GLOW: AI-Simulated Students Improve GTA Readiness

Ethan Dickey, Ashok Saravanan, Alexander Siladie, Houyame Lkhider, Quiondriya Gee, Andres Bejarano

Rubric-anchored, persona-driven simulations provide repeatable office-hour rehearsal at cohort scale and produce analytics that help programs target coaching.

Demo: learn-loop.org

PURDUE UNIVERSITY
Department of Computer Science

Context and problem

In large CS programs, GTAs are the front line for office-hour questions that are simultaneously conceptual and emotional. Yet onboarding often emphasizes policies and logistics, leaving limited opportunities to rehearse the "hard conversations" (conflict, distress, disengagement) that shape student experience. Traditional role-plays are instructor-intensive and difficult to scale, while simplistic checklists risk drifting from pedagogical intent.

Intervention: GLOW

GLOW (Graduate Learning Orientation Workshop) is a chat-based practice environment where GTAs rehearse short office-hour conversations with AI student personas. After each simulation, GTAs receive immediate, domain-level feedback aligned to an educator-authored, behaviorally anchored rubric and are invited to retry a comparable scenario. The goal is rehearsal – practice plus targeted feedback – rather than one-shot assessment.

Rubric and Personas

Rubric domains: student-driven learning · conceptual understanding · time management · adaptation to student needs · professional communication.

Personas: Confused · Passive · Aggressive · Distressed (derived from TA-training literature and local focus groups; refined through pilot runs with experienced GTAs and instructional staff until judged realistic).

System Design

Each simulation uses one LLM to generate persona-consistent student dialogue and a second LLM to apply rubric anchors and produce interpretable feedback. This two-step design keeps pedagogy in the human-authored rubric while using the LLMs for generation and classification.

Fig. 1 (top): Representative excerpt from an AI-simulated Aggressive persona office-hour chat (anonymized).
Fig. 2 (bottom): Representative rubric-anchored feedback output (selected elements) from a separate session.

	N=266, deployed in F25 GTA training	Mean score (all attempts): 69%	Completion rate: 89%	1st attempt pass (≥80%): 51%

"Student responses felt fairly realistic... evaluations were very specific to our responses."

"... interacting with 'corner-case' students helped me a lot on thinking how I should interact with them professionally"

Fig. 3 (right): admin dashboard with live metrics about Fall 2025 GTA training at Purdue University (N=266).


Methods and Key Findings

Design: We integrated GLOW into mandatory GTA training in a research-intensive CS department (two cohorts; N=266). Participants completed multiple simulations against a preset benchmark, while a randomly assigned control group completed effort-matched discussion and quiz activities covering similar policies and scenarios. Evidence includes simulation analytics (attempts, rubric scores, persona outcomes) and post-training reflections to contextualize how GTAs experienced the practice loop. [*Fall 25: gpt-4.1 for personas and gpt-5 for grading*]

Key findings: Across cohorts, GTAs engaged with the simulator at high rates and generally met preset benchmarks within one to three attempts. Score trajectories improved across retries, consistent with GTAs using rubric feedback to adjust strategy rather than treating the simulation as a single assessment. Persona-level outcomes were diagnostic: Confused cases were most tractable, while Aggressive cases were harder and tended to depress communication/adaptation performance – pinpointing where targeted coaching should focus.
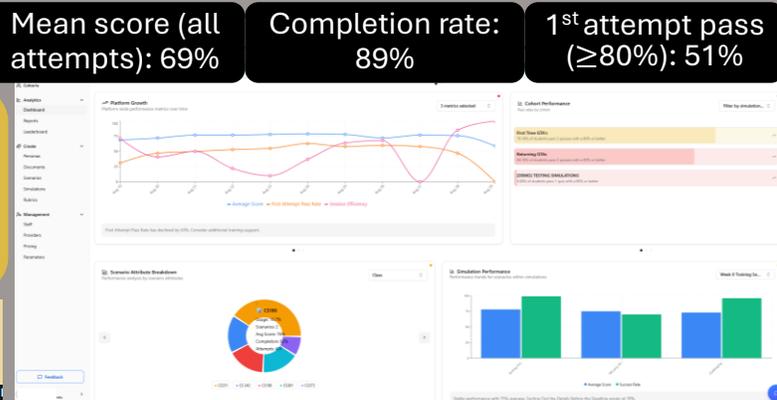
Implications for GTA Programs

Rubric-anchored simulation yields actionable precision: attempt counts, domain scores, and persona outcomes make growth observable for both individuals and program leaders. Rather than replacing human mentoring, the analytics help concentrate scarce coaching time on the highest-leverage skills (especially de-escalation and boundary-setting). Because the rubric domains are discipline-agnostic, the approach is portable to other STEM and peer-tutoring contexts.

Boundaries

These observations capture the onboarding window and do not claim long-term classroom impact. Simulations are text-based and cannot represent nonverbal cues; randomized practice introduces controlled heterogeneity but also noise. Self-report can inform readiness but is not a substitute for observation.

Demo + Materials + Contact

Live demo + scenario examples + full rubric + implementation notes (QR).
Handouts available with sample transcripts and rubric.
Email: dickeye at purdue.edu