



Adapting Intelligent Tutoring Systems for Technical Workforce Training: A Case Study of Two Cybersecurity Tutor Designs

Marshall An^(✉) , Leah Teffera , Mahboobeh Mehrvarz , Can Kultur ,
Bruce M. McLaren , and Christopher Bogart

Carnegie Mellon University, Pittsburgh, PA 15213, USA
{haokanga,lteffera,bmclaren,cbogart}@andrew.cmu.edu, ckultur@cs.cmu.edu

Abstract. Intelligent Tutoring Systems (ITSs) have a strong evidence base, yet their use in technical workforce training remains limited. This paper documents an iterative co-design process with community college faculty that produced two ITS adaptations for cybersecurity workforce training. One tutor targets DDoS incident response and embeds a lightweight interactive simulation to make infrastructure state and mitigation effects observable during guided practice. The other targets security controls and scaffolds contrastive reasoning to support conceptual classification across related taxonomies emphasized in certification exam guides. We evaluated each tutor against a non-ITS comparison activity using randomized controlled trials within an authentic workforce training pathway. Results provide unit-specific evaluative evidence that ITS effectiveness was not uniform: the simulation-based tutor yielded a learning advantage concentrated in knowledge components requiring causal reasoning about mitigation impacts, whereas the contrastive-reasoning tutor did not outperform its comparison condition. Because simulation is confounded with instructional topic and implementation, we treat these as unit-specific case findings rather than a causal comparison of tutor types. Together, this work contributes a documented co-design process and evaluative evidence illustrating how alignment between tutor affordances and the cognitive demands of targeted skills may shape when ITS support adds value in workforce settings, while also surfacing design tensions—such as perceived repetitiveness under mastery-oriented task-loop adaptivity—that warrant further investigation.

Keywords: Intelligent tutoring systems · Workforce training · Simulation-based learning · Cybersecurity education

1 Introduction

Community colleges in the United States, accessible to tens of millions of workers, play a vital role in workforce training, particularly for technical job roles. They prove effective in technical education, with returns comparable to four-year

programs [31]. However, these institutions face significant resource constraints [21, 29, 34], motivating interest in effective and scalable teaching methods such as Intelligent Tutoring Systems (ITSs).

Despite ITSs’ strong record of improving average learning outcomes across domains [27, 35], technical workforce training remains comparatively underrepresented in the ITS literature [17]. Although prior work has argued that ITS benefits may extend to workplace training and continuing professional development, empirical evidence with adult learners in workforce-oriented settings remains limited [17, 18]. This gap is reflected in a recent systematic review that identified only 10 empirical studies on professional training out of 782 screened records over a five-year period [30]. Consequently, there is limited design guidance on how to adapt ITSs to realistic workforce-training constraints and on which design choices, if any, are likely to yield measurable learning benefits in these contexts.

In this paper, we document the iterative co-design, implementation, and evaluation of two ITSs for cybersecurity workforce training, developed in collaboration with community college faculty. An earlier experience report detailed the co-design process and formative classroom pilots for these tutors [6]. The present study extends that work by evaluating the refined tutors in a larger workforce-training cohort and examining how learning outcomes varied across tutor designs and targeted knowledge components (KCs) [24]. One tutor targets DDoS incident response and embeds a lightweight interactive simulation to support decision-making and causal reasoning about the effects of mitigation actions; the other tutor targets security controls and scaffolds contrastive reasoning to support conceptual classification across related taxonomies required by industry-aligned certification exams. We do not aim to isolate simulation as a standalone design feature; rather, we examine how two distinct ITS designs were adapted to different instructional challenges under practical workforce-training constraints. We evaluated each tutor against a non-ITS comparison activity using randomized controlled trials within an authentic workforce training pathway.

Our results provide unit-specific evidence that ITS effectiveness was not uniform across the two instructional units in the targeted technical workforce training context. The simulation-based tutor yielded learning gains on KCs requiring reasoning about the causal effects of mitigation actions, whereas the contrastive-reasoning tutor did not outperform its comparison condition. Because simulation is confounded with instructional topic and tutor implementation, we interpret cross-unit differences descriptively rather than as a causal estimate of simulation as an isolated feature. The contribution of this work is therefore a documented case study of how two ITS adaptations were co-designed, implemented, and evaluated under authentic workforce-training constraints, with implications for aligning ITS affordances with the cognitive demands of targeted KCs. We also discuss unresolved design tensions, such as perceived repetitiveness under mastery-oriented task-loop adaptivity, and directions for future investigation.

2 Related Work

2.1 Operationalizing ITS Adaptivity for Workforce Tasks

Research shows ITSs can outperform traditional instruction [27] and achieve effects similar to human tutoring [35]. One evidence-based approach to designing effective ITSs is to incorporate adaptive learning following the “Adaptivity Grid” framework [4], which delineates three forms of adaptivity: (i) step-loop adaptivity (within a problem), where the system responds to individual student actions within an instructional task; (ii) task-loop adaptivity (between problems), where the system personalizes subsequent tasks for the learner; and (iii) design-loop adaptivity (between material versions), where course designers update the content and system based on data-driven reflection. This framework has guided many empirical studies examining how different forms of adaptivity affect learning. In particular, research shows that task-loop adaptivity can accelerate learning by tailoring subsequent practice based on individual progress [11, 16, 26].

Step-loop and task-loop adaptivity in ITSs are typically implemented through representations of target skills as *condition–action* pairs. This requirement is grounded in ACT-R, a cognitive architecture in which cognition is modeled through production rules that specify the conditions under which an action should be taken [7, 8]. For example, in ITSs developed using Cognitive Tutor Authoring Tools (CTAT) [1, 2], this condition–action structure is directly reflected in tutor implementations. In model-tracing tutors, productions are explicitly encoded as IF–THEN rules that specify when an action is correct and what behavior is expected; in example-tracing tutors [3], the same structure is represented implicitly through behavioral graphs that enumerate acceptable subsequent action(s) given the current problem-solving state (conditions).

Applying these adaptivity best practices in technical workforce domains therefore entails meticulous design effort. In particular, it requires job-aligned scenarios with sufficient fidelity to instantiate the *conditions* under which production rules should activate, while also making learner *actions* and their effects observable for feedback and adaptive practice. In many real-world IT job tasks, however, these conditions and action consequences depend on costly, risky, or operationally complex infrastructure that is difficult to provision for instruction at scale, as illustrated in subsequent sections. These constraints can limit the ability to encode job-relevant skills as production rules and to operationalize step- and task-loop adaptivity in realistic settings. This motivated us to incorporate lightweight simulations that represent key system states and action consequences within the tutor.

2.2 Simulation and Integration with ITSs

Simulation has a well-established record of supporting learning [12, 19], but its integration into ITS-based workforce training remains limited. Recent reviews describe a growing body of simulation-based training research while noting that

adaptive tutoring is largely absent from these systems [30]; earlier work similarly found only isolated ITS prototypes combining domain instruction with simulation components [33].

3 Iterative Instructional Co-design

We adopted iterative co-design with community college faculty to develop ITS adaptations that were instructionally meaningful, feasible within institutional constraints, and aligned with workforce-relevant learning objectives. Co-design is a collaborative design method that engages stakeholders as partners throughout design processes, rather than as “sources of information” [32].

3.1 Instructional Design Context

The target learner population was drawn from the Community College of Allegheny County, a large public two-year college in the United States. We grounded our co-design in a core cybersecurity course that serves both workforce certificate and associate degree pathways and functions as a gateway to more advanced topics. The course objectives are aligned with CompTIA Security+ [20], a widely recognized entry-level cybersecurity credential, and with U.S. Department of Defense (DoD) directive 8140/8570.01-M requirements [14].

3.2 Identifying High-Priority Instructional Challenges

Through curricular analysis and faculty reflection, we identified two recurring instructional challenges. One challenge concerned developing practical cybersecurity skills: delivering realistic, hands-on experiences with cloud infrastructure (e.g., detecting, analyzing, and mitigating DDoS attacks) is typically expensive and infeasible for community colleges, limiting opportunities for experiential practice. A second challenge concerned comparing closely related taxonomies: business-as-usual textbook materials emphasized definitions presented in isolation and offered little instructional support for contrasting categories, despite assessment tasks that require such distinctions. These challenges motivated two corresponding ITS designs: a DDoS incident response tutor that provides experiential learning and supports causal reasoning through lightweight simulations of workplace-relevant tasks, and a security-control tutor that scaffolds contrastive reasoning through tailored classification exercises. Both were refined through formative classroom pilots before evaluation.

3.3 DDoS Incident Response Tutor for Causal Reasoning

The first tutor targets a workforce-relevant competence that faculty identified as especially challenging to teach in their community-college classes: incident response to DDoS attacks for IT infrastructure. To provide experiential learning without requiring costly lab infrastructure, we designed a lightweight interactive

simulation embedded within an ITS (Fig. 1). The tutor walks learners through a DDoS incident response workflow aligned with the NIST Incident Response Life Cycle [13], including detecting anomalous traffic and infrastructure utilization, diagnosing possible causes, selecting and applying mitigation techniques (e.g., scaling and rate limiting), and reflecting on outcomes.

The tutor targets 18 KCs spanning the incident response workflow. During co-design, we distinguished between two categories of KCs that differed in whether they required learners to interpret dynamic system-state changes. Simulation-Independent General Knowledge (SIGK) KCs involve procedural and conceptual knowledge that can be assessed without observing dynamic infrastructure behavior, such as correlating user-reported connectivity failures to identify a potential DDoS attack, or prioritizing infrastructure monitoring following user-reported connectivity failures. Simulation-Dependent Causal Reasoning (SDCR) KCs, by contrast, require learners to observe and interpret how infrastructure metrics change in response to mitigation actions, such as recognizing that infrastructure scaling does not directly reduce traffic, or recognizing that rate limiting reduces utilization by blocking malicious traffic. These relationships are difficult to support through static text alone because learners must connect a selected mitigation action to its observable effects on traffic and utilization. This distinction motivated embedding a lightweight simulation within the tutor rather than relying on static instructional materials alone. Without observable state changes, the relevant production-rule conditions for SDCR KCs are harder to instantiate, limiting assessment and feedback on learners’ causal reasoning. The simulation also enabled richer step-loop support than static materials alone could readily provide: because learners could observe traffic and utilization metrics updating in real time, multi-level hints could guide their interpretation before revealing the correct answer. For example, hints could prompt learners to examine whether traffic volume changes after scaling, addressing a misconception observed during formative pilots: some learners assumed that provisioning additional resources directly reduces incoming traffic, when scaling instead allows the infrastructure to handle the same traffic volume more effectively. In this way, the tutor grounded targeted support in observable system behavior rather than abstract assertion.

We particularly document the design and implementation of task-loop adaptivity in this tutor because its scenario-based, multi-step incident-response procedure, grounded in the NIST Incident Response framework, poses distinct challenges for adaptive practice. A formative pilot suggested that simulation alone may not reliably support all learners: most learners (4 out of 5) progressed through the DDoS tutor with no substantial learning gains. Examination of fine-grained interaction logs collected through DataShop [23] revealed that some learners exhibited rapid, repeated incorrect attempts, a behavioral pattern consistent with rapid-guessing behaviors documented in the “gaming the system” literature [9, 10]. These observations, discussed with course faculty during co-reflection sessions, informed our decision to add task-loop adaptivity to (1) provide targeted, repeated practice on unmastered KCs and (2) discourage gaming

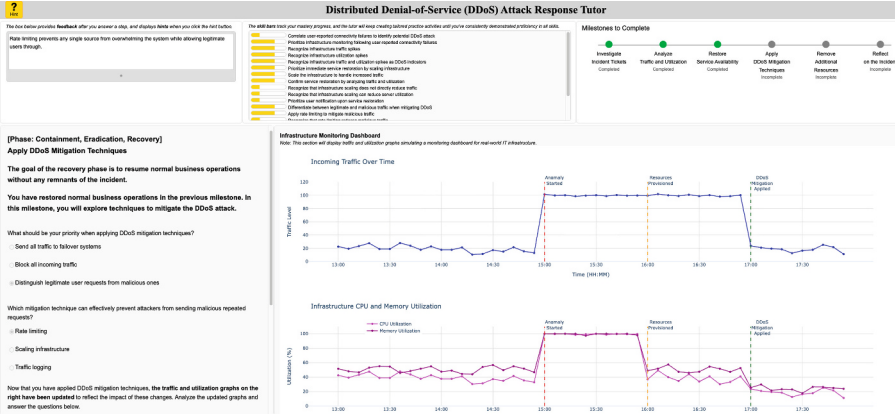


Fig. 1. DDoS Incident Response Tutor. Top left: hint box offering multi-level hints and immediate feedback. Top right: milestone tracker segments the complex DDoS response workflow into milestones and indicates student progress for reflection. Bottom left: steps for the current milestone guide the student through the procedure. Bottom right: simulation dashboard presenting interactive network-traffic and resource-utilization graphs that update in response to mitigation techniques applied by students.

behaviors by preventing learners from bypassing core practice and progressing without demonstrated understanding.

However, implementing task-loop adaptivity in the DDoS tutor posed a challenge in representing the conditions and action consequences discussed in Sect. 2.1. Unlike many ITS domains where unmastered KCs can be practiced through independent problems drawn from a pool, incident-response skills are embedded in a scenario whose steps are tightly coupled to evolving system state and prior actions, such as diagnosis, mitigation, and recovery decisions. Presenting individual KC-linked steps in isolation would remove the contextual information needed to define the relevant production-rule conditions, whereas repeating the entire procedure until all KCs were mastered would risk unnecessary overpractice. Our solution was to repeat the scenario while dynamically adapting the interface based on estimated skill mastery, operationalized through Bayesian Knowledge Tracing (BKT) [15] over KC-linked learner performance. After the initial round, the system provides additional rounds of practice in which steps linked to mastered skills are auto-completed, allowing learners to focus on unmastered steps until all reach mastery.

This adaptive interface design draws inspiration from Huang et al. [22], but differs in operating at the task-loop level by auto-completing mastered steps, rather than at the step-loop level by providing additional scaffolded steps in response to errors. Because this implementation differs from conventional task-loop adaptivity based on selecting additional independent tasks, it warranted particular scrutiny during evaluation, which we return to in the Results and Discussion.

3.4 Security Control Tutor for Contrastive Reasoning

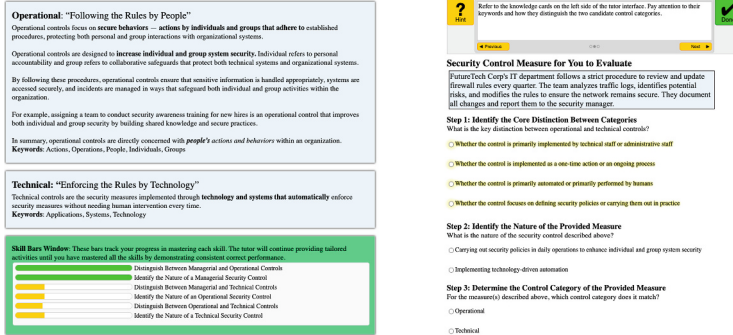


Fig. 2. Security Control Tutor. Bottom-left: Skill bars track mastery progress for each skill, enabling adaptive practice selection. Bottom-right: Contrastive reasoning structured into a scaffolded three-step comparison process. Top-left: Introduction to taxonomies using plain language with comparative scaffolding, supplemented by visual aids. Top-right: Hint box providing multi-level hints, and immediate misconception-targeted feedback.

The second tutor addresses a recurring pedagogical limitation of the textbook (CompTIA Security+ Cert Guide [20]) that the faculty and learning scientists collaboratively identified: it confines cybersecurity taxonomies to isolated subsections without comparison across them. For example, in Chap. 1 each control type (e.g., preventive, detective, corrective) is introduced in its own subsection, with definitions and examples presented independently. This gap is particularly consequential given its misalignment with target assessments—the Cert Guide’s evaluation instruments rely heavily on classification-based multiple-choice questions (MCQs). To address this gap, the Security Control Tutor (Fig. 2) was developed to scaffold contrastive reasoning aligned with Cert Guide assessment demands by presenting exercises in which learners compare and justify classifications across categories within authentic scenarios.

The tutor targets 11 KCs spanning security-control taxonomies. Rather than asking learners to memorize definitions alone, the scaffolded comparison process guides them through three steps: identifying the core distinction between relevant categories, characterizing the nature of the described control measure, and then determining the matching control category. For example, faculty identified distinguishing managerial from operational controls as a common learner struggle. In such scenarios, the tutor first guides learners to identify the core distinction as whether the measure focuses on defining security policies or on carrying them out in practice. Learners then analyze the provided scenario and use that distinction to determine whether the control is managerial or operational. Multi-level hints and misconception-targeted feedback were designed to support these distinctions explicitly.

As with the first tutor, formative classroom pilots motivated the incorporation of task-loop adaptivity using BKT; meanwhile, adaptivity in this tutor followed a more conventional pattern, selecting subsequent problems from a pool targeting unmastered KCs.

4 Evaluation Design and Participants

A pragmatic challenge in evaluating ITSs within authentic community college settings is that individual cohorts are often small, which can limit statistical power for controlled comparisons within a single course offering. Across three formative classroom deployments, tutor use supported iterative design refinement, but the number of learners who both consented to research use and completed all required activities ranged from only 4 to 9 per deployment. These sample sizes were consistently insufficient to support a definitive randomized evaluation within the available study timeframe.

To enable a more adequately powered randomized evaluation while maintaining workforce relevance, we recruited participants from a complementary workforce training pathway: an 8-month technician training program for government personnel in which cloud computing and security-related technical skills are core instructional components. Although this population differs from community college students in organizational context and prior experience, both groups consist of adult learners in technical workforce training pathways oriented toward industry-aligned skill development. Importantly, while the surrounding curricula differ, the DDoS incident response and security control classification skills targeted by the ITSs were validated as required learning objectives by instructors from both the community college and technician program contexts.

The evaluation comprised two concurrent randomized experiments, one for each instructional unit. In the DDoS unit, the experimental condition used the ITS, whereas the control condition used equivalent static content without simulation or adaptive repeated practice. In the Security Control unit, the experimental condition used the ITS, whereas the control condition used equivalent content without contrastive-reasoning scaffolds or adaptive repeated practice. To use instructional time efficiently and ensure exposure to both topics and conditions, we used a counterbalanced assignment scheme: each participant experienced the experimental condition for one unit and the control condition for the other, with random assignment to one of two sequences. We assessed learning with unit-specific pre- and post-tests and collected open-ended post-activity survey responses about helpful, frustrating, or unclear tutor features.

Participants came from one cohort of 36 enrolled learners. Of these, 27 consented and completed all required DDoS-unit activities, and 25 also completed all required Security Control activities. The sample was predominantly male (96.3%, $n = 26$; female: 3.7%, $n = 1$). This gender imbalance reflects the composition of the underlying training cohort and broader patterns of gender representation in technical training pipelines. Many reported substantial prior technical exposure: 66.7% ($n = 18$) held at least one cloud or security-related certification,

and most reported prior programming experience and multiple years of service. Given the small sample and skewed gender distribution, demographic variables were not included as covariates or interaction terms; results should therefore be interpreted as instructional effects within this workforce training population rather than as demographic generalizations.

5 Results

Table 1. Descriptive Statistics for Pretest and Posttest Scores by Tutor and Condition

Tutor	Condition	N	Pre-test M (SD)	Post-test M (SD)	Gain M (SD)
DDoS Incident Response Tutor (10-item quiz)					
	Experimental	14	7.71 (2.20)	9.43 (0.85)	1.71 (2.30)
	Control	13	7.08 (1.50)	7.54 (2.79)	0.46 (2.30)
	Total	27	7.41 (1.89)	8.52 (2.21)	1.11 (2.34)
Security Control Tutor (12-item quiz)					
	Experimental	14	7.86 (3.48)	9.50 (1.91)	1.64 (2.27)
	Control	11	7.82 (1.83)	9.64 (1.86)	1.82 (2.56)
	Total	25	7.84 (2.82)	9.56 (1.85)	1.72 (2.35)

5.1 Learning Effects Relative to Control Within Each Unit

Table 1 presents pre-test, post-test, and gain scores for all groups. To assess learning effects relative to the control condition within each unit, we estimated separate OLS regression models predicting post-test performance for each tutor, controlling for pre-test scores.

For the DDoS unit, the overall model was statistically significant ($F(2, 24) = 4.44$, $p = .023$, adjusted $R^2 = .209$). Instructional condition was a significant predictor of post-test performance: learners in the experimental condition outperformed those in the control condition ($\beta = 1.68$, $SE = 0.77$, $t = 2.18$, $p = .039$, 95% CI [0.09, 3.26]). Pre-test performance showed a positive but non-significant association with post-test scores ($\beta = 0.34$, $SE = 0.21$, $p = .118$, 95% CI [-0.09, 0.76]).

In contrast, the contrastive-reasoning tutor did not outperform its comparison activity. While the overall model was also statistically significant ($F(2, 22) = 5.07$, $p = .016$, adjusted $R^2 = .253$), the instructional condition was not a significant predictor of post-test performance ($\beta = -0.15$, $SE = 0.64$, $t = -0.23$, $p = .817$, 95% CI [-1.49, 1.19]). Pre-test performance was the only significant predictor ($\beta = 0.37$, $SE = 0.12$, $t = 3.18$, $p = .004$, 95% CI [0.13, 0.61]).

Taken together, these results indicate that, although both tutors were designed and refined in accordance with the “Adaptivity Grid” framework [4], incorporating step-loop, task-loop, and design-loop adaptivity, only the simulation-based DDoS tutor showed a measurable learning advantage over its control condition. Across the two units, observed benefits appeared to depend on how the tutor’s affordances aligned with the specific cognitive demands of the targeted KCs rather than on adherence to general ITS best practices alone.

5.2 Within the Simulation-Based Tutor: Effects on Causal-Reasoning and General-Knowledge KCs

To investigate the specific value of the simulation component, we used the SIGK–SDCR distinction established *a priori* during co-design. This distinction was central to the tutor design because SDCR KCs require learners to identify the causal effects of mitigation actions on infrastructure states, whereas SIGK KCs can be assessed without observing dynamic infrastructure behavior. The SDCR category therefore aligns directly with the affordances of simulation for observing and explaining causal effects, drawing on the established educational principle that observing causal effects supports learning [5]. Accordingly, the simulation was designed to let learners observe how infrastructure metrics changed in response to their actions through visualizations with reasonable fidelity to real-world cloud monitoring tools (e.g., AWS CloudWatch, Azure Monitor). To further promote causal reasoning, whenever learners selected a mitigation action, the tutor prompted them to reflect on and identify the observed effects on traffic and system utilization as required steps in the workflow.

The SDCR category comprised four quiz items, and the SIGK category comprised six quiz items. To examine whether learning effects were more evident for SDCR than SIGK items, we estimated separate OLS regression models for each KC category, controlling for pre-test scores. For SDCR items, the overall model was statistically significant, $F(2, 24) = 4.08$, $p = .030$, with $R^2 = .254$ (adjusted $R^2 = .191$). Instructional condition was a significant predictor of post-test performance: learners in the experimental condition outperformed those in the control condition ($\beta = 1.14$, $SE = 0.43$, $t = 2.65$, $p = .014$, 95% CI [0.25, 2.03]). Pre-test performance was not a significant predictor ($\beta = 0.09$, $SE = 0.22$, $p = .691$). In contrast, for SIGK items, the overall model did not reach conventional levels of statistical significance, $F(2, 24) = 3.01$, $p = .068$, with $R^2 = .201$ (adjusted $R^2 = .134$). The effect of instructional condition was not statistically significant ($\beta = 0.67$, $SE = 0.44$, $t = 1.52$, $p = .141$, 95% CI [−0.24, 1.57]). Pre-test performance showed a positive but non-significant association with post-test performance ($\beta = 0.27$, $SE = 0.15$, $p = .078$).

Taken together, these results suggest that the simulation-based tutor’s measured advantage was more evident for KCs that required reasoning about the causal effects of mitigation actions, whereas evidence of added value for general incident-response knowledge was weaker in this sample. In the next subsection, we analyze learners’ open-ended survey responses to contextualize these patterns and to surface perceived mechanisms and design trade-offs.

5.3 Survey Feedback Analysis

To complement the quantitative results and interpret the associated design trade-offs, we analyzed learners' post-activity survey responses. The goal was not to make standalone claims of effectiveness, but to identify recurring themes that help explain the observed learning patterns and implementation challenges.

A dominant theme concerned the value of the interactive simulation for understanding mitigation effects. Learners frequently described the visualizations as helping them reason about how traffic patterns and resource utilization changed in response to their decisions. One learner noted that "the graphs definitely helped visualize and better understand the concepts in the Detection and the Recovery phases," and that observing infrastructure state before and after mitigation techniques was "extremely helpful." Another learner similarly described the visual aids as supporting their understanding of the relationships among network traffic, CPU and memory utilization, and scaling. These responses align with the quantitative finding that the simulation-based tutor's measured advantage was more evident for items requiring causal reasoning about the impact of actions, suggesting that learners valued the simulation's support for action-effect reasoning rather than engaging with it only at a surface level.

At the same time, feedback also revealed interface-level trade-offs related to the simulation's prominence. Some learners felt that the visualizations competed with the text-based procedural steps. One observed that "the graphics were good, [but] they just didn't need to be the focus of the layout. They support the written material, not the other way around."

Learners also reported mixed perceptions of task-loop adaptivity. Some participants appreciated repeated practice and explicitly endorsed repetition as supporting mastery (e.g., "I liked the repetition since I agree that it's the best way to master a subject"), whereas others reported frustration with perceived repetitiveness and time cost (e.g., "the repetitive nature of the tutor, very time-consuming"). We interpret this cautiously: on one hand, repetition is to some extent inherent to task-loop adaptivity, which by design provides additional practice for KCs that have not yet reached mastery according to the model; meanwhile, learner feedback highlights a persistent design challenge—how to preserve the instructional benefits of adaptive repetition while improving how it is experienced by learners within realistic constraints on development effort and instructional time. Feedback on step-loop adaptivity was generally positive, with learners noting the usefulness of targeted feedback and hints.

In contrast, survey responses related to the Security Control tutor were generally less differentiated, with fewer comments pointing to specific learning-related affordances, mirroring the lack of condition differences in the quantitative results.

Overall, the survey feedback provides converging evidence that perceived learning support depended on alignment between tutor affordances and the targeted KCs, while also surfacing a tension between structured adaptive practice and learners' perceptions of repetition and time cost.

6 Discussion

Although this study does not isolate simulation as an independent causal factor, the observed pattern is consistent with situated and experiential learning theories, which emphasize contextually meaningful activity when learning depends on understanding action consequences within a system (e.g., [25,28]). In the DDoS tutor, the simulation externalized system states such as traffic and utilization, made mitigation effects observable, and required learners to articulate these effects. This design aligns with the SDCR results: the experimental advantage was concentrated in KCs requiring causal reasoning about mitigation impacts, whereas evidence of added value for general incident-response knowledge was weaker. The implication is not that simulation substitutes for hands-on experience, but that a minimal simulation embedded in an ITS can feasibly represent the conditions and consequences needed for step-level feedback and adaptive practice in resource-constrained workforce settings. Future work should test whether this alignment between “observable action effects” and targeted KCs generalizes across additional technical topics and alternative implementations.

6.1 Limitations and Future Work

Several limitations constrain interpretation and motivate targeted future work. First, simulation is confounded with instructional topic and implementation: the simulation-based tutor targeted DDoS incident response, whereas the contrastive-reasoning tutor targeted security control classification. Cross-unit differences therefore cannot be interpreted as a causal estimate of simulation as an isolated feature. Future work should evaluate alternative implementations within a single topic, varying interactive simulation versus static representations while holding learning objectives, assessments, and adaptive logic constant.

Second, the evaluation sample came from a government technician training program rather than the community college context for which the tutors were co-designed. Participants reported relatively high prior technical experience and certification, which may have reduced the marginal benefit of some ITS adaptations, especially for conceptual and taxonomic knowledge. Replication with community college learners is needed to assess whether the same design patterns hold in introductory workforce pathways.

Third, learning outcomes were assessed with unit-aligned pre- and post-tests over a short time horizon. These measures are appropriate for immediate learning effects but do not capture retention or transfer to novel scenarios, which are especially important for incident response skills requiring adaptation to various system states and constraints. Future evaluations should incorporate delayed post-tests and transfer tasks.

Fourth, learner feedback revealed mixed perceptions of task-loop adaptivity, especially repetitiveness and time cost. This motivates further tuning of BKT parameters, such as mastery thresholds and slip/guess assumptions, and more meaningful variation across practice rounds, such as differing system parameters or constraints, to preserve instructional value while improving learner experience.

Finally, the modest sample size and skewed gender distribution limited statistical power and precluded analyses of subgroup differences. While this reflects realistic constraints of authentic workforce training cohorts, future multi-cohort studies are needed to examine robustness across populations and whether specific design features differentially benefit learners with varying backgrounds or prior experience.

Together, these limitations highlight that adapting ITSs for workforce training requires not only adherence to established tutoring principles but also careful alignment among instructional affordances, targeted KCs, learner populations, and practical constraints. Addressing these limitations through targeted follow-up studies will support the development of more generalizable design principles for workforce-oriented ITSs.

6.2 Conclusion

This paper contributes a documented case study of adapting ITSs for technical workforce training under practical constraints, resulting in two tutor designs and their randomized evaluation in authentic workforce training settings. Across the two instructional units, ITS effectiveness was not uniform: the simulation-based incident-response tutor outperformed its control, with advantages concentrated in KCs requiring causal reasoning about mitigation impacts, whereas the contrastive-reasoning security-control tutor did not outperform its comparison condition. Although the study does not isolate simulation as an independent causal factor, the results and learner feedback support a design-alignment interpretation: workforce-oriented ITS benefits may depend on whether the tutor introduces affordances that directly support the cognitive work required by the targeted KCs, especially when job-relevant conditions and action consequences are otherwise difficult to represent. These findings motivate future research that systematically varies simulation within topics, expands outcome measures to retention and transfer, and refines adaptive repetition designs to better fit workforce training constraints.

Acknowledgments. This material is based upon work partially supported by the National Science Foundation, under Grant No. 2222762. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Aleven, V., McLaren, B., Sewall, J., Koedinger, K.R.: Example-tracing tutors: a new paradigm for intelligent tutoring systems (2009)
2. Aleven, V., McLaren, B.M., Sewall, J., Koedinger, K.R.: The Cognitive Tutor Authoring Tools (CTAT): preliminary evaluation of efficiency gains. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 61–70. Springer, Heidelberg (2006). https://doi.org/10.1007/11774303_7

3. Aleven, V., et al.: Example-tracing tutors: intelligent tutor development for non-programmers. *Int. J. Artif. Intell. Educ.* **26**, 224–269 (2016)
4. Aleven, V., McLaughlin, E.A., Glenn, R.A., Koedinger, K.R.: Instruction based on adaptive learning technologies. *Handbook Res. Learn. Inst.* **2**, 522–560 (2016)
5. Alexander, D., Haysom, J., Bowen, M.: Predict, observe, explain: activities enhancing scientific understanding, NSTA Press (2010)
6. An, M., Mehrvarz, M., Teffera, L., Kisow, M., McLaren, B.M., Bogart, C.: Partnering with community college faculty to co-design intelligent tutoring systems for cybersecurity workforce training. In: *Proceedings of the 57th ACM Technical Symposium on Computer Science Education V. 1*, pp. 52–58 (2026)
7. Anderson, J.R.: Automaticity and the ACT theory. *Am. J. Psychol.*, 165–180 (1992)
8. Anderson, J.R.: ACT: a simple theory of complex cognition. *Am. Psychol.* **51**(4), 355 (1996)
9. Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z.: Off-task behavior in the cognitive tutor classroom: when students “game the system”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 383–390 (2004)
10. Baker, R.S., Roll, I., Corbett, A.T., Koedinger, K.R.: Do performance goals lead students to game the system? In: *AIED*, pp. 57–64 (2005)
11. Bernacki, M.L., Greene, M.J., Lobczowski, N.G.: A systematic review of research on personalized learning: personalized by whom, to what, how, and for what purpose (s)? *Educ. Psychol. Rev.* **33**(4), 1675–1715 (2021)
12. Chernikova, O., et al.: Personalization through adaptivity or adaptability? A meta-analysis on simulation-based learning in higher education. *Educ. Res. Rev.* **46**, 100662 (2025)
13. Cichonski, P., Millar, T., Grance, T., Scarfone, K., et al.: Computer security incident handling guide. *NIST Spec. Publ.* **800**(61), 1–147 (2012)
14. CompTIA Partners: CompTIA Security+ Certification Overview (2025). <https://partners.comptia.org/certifications/security>. Accessed 14 June 2025
15. Corbett, A.T., Anderson, J.R.: Knowledge tracing: modeling the acquisition of procedural knowledge. *User Model. User-Adap. Inter.* **4**, 253–278 (1994)
16. Eglington, L.G., Pavlik, P.I., Jr.: How to optimize student learning using student models that adapt rapidly to individual differences. *Int. J. Artif. Intell. Educ.* **33**(3), 497–518 (2023)
17. Gupta, A.: Designing and deploying scalable intelligent tutoring systems to enhance adult education, Drexel University (2025). Ph.D. thesis
18. Gupta, A., Siddiqui, M., Smith, G., Reddig, J., MacLellan, C.: Intelligent tutors for adult learners: an analysis of needs and challenges. *arXiv preprint arXiv:2412.04477* (2024)
19. Hallinger, P., Wang, R.: The evolution of simulation-based learning across the disciplines, 1965–2018: a science map of the literature. *Simulat. Gam.* **51**(1), 9–32 (2020)
20. Heuermann, L.: *CompTIA Security+ SY0-701 Cert Guide*. Pearson IT Certification (2024)
21. Holzer, H.J., Lipson, R., Wright, G.: Community colleges and workforce development: Are they achieving their potential? (2023)
22. Huang, Y., et al.: A general multi-method approach to data-driven redesign of tutoring systems. In: *LAK21: 11th International Learning Analytics and Knowledge Conference*, pp. 161–172. ACM, Irvine CA USA (2021). <https://doi.org/10.1145/3448139.3448155>, <https://dl.acm.org/doi/10.1145/3448139.3448155>

23. Koedinger, K.R., Baker, R.S., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J.: A data repository for the EDM community: the PSLC DataShop. *Handbook Educat. Data Min.* **43**, 43–56 (2010)
24. Koedinger, K.R., Corbett, A.T., Perfetti, C.: The knowledge-learning-instruction framework: bridging the science-practice chasm to enhance robust student learning. *Cogn. Sci.* **36**(5), 757–798 (2012)
25. Kolb, D.A.: *Experiential Learning: Experience as the Source of Learning and Development*, FT Press (2014)
26. Kularbphettong, K., Kedsiribut, P., Roonrakwit, P.: Developing an adaptive web-based intelligent tutoring system using mastery learning technique. *Proc. Soc. Behav. Sci.* **191**, 686–691 (2015)
27. Kulik, J.A., Fletcher, J.D.: Effectiveness of intelligent tutoring systems: a meta-analytic review. *Rev. Educat. Res.* **86**(1), 42–78 (2016). <https://doi.org/10.3102/0034654315581420>, <https://journals.sagepub.com/doi/10.3102/0034654315581420>
28. Lave, J., Wenger, E.: *Situated learning: Legitimate peripheral participation*, Cambridge University Press (1991)
29. Legusov, O., Raby, R.L., Mou, L., Gómez-Gajardo, F., Zhou, Y.: How community colleges and other TVET institutions contribute to the united nations sustainable development goals. *J. Furth. High. Educ.* **46**(1), 89–106 (2022)
30. Odéen, M., Sellberg, C., Praetorius, G., Lindwall, O., Englund, L., Andersson, A.: Intelligent learning systems for simulator-based professional training: a systematic literature review. In: *International Conference in Methodologies and intelligent Systems for Technology Enhanced Learning*, pp. 109–119. Springer Nature Switzerland Cham (2024)
31. Oden Choi, J., et al.: Bridging the community college cybersecurity classroom and workplace with the cyberSim lab. In: *Proceedings of the 56th ACM Technical Symposium on Computer Science Education*, vol. 1, p. 875–881. SIGCSETS 2025, Association for Computing Machinery, New York, NY, USA (2025). <https://doi.org/10.1145/3641554.3701834>
32. Sarmiento, J.P., Wise, A.F.: Participatory and co-design of learning analytics: an initial review of the literature. In: *LAK22: 12th International Learning Analytics and Knowledge Conference*, pp. 535–541. LAK22, Association for Computing Machinery, New York, NY, USA (2022). <https://doi.org/10.1145/3506860.3506910>, <https://doi.org/10.1145/3506860.3506910>
33. Sehrawat, A., Keelan, R., Shimada, K., Wilfong, D.M., McCormick, J.T., Rabin, Y.: Simulation-based cryosurgery intelligent tutoring system prototype. *Technol. Cancer Res. Treat.* **15**(2), 396–407 (2016)
34. Sergeev, A., Kinney, M.B., Kuhl, S.A., Alaraje, N., Highum, M., Mehandiratta, P.: University, community college, and industry partnership: Revamping robotics education to meet 21st century workforce needs—NSF-sponsored project final report. In: *2019 ASEE Annual Conference & Exposition* (2019)
35. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educat. Psychol.* **46**(4), 197–221 (2011)