

What to Do About Grades? — Session Summary

Illinois CS Summer Teaching Workshop · Lunch discussion · June 11, 2026 · Zoom Facilitator: Jérémie Lumbroso (University of Pennsylvania) · lumbroso@seas.upenn.edu Prepared in collaboration with **Claude Fable 5** (Anthropic), who also collaborated on the session materials.

*This summary is released for the benefit of attendees under the **Chatham House Rule** announced at the session: Ideas are reported freely; no statement is attributed to any participant. If you would like attribution for something you said, email the facilitator with subject "Attribution request" and it will be added.*

Provenance markers: 🗨️ = said aloud during the session; 💬 = contributed in the Zoom chat. Resource links shared in the chat are listed as offered for circulation.

1 · What the room said a grade is for

Asked for one word, the room produced (🗨️ Slido word cloud + 💬 chat): **feedback · assessment (xmany) · evaluation · measurement · motivation · judgment · fairness · administration · competence · high stakes · certificate of completion · deadline.**

The facilitator's frame: A grade is **one number doing six jobs** (feedback, motivation, certification, sorting, accountability, protecting rigor) — and most grading debates are two people defending *different jobs*, each correct about their own. The session's method was to keep asking "which job are we talking about?"

2 · The doubts the room already carries

Invited to share doubts about grading anonymously, participants offered (🗨️ read aloud from anonymous Slido submissions, plus 💬 chat):

- Extended deadlines anecdotally led to **more** missed submissions; students treat deadlines as fixed and procrastinate past even generous tolerances. 🗨️
- Exams aren't a great reflection of knowledge — and take-home work "can no longer reliably be used as proof of learning." 🗨️
- The correlation between grades and actual ability is "potentially very weak, especially for those who grade-grub." 🗨️
- The thin line: Weights high enough to motivate completion, but not so high that students "stress and cheat to get the perfect score." 🗨️
- Grades demotivate as easily as they motivate. 🗨️
- "The grade is more aligned with effort than with any sense of skill or mastery — I know they worked hard; I don't know what they can do." 🗨️
- "My assessments aren't very effective and don't provide a good picture of the student" — a doubt the facilitator said had haunted him too. 🗨️
- "Grading is likely at odds with the ideal learning experience... it represents a constraint on the teaching side and an influence on the student side." 🗨️
- "I'm still a bit scared of ungrading, but a more **dialogical** approach to grading is what I need to try next." 🗨️
- "I don't have a lot of doubts, but I feel the constraints of fixed-length semesters." 💬
- "Is it so wrong that grade outcomes be aligned with effort (at least a bit)?" 💬 — with a counterpoint that some students would love *claimed* effort to be the larger share of the grade. 💬
- A new and demoralizing pattern reported: "It's not enough that I get an A — others should get a *lower* grade." 💬
- A natural experiment: A course that made its final optional (max of final and prior exam average) saw **pushback from students who felt betrayed that their study efforts were now**

"meaningless" — grade-investment behaving like a sunk-cost asset, not a learning instrument. 🗨️

- Rubrics cut both ways: For proofs and auto-graded work, "a rubric seems exclusively positive by making expectations clear." 🗨️

3 · What the facilitator reported trying (the "seeds")

- **The lamppost parable** as the session's frame: We measure where the light is, not where the keys are — and the story is about Goodhart's law, not about the drunk. 🗨️ (One participant: The lamppost comic is on their office door, referenced "almost every conversation." 🗨️ Another: "I never realized the guy was drunk in the story." 🗨️)
- **A Jevons paradox of grading:** Every accommodation (retakes, drops, extensions, regrade paths) made grade-points cheaper to pursue — and grade-seeking *increased*: More point-conversations, more "ChatGPT tells" (infinite-retake problem sets submitted with "insert your name here" still in them). 🗨️
- **Fall '25 (discrete math, career-pivot master's students):** flat A announced before the final; choice between a traditional exam and a *dialogic* exam (self-assessed inventory of every concept). About two-thirds chose the dialogic exam; a third still wanted to measure themselves against the traditional one. Honest cost: Course-evaluation complaints from students who had enrolled expecting a credential that would distinguish them, and for whom the rules changed mid-game. 🗨️
- **Spring '26 (graduate algorithms, ~110 students):** universal A in the syllabus from day one, with two tracks — one explicitly excused from attending, one committing to every session. ~40 students chose the showing-up track; per-lecture anonymous surveys drove weekly iteration; the facilitator reported working harder than ever (~20 hours/lecture) but having "real conversations... instead of being asked what the rubric was." The final was an optional evening **thinkathon** ("your A is secure"). 🗨️
- **Three theoretical claims** offered: (1) a grade is a *channel* carrying very few bits; (2) what a rubric can measure is the *script* — the sophisticated, checkable curriculum — so rubrics structurally cannot see original ("own-coded") work; (3) the *arbiter paradox*: Problem-posing education (Freire) makes students co-investigators who know things we don't, and a fixed rubric cannot grade what its author couldn't anticipate. 🗨️
- **Cybernetic vs. epistemic AI use** (the facilitator's framework): AI used to satisfy an external loop (grade, deadline, boss) vs. AI used to deepen one's own understanding. Neither is bad; the skill is knowing which one you're doing. Grading economies conscript students into the cybernetic mode. 🗨️ (Chat echo: "a constructivist lens is fairly at odds with the banking model... direct instruction is closer to providing 'knowledge nutrients' than shoving knowledge into brains." 🗨️)
- **Two portable mechanisms:** the thinkathon (a protected, ungraded, attention-undivided block — inspired by a hackathon's intrinsic energy); and the **metaphor exercise** (students generate and self-grade metaphors for an algorithm against a class-built checklist — the winning "Flagpole" metaphor for the majority-vote algorithm was honest even about the algorithm's failure mode, then used to teach peers). 🗨️

4 · The discussion: Where the room pushed back, agreed, and built

On rigor and standards ("is the flat A giving up?")

- One participant opened by doubting the words themselves: People talk as if clear, rigorous standards exist and the A reflects them — "I'm doubting that that's what it's all about." 🗨️
- Another recalled a professor (later named with appreciation in chat 🗨️) who gave **everyone an A in an extremely hard honors course** — every possible reading and exercise provided; trying was the condition; the reputation survived. 🗨️
- Skeptical reply: That works for honors students; "this would be a total disaster in any *mandatory* class," where some students "given the option would do none of the work." 🗨️

- Counter-reply: The mandatory class exists because the content matters; the students who don't care are *already* memorizing, forgetting, or cheating their way through under grading — "would they really do nothing... or would they actually get something out of it?" 🗨️
- A grade-appeal observation: At an institution whose culture is to appease appeals with a point or two, "if you gave an A, there would be no challenging anything. **I don't think it's giving up.**" 🗨️
- "+1 to doubting rigor and standards — we need to constantly re-evaluate who is defining them and who they're keeping out." 🗨️ A recent dissertation on the field's conflicting conceptions of "rigor" was shared: <https://doi.org/10.1145/3776542> 🗨️
- Differentiated rigor: With thousands of majors, "not every CS student needs the same level of rigor — you need different pathways" (research-bound vs. tools-competent), which is also a resource problem. 🗨️

On motivation without grades

- Grades function as deadlines and external scaffolding for students who are hard to motivate — remove them and what fills the vacuum? 🗨️
- One practice: Tie every homework explicitly to learning goals and career relevance instead of points — "I can't say it's very successful given the integrity issues we still have, but that's the hope." 🗨️
- The motivation-contagion worry (facilitator): Mixing students who want to be there with students conscripted by requirements may damage the classroom for both — "motivation is socially contagious." 🗨️
- **The cross-course externality (perhaps the sharpest objection of the hour):** "Even if students have intrinsic motivation for *my* course, given the pressures of being a student, a guaranteed A makes it *rational* to invest effort in other courses where the A is not guaranteed." 🗨️ Several endorsements; corroborated by credit/no-credit workshops being deprioritized even when they help with other classes. 🗨️ (The facilitator notes, post session: This is the strongest argument that single-course ungrading is a coordination problem — see §5.)
- Related skepticism: "I don't know if I agree that all students are in college to learn as much as to get a degree (both valid!)... many students give the assignment entirely to [AI] even when the autograder gives all the tests and instant feedback." 🗨️
- "The difference between prisoners and volunteers." 🗨️

On sorting, signaling, and the institutional mission

- The steelman for grades: Universities provide (1) structure and motivation — the "personal trainer" function — and (2) **signaling to employers**; students come precisely because faculty know the discipline ("I hire a plumber for the plumber's expertise") — "I love everything you're talking about, except for those two things." 🗨️
- "Signaling competency is one of the value propositions of universities." 🗨️ "Universities let businesses know that those students are vetted." 🗨️
- The facilitator's challenge: Curricula lag the world structurally — we measure yesterday's topics because those are the ones we've built assessments for; at the *average* institution (not the well-resourced one), is evaluation valuable when it evaluates the wrong thing? 🗨️
- Strong push-back: Curriculum design is fundamentals, not laziness — "assembly changes how you think about high-level code"; programming languages courses teach the structure of languages, not languages. 🗨️
- The productive synthesis several people then built: The real question is *what kind of rigor the assessment should target* — assess "can they write assembly" or "has assembly changed how they think"? One participant later conceded in chat: "Fair criticism. I assess can they write assembly, not whether it has changed their thinking." 🗨️ → "It is a means to an end and not the end, in thinking about the rigor of the assessment." 🗨️

- Institutional-level fixes exist: MIT's first-year P/NR policy (instituted at the **registrar** level, partly to reduce transition harm) — "a lot of the ideas we've discussed today fall on the shoulders of individual instructors" instead. ☹️
- The freedom-to-experiment constraint: Non-tenure-track and teaching faculty are "hindered by guarding against poor course evals" — radical grading experiments carry asymmetric career risk. ☹️
- Norm-referenced grading (curves) drew unanimous chat disapproval — "incentivizes cheating if they are forced to compete for grades." ☹️

Practices and resources shared for circulation ☹️

- A 0%-homework-weight syllabus design (homework feeds learning, not the grade; resubmissions guarantee project credit for engaged students):
<https://courses.cs.washington.edu/courses/cse373/26sp/#how-is-this-course-graded>
- A discrete-math course with zero homework weight (small weights only on prep problems and tutorial attendance, expected grade 100%).
- MIT first-year grading: <https://registrar.mit.edu/classes-grades-evaluations/grades/grading-policies/first-year-grading>
- Grading rubrics that include **penalties for regurgitation**.
- "Need for Cognition" and cognitive forcing functions vs. AI overreliance: Buçinca, Malaya & Gajos 2021, <https://doi.org/10.1145/3449287>
- A self-iteration practice: Leave notes-to-future-self in the slide deck immediately after each lecture, to be found while prepping next term.
- An observed AI "tell": Students teaching the question prompt back to the instructor without advancing the conversation. ☹️

5 • Threads left open (named, not resolved)

1. **The cross-course externality.** A single ungraded course inside a graded curriculum may rationally lose effort to graded competitors. Is ungrading an instructor-level or registrar-level intervention? (MIT's P/NR suggests the latter is possible.)
2. **Equity.** Who thrives and who drifts without grades — voted for as a track, barely discussed for time. ☹️ / 📝
3. **The mandatory-course problem.** The honors flat-A existence proof may not transfer to required courses; the counter-claim (the disengaged were already lost under grading, just invisibly) needs evidence.
4. **Differentiated rigor / pathways** — promising reframe, resource-hungry.
5. **Course evaluations as the instructor's own grade** — the Jevons dynamic applied to *us*: Those most exposed to evals can least afford to experiment.

Corrections welcome — at least one characterization of a colleague's grading model was flagged as inaccurate during the session and will be corrected in any future version of these materials.

Recorded locally with consent announced at the start; processed under the Chatham House Rule. Slides and a longer write-up: lumbroso@seas.upenn.edu · system3.blog (forthcoming).