Classroom Evaluation of a Gamified Adaptive Tutoring System

Ying Tang
Dept. of Electrical and Computer Engineering
Rowan University
Glassboro, NJ, USA
tang@rowan.edu

Ryan Hare
Dept. of Electrical and Computer Engineering
Rowan University
Glassboro, NJ, USA
harer6@students.rowan.edu

Sarah Ferguson
College of Education
Rowan University
Glassboro, NJ, USA
fergusons@rowan.edu

***Abstract*—This Research to Practice Work-in-Progress Paper builds on prior developments of a gamified adaptive tutoring system that automates and personalizes a student's learning process without instructor intervention. To address the continued expansion of general education, as well as the grand challenge of personalized learning, automated learning systems are becoming common within higher education. Our personalized learning system uses an uses a structured, general-purpose game model that enables us to both track and control student progress through the sections of the game. While students play through a system-integrated game, a back-end AI component adaptively chooses both where the student is directed and what help they receive to optimize their learning. The end result is a fully integrated game system that can measure student performance using integrated tests, leveraging that information to adjust game content, address learner misconceptions, and lead to a faster and more effective learning session. As part of continued research, we present results from comparison testing of our educational game system in tandem with relevant course material.**

With our preliminary results, we focus on demonstrating the system's ability to provide appropriate content to players based on expert opinion. We show the educational utility of the game system, demonstrating an increase in student performance post-intervention on relevant content tests. We also show results from self-efficacy surveys administered to students to test their opinion of their own abilities. By sharing our testing and verification, we demonstrate the effectiveness of our intelligent educational game system.

Keywords—Gamification, Educational Software, Learning Technology, Higher Education, Electrical Engineering

I. Introduction

With an ever-growing population of students taking part in generalized educational systems, it becomes increasingly clear that one-size-fits-all approaches cannot possibly address the needs of all students [1]. At the same time, larger classroom sizes make it difficult for instructors to provide personalized tutoring or individual attention to each student. This problem is growing so widespread that continued advancement in personalized student education was chosen as one of the 14 grand challenges in engineering [2].

The recent advent of intelligent tutoring systems (ITSs) offers one solution to the personalized learning problem. These systems are often structured to provide lessons to students on an as-needed basis, allowing students to explore at their own pace while simultaneously offering a wealth of learning materials to help with learning [3]. Additionally, recent ITSs are expanding into the fields of machine learning and data science, using student data to learn trends, automate decisions, or improve the overall learning experience [4]. In addition to the challenge of personalized instruction, instructors also struggle to engage students, ensure classroom participation, and allow for reinforcement of concepts, especially with the prevalence of online learning [5]. Even ITSs cannot fully address this issue, as many web-based ITSs focus on reading study materials and answering questions. As such, a recent focus in engineering education has been the combination of ITSs with serious games [6].

Serious games (SGs) are virtual or physical games that are designed with a central purpose other than entertainment, typically education or training [7]. These games are beneficial for educators by improving student engagement through exciting visuals or expansive virtual environments [8]. Furthermore, SGs provide a highly variable environment in which to both test the student and provide just-in-time support.

Inspired by recent advancements in SGs, we present results from the ongoing development of *Gridlock* [1], an ITS integrated with an existing narrative game. We refer to the combined system as the personalized instruction and need-aware gaming (PING) system. The narrative game, *Gridlock,* puts students in a virtual environment to repair a traffic light as part of early education in digital logic and digital systems design. The game is designed to run in tandem with an in-classroom lab assignment. As students progress through the game, the PING system tracks their progress using a Learning Attributed Petri net graph [7] and tests their relevant knowledge. Artificial intelligence (AI) agents in the game then control student progress and provide hints, enabling or disabling certain areas depending on their estimated knowledge. Within areas of the game, students then receive individualized support based on their measured performance and estimated knowledge.

This work was supported in part by the National Science Foundation under Grant 1913809 and in part by the U.S. Department of Education Gruaduate Assistance in Areas of National Need (GAANN) Grant Number P200A180055.

With our preliminary results and pre-post-intervention testing, we focus on answering the following research questions:

RQ1. Can the game provide adaptive content appropriate to a student's indicated needs?

RQ2. Does usage of the game impact student knowledge retention?

RQ3. Does usage of the game impact students' opinions of their own abilities?

II. Gridlock and PING System Overview

As stated, Gridlock is a game designed to instruct students in the basics of digital logic and digital systems design; both core topics in electrical and computer engineering. The game is run in tandem with an in-classroom lab assignment where students design and program a traffic light controller to demonstrate their abilities in sequential logic design. To engage students in a more real-world problem-solving process, students first witness a simulated car accident in the virtual environment caused by a failed traffic light (shown in Fig. 1). After that, students are tasked to explore the environment, learn core concepts relevant to the assignment, and then design their own traffic light controller.

Fig. 1: The introduction of Gridlock where students witness a virtual traffic accident.

After witnessing the traffic accident, students are given an initial quiz to establish their baseline knowledge. The quiz covers seven topics identified by the development team as being necessary to solve the problem. Students then enter into the traffic control office and meet the AI, a virtual character in charge of the traffic lights in the game. The AI instructs them on the problem and how to progress through the game. Students then enter in to the computer room, shown in Fig. 2.

In this room, students interact with different computers to explore so-called "subject-specific learning blocks". Each block pertains to one of the aforementioned topics. Depending on the student's performance in the initial quiz, certain computers are already "completed" when the student enters the room. Upon interacting with a computer, the system selects a topic the student has not yet completed and moves them into the relevant subject-specific learning block.

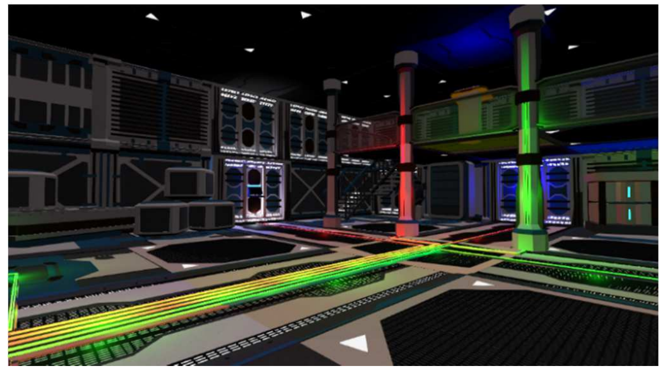

Fig. 2: The "computer room" where students must interact with computers. Students progress forward once all lights have been turned green.

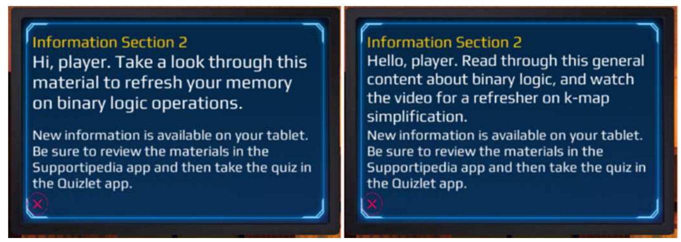

Fig. 3: Example assistance dialogue given to students. In this case, two different students at the same point in the game.

Within a learning block, students are first given personalized dialogue chosen by the system from a pool of possible options, as shown in Fig. 3. Students are then given help materials which are also selected by the system from a pool of options. These materials are composed of videos, images, presentation slides, text, equations, and other useful information pertaining to the chosen learning block. Once the student has reviewed the presented material, they are tested again on that specific learning block's material to determine if they complete the block or not. Once all blocks have been completed, the students progress to submit their final traffic light design.

A. System Framework

The PING system takes the place of a human tutor to assist students on the content presented by the attached game. Fig. 4 shows a high-level overview of the components of the PING system for general-purpose applications. As the student interacts with the game interface, their actions are recorded and measured to build a model of that student's knowledge. That model, in turn, is used by the AI component to select both pathing and assistance actions to provide to the student. In *Gridlock*, the learning blocks are shown when interacting with computers in the computer room, but other games may use different rooms, game levels, or other variations to distinguish content-specific learning blocks.

In the system, data collection occurs constantly to keep an accurate model of the student's current level of knowledge. In *Gridlock*, this data collection manifests as results from content tests, game actions (such as key presses and mouse movement), and estimated emotion state data extracted from webcam images. We collect one feature vector of these values for each subject-specific learning block to inform our decision-making.

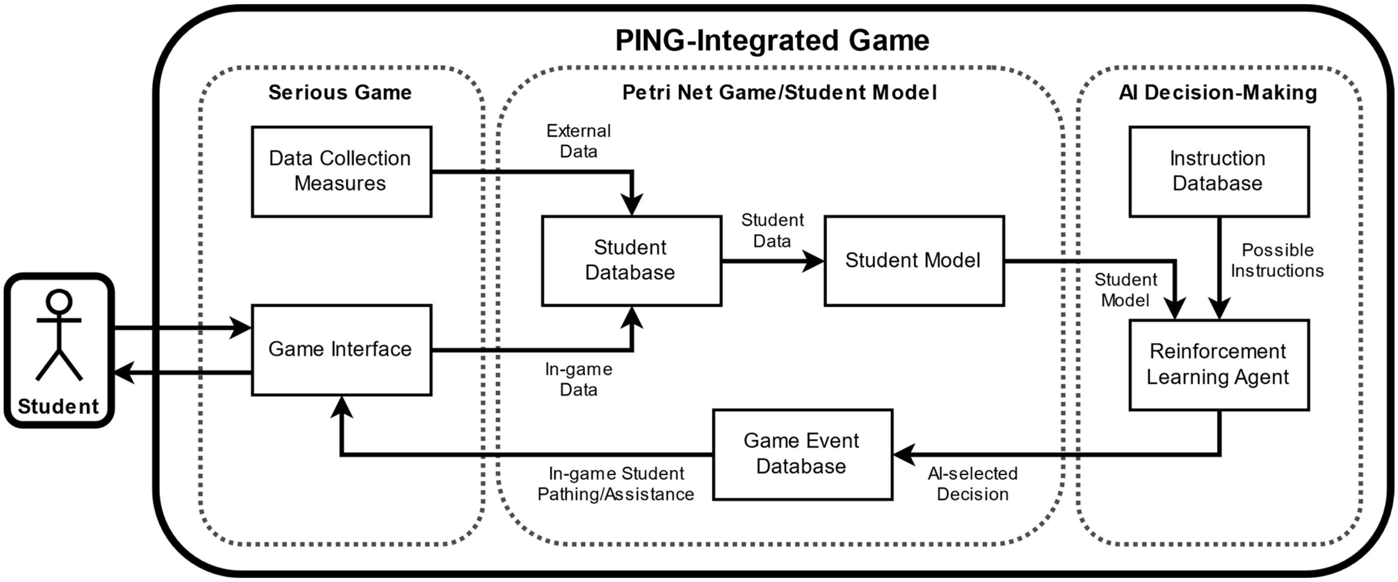

Fig. 4: Top-level overview of the PING system.

However, the system is designed for flexible data inputs and student models based on the desired application.

The student tracking module models the game as a Petri net, a set of connected states that the student moves between [7]. The system can adaptively adjust what states the student moves to as determined by their performance in the game [7]. The system can then prioritize blocks that the student needs help with. Within blocks, as well, the system also adjusts what help the student receives, picking from a pool of possible help actions. The student then repeats in a cycle of assistance and testing within each block until they demonstrate mastery of that specific subject.

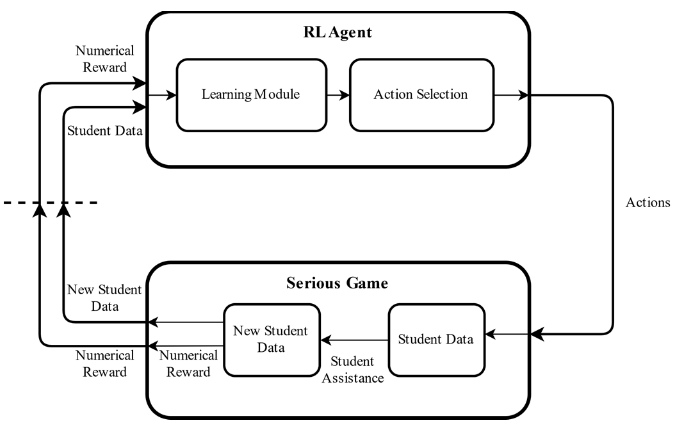

Fig. 5: The cycle of reinforcement learning.

Once the model has been successfully incorporated, the game needs an artificial intelligence (AI) agent to make decisions on both what blocks to visit and what help to provide. Creating a deterministic or rule-based system for this would prove time-consuming and might not handle all edge cases. Instead, we opted to adapt reinforcement learning (RL) to adaptively learn and update system behavior. Fig. 6 shows the typical cycle of reinforcement learning. As shown, when the agent observes a student, it will choose an action to help that student. It will then observe a result and receive a reward based on if the student improved. If the reward is positive, then the provided help must be "good" and will be prioritized in future interactions with similar students.

III. Game Assessment and Testing

A. *Adaptivity Evaluation*

One of the main aspects of the PING system we intended to assess was the ability of the system to provide appropriate support to students. To verify the system's performance, we extracted several student feature vectors from our in-classroom testing of *Gridlock* and the PING system, along with the corresponding actions chosen by the system. We then had an expert in the relevant material observe the student's performance and select assistance to provide. Table 1 shows results from a focus group of 6 students. For each student, their performance was captured from their initial testing in the game. For 4 of the 6 students, the system's decision reflected the expert opinion.

Table 1: Expert-identified student assistance compared to system selection for 10 students.

Student	Expert-identified Help Category	System-selected Help Category
1	D-flip-flop operations	D-flip-flop operations
2	**Verilog syntax**	**General Verilog assistance**
3	D-flip-flop operations	D-flip-flop operations
4	Logic gates (w/ video)	Logic gates (w/ video)
5	Binary logic	Binary logic
6	**Finite state machines**	**Finite state machines and state registers (w/ video)**

For student 2, the expert opinion was to provide assistance on the specific syntax of Verilog code, the software language used in the accompanying lab assignment. This help was selected because the student showed decent performance in all subjects except for Verilog code syntax. Furthermore, the questions provided to this student in the coding section all dealt with Verilog syntax. The system, meanwhile, chose to provide a general-purpose assistance action instead. This specific instance could indicate the system, at this time, is in need of additional training to learn more optimal behavior. Alternatively, it could indicate that the system learned to prioritize a more general-purpose action compared to a specific one. For student 6, the expert suggested basic assistance, while the system suggested advanced assistance with a video on the same topic. In both cases, the chosen topic was the same, but the system chose a more advanced help action. In this case, it's possible that the video proved more effective as observed by the system, and the system prioritized the video over the standard assistance.

From this initial testing, system performance demonstrates overall positive behavior. Deviations from expert opinions may be related to the structure of the actions in the game. Since the system tends to prefer general-purpose assistance over specific assistance, it could indicate that the general-purpose help categories are effective for all students, regardless of areas of difficulty. In this case, the current implementation may benefit from additional content added to the specific help categories to ensure students receive appropriate assistance.

B. *Educational Impact*

To verify educational merit, the game was used in conjunction with a relevant lab assignment in both Introduction to Digital Systems at Rowan University and Computer Architecture at Mercy College. Students were divided into two groups: a control group which applied the standard lab assignment and a treatment group which integrated the game. First, students were assigned an optional pre- and post-lab content test which tested them on their ability to design a sequential state machine, similar to the traffic light problem. Average student scores are shown in Table 2.

Table 2: Pre- and post-lab content test results from control and treatment groups. Standard deviations listed in parentheses.

Measure	Treatment Group (n = 20)	Control Group (n = 11)
Pre-lab	4.76 (2.17)	4.18 (1.60)
Post-lab	4.57 (3.27)	3.24 (1.58)

As the testing was optional, overall student participation was limited to a subset of overall participants. However, results still show a moderate increase (Cohen's $d = 0.52$) in post-test scores in the treatment group compared to the control. Both groups showed a decrease in average score from pre- to post-test. Both problems are structured similarly, but numerical differences could have led to a more difficult post-test. The overall decrease in the treatment group was less ($d = -0.19$) compared to the control ($d = -0.94$). So, while both groups of students demonstrated a slight loss in performance from pre- to post-test, the treatment group demonstrated a less severe decrease, indicating a positive impact on knowledge retention when using the game system.

Table 3: Self-efficacy scores from students in the treatment group who participated in the post-lab efficacy survey (n = 20). Standard deviations listed in parentheses.

Measure	Pre-Lab Scores	Post-Lab Scores
Engineering Interest (scale 1-3)	2.47 (0.44)	2.57 (0.43)
Engineering Ability (scale 1-7)	4.59 (0.69)	4.68 (0.71)
Sense of Belonging (scale 1-7)	4.76 (0.53)	4.66 (0.53)

Another metric used in our testing is a self-efficacy survey also administered pre- and post-lab, where students were asked questions regarding their interest in engineering and their opinion of their own ability. Table 3 shows average ratings from pre- and post-lab surveys of students in the treatment group. Among the treatment group, there was little significant difference between the pre- and post-lab scores. Results indicate that the game in its current form and the lab as a whole do not have an appreciable effect on students' interest in engineering, confidence in ability, or sense of belonging among their peers.

IV. Conclusions

In this paper, we address the continued push toward widespread personalized and automated student education with the personalized instruction and need-aware gaming (PING) system. We provide developmental insights from *Gridlock*, a game integrated with the PING system that deals with digital logic design, including layout of the game and integration with the PING system. Our resulting game system achieves positive results in its ability to adapt to unique students, demonstrated from a comparison test between expert opinion and system decision-making. We also demonstrate the full system's educational utility through course testing at two universities, evaluating the game system through content knowledge tests and participant surveys. Throughout initial testing, treatment groups demonstrated increased post-intervention content knowledge and high interest in both the design problem and the game scenario. Additionally, participants consistently reported that the game provided greater or equivalent educational utility compared to a standard textbook.

Future expansions of this research are focused on improving the self-learning artificial intelligence agents in the game as the current implementation still requires significant amounts of student data to learn optimal behavior. Furthermore, the educational game component still requires additional adjustments to ensure that all in-game assistance is appropriate and helpful to students. Finally, due to low student participation and external factors impacting in-classroom testing, additional in-classroom testing is also planned to gather additional comparison data and further verify the game's educational utility.

References

[1] Ying Tang, Joleen Liang, Ryan Hare, and Fei-Yue Wang, “A Personalized Learning System for Parallel Intelligent Education,” IEEE Trans. on Computational Social Systems, Vol. 7, No. 2, pp. 352-361, 2020.

[2] National Academy of Engineering. 2016. Grand Challenges for Engineering: Imperatives, Prospects, and Priorities: Summary of a Forum. Washington, DC: The National Academies Press. https://doi.org/10.17226/23440.

[3] A. Anohina, "Advances in Intelligent Tutoring Systems: Problem-solving Modes and Model of Hints," International Journal of Computers, Communication and Control, vol. 2, no. 1, pp. 48-55, 2007.

[4] Elham Mousavinasab, Nahid Zarifsanaiey, Sharareh R. Niakan Kalhori, Mahnaz Rakhshan, Leila Keikha & Marjan Ghazi Saeedi (2021) Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods, Interactive Learning Environments, 29:1, 142-163, DOI: 10.1080/10494820.2018.1558257

[5] Martin, F. & Bolliger, D.U. (2018). Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. Online Learning 22(1), 205-222. doi:10.24059/olj.v22i1.1092.

[6] Beyyoudh, M., Idrissi, M.K., Bennani, S. (2020). A New Approach of Integrating Serious Games in Intelligent Tutoring Systems. In: Serrhini, M., Silva, C., Aljahdali, S. (eds) Innovation in Information Systems and Technologies to Support Learning Research. EMENA-ISTL 2019. Learning and Analytics in Intelligent Systems, vol 7. Springer, Cham. https://doi.org/10.1007/978-3-030-36778-7_10

[7] J. Liang, Y. Tang, R. Hare, B. Wu and F. -Y. Wang, "A Learning-Embedded Attributed Petri Net to Optimize Student Learning in a Serious Game," in IEEE Transactions on Computational Social Systems, doi: 10.1109/TCSS.2021.3132355

[8] Geoffrey Hookham and Keith Nesbitt, “A Systematic Review of the Definition and Measurement of Engagement in Serious Games,” ACSW 2019: Proceedings of the Australasian Computer Science Week MulticonferenceJanuary 2019 Article No.: 42, Pages 1-10. https://doi.org/10.1145/3290688.3290747.