World Journal of Educational Research
ISSN 2375-9771 (Print) ISSN 2333-5998 (Online)
Vol. 13, No. 2, 2026
www.scholink.org/ojs/index.php/wjer

Original Paper

Reconstructing the Assessment System of the Undergraduate Course "Plant Physiology" in the AI Era: A Framework for Authentic and AI-Resilient Task Design

Jiantao Peng[1]

[1] College of Tobacco Science, Guizhou University, Guiyang, Guizhou 550025, China

Received: January 29, 2026 Accepted: March 29, 2026 Online Published: April 24, 2026

doi:10.22158/wjer.v13n2p127 URL: http://dx.doi.org/10.22158/wjer.v13n2p127

Abstract

The proliferation of generative artificial intelligence (GenAI) has fundamentally challenged the validity of traditional assessments in higher education. In data-rich, laboratory-based disciplines such as undergraduate plant physiology, conventional product-heavy assessments—like essay writing and closed-book examinations—often fail to capture underlying scientific reasoning and are highly susceptible to GenAI simulation. Drawing upon frameworks of constructive alignment, evaluative judgment, and authentic assessment, this paper argues for a systemic shift from misconduct management to "AI-resilient" structural task design. We propose replacing the traditional assessment paradigm with an integrated, seven-part ecology of multimodal tasks. This reconstructed system distributes evidence of learning across the semester through in-class rapid reasoning exercises, process-oriented laboratory notebooks, collaborative inquiry projects, interactive oral defenses (mini-vivas), and reflective AI-use portfolios. By prioritizing visible reasoning, structural verification of knowledge, and critical AI literacy over unsupervised textual output, this framework ensures that students develop and demonstrate authentic disciplinary competence. Ultimately, this approach not only safeguards academic integrity but also enhances the educational validity of plant science curricula, offering a scalable template for assessment reform across the biosciences.

Keywords

Generative AI, assessment redesign, plant physiology, authentic assessment, oral defense, evaluative judgment, higher education

1. Introduction

Generative artificial intelligence (GenAI) has shifted the higher education assessment debate from

misconduct management to validity. As large language models generate plausible essays, summaries, laboratory write-ups, and examination responses with limited student engagement (Ahangama, 2026), the core issue extends beyond academic dishonesty. The deeper concern is that conventional assessment formats no longer provide robust evidence of student knowledge, capability, and justification. Research highlights the difficulty of distinguishing AI-generated work from student writing (Farazouli et al., 2024), and scholars increasingly frame AI and assessment as a "wicked problem" rather than a technical issue with a singular solution (Corbin et al., 2025a).

Retreating to surveillance or traditional final examinations is not the only defensible response. Evidence suggests the pedagogical justification for high-stakes final examinations is weak, with significant drawbacks for authenticity, wellbeing, equity, and long-term learning (French et al., 2024). Assessment must be conceptualized not merely as certification, but as a design challenge affecting feedback, future learning, and evaluative judgment (Boud & Falchikov, 2006; Carless, 2007). Thus, the AI era intensifies the need for improved assessment design rather than justifying conservatism.

These implications are particularly pronounced in plant physiology. Unlike courses where success is demonstrated through polished exposition, plant physiology depends on interpretive and experimental capacities that are difficult to infer from finished text alone. Students must explain processes across scales, select variables, justify controls, interpret data, troubleshoot methods, and defend causal explanations under uncertainty. Plant-science education research demonstrates that inquiry-rich, research-infused, and communication-centered approaches significantly improve knowledge, reasoning, writing, and disciplinary interest (Beckmann et al., 2015; Hiatt et al., 2021; Ward et al., 2014).

This paper argues that undergraduate "Plant Physiology" courses must reconstruct their assessment systems into integrated ecologies of authentic and AI-resilient tasks. Because authenticity or policy alone cannot guarantee integrity (Ellis et al., 2020; Fawns et al., 2025), the objective is not a single "AI-resistant" task. Rather, the course should structurally redesign assessment so students repeatedly generate evidence of reasoning across multiple modalities while learning to use AI tools critically and ethically.

2. Conceptual Framing: From Constructive Alignment to AI-Resilient Assessment

In the GenAI era, course assessment should be treated as a coherent system aligned with higher-order learning outcomes. This paper proposes redesigned framework synthesizing research on constructive alignment, assessment for learning, evaluative judgment, authentic assessment, and AI literacy.

This approach is grounded in constructive alignment (Biggs, 1996), which requires designing activities and assessments that directly elicit intended learning outcomes. Assessment must align with long-term learning (Boud & Falchikov, 2006), with tasks functioning as part of an iterative learning process where feedback acts as feedforward (Carless, 2007). Broad conceptual frameworks must be translated into discipline-specific solutions, such as an Assessment Design Matrix (ADM) that explicitly links graduate competencies, assessment strategies, and assessment methods (Ahangama, 2026).

A second foundational concept is evaluative judgment—the capability to make decisions about the quality of work (Tai et al., 2018). In a GenAI context, students must judge not only their own work but also the quality and biases of AI outputs (Bearman et al., 2024). AI literacy—comprising technical understanding, ethics, and critical evaluation (Lintner, 2024)—must become an integrated component of scientific literacy.

A third foundation is authenticity. While Gulikers et al. (2004) locate authenticity in the alignment between assessment and professional practice, recent scholarship warns against treating it as a universal remedy (Fawns et al., 2025; McArthur, 2023). Data on contract cheating reveals that highly "authentic" tasks are still routinely outsourced (Ellis et al., 2020).

This necessitates a fourth foundation: AI-resilient, rather than "AI-proof," assessment. Corbin et al. (2025b) distinguish between discursive changes (rules prohibiting AI) and structural changes (mechanics requiring real-time interpretation). For plant physiology, policy must be backed by task structures that generate verifiable evidence of understanding, requiring a systemic approach to cognitive offloading and creativity (Peters et al., 2025).

3. Why Plant Physiology Requires Whole-Process Assessment Redesign

Given its empirical and interpretive core, plant physiology is well-suited for process-oriented and multimodal assessment reform. Core topics—water regulation, carbon assimilation, and stress responses—are addressed through observation, measurement, modeling, and argumentation.

Historically, science assessments have over-weighted end-point products: polished lab reports, literature reviews, and closed-book examinations. These formats create distortions by rewarding retrospective neatness or short-term retrieval over iterative inquiry. GenAI magnifies these weaknesses by easily simulating competence without requiring the underlying scientific process.

Conversely, plant-science pedagogy advocates for inquiry-centered models. The "Plant Detectives" model organizes learning around unknown cases and collaborative inquiry (Beckmann et al., 2015), while course-based undergraduate research experiences shift awareness and interest in plants (Hiatt et al., 2021; Ward et al., 2014). GenAI acts as a catalyst exposing the mismatch between traditional product-heavy assessments and the discipline's inquiry-based goals.

A reconstructed course should explicitly assess cross-scale causal reasoning, experimental design, data interpretation, diagnostic problem-solving, and the transparent, critical use of AI tools.

4. Principles for Reconstructing the Assessment Process

4.1 Assess Visible Reasoning, not only Polished Output

Assessment should prioritize visible reasoning and process-oriented evaluation (Bannister et al., 2025). Product-only assessment must be supplemented by checkpoints, oral explanations, and tasks requiring real-time justification of decisions.

4.2 Build a Distributed Evidence Base across the Semester

Heavy dependence on a single final task is pedagogically weak (French et al., 2024). Assessment systems should support learning across time (Boud & Falchikov, 2006; Carless, 2007). A semester-long sequence of low-, medium-, and high-stakes tasks provides a more robust, distributed evidence base.

4.3 Combine Collaboration with Individual Accountability

Tasks requiring genuine collaboration offer greater resistance to AI manipulation (Bannister et al., 2025). However, certification must remain individually defensible. Pairing group inquiry with individual oral defenses preserves authenticity while protecting validity (Davey et al., 2025; Ward et al., 2024).

4.4 Use Local, Novel, and Messy Evidence

Students should evaluate unseen and situated evidence—class-generated datasets, local observations, or experimental anomalies. This diminishes the temptation to plagiarize and prevents cognitive offloading to AI (Peters et al., 2025). Furthermore, balancing detailed guidance with student autonomy helps avoid inadvertently facilitating GenAI prompt formulation (Bannister et al., 2025).

4.5 Make AI Critique Assessable

AI should be treated as an object of disciplinary judgment (Chan, 2023). Evaluative judgment requires students to recognize quality in AI-mediated processes (Bearman et al., 2024). Students should be assessed on their ability to accept, reject, verify, or correct AI outputs.

4.6 Design Inclusion Structurally, not Retrospectively

Oral and authentic assessments are not automatically inclusive (Fawns et al., 2025; McArthur, 2023). Equity depends on explicit criteria, structured practice, and scaffolding to reduce anxiety and support participation.

5. A Proposed Reconstruction of the Entire Assessment Process

The proposed course replaces the traditional package of practical reports, a term paper, and a final examination with a cumulative, seven-part structure (Table 1). This aligns with structural redesign recommendations in AI-assessment literature (Corbin et al., 2025b).

Table 1. Proposed Assessment Reconstruction: Components, Weighting, and AI-Resilience

Component	Weight	What Students Do	Main Evidence of Learning	Why it is AI-Resilient
Diagnostic concept + AI-use baseline	Hurdle	Complete a short concept inventory and AI policy quiz in week 1.	Starting misconceptions, baseline AI understanding.	Establishes expectations and identifies support needs.
Weekly in-class	10%	Interpret unseen	Real-time	Time-limited,

rapid reasoning tasks		graphs, images, or stress scenarios in class.	reasoning and concept application.	supervised, contingent on the day's materials.
Laboratory notebook + raw-data dossier	15%	Maintain methods, observations, troubleshooting notes, and raw data.	Process trace of experimental work.	Hard to outsource; grounded in actual, localized lab activity.
Two mini-vivas after laboratory blocks	15%	Individually explain methods, errors, results, and revisions.	Understanding of one's own experimental decisions.	Follow-up questioning exposes shallow or borrowed knowledge.
Group inquiry project	20%	Design and run an investigation (e.g., drought, salinity).	Proposal, dataset, figures, poster, brief team presentation.	Mirrors authentic research; balanced by an individual defense.
Individual integrative oral exam	25%	Defend project decisions and interpret a new dataset under questioning.	Mechanistic explanation, transfer, communication, judgment.	Live explanation makes generic AI substitution ineffective.
Reflective portfolio on feedback and AI use	15%	Submit revision memo, AI-use disclosure, and self-evaluation.	Evaluative judgment, self-regulation, ethical AI use.	Assesses how students judge and use tools, not just the final product.

6. Examples of "AI-Resistant" Task Design in Plant Physiology

The following examples demonstrate assessment tasks that are difficult to outsource and reliant on contextual judgment.

6.1 Photosynthesis Limitation Diagnosis

Students receive an unseen figure set (e.g., net photosynthetic rate, stomatal conductance) from control and stressed plants. They must identify the primary limitation, provide supporting evidence, and propose confirmatory measurements. An oral follow-up requires them to state what result would *falsify* their interpretation.

6.2 Water-relations Emergency Brief

Students receive a scenario where a greenhouse trial shows wilting, but midday leaf water potential data contradicts the symptoms. They must produce a two-minute live emergency brief for the greenhouse manager explaining the physiological anomaly and next diagnostic steps.

6.3 Mineral Nutrition Diagnostic Memo Plus Viva

Students analyze symptom images, partial tissue-analysis data, and treatment history to identify a deficiency. During a mini-viva, they must explain why they rejected alternative diagnoses, foregrounding comparative judgment.

6.4 Hormone-signaling Experimental Design Studio

Working in groups, students sketch an experimental design distinguishing ABA-mediated growth from ethylene-mediated stress. They then complete an individual oral defense justifying one specific design choice and anticipating one confound.

6.5 Abiotic Stress Case Conference

Students analyze a complex, imperfect dataset (e.g., missing variables, anomalous replicates). By building ambiguity into the task, the instructor assesses whether students can make disciplined judgments under uncertainty.

6.6 AI Critique and Verification Memo

Students receive an AI-generated explanation of a physiological phenomenon containing subtle flaws (e.g., causal leaps, fabricated references). They must annotate the text, verify claims, and rewrite the explanation accurately, directly assessing AI literacy.

7. Rubrics, Feedback, Moderation, Inclusion, and Workload

A reconstructed assessment system requires reconstructed rubrics. Dimensions should reflect conceptual accuracy, causal reasoning, use of evidence, and transparency of AI use, rather than fluency alone. Educators can utilize frameworks like the PANDORA GenAI Susceptibility Rubric to evaluate their assessments' vulnerability to AI manipulation based on intellectual complexity and creative opportunity (Bannister et al., 2025).

Feedback must shift from correcting outputs to refining scientific judgment. Weekly tasks require rapid debriefs, and project checkpoints should necessitate response-to-feedback notes. For quality assurance, oral assessments should use shared question banks, structured rubrics, and assessor calibration (Tan, 2022; Ward et al., 2024).

Inclusion requires explicit attention. Practice and feedback are key facilitators of oral performance; studies show no significant performance differences across demographics when students are carefully scaffolded (Davey et al., 2025).

Finally, while workload concerns are valid (Corbin et al., 2025a), redesigns can redistribute labor productively. Short in-class tasks replace the grading of long take-home essays, and notebook assessments can be streamlined through selective review (Fleming, 2023), shifting faculty effort from

post hoc policing to educational judgment.

8. Discussion

The proposed reconstruction offers three major strengths: it improves validity by aligning assessment with mechanistic reasoning and experimental thinking; it improves integrity by structurally verifying understanding; and it enhances educational relevance by developing critical AI literacy.

However, authenticity alone is insufficient; even authentic-seeming assessments can be outsourced (Ellis et al., 2020), and authenticity must connect to deeper questions of work and equity (Fawns et al., 2025; McArthur, 2023). The design works only as a holistic system where distributed evidence, process documentation, oral defense, and AI literacy reinforce one another.

Future empirical studies should examine how this reconstructed course affects student learning, anxiety, misconduct patterns, and AI literacy. As AI literacy instruments develop (Lintner, 2024), plant physiology could serve as a model for designing domain-specific measures of AI-supported scientific judgment.

9. Conclusion

The GenAI era has made assessment reform an imperative. It has exposed the fragility of systems that over-rely on unsupervised, polished, end-point textual production. For undergraduate plant physiology, the appropriate response is not to ban tools or restore exclusionary final examinations, but to redesign the course so students must repeatedly observe, record, interpret, justify, revise, and ethically augment their work. A reconstructed assessment ecology integrating in-class reasoning, process evidence, oral defense, and reflective AI-use documentation is more faithful to disciplinary practice, more defensible against AI misuse, and ultimately more educationally valuable.

Acknowledgement

This research was funded by 2022 Guizhou Provincial Higher Education Teaching Content and Curriculum System Reform Project: “Research on Instructional Design Based on Objective-Based Teaching Approach—A Case Study of the Plant Physiology (Bilingual) Course”.

References

Ahangama, N. (2026). Designing assessments in the generative AI era: A tailored assessment framework for ICT tertiary education. *International Journal of Educational Technology in Higher Education*, *23*, Article 9. https://doi.org/10.1186/s41239-026-00582-0

Bannister, P., Santamaría Urbieta, A., & Brufau Alvira, N. (2025). Appraising higher education assessment validity: Development of the PANDORA GenAI Susceptibility Rubric. *Journal of Applied Learning & Teaching*, *8*(1), Article 20. https://doi.org/10.37074/jalt.2025.8.1.20

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for

a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, *49*(6), 893-905. https://doi.org/10.1080/02602938.2024.2335321.

Beckmann, E. A., Estavillo, G. M., Mathesius, U., Djordjevic, M. A., & Nicotra, A. B. (2015). The plant detectives: Innovative undergraduate teaching to inspire the next generation of plant biologists. *Frontiers in Plant Science*, *6*, Article 729. https://doi.org/10.3389/fpls.2015.00729

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347-364. https://doi.org/10.1007/BF00138871.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, *31*(4), 399-413. https://doi.org/10.1080/02602930600679050

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, *44*(1), 57-66. https://doi.org/10.1080/14703290601081332

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, *20*, Article 38. https://doi.org/10.1186/s41239-023-00408-3

Corbin, T., Bearman, M., Boud, D., & Dawson, P. (2025a). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*, *50*(6). https://doi.org/10.1080/02602938.2025.2553340

Corbin, T., Dawson, P., & Liu, D. (2025b). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*, *50*(7), 1087-1097. https://doi.org/10.1080/02602938.2025.2503964

Davey, S. K., Birbeck, D., Nallaya, S., Sallows, G., & Della Vedova, C. B. (2025). Utilising one-on-one interactive oral assessments as the major final assessment within a bioscience course. *Assessment & Evaluation in Higher Education*, *50*(7), 1154-1171. https://doi.org/10.1080/02602938.2025.2502577.

Ellis, C., van Haeringen, K., Harper, R., Bretag, T., Zucker, I., McBride, S., Rozenberg, P., Newton, P., & Saddiqui, S. (2020). Does authentic assessment assure academic integrity? Evidence from contract cheating data. *Higher Education Research & Development*, *39*(3), 454-469. https://doi.org/10.1080/07294360.2019.1680956

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2024). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots' impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, *49*(3), 363-375. https://doi.org/10.1080/02602938.2023.2241676

Fawns, T., Bearman, M., Dawson, P., Nieminen, J. H., Ashford-Rowe, K., Willey, K., Jensen, L. X., Damša, C., & Press, N. (2025). Authentic assessment: From panacea to criticality. *Assessment & Evaluation in Higher Education*, *50*(3), 396-408. https://doi.org/10.1080/02602938.2024.2404634

Fleming, E. J. (2023). An approach to improve laboratory notebook feedback and decrease instructor

fatigue from grading undergraduate laboratory notebooks. *Journal of Microbiology & Biology Education*, *24*(1), e00018-22. https://doi.org/10.1128/jmbe.00018-22.

French, S., Dickerson, A., & Mulder, R. A. (2024). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education*, *88*(3), 893-918. https://doi.org/10.1007/s10734-023-01148-z

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, *52*(3), 67-86. https://doi.org/10.1007/BF02504676.

Hiatt, A. C., Hove, A. A., Antonio, T. M., Boyd, A. E., Clarke, H. D., Horton, J. L., & Murrell, Z. E. (2021). Authentic research in the classroom increases appreciation for plants in undergraduate biology students. *Integrative and Comparative Biology*, *61*(3), 969-980. https://doi.org/10.1093/icb/icab089

Lintner, T. (2024). A systematic review of AI literacy scales. *npj Science of Learning*, *9*, Article 50. https://doi.org/10.1038/s41539-024-00264-4

McArthur, J. (2023). Rethinking authentic assessment: Work, well-being, and society. *Higher Education*, *85*(1), 85-101. https://doi.org/10.1007/s10734-022-00822-y

Peters, M., & Angelov, D. (2025). Redefining assessment tasks to promote students' creativity and integrity in the age of generative artificial intelligence. *International Journal for Educational Integrity*, *21*, Article 25. https://doi.org/10.1007/s40979-025-00201-x

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, *76*(3), 467-481. https://doi.org/10.1007/s10734-017-0220-3

Tan, C. P. (2022). Developing interactive oral assessments to foster graduate attributes in higher education. *Assessment & Evaluation in Higher Education*, *47*(8), 1183-1199. https://doi.org/10.1080/02602938.2021.2020722

Ward, J. R., Clarke, H. D., & Horton, J. L. (2014). Effects of a research-infused botanical curriculum on undergraduates' content knowledge, STEM competencies, and attitudes toward plant sciences. *CBE—Life Sciences Education*, *13*(3), 387-396. https://doi.org/10.1187/cbe.13-12-0231

Ward, M., O'Riordan, F., Logan-Fleming, D., Cooke, D., Concannon-Gibney, T., Efthymiou, M., & Watkins, N. (2024). Interactive oral assessment case studies: An innovative, academically rigorous, authentic assessment approach. *Innovations in Education and Teaching International*, *61*(5), 930-947. https://doi.org/10.1080/14703297.2023.2251967