Journal of Data Science 24 (1), 286–291
January 2026
DOI: 10.6339/26-JDS1208I
Education in Data Science

# Rejoinder: "Addressing the Challenges of AI-Generated Assignment Submissions in Education: Insights and Strategies"☆

Shuyi Wang[1,*], Longxin Xu[1], Junchen Liu[1], and Yujia Zhai[1]
[1]*School of Management, Tianjin Normal University, China*

**Keywords** *academic integrity; AI in education; assessment validity; Human-AI collaboration; pedagogical strategies*

We are deeply grateful to all the discussants for their thoughtful, constructive, and insightful comments on our paper (Wang et al., 2025). The diversity and depth of perspectives from Drs. Chan and Pua; Dr. Columbus; Dr. Furfaro; Drs. Huo and Ni; Dr. Lock; Dr. Loux; Dr. Nicosia; and Drs. Zaghi and Harel have significantly enriched the discourse on AI in education. Their contributions not only validate the timeliness of our work but also extend its scope in meaningful directions. We organize our response around five major themes that emerged from these discussions.

## 1 From Academic Integrity to Epistemic Validity

Dr. Columbus offers a reframing of the AI assessment challenge: generative AI has not fundamentally disrupted assessment so much as it has exposed long-standing weaknesses in how learning outcomes are operationalized. This diagnosis aligns with established validity theory in educational measurement: tasks can appear "valid" on the surface yet fail to support warranted inferences about the intended construct (Cronbach and Meehl, 1955; Messick, 1995; Kane, 2013; AERA, APA, & NCME, 2014). We wholeheartedly agree with this characterization. Her distinction between "surface validity" and "construct validity" illuminates why many traditional assignments—despite appearing to measure understanding—actually reward fluency and template conformity, precisely the features that AI excels at producing.

In the validity literature, "construct validity" is explicitly about the evidence-based justification of score interpretations, not the plausibility or polish of responses (Cronbach and Meehl, 1955; Messick, 1995). Similarly, Drs. Zaghi and Harel argue persuasively that the central question should shift from "Did AI write this?" to "What did the student do?" We embrace this reorientation. As they note, in real workplaces, professionals are increasingly expected to use AI tools appropriately with human accountability, rather than to avoid them; workforce and industry reports explicitly foreground AI literacy, responsible use, and critical judgment as job-relevant competencies (World Economic Forum, 2025; Microsoft, 2025). This insight reinforces our original argument that detection-based approaches are fundamentally misaligned with educational goals, because detector outputs do not provide a stable or equitable basis for high-stakes assessment decisions (Sadasivan et al., 2023; Liang et al., 2023).

We acknowledge that our paper focused primarily on practical strategies while perhaps under-emphasizing this deeper epistemic dimension. The discussants have helped us see that the

☆Main article: https://doi.org/10.6339/25-JDS1208.
*Corresponding author. Email: nkwshuyi@gmail.com.

AI challenge presents an opportunity—what Dr. Columbus calls an "epistemic stress test"—to finally confront long-standing ambiguities about what assessment is truly for.

## 2 The Dynamic Nature of AI Capabilities

Drs. Huo and Ni raise a critical point that deserves emphasis: AI capability is a moving target. They correctly observe that "any pedagogical strategy that relies on current AI weaknesses is therefore inherently short-lived." This insight has profound implications for assessment design: sustainable solutions must focus on skills that remain distinctively human regardless of AI advancement. This "moving target" claim is empirically supported by year-over-year benchmarking syntheses documenting rapid capability gains across model generations (AI Index Steering Committee, 2025).

Drs. Chan and Pua's classroom experiences vividly illustrate the risks of failing to do so. Their economics students, when asked about the consistency of demand curves with utility theory, simply stated conclusions about "diminishing marginal utility" without providing supporting analysis—a response that mirrored ChatGPT's output. This demonstrates the danger of students accepting AI-generated answers at face value (plausible-sounding but superficial) rather than engaging in the deep reasoning required to verify them. This concern is consistent with documented risks that generative systems can produce fluent but incorrect outputs and that outputs are not fully predictable, which makes "verification and justification" a necessary learning objective rather than an optional add-on (National Institute of Standards and Technology (NIST), 2023; Code.org, 2024a).

We appreciate their candid assessment of our strategies. Regarding their concern about the scalability of oral assessments, we advocate for strategic sampling—focusing verification on high-scoring submissions or using random selection. The goal is to create sufficient "verification uncertainty" to discourage wholesale delegation to AI, maintaining assessment integrity without imposing an unmanageable workload on instructors. This "sampling for verification" logic is consistent with deterrence evidence emphasizing the certainty of detection/apprehension as a more reliable deterrent lever than severity alone (Nagin, 2013).

## 3 Practical Implementation: Strategies and Limitations

Several discussants offered valuable concrete implementations of our proposed strategies. Dr. Nicosia's R-based statistics assignment at Université Laval exemplifies how to make AI use "mandatory but constrained." His empirical data (Table 2 in his discussion) suggests high student engagement and learning efficacy. By requiring students to build S3-based contextual wrappers, students must make deliberate decisions about information architecture and prompting—precisely the skills we argue should be cultivated. While we initially shared a concern that unsupervised assignments might lead to total AI delegation, Nicosia's detailed criterion-referenced rubric (Table 1 in his discussion)—focusing on software quality and integration logic—provides a robust countermeasure. Furthermore, his initiative to consolidate student work into a shared `contextR` package validly demonstrates how AI-assisted assessment can transcend the classroom to create community value.

Dr. Lock's endorsement of flipped classrooms and frequent low-stakes in-class assessments aligns well with our framework. We are honored that she sees convergence between our recommendations and the next revision of the ASA-endorsed College GAISE guidelines (GAISE College Report ASA Revision Committee, 2025). The current College GAISE revision materials emphasize principles that closely match our argument in an AI era—conceptual understanding over manual calculation, authentic data with context, multivariable thinking, effective written/oral communication, ethical practice, active learning, varied assessments, and inclusive course design (GAISE College Report ASA Revision Committee, 2025, 2016). Dr. Lock's observation that "AI has created something of a 'John Henry' moment" captures the urgency of curricular adaptation.

Dr. Furfaro raises legitimate concerns about the scalability of oral assessments in large classes. We suggest that random sampling combined with targeted verification can create positive peer pressure. Furthermore, she proposes a valuable direction: developing LLM-based tools in controlled environments where instructors can review full student-AI conversations. Building on this suggestion, we propose leveraging educational AI platforms that provide auditable, instructor-visible conversation transcripts (and moderation flags) by design—for example, Code.org's AI Chat Lab teacher view (full student–AI chat logs, including flagged interactions; 90-day retention) and Khan Academy's Khanmigo (teachers can view students' full chat transcripts via the Teacher Dashboard) (Code.org, 2024b; Khan Academy, 2024). This would operationalize her vision, enabling transparent AI use where instructors can verify which parts of the work were human-led, while making the "audit trail" a normal component of the learning process rather than an after-the-fact investigation.

## 4 Differentiated Approaches by Course Level

Dr. Loux makes a crucial distinction we had not adequately addressed: the role of AI differs fundamentally between introductory and advanced courses. In introductory courses, AI tools like GitHub Copilot may actually lower barriers to entry, allowing novice students to create meaningful outputs—data visualizations, simple analyses—that would otherwise be beyond their reach. This can build confidence and spark interest.

This insight resonates with our own teaching experiences. When students who have never programmed can suddenly create working applications through AI-assisted coding, their engagement and enthusiasm increase dramatically. In such contexts, AI serves as an enabler rather than a shortcut—these students are not circumventing learning they should have done; they are accessing capabilities they could never have developed in the time available. At the same time, empirical security analyses show that code-generation assistants can produce insecure suggestions, strengthening the case that "verification and justification" must be explicitly taught alongside AI-assisted productivity (Pearce et al., 2022).

In advanced courses, by contrast, students should be developing independent skills while learning to use AI as "a new tool to help with mundane tasks, not to replace human engagement." Dr. Loux's concept of "judicious use" captures this well—acknowledging that bright-line rules in syllabi cannot capture the nuance required.

## 5 Redefining Competency in Human–AI Collaboration

The most forward-looking contributions concern what competency means in an AI-saturated world.

Drs. Zaghi and Harel identify four "deep fundamentals" that assessment must now target: core conceptual invariants, a sense of scale and magnitude, sensitivity to assumptions, and judgment about tool appropriateness. They further propose four future-facing skill clusters, including collaborating with AI as a model to be tested and communicating uncertainty. We find this framework compelling. Their observation that students must learn to treat AI "neither as an authority nor as a toy to be tricked" resonates deeply with our pedagogical philosophy.

Drs. Huo and Ni offer concrete assignment strategies that operationalize these exact competencies. We are particularly drawn to their proposals for "Model Defense Briefs" (shifting weight from code to justification) and "Forensic AI-Auditing" (asking students to diagnose flawed AI outputs). Their "design-before-prompt" protocol—requiring students to articulate an analytical plan before engaging AI—is an excellent method to ensure that human reasoning remains the "architect" while AI serves as the "builder."

One theme emerges across all discussions: accountability. As Drs. Zaghi and Harel note, professionals cannot hide behind "the AI said so." This represents a fundamental shift: students must learn to take ownership of AI-assisted outputs, understanding that the final judgment—and its consequences—remain theirs alone. Recent real-world assessment evidence underscores why accountability cannot be outsourced: in a blind "infiltration" study injecting fully AI-written exam submissions into a university examination system, 94% were undetected by markers (Scarfe et al., 2024).

## 6 Institutional and Cultural Contexts

While our original paper and most discussions focus on classroom-level interventions, several discussants implicitly raise broader institutional questions. Dr. Lock's reference to the GAISE report reminds us that sustainable change requires coordination across professional communities. Individual instructors innovating in isolation face significant headwinds; systematic reform requires institutional support, revised assessment policies, and—critically—student buy-in.

Our experience suggests that students often resist pedagogical innovations that increase their cognitive load, particularly when they have been conditioned by years of passive learning (Seidel and Tanner, 2013; Deslauriers et al., 2019). The flipped classroom model, while educationally sound, can generate initial resistance; evidence also shows that students can feel they learn less under active-learning approaches even when measured learning improves, which makes expectation-setting and explanation part of the instructional design (Deslauriers et al., 2019).

Explaining to students why these approaches serve their long-term interests—why the skills being cultivated will remain valuable even as AI capabilities advance—is itself a pedagogical task that deserves attention. Cultural context also matters. Educational traditions vary significantly across regions, and strategies that work well in one context may require adaptation elsewhere. The discussions have been predominantly grounded in North American and Chinese educational contexts; we would welcome future work exploring how these frameworks translate to educational systems with different traditions, resources, and constraints.

## 7 Concluding Remarks

The eight discussions have collectively advanced our understanding of AI in education far beyond our original contribution. Where we offered practical strategies, our discussants have provided theoretical grounding, empirical validation, boundary conditions, and extensions. We are particularly grateful for the constructive spirit of engagement—even where discussants expressed skepticism about specific recommendations, they did so in ways that clarified the conditions under which different approaches might succeed or fail.

Several key insights emerge from this collective dialogue:

First, the AI challenge in education is fundamentally an assessment validity challenge, in the classical sense that claims about learning require defensible inferences from observed work to underlying competence (Cronbach and Meehl, 1955; Messick, 1995; Kane, 2013). Detection-based approaches address symptoms rather than causes; moreover, detector reliability and fairness problems make them a weak foundation for high-stakes educational decisions (Sadasivan et al., 2023; Liang et al., 2023; Scarfe et al., 2024).

Second, pedagogical strategies must be robust to AI capability improvements. Approaches that exploit current AI limitations will become obsolete; approaches that cultivate durable human competencies (planning, justification, verification, judgment) remain relevant as model capabilities advance (AI Index Steering Committee, 2025).

Third, context matters enormously. The appropriate role of AI differs across course levels, disciplines, and educational goals. Blanket policies—whether permissive or restrictive—will inevitably be ill-suited to many contexts.

Fourth, human-AI collaboration is not merely a concession to technological reality but an educational opportunity. Students who learn to work effectively with AI—maintaining critical judgment while leveraging AI capabilities—will be better prepared for professional life than those trained in AI-free environments, consistent with workplace skills framings that foreground AI literacy and responsible tool use (World Economic Forum, 2025).

Finally, accountability must remain with humans. Whatever role AI plays in the learning process, students must understand that they bear responsibility for the outputs they submit and the judgments they make-an accountability principle also emphasized in AI risk governance frameworks that treat human oversight as a core control for AI-related risk (National Institute of Standards and Technology (NIST), 2023).

The discussants have convinced us that we are at an inflection point in educational history—what Dr. Lock aptly calls a "John Henry moment." The question is not whether AI will transform education but how educators will shape that transformation. We hope that our paper and these rich discussions contribute to a future in which AI enhances rather than undermines the development of thoughtful, capable, and responsible professionals.

## References

AERA, APA, & NCME (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf.

AI Index Steering Committee (2025). The AI index 2025 annual report. *Technical report*, Institute for Human-Centered AI, Stanford University, Stanford, CA. Available at https://doi.org/10.48550/arXiv.2504.07139.

Code.org (2024a). AI Chat Lab FAQ. Available at https://support.code.org/hc/en-us/articles/30162711193741-AI-Chat-Lab-FAQ. Accessed December 21, 2025.

Code.org (2024b). Viewing student AI chat history as a teacher. Available at https://support.code.org/hc/en-us/articles/30681531276045-Viewing-Student-AI-Chat-History-as-a-Teacher. Accessed December 21, 2025.

Cronbach LJ, Meehl PE (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4): 281–302. https://doi.org/10.1037/h0040957.

Deslauriers L, et al. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *PNAS*, 116(39): 19251–19257. https://doi.org/10.1073/pnas.1821936116.

GAISE College Report ASA Revision Committee (2016). Guidelines for assessment and instruction in statistics education college report 2016. *Technical report*, American Statistical Association. Available at https://www.amstat.org/asa/files/pdfs/gaise/gaisecollege_full.pdf.

GAISE College Report ASA Revision Committee (2025). Guidelines for assessment and instruction in statistics education college report 2025. *Technical report*, American Statistical Association. [Forthcoming]. Draft available at https://amstat.quarto.pub/college-gaise/.

Kane MT (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1): 1–73. https://doi.org/10.1111/jedm.12000.

Khan Academy (2024). How do I view my students' Khanmigo chat history? Available at https://support.khanacademy.org/hc/en-us/articles/15127248640525-How-do-I-view-my-students-Khanmigo-chat-history. Accessed December 21, 2025.

Liang W, et al. (2023). GPT detectors are biased against non-native English writers. Available at https://arxiv.org/abs/2304.02819.

Messick S (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9): 741–749. https://doi.org/10.1037/0003-066X.50.9.741.

Microsoft (2025). Work trend index 2025. *Technical report*, Microsoft WorkLab. Available at https://www.microsoft.com/en-us/worklab/work-trend-index/.

Nagin DS (2013). Deterrence in the twenty-first century. In: Tonry M (ed.), *Crime and Justice*, volume 42, 199–263. https://prohic.nl/wp-content/uploads/2020/11/2020-06-02-DeterrenceMetaNagin.2013.pdf.

National Institute of Standards and Technology (NIST) (2023). Artificial intelligence risk management framework (AI RMF 1.0). *Technical Report NIST.AI.100-1*, National Institute of Standards and Technology. Available at https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

Pearce H, et al. (2022). Asleep at the keyboard? Assessing the security of GitHub Copilot's code contributions. Available at https://arxiv.org/abs/2108.09293.

Sadasivan V, et al. (2023). Can AI-Generated text be reliably detected? Available at https://arxiv.org/abs/2303.11156.

Scarfe P, Watcham K, Clarke A, Roesch E (2024). A real-world test of artificial intelligence infiltration of a university examinations system: A "Turing Test" case study. *PLoS ONE*, 19(6): e0305354. https://doi.org/10.1371/journal.pone.0305354.

Seidel SB, Tanner KD (2013). "What if students revolt?" —considering student resistance: Origins, options, and opportunities for investigation. *CBE—Life Sciences Education*, 12(4): 586–595. https://doi.org/10.1187/cbe-13-09-0190.

Wang S, Xu L, Liu L, Zhai L (2026). Addressing the challenges of AI-generated assignment submissions in education: Insights and strategies. *Journal of Data Science*, 24(1): 1–7. https://doi.org/10.6339/25-JDS1208

World Economic Forum (2025). The future of jobs report 2025. *Technical report*, World Economic Forum. Available at https://www.weforum.org/reports/the-future-of-jobs-report-2025/.