

# Assessment in Online Courses in the Age of AI Agents

University of Georgia • Office of Online Learning • March 2026

**The Core Challenge:** AI agents — tools that can browse, reason, write, and execute tasks autonomously within a browser — can now complete many traditional online assessments without meaningful student engagement. Course context determines the right response. Three course types matter: **high-stakes/low-interest required courses** (such as CBK requirements), where student motivation to circumvent learning is highest; **major and program courses**, where professional identity and intrinsic interest reduce that motivation; and **exploratory or elective gen ed courses**, where student choice provides moderate engagement. **Class size** is the second key dimension, determining what is logistically feasible.

**UGA's QEP:** UGA's Quality Enhancement Plan focuses on building a community that embraces *active learning*. AI-era assessment reform supports that goal directly. The approaches in this handout require students to demonstrate understanding by doing, reflecting, and engaging — not by retrieving information. Responding to AI by redesigning assessment is an opportunity to advance UGA's active learning commitments.

**UGA's Starting Position:** UGA ranked in the **top 20 nationally for online programs** (U.S. News, 2026) and has been steadily growing its online graduate offerings with strategic central investment. UGA has also launched the **Comprehensive Learner Record** to capture richer evidence of student growth and established the **Leadership Council on AI** to coordinate campus-wide AI strategy. The **Center for Teaching and Learning** supports faculty course redesign, and the **Office of Online Learning** provides instructional design capacity across programs. UGA is well positioned to respond to the evolving demands of assessment in the age of AI.

**Graduate and Professional Programs:** The matrix and strategies in this handout apply across undergraduate and graduate contexts, but the implications differ. Graduate and professional students (MBA, MPH, MSW and others) are often working professionals with stronger intrinsic motivation — circumvention risk is lower, but the stakes of genuine mastery are higher. Discipline-specific AI literacy is particularly important at the graduate level, where students are entering or advancing in fields where AI tool use is already standard. Assessment design in these programs should reflect professional norms and the AI-integrated environments students will work in.

## BEFORE ASSESSMENT: DESIGNING COURSES THAT REQUIRE ENGAGEMENT

### Spiral the Core Concepts

Return to fundamental principles across multiple modules. A student who outsources one assignment encounters the same concept again in a new context that requires fresh judgment. Repeated contact reduces the value of delegation to an AI.

### Design AI-Resistant Peer Evaluation

Have students use a UGA AI tool to evaluate peers' products or processes, then score how well the AI did. To evaluate the AI's performance, students must do the underlying evaluation themselves — making further delegation circular and self-defeating. For this to work, rubric criteria must require discipline-specific judgment, contextual reasoning, or evaluation of process rather than product. Rubrics that read like checklists are effectively multiple choice tests that AI can navigate by pattern-matching.

### Use Apply-a-Concept Activities

Tasks that ask whether a principle explains a situation require recognition and judgment, not production. They are harder to outsource because evaluating fit between concept and context requires having internalized the concept. Also, more personalized or unique contexts currently decrease AI's ability to perform accurately.

### Design Cumulative Low-Stakes Checks

Activities that build on each other make gaps in understanding visible in later work. Students who do not engage genuinely early will encounter compounding difficulty — and students who do engage are rewarded for it.

### Co-Scaffold AI Literacy and Disciplinary Learning

Design activities that develop disciplinary knowledge and AI literacy together. Ask students to evaluate AI-generated claims using discipline-specific standards, identify where AI reasoning fails in their field, or document how they directed and assessed AI assistance. Model expert AI use. These skills transfer directly to professional practice.

*These course design strategies create the conditions that make assessment more meaningful and AI circumvention less rewarding and they are consistent with UGA's active learning QEP.*

# Assessment in Online Courses in the Age of AI Agents

University of Georgia • Office of Online Learning • March 2026

## FRAMEWORK: ASSESSMENT STRATEGIES BY COURSE CONTEXT

**Circumvention risk** reflects the combination of stakes (grade/requirement pressure) and student interest. High-stakes, low-interest required courses — such as CBK requirements — represent the highest risk. Strategies in those cells address both resistance to AI circumvention and relevance-building, since motivation to circumvent decreases when students see the purpose of the work.

Course Type → Class Size ↓	High-Stakes / Low-Interest Gen Ed Required core; CBK-type; diverse populations	Major / Program Course UG and graduate/professional; discipline-specific mastery	Exploratory / Elective Gen Ed Student-chosen; moderate intrinsic interest
<b>Small (&lt;30)</b>	<p><b>▲ High Risk</b></p> <p><b>Oral discussion checkpoints</b> tied to submitted work — confirm ownership and raise relevance through conversation.</p> <p><b>Locally-situated case studies</b> connecting required content to students' own programs or career goals.</p> <p><i>Design goal: make the "why" visible. Circumvention drops when students see personal stakes.</i></p>	<p><b>▼ Lower Risk</b></p> <p><b>Oral exams / vivas</b> — assess depth; AI cannot substitute for the student.</p> <p><b>Iterative projects with staged deliverables</b> and instructor feedback loops.</p> <p><b>Co-scaffolded AI tasks</b> — direct AI to solve a disciplinary problem, then critically evaluate its output against professional standards.</p> <p><i>AI-use policy: discipline-normed; mirrors professional practice. Graduate programs should reflect field-specific AI norms.</i></p>	<p><b>— Moderate Risk</b></p> <p><b>Reflective portfolios</b> documenting growth; authentic voice is hard to fabricate over time.</p> <p><b>Student-directed inquiry projects</b> anchored in genuine curiosity.</p> <p><i>AI-use policy: transparent co-use with required process reflection.</i></p>
<b>Medium (30–100)</b>	<p><b>▲ Highest Risk</b></p> <p><b>Staged group projects</b> with individual reflection at each stage — separates process from product; harder to fully outsource.</p> <p><b>Scenario-based application quizzes</b> using novel, locally-relevant contexts each term.</p> <p><i>Design goal: peer accountability raises engagement in required courses. TA-supported rubrics are essential.</i></p>	<p><b>▼ Lower Risk</b></p> <p><b>Authentic case-based assessments</b> with novel parameters each term.</p> <p><b>Staged groupwork</b> — shared deliverable plus individual reflection at each stage.</p> <p><b>AI literacy documentation</b> — students submit a log of AI interactions alongside their work, evaluated for quality of prompting and critical judgment.</p> <p><i>AI-use policy: require process documentation; AI as tool, not author.</i></p>	<p><b>— Moderate Risk</b></p> <p><b>Peer-reviewed creative or analytical projects</b> with structured reflection.</p> <p><b>Redesigned open-note assessments</b> emphasizing judgment and synthesis.</p> <p><i>AI-use policy: define permitted tools; assess metacognitive reflection separately.</i></p>
<b>Large (&gt;100)</b>	<p><b>▲ Highest Risk</b></p> <p><b>Embedded synchronous checkpoints</b> — brief live or recorded responses confirm engagement.</p> <p><b>AI-evaluation tasks</b> — students critique or correct AI-generated responses using course concepts; assesses mastery through discernment, not production.</p> <p><b>Peer-assessed process portfolios</b> using structured rubrics — scalable; shifts focus to learning over product.</p> <p><i>Design goal: anonymous large required courses are highest circumvention risk at UGA. Faculty workload is critical.</i></p>	<p><b>— Moderate Risk</b></p> <p><b>Proctored or synchronous capstone assessments</b> at key milestones.</p> <p><b>AI-integrated problem sets</b> where students extend or evaluate AI-generated work against disciplinary standards.</p> <p><i>AI-use policy: structured transparency; require documentation of AI interaction. Graduate programs: align with professional licensure and accreditation expectations.</i></p>	<p><b>— Moderate Risk</b></p> <p><b>Peer-assessed portfolios</b> with structured rubrics — scales reflection without full faculty grading load.</p> <p><b>Application-focused redesigned tests</b> with open-AI transparency requirements.</p> <p><i>AI-use policy: focus on process artifacts alongside final product.</i></p>

**On Faculty Workload:** Several of the stronger assessment approaches in this matrix require more faculty time to assess. Recommendations for addressing this structurally are on page 5.

**On Accommodations and Accessibility:** Synchronous and oral components create barriers for students with disabilities, working students, caregivers, and students across time zones. Assessment redesign should account for these constraints from the start.

# Assessment in Online Courses in the Age of AI Agents

University of Georgia • Office of Online Learning • March 2026

## ASSESSMENT APPROACHES: KEY FEATURES

### Project-Based & Authentic Assessment

Tasks anchored in real-world problems with parameters that change each term. Generic prompts are AI-vulnerable; locally-situated, novel problems are not. Graduate programs should anchor projects in professional practice contexts.

All sizes High design cost

### Staged Groupwork with Individual Reflection

Groups produce shared deliverables at multiple stages. Each student submits a structured reflection documenting their contribution and learning at each stage. This creates a record of individual thinking that is difficult to fabricate and scales with TA support.

Medium–Large Requires rubric design

### Portfolio & Reflective Assignments

Students curate evidence of growth with metacognitive commentary over time. Authentic voice accumulated across a semester resists AI substitution. Feeds naturally into UGA's Comprehensive Learner Record, particularly when students document growth in AI literacy alongside disciplinary skills.

Small–Medium Long time horizon

### Oral Exams / Vivas

Synchronous conversation — live or recorded — in which students explain, defend, or extend their work. Brief checkpoints of 5–10 minutes substantially increase assessment validity. Most practical in small courses or as a sampling strategy in larger ones. Requires accommodation planning for students with disabilities or scheduling constraints.

Small–Medium Scheduling intensive

### Redesigning Traditional Tests for the AI Era

Shift from recall to application, analysis, and judgment. Use novel scenarios, ask students to evaluate AI-generated responses, or add real-time constraints. Open-note, open-AI formats with transparency requirements assess higher-order thinking directly. Asking students to critique AI output using disciplinary standards simultaneously assesses content mastery and AI literacy.

All sizes Faculty development needed

### Co-Scaffolded Assessment Tasks

Assessments that require students to demonstrate disciplinary knowledge and AI literacy together — for example, directing an AI to solve a discipline-specific problem and then evaluating its output against course concepts, or documenting the prompts and judgments made while using AI to complete a task. These tasks make AI engagement visible and assessable. In graduate and professional programs, they mirror workflows where AI use is expected and critical evaluation is essential.

All sizes High transferability

# Assessment in Online Courses in the Age of AI Agents

University of Georgia • Office of Online Learning • March 2026

## MANAGING THE GRADING LOAD

### Using AI to Help Grade

- **Apply rubrics at scale.** Paste a rubric and a batch of student submissions into an AI tool and ask it to score each against the rubric criteria. Review and adjust — do not accept scores without verification.
- **Generate first-draft feedback.** Ask AI to draft individualized feedback for each submission based on the rubric. NotebookLM is particularly useful here — upload your rubric and student submissions as source documents and prompt it to draft feedback anchored to your rubric criteria. Because NotebookLM grounds responses in uploaded sources, feedback stays tied to your rubric language rather than generating generic comments. Edit for accuracy and tone before returning to students.
- **Summarize long submissions.** For portfolios or reflective assignments, use AI to produce a summary of each student's key claims before reading in full. Prioritize where to focus close reading.
- **Flag outliers.** Ask AI to identify submissions that are significantly stronger, weaker, or inconsistent with prior work — useful for targeting grading attention and detecting potential integrity issues.
- **Check for consistency.** After grading a sample manually, ask AI to apply the same standards to the remainder and compare results. Use discrepancies to recalibrate.
- **Generate common feedback themes.** After grading, ask AI to analyze all submissions and identify the most common errors or gaps. Use this to inform a class-wide response rather than repeating the same feedback individually.

### Structural Approaches That Distribute the Load

- **Peer evaluation.** Have students use a UGA AI tool to evaluate peers' products or processes, then score how well the AI did. To assess the AI's performance, students must do the underlying evaluation themselves — making further delegation circular and self-defeating. Faculty spot-check a sample of peer grades rather than grading every submission. For this to work, rubric criteria must require discipline-specific judgment, contextual reasoning, or evaluation of process rather than product. Rubrics that read like checklists are effectively multiple choice tests that AI can navigate by pattern-matching.
- **Group work with individual reflection.** Grade the group deliverable once and the individual reflections separately. One product review covers multiple students; the reflection grades are short and structured.
- **Sampling strategies.** In large courses, randomly select a subset of submissions to grade in depth each week. Communicate this to students — the possibility of being selected maintains engagement without requiring full coverage every time.
- **TA calibration.** Before TAs begin grading, have them independently score the same 3–5 submissions and compare results. Invest time in calibration upfront to reduce inconsistency and rework later.

### Design Choices That Reduce Grading Volume

- **Grade checkpoints, not just final products.** Brief staged check-ins are faster to grade than full submissions and catch problems early. Feedback at stage two reduces the volume of revision at stage four.
- **Use completion grading for low-stakes work.** Reserve detailed rubric grading for high-stakes submissions. Low-stakes checks can be graded on completion, with AI flagging anything that appears off-task and creating reports about student response patterns.
- **Build reusable rubrics.** Invest time once in building a well-structured rubric that AI can apply consistently across terms. A rubric that works with AI grading tools saves significant time at scale.

---

# Assessment in Online Courses in the Age of AI Agents

University of Georgia • Office of Online Learning • March 2026

---

## ACTIONABLE RECOMMENDATIONS FOR UGA

### 1. Develop Tiered Assessment Guidelines

- Create context-specific guidance organized by class size and course type — not a single blanket policy.
- Provide worked examples from multiple disciplines, including graduate and professional programs, through the CTL.

### 2. Invest in Faculty Development and Reduce Workload Barriers

- Expand CTL course redesign institutes to include AI-era assessment design.
- Build peer networks of faculty who have successfully redesigned assessments.
- Prioritize large-enrollment gen ed instructors — highest impact, highest vulnerability.
- Invest in TA training and rubric infrastructure so that stronger assessment approaches do not simply increase individual faculty burden. Workload needs to be addressed structurally.
- Consider course release, stipends, or other incentives for faculty undertaking significant assessment redesign.

### 3. Connect Assessment Reform to the CLR and QEP

- Portfolio and reflective assessments feed directly into UGA's Comprehensive Learner Record.
- Align assessment redesign incentives with CLR adoption goals.
- AI-era assessment redesign advances UGA's active learning QEP — both require students to engage, apply, and reflect rather than retrieve.
- Co-scaffolded AI literacy is a transferable competency that fits within the CLR framework. Students who can critically direct and evaluate AI in a disciplinary context are more workforce-ready.

### 4. Address Equity, Accessibility, and Student Voice

- Build flexibility into new assessment models from the start and coordinate with Student Affairs and Disability Services before rollout.
- Assessment redesign should not disadvantage working students, caregivers, or students across time zones.
- Include students in developing AI-era assessment norms — not just in evaluating them after the fact.
- Some students may be conscientious AI objectors; please consider alternative assignments for that group.

### 5. Pilot, Assess, and Share

- Identify 2–3 programs to pilot redesigned assessment approaches in AY 2025–26, spanning undergraduate and graduate contexts.
- Adapt UGA's Office of Assessment methods to evaluate impact on student learning outcomes.
- Share results. Our scale and infrastructure position us to shape national and system-wide practice.

# Assessment in Online Courses in the Age of AI Agents

University of Georgia • Office of Online Learning • March 2026

## A CONSTRUCTIVE VISION: THE LEARNING AGENT

### Proof of Concept

UGA Online has developed a prototype that uses AI as the assessment engine itself. The **Learning Agent**:

- Ingests the instructor's **learning outcomes** for a course
- Allows the instructor to specify what **mastery looks like** in terms of both knowledge and observable student behavior
- Engages the student in a **Socratic dialogue** that probes depth of understanding rather than surface recall
- Documents the conversation and **identifies evidence of mastery**, flagging specific student responses that demonstrate outcome achievement
- Submits a summary, highlights, and full transcript to a **course dropbox for faculty verification** — the instructor makes the final determination

**Example:** A student selects the outcome "evaluate the quality and credibility of data-based claims." The agent presents a headline — "*Product X reduces cold duration by 50%*" — and works through sample size, generalizability, methodology, and corroboration with the student. The agent follows the student's reasoning and probes further at each step. A student delegating to another AI cannot navigate the conversation without genuine understanding of the material.

**View the proof of concept:** [https://kaltura.uga.edu/media/t/1\\_mapn0z0w](https://kaltura.uga.edu/media/t/1_mapn0z0w)

**Why this matters:** The Learning Agent addresses the highest-risk cells in the matrix — particularly large, high-stakes, low-interest required courses — but works across all three course types and enrollment sizes, including graduate and professional programs. It scales assessment without proportionally increasing faculty grading time and generates verifiable evidence of student thinking that supports both the QEP and the CLR. *Assessment validity and reliability will need to be established as the tool matures, including consistency across diverse student populations.*

## DISCUSSION QUESTIONS

- Q Where on the matrix is UGA most vulnerable right now? Where are we already well-positioned?
- Q How should assessment design differ between undergraduate, graduate, and professional programs at UGA?
- Q What do faculty need — in time, support, or incentives — to redesign assessments in large gen ed online courses?
- Q How do we ensure that assessment redesign does not disadvantage working students, caregivers, and non-traditional students?
- Q How do we define AI literacy as a disciplinary competency — and should it look different in a biology course than a business course?
- Q What role should students play in shaping AI-era assessment norms at UGA?
- Q What would faculty need to trust and adopt a tool like the Learning Agent — and where would it have the most immediate impact?