

GLOW: AI-Simulated Students Improve GTA Readiness

Ethan Dickey
dickeye@purdue.edu
Purdue University
West Lafayette, IN, USA

Ashok Saravanan
Purdue University
West Lafayette, IN, USA

Alexander Siladie
Purdue University
West Lafayette, IN, USA

Houyame Lkhider
Purdue University
West Lafayette, IN, USA

Quiondriya Gee
Purdue University
West Lafayette, IN, USA

Andres Bejarano
abejara@purdue.edu
Purdue University
West Lafayette, IN, USA

Abstract

GLOW (Graduate Learning Orientation Workshop) is a chat-based practice environment in which Graduate Teaching Assistants (GTAs) rehearse office-hour conversations with AI student personas and receive feedback aligned to a human-authored, behaviorally anchored rubric. Deployed during mandatory GTA onboarding for two cohorts in a large research-intensive computer science department ($N = 266$), GLOW enabled authentic, repeatable practice at scale, showed consistent improvement across retries, and was associated with higher self-reported confidence for difficult cases than an effort-matched discussion-and-quiz condition. Persona-level analytics – especially lower success with Aggressive students – pinpointed where coaching should focus. We argue that rubric-anchored simulation is a pragmatic, scalable way to support GTA readiness without sacrificing pedagogical grounding.

Keywords

Graduate Teaching Assistants (GTAs), AI-simulated student interactions, Behaviorally anchored rubric, Persona-based training analytics, Scalable TA training

ACM Reference Format:

Ethan Dickey, Ashok Saravanan, Alexander Siladie, Houyame Lkhider, Quiondriya Gee, and Andres Bejarano. 2026. GLOW: AI-Simulated Students Improve GTA Readiness. In *Proceedings of the 57th ACM Technical Symposium on Computer Science Education V.2 (SIGCSE TS 2026)*, February 18–21, 2026, St. Louis, MO, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3770761.3777354>

1 Introduction

Graduate TAs (GTAs) are central to the undergraduate experience in large CS programs: they guide labs, hold office hours, and serve as the front line for questions that are simultaneously conceptual and emotional. In our department alone, more than 250 GTAs are assigned annually across introductory and upper-division computing courses, including programming, systems, and data structures, which makes consistent preparation and feedback a nontrivial coordination problem. Yet most orientations concentrate on policies

and grading tools, offering limited opportunities to practice the conversations that matter most – deescalating conflict, responding to distress, probing for understanding, and maintaining professional rapport under time pressure [1, 2]. Traditional role-plays are instructor-intensive and difficult to scale, while automated checklists risk optimizing for superficial behaviors that drift from pedagogical values [4–7]. At the same time, evidence from AI-supported CS instruction shows that compact, feedback-rich AI activities can scale practice without diluting instructional intent [3].

GLOW addresses this gap by coupling authentic simulation with a behaviorally anchored rubric authored by educators. The platform presents short, persona-driven office-hour chats (Confused, Passive, Aggressive, Distressed), returns domain-level feedback (student-driven learning, conceptual understanding, time management, adaptation, professional communication), and invites immediate retry on a comparable scenario. Our study asks whether this tight practice loop improves readiness during onboarding, what patterns appear across personas and domains, and how analytics can guide program-level coaching. Most GTA orientations emphasize policy and logistics while offering little opportunity to practice the hard conversations that shape students’ experiences: distress, grade disputes, or persistent confusion. Our contribution is an operational, program-level solution that integrates (1) authentic simulation with parametrized personas, (2) a set of instructional domains with behavioral anchors, and (3) analytics that make growth observable for individuals and for the program. Rather than replacing human mentoring, GLOW concentrates it: evidence from simulations directs where limited coaching time yields the most impact.

2 GLOW in Practice

A GLOW session¹ feels like a short office hour chat. GTAs interact with personas on a wide range of real assessments pulled from active courses, then receive immediate feedback in five domains: student-driven learning, conceptual understanding, time management, adaptation to student needs, and professional communication. The four core personas (Confused, Passive, Aggressive, Distressed) were derived from recurring patterns in prior TA-training literature and from local faculty and staff focus groups that surfaced common office-hour challenges, such as grade disputes, emotionally charged conversations, and students who disengage or defer excessively to the TA. Short pilot runs with experienced GTAs, teaching faculty, and instructional specialists refined personas until judged realistic.



This work is licensed under a Creative Commons Attribution 4.0 International License. *SIGCSE TS 2026, St. Louis, MO, USA*

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2255-4/2026/02
<https://doi.org/10.1145/3770761.3777354>

¹As of Nov 2025, GLOW is accessible at <https://glow.cs.purdue.edu>

The educator-authored rubric, designed to be pedagogically sound and AI-friendly, provides stable, interpretable domain feedback across scenarios. The loop is intentionally tight: prepare, simulate, read targeted feedback, and retry a comparable case. This structure supports rehearsal, not just assessment, and scales to large cohorts without additional faculty and staff load. Each simulation uses an LLM with persona prompts encoding affect and context; a second LLM applies rubric anchors to generate targeted feedback. This two-step design keeps pedagogy in the human-authored rubric while using the LLM only for generation and classification.

3 Study Overview

We integrated GLOW into the department’s mandatory GTA onboarding. Participants completed multiple simulations against a preset benchmark, while a control group engaged in effort-matched discussions and quizzes with similar policies and scenarios. In total, the broader orientation served 266 GTAs; for this poster we focus on simulation analytics, rubric-based scores, and surveys that capture scenario-specific confidence, perceived usefulness, and satisfaction.

4 Findings

Across two onboarding cohorts, GTAs engaged with the simulator at high rates and generally met preset benchmarks within one to three attempts. Score trajectories improved with retries, indicating that GTAs used feedback to adjust strategies in subsequent runs rather than treating the simulation as a single assessment. Qualitative comments emphasized the realism of the conversations and the usefulness of concrete, rubric-mapped suggestions (e.g., prompting students to articulate their plan before offering hints, or explicitly acknowledging frustration to reset tone).

Persona-level outcomes were diagnostic. Confused cases were more tractable, often rewarding clear scaffolding and pacing. Aggressive cases were markedly harder: GTAs received lower scores in professional communication and adaptation when the persona challenged their authority or escalated emotionally. This pattern signaled need for targeted coaching modules (e.g., language for boundary-setting, “naming and reframing”) and scenario refinements that stage deescalation opportunities earlier in the dialogue.

New and returning GTAs performed comparably on adaptation criteria once they had completed the simulation loop, suggesting that structured practice narrows experience gaps early. This finding tempers the assumption that tenure alone predicts readiness for affect-laden interactions and supports allocating practice opportunities to all GTAs rather than prioritizing new staff exclusively.

Self-report measures aligned with performance trends. Relative to an effort-matched discussion/quiz control, GTAs who used the simulation reported higher confidence for difficult scenarios (distress, grade disputes, confrontational behavior) after training. Satisfaction and perceived usefulness were positive; respondents highlighted immediate feedback and the chance to retry as key features.

5 Discussion

Taken together, the results indicate that rubric-anchored simulation offers three advantages in GTA onboarding. First, it creates repeatable, authentic practice that surfaces specific behaviors to improve. Second, persona variance exposes systematic weaknesses (notably deescalation and tone management), allowing programs to focus

scarce coaching time where it matters most. Third, generated analytics (attempt counts, domain scores, and persona outcomes) make growth visible to learners and program leads, enabling iterative refinement of training materials.

We note some important boundaries. Our observations capture the onboarding window; they do not claim long-term classroom impact. While quizzes required to “pass” were standardized, practice sessions were randomized, which introduces controlled heterogeneity but also noise. Notably, the simulations were text-based, failing to capture the nuance of body language (tone was mimicked with text formatting). Finally, while confidence gains are encouraging for readiness, self-report is not a substitute for observation.

6 Implications

For program leaders, rubric-anchored simulation yields actionable precision: dashboards and persona/domain breakdowns identify where GTAs struggle and where they are improving, enabling targeted workshops and mentored role-plays rather than generic refreshers. Because the rubric domains are discipline-agnostic, the approach is portable to other STEM and peer-tutoring settings.

7 Next Steps

Immediate next steps include (i) linking onboarding data to mid-semester observations and end-of-term feedback; (ii) developing micro-modules on deescalation and boundary setting for Aggressive personas; and (iii) iterating scenario prompts to better separate mid- and high-intensity cases while maintaining comparability.

Acknowledgments

Dickey, Saravanan, Siladie, and Bejarano acknowledge funding by Purdue’s Innovation Hub (IH-AI-23002).

References

- [1] Umair Z. Ahmed, Shubham Sahai, Ben Leong, and Amey Karkare. 2025. Feasibility Study of Augmenting Teaching Assistants with AI for CS1 Programming Feedback. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1* (Pittsburgh, PA, USA) (SIGCSETS 2025). Association for Computing Machinery, New York, NY, USA, 11–17. doi:10.1145/3641554.3701972
- [2] Ildar Akhmetov, Sadaf Ahmed, and Kezziah Ayuno. 2024. How We Manage an Army of Teaching Assistants: Experience Report on Scaling a CS1 Course. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) (SIGCSE 2024). Association for Computing Machinery, New York, NY, USA, 32–38. doi:10.1145/3626252.3630871
- [3] Ethan Dickey, Andres Bejarano, Rhianna Kuperus, and Bárbara Fagundes. 2025. Evaluating the AI-Lab Intervention: Impact on Student Perception and Use of Generative AI in Early Undergraduate Computer Science Courses. arXiv:2505.00100 [cs.CY] <https://arxiv.org/abs/2505.00100>
- [4] Runda Liu, Shengqi Chen, Jiajie Chen, Songjie Niu, Yuchun Ma, and Xiaofeng Tang. 2025. Iterative Design of a Teaching Assistant Training Program in Computer Science Using the Agile Method. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1* (Pittsburgh, PA, USA) (SIGCSETS 2025). Association for Computing Machinery, New York, NY, USA, 680–686. doi:10.1145/3641554.3701829
- [5] Emily Marshman, Ryan Sayer, Charles Henderson, Edit Yerushalmi, and Chandralekha Singh. 2018. The Challenges of Changing Teaching Assistants’ Grading Practices: Requiring Students to Show Evidence of Understanding. *Canadian Journal of Physics* 96, 4 (2018), 420–437. doi:10.1139/cjp-2017-0030
- [6] Mia Minnes. 2022. Designing TA Training for Computer Science Graduate Students: Remote and Self-paced Options for A Supported Introduction to Reflective Teaching. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - Volume 1* (Providence, RI, USA) (SIGCSE 2022). Association for Computing Machinery, New York, NY, USA, 752–758. doi:10.1145/3478431.3499342
- [7] Emma Riese, Madeleine Lorås, Martin Ukrop, and Tomáš Effenberger. 2021. Challenges Faced by Teaching Assistants in Computer Science Education Across Europe. In *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1* (Virtual Event, Germany) (ITiCSE ’21). Association for Computing Machinery, New York, NY, USA, 547–553. doi:10.1145/3430665.3456304