach students. Holmes, Bialik, and Fadel [18] argue along similar lines, LLM in education should augment rather than replace human cognition, and assessment should target exactly the lines that separate human from machine reasoning.

This principle is the steepest implementation difficulties. Good multi-LLM debate tasks require faculty expertise that is short right now, and the approach remains untested at scale. It still points worth attention, assessment that judges students on managing and critiquing LLM output rather than on producing output that LLM could have produced just as well.

5. Implementation

5.1 Institutional-Level Changes

Three structural changes needs attention. Every program should audit its assessment architecture for courses in which LLM can functionally replace learning. The Brown case provides the diagnostic, that a wide gap between take-home and in-person performance in the same course. The assessment has drifted away from learning. Programs should set a floor on the share of assessment completed without LLM access, just as a degree requires a certain number of in-person courses. Universities should fund teacher development for AI-resilient assessment design. Most teachers have received no training in making productive struggle visible and cannot be expected to possess that expertise independently.

5.2 Classroom-Level Changes

Instructors can behave right now. Count drafts, revisions, and reflective journals for at least 30% of the course, so process is assessed alongside product. Brief oral assessments is needed, even one five-minute session per student per semester, because the prospect of explaining reasoning aloud changes the incentive structure of home work. And explain the design out loud, so students can see why the course is built this way, what the research says about struggle and learning, and how productive failure is supposed to work.

5.3 Systemic Constraints

No classroom-level change fully solves the incentive problem, because the grade is not a classroom-level invention. Grades involves a larger system of graduate admissions, employer screening, and credential signaling. Until the whole system changes toward competency-based credentialing, the incentive to optimize grades rather than learning would become powerful. Competency-based models are spreading, and the growth of Western Governors University [19] shows real demand for alternatives to the mid-stage visible model, though they remain marginal. The design principles proposed here are no substitute for systemic reform. They are temporary methods, and to keep the system somewhat less misaligned while the larger structure evolves.

6. Limitations

This paper is just in theory and the proposals should require further empirical validation. The four principles hinge on existing case evidence that is more presume rather than conclusive. The AI-Off checkpoint principle, for instance, has been taken up by several universities but never evaluated thoroughly in a broader context. The experiments will not the same across many students for different majors. Those with strong self-regulation may favour AI-Off checkpoints while those with weaker learning abilities may experience difficulties to accept the system. The link between struggle and learning is not simple linear, and too much struggle may produce cognitive overload [8]; too little may produce shallow processing. Balancing the right level for each student requires finer-grained rethinking.

7. Conclusion

This paper shows when an incentive structure is disrupted, what chaos follows for students. When 82 percent of students think LLM will not do good to their study but still use it anyway. The problem is not related to student moral standards but is about whole education system design.

Making productive study become the standard and a must option, four states are put forward in this paper. Struggle must be structurally remain necessary. Effort and hard work must be visible in the full learning stages, and process-oriented assessment must be used. Difficulty should be make aloud rather than just a grading in the end. And assessment must change from just answers to evaluating the whole process, for instance the multi-AI debate tasks may help. Each change may get costs, and none is a full repair. If the teachers just classify students as good or bad in a course, AI is trivially achievable but a bit of meaningless. Deep learning is required and the incentive structure must make the effort link visible.

Acknowledgments

This work is supported by the Beijing Digital Education Research Project under No. BDEC2026619050.

References

[1] E. Whitford. Brown Professor Suspects Most of His Class Used AI to Cheat. Inside Higher Ed, 2026.
[2] University of Pittsburgh. SERU Survey on AI Use. Student Experience in the Research University Survey, Spring 2024.
[3] Student Voice and Inside Higher Ed. Survey: College Students' Views on AI. Inside Higher Ed, 2025.
[4] L. Schumann. That Chart from Brown Isn't Really About Cheating. Inside Higher Ed, 2026.
[5] E. F. Risko and S. J. Gilbert. Cognitive Offloading. Trends in Cognitive Sciences, 2016, 20(9): 676-688.
[6] E. L. Deci and R. M. Ryan. The 'What' and 'Why' of Goal Pursuits: Human Needs and the Self-Determination of Behavior. Psychological Inquiry, 2000, 11(4): 227-268.
[7] R. M. Ryan and E. L. Deci. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. American Psychologist, 2000, 55(1): 68-78.
[8] J. Sweller. Cognitive Load during Problem

Solving: Effects on Learning. Cognitive Science, 1988, 12(2): 257-285.
[9] S. H. Fraundorf, A. Kirk-Johnson, and B. M. Galla. The Role of Perceived Effort in Memory and Learning. Cognitive Psychology, 2019, 114: 101237.
[10] M. Kapur. Productive Failure. Cognitive Science, 2008, 32(6): 1008-1042.
[11] M. Kapur. Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success. Educational Psychologist, 2016, 51(2): 289-299.
[12] T. Sinha and M. Kapur. When Problem Solving Followed by Instruction Works: Evidence for Productive Failure. Review of Educational Research, 2021, 91(5): 669-698.
[13] L. S. Vygotsky. Mind in Society. Harvard University Press, 1978.
[14] K. Loibl and N. Rummel. Knowing What You Don't Know Makes Failure Productive. Learning and Instruction, 2014, 34: 38-49.
[15] K. Palmer. To AI-Proof Lawyers, Some Law Schools Restrict Technology. Inside Higher Ed, 2026.
[16] D. F. Chambliss. A Practical, Scalable Template for Oral Exams. Inside Higher Ed, 2026.
[17] E. R. Mollick and L. Mollick. Using AI to Implement Effective Teaching Strategies in Classrooms. SSRN Electronic Journal, 2023.
[18] W. Holmes, M. Bialik, and C. Fadel. Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign, 2019.
[19] Western Governors University. WGU President on the State of Online Education. Inside Higher Ed, 2026.