Shih, JL. et al. (Eds.) (2023). Proceedings of the 31st International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Design and Implementation of an Educational Game for Teaching Artificial Intelligence to High School Students

Ning WANG[a*], Ryan MONTGOMERY[b], Eric GREENWALD[b] & Maxyn Leitner[a]
[a]*Institute for Creative Technologies, University of Southern California, USA*
[b]*Lawrence Hall of Science, University of California, Berkeley, USA*
*nwang@ict.usc.edu

Abstract: Artificial intelligence (AI) is becoming ubiquitous in our daily lives. In more and more fields, AI systems are transforming how knowledge is constructed, discoveries are realized, and how solutions are developed and tested. These changes have profound implications for the future workforce and citizenry. Yet, learning AI remains a niche subject largely reserved for advanced post-secondary educational contexts. While there is growing attention to broadening AI educational opportunities and, especially, to providing learning experiences for younger students, relatively little is currently known about how to most effectively provide AI education to K-12 (kindergarten through 12th grade) students. In this paper, we discuss the design and present findings from an implementation study of an educational game for high-school AI education called ARIN-561. Drawing on an integrated analysis of gameplay log data, pre/post knowledge, and disposition surveys for nearly 1,000 high school students, we present findings on the efficacy of the educational game and its constituent activities in advancing AI learning goals. We explore possible interactions between learning outcomes, incoming math knowledge, prior gaming experience, and other factors that can inform future learning design and shed light on what can position youth for success in game-based AI learning experiences.

Keywords: K-12 AI education, youth AI education, educational game

1. Introduction

Artificial Intelligence (AI) is a foundational technology permeating every aspect of our daily lives. Rapid advances in the design and implementation of AI systems have led to the ever-expanding role for AI in society. It is also profoundly transforming our workforce around the globe. While some of today's youth will become the future AI workforce and a majority of them will join a workforce that utilizes AI, all will become end-users, such as consumers of AI (Gardner-McCune et al., 2019). It is critical, therefore, to prepare future generations with basic knowledge of AI, not just through higher education, but beginning with childhood learning.

While AI's impact on society is deepening and expanding in myriad ways, and innovative educational opportunities are being rapidly developed, there has been little research into how students, especially pre-college students, construct an understanding of and gain practice with core ideas in the field. As a result, there is yet little possibility of grounding the design of learning experiences in evidence-based accounts of how youth learn AI concepts, how understanding progresses across concepts, or what concepts are most appropriate for what age-levels. AI is built on a foundation of philosophy, psychology, and mathematics, and it centers around using algorithms to solve real-world problems (Russell and Norvig, 2016). This provides a theoretical foundation to connect AI learning to existing Science, Technology, Engineering, and Mathematics (STEM) subjects in K-12 classrooms. Given the packed schedule of existing courses of K-12 students, it becomes a more realistic approach to embed AI education in K-12 classrooms. Such an approach to AI instruction offers a rich context to learn scientific and mathematical concepts already taught in K-12 (Wang and Johnson, 2019) and to apply them to problem-solving.

One technology-based approach to bring AI to the K12 classroom that has shown promise in other STEM disciplines is digital game-based learning. Decades of research evidence point to the efficacy of game-based learning in promoting student learning (Plass et al., 2020). However, there is little research into using game-based learning for AI education for youth (Lee et al., 2021), given that the research field of K-12 AI education is still in its infancy. In this paper, we will discuss the design and initial implementation study of an educational game, called ARIN-561, for teaching high-school students about AI. We conducted an evaluation study at high schools in the United States. Results indicate the potential of ARIN-561 to build AI knowledge, especially for students who have background knowledge with the relevant mathematical concepts typically taught at the high school level.

2. Related Work

AI education has long been absent from K-12 classrooms. Recent efforts are beginning to investigate the integration of AI into K-12 schools, including defining AI literacy (Long and Magerko, 2020) and developing curricula and guidelines (Gardner-McCune et al., 2019; MIT AI Education Initiative, 2021). Researchers in youth AI education have been experimenting with teaching AI, including machine learning (Rodríguez-García et al., 2021; Zhou et al., 2021) and ethics (Forsyth et al., 2021), within the context of computational thinking (Ritter et al., 2019) through conversational agents (Lin et al., 2020), dance (Payne et al., 2021), and game-based learning (Lee et al., 2021). Discussions on youth AI education are heating up in Europe (Kandlhofer et al., 2019; AI+, 2021), China (Peterson et al., 2021), Israel (Shamir and Levin, 2020), and around the world (Youjun et al., 2018; Yukun and Tang, 2018). For example, researchers in Thailand have designed an agricultural-based AI challenge to foster middle-school students' learning of the machine learning process in the form of a game (Sakulkueakulsuk et al., 2018), where students build machine learning models to classify ripe or unripe mangoes. In Australia, researchers have designed and implemented classroom activities for teaching fundamental concepts of AI to Year 6 students to demystify AI through activities such as an unplugged activity on facial recognition and a simple robotic exercise that introduces the concept of machine learning (Ho et al., 2019).

The work presented here aims to uncover how to design an educational game to meet the challenges of teaching AI to K-12 students. This work builds upon explorations into how K-12 students approach AI concepts, what obstacles they face, and how to guide them through obstacles (Greenwald et al., 2021). This work also draws upon previous investigations into linking AI to the K-12 math curriculum to identify AI concepts suitable for high school students (Wang and Johnson, 2019), as well as work investigating the learning of computational thinking (Lee et al., 2011) and seminal research into comprehension of mathematical representations (e.g., Curcio, 1987; Friel et al., 2001).

3. ARIN-561 Game-Based Learning Environment

ARIN-561 is a 3D role-playing game designed to teach high-school students AI concepts, prompt them to apply their math knowledge, and develop their AI problem-solving skills. In the game, students play as a space-faring scientist who has crash landed on an alien planet, named ARIN-561 (Figure 1). In order to safely return home, the scientist begins exploring the planet to gather resources needed to repair the broken ship while uncovering the mystery of the planet. The activities for survival and for exploration form the basis for the tasks the students carry out in the game. The game currently covers three classical search algorithms: breadth-first search (BFS), depth-first search (DFS), and greedy search. Each topic consists of two modules: a tutorial module (e.g., Figure 1 bottom left) and a transfer module (e.g., Figure 1 bottom right). Embedded in all the tutorial and transfer modules are quizzes that help students pause and self-assess (Figure 1 top right). In-game challenges, such as searching for missing spaceship parts or cracking passwords, serve as natural opportunities for the introduction of search as a topic. The essential concepts such as space and time complexity also lend opportunities to connect math knowledge familiar to high school students and these AI concepts that are usually taught in higher education. The integrated educational content in ARIN-561 leverages this opportunity by supporting the students' application of math knowledge to the evaluation of each algorithm as they progress through the game. In additional to the learning modules, students can also explore the game environment for "off-task" activities (Sabourin et al., 2011), such as gathering minerals around the planet.

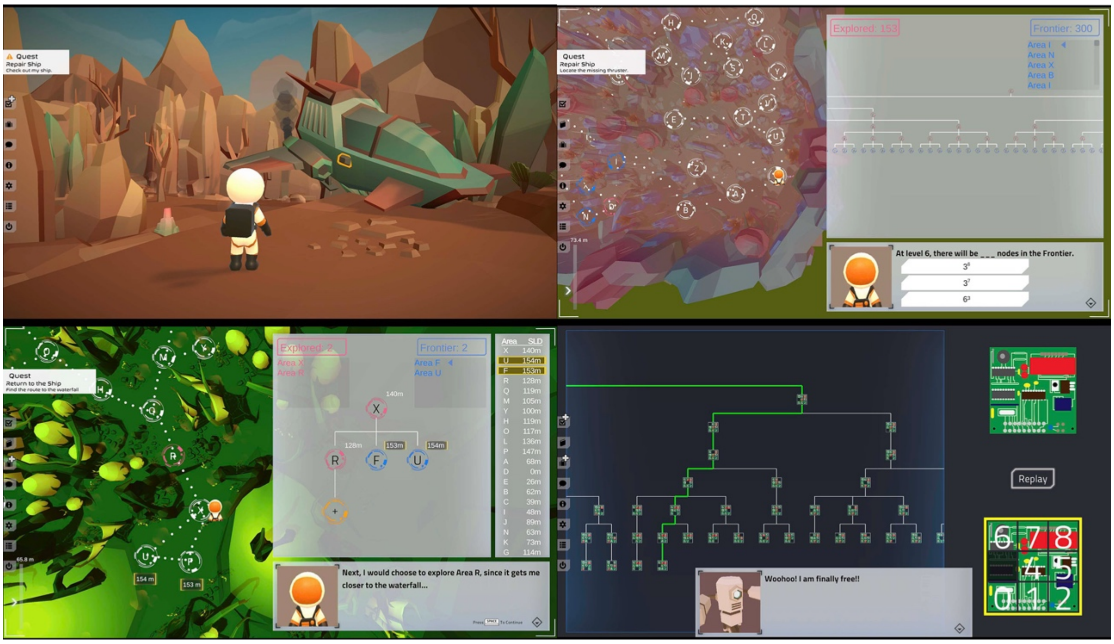

Figure 1. Screen capture from ARIN-561. Top-left: The player crash landed on a foreign planet. Top-right: student is presented with a quiz question about estimating the complexity of search algorithms. Bottom left: student think-alouds through the greedy search algorithm. Bottom-right: the student solves an 8-puzzle using one of the search algorithm to fix their companion robot's circuit board.

3.1 Learning Experience Design in ARIN-561

The design of ARIN-561 is guided by lessons learned from pilot studies on student AI problem-solving (Greenwald et al., 2021). We also developed additional design principles based on our observations of the characteristics unique to AI learning, and how lessons learned from related fields, such as computer science education (Lee et al., 2011), could be used to inform the design here.

Facilitate Abstraction In a cognitive interview study, researchers explored how K-12 students, particularly those in high school, approach AI concepts, what obstacles they face, and how to guide them through obstacles (Greenwald et al., 2021). In the study, students were presented a set of AI problems in a wide range of topics, such as various machine learning algorithms. Students' think-alouds as they attempted to solve the AI problems (using paper and pencil) shed light on the critical step most students struggled with — problem formulation, or the *Abstraction* phase in computational thinking, i.e., *Abstraction*, *Automation*, and *Analysis* (Lee et al., 2011). This is an initial step where students formulate a problem described in natural language (e.g., find the shortest path) into one that can be solved by a computer, such as creating variables (e.g., distance), determining end conditions (e.g., search ends when distance can't be minimized). Strategies employed by the expert interviewer provided a basis for pedagogical design in ARIN-561.

One of the design decisions made to facilitate Abstraction was to display the real-world problems and the abstract representations side-by-side and to update both synchronously (e.g., the lower-left screen in Figure 1). In an ARIN-561 route-planning problem where students are tasked to use search algorithms to plan a route to a waterfall, a map is placed on one side of the screen while a search tree is displayed over and next to the map. As students direct the algorithm by exploring locations, connected via roads, on the map, the search tree updates accordingly step-by-step and illustrates how locations are represented as nodes and roads are represented as edges in a tree data structure, and how route-planning on a physical map is computationally solved as the expansion of a search tree. When a goal node is reached in the search tree, it is highlighted both on the search tree as a path across edges from root to the goal node, and on the map as a route reaching the waterfall connected via roads.

Learning Transfer In education, transfer of learning occurs when learning in one context enhances or undermines a related performance in another context (Perkins et al., 1992). During learning transfer, students apply learning in one discipline across multiple situations. Transfer of learning is particularly important for AI education, as AI can be considered as a discipline of using algorithms to solve real-world problems. When students learn how an AI algorithm can be used to solve illustrative problems in one domain, it is critical to also guide them through problems from a different context to help them build the connection — the abstract representation of the algorithm that can be applied to formulate solutions to seemingly different and unrelated problems.

For each algorithm covered in ARIN-561, we developed a tutorial problem and a transfer problem. The two sets of problems are different enough to arguably be considered as far transfers (instead of near transfers) (Perkins et al., 1992). Tutorial problems are chosen from domains familiar to the students, such as finding a route from point A to B on a map. In a typical tutorial module, students are scaffolded through the abstraction, automation, and analysis processes (Lee et al., 2011) through the player character's think-aloud and their dialogue with the companion robot. In the abstraction phase, the students are guided to create an abstract representation of the practical problem. After students demonstrate their understanding by correctly expanding the tree for several levels, they are provided with the option to automate the process. In the automation phase, students can watch the search tree continue to expand automatically, on the same interface — physical map and abstract search tree placed side-by-side. Students can also pause and step through the tree expansion one step at a time to examine the process closely. The automated expansion animation helps illustrate the characteristics of the search algorithms, e.g., expanding in a breadth-first or depth-first fashion. In the analysis phase, students are guided by the game narrative to examine the solution (e.g., the route found) and to evaluate the process through which the solution was generated (e.g., time and space complexity of the search algorithm). The subsequent transfer problem module presented students with a different problem, such as cracking a password or solving an 8- puzzle. Students were guided by similar but much abbreviated scaffolds through the *Abstraction*, *Automation*, and *Analysis* processes in the transfer phase.

Comparative Explanation AI is human ideas represented mathematically and realized computationally. From Classical Search to Local Search, from Propositional to First-Order Logic, from Decision Trees to Genetic Algorithms, AI algorithms build on each other: a new algorithm is often created by modifying an existing one, to solve problems that the existing one was not suited or able to solve. This insight creates both challenges and opportunities for AI education. The evolutionary characteristics of AI algorithms provide a basis for pedagogy that leverages students' prior knowledge (of an algorithm they are already familiar with) while constructing the new ones. By directly comparing the new and old algorithms, for example, students not only learn the new, but also reinforce the learning of the old. Such comparisons are not just algorithmic, but also the contextual in terms of application. Understanding the pros of the new and cons of the old in what problems they are or are not suited to address is a key to using AI for problem solving. The approach of prior knowledge activation is not new (Alvermann et al., 1985), nor is the issue of activating inaccurate prior knowledge (van Loon et al., 2013). The explicit and direct comparison between the new and old, when discussing the new, may offer an opportunity for students to reexamine their misconceptions of the old.

In ARIN-561, game modules are organized by learning topics, such as BFS and DFS. After scaffolding students through the first AI algorithm (such as BFS), each new AI algorithm (e.g., DFS) is introduced through an example problem that the previous algorithms fail to solve (e.g., computer runs out of memory when using BFS for route planning). The students are then guided through the Analysis phase to uncover why the previous algorithm failed (e.g., storing too many nodes in computer memory) and how to modify it to address its weakness (e.g., prioritizing expanding child nodes instead of sibling nodes in the search tree). Such modification thus results in the birth of the new algorithm (e.g., DFS). The direct comparisons are not only realized in the explanations through the game narrative, but also illustrated on the user interfaces across the learning of different algorithms.

4. Methods

4.1 Recruitment

Twenty-three math, science, and computer science teachers from a school district in a major metropolitan area in the United States participated in the study. 1274 high school-aged youth from classes taught by participating teachers were recruited for the study.

4.2 Procedure

Participating teachers were provided an overview of the game, learning goals, and study procedure a few months before the study began. A few weeks prior to the study, students were given an online parental consent form and a youth assent form. Only students who consented participated in the study. The study was carried out over 4 class sessions, each lasting 45-55 minutes long, with at least 2 class sessions dedicated to individual gameplay for students. During the first session, students were first assigned IDs to protect their identity throughout the study, and then completed the pre-survey online. At the end of the first session, students logged into the ARIN-561 game online via a web browser. Any technical difficulties encountered were addressed during the first session, via support from the research team. During the second and third sessions, students continued to interact with ARIN-561 at their own pace. Game progression, play time, and answers to in-game questions were recorded for each participant. During the fourth session, students completed the post-survey online.

With restricted access to school campuses due to COVID-19, the study was carried out entirely by the participating teachers. The research team did not participate in the data collection. Additionally, because students were not required to answer all the questions on the pre- and post-surveys, there are missing data at the item level for some students.

4.3 Measures

The pre-survey consisted of items about students' demographic background, AI Use Type, Interest in AI, AI Knowledge (15 questions), Math Self-efficacy [Liu and Koirala, 2009], and Math Knowledge. All scales except the Math Self-efficacy were developed by the research team. The AI Use Type included items such as "When I think about how I'd like to interact with AI in the future, I expect that: I will use AI systems in my everyday life as a consumer, and I expect to USE AI systems as a part of my job.” The Interest in AI scale included questions such as “Outside of school I try to learn a lot about AI.” The assessment of AI knowledge and math knowledge specifically focused on the content covered in ARIN-561, in the format of multiple-choice questions. The AI questions were set in the context of solving AI problems similar to those encountered in the game. The questions assessed students’ understanding of, for example, pros and cons of the search algorithms, search algorithms most applicable to specific types of problems, etc. In the post-survey, the same items on interest in AI and AI knowledge from the pre-survey were included. In addition to the surveys, game logs from ARIN-561 were collected. The logs included the in-game click-stream data and responses to in-game quizzes.

5. Results

Of the 1274 participating students, 1014 completed the post-survey. The research team was able to match pre-, post- surveys, and game logs for 764 students. Other than normal attrition (e.g., students absent at either pre, post administration, or game play class), additional data loss was primarily due to errors in student ID entries on the survey platform, which resulted in mismatches of student IDs between both surveys and game logs. We conducted ANOVA analyses to ensure the final sample of 764 students was not significantly different from the full participant sample in terms of background, such as gender, race/ethnicity, and prior mathematical knowledge.

The participants’ average age was 16, with 18% 12th graders, 30% 11th graders, 23% 10th graders, and 29% 9th graders. A total of 46% of the students identified as male, 48% identified as female and 6% identified as other categories or preferred not to disclose. 27% of the students speak English at home, 67% speak both English and a second language at home, and 6% speak only a language other than English at home. Spanish is reported as the non-English language for those students. Interestingly, even though ARIN-561 and the surveys are offered in both English and Spanish, and the teachers were briefed about the language choice prior to the study, all the students chose to use the English version of the surveys and the game.

5.1 AI Learning Gain

We conducted a paired sample t-test on the AI knowledge scale from pre- and post- surveys to examine if playing the game resulted in gains in AI knowledge. Table 1 summarizes the pre/post changes in AI Knowledge and in sub-constructs directly relevant to modules in the game (additional AI knowledge items covered in broader topics such as search tree representations). Results show that students who participated in the study demonstrated statistically significant gain in AI knowledge, with a mean difference of 0.37 on a 34-point scale (p = .011). The AI knowledge scale include 3 sub-scales for each of the search algorithms covered in the game (BFS, DFS, and Greedy search). Additional paired-sample t-tests revealed a statistically significant gain for BFS learning (mean difference of .30* on an 11-point scale, p = .001), a smaller and not statistically significant change for DFS learning (p = .088), and a nearly flat outcome for the items focused on the Greedy search algorithm.

Table 1. *Paired-sample t-test results on AI learning gains (pre/post), breaking down by Overall (all items in the scale), BFS, DFS, and Greedy Search learning sub-scales.*

Metric	Pre-Test	Post-Test	Max-Score	T-statistics	p-value	Effect Size
Overall	14.16	14.53	34	2.54	0.011	0.105*
BFS	5.30	5.60	11	3.25	0.001	0.149*
DFS	4.36	4.50	9	1.71	0.088	0.081
Greedy	2.09	2.06	3	-0.48	0.629	-0.022

5.2 Student Background and AI Learning

In the pre-survey, we gathered data on students' demographic background, such as gender, grade level, language spoke at home, and video game experiences. ANOVA tests show that pre/post AI learning gains did not differ significantly between students of different gender, grade-level, and language spoken at home. Learning gains differed however between students with different prior gaming experience (Figure 2). The participants reported a wide range of gaming experiences. Given the detailed categorization of gaming experiences, we grouped the students into two groups: those who don't play video games or play 1-2 hours per week (60% of the sample), and those who play 3 or more hours per week (40%). Students who play video games less than 2 hours per week had significantly lower gain on overall AI knowledge ($M_{<2h}$ = .0132, $M_{>3h}$ = .8428, p = .006), including sub-scales on BFS learning ($M_{<2h}$ = .0132, $M_{>3h}$ = .8428, p = .003), Greedy learning ($M_{<2h}$ = −.18, $M_{>3h}$ = .11, p = .019) but not DFS learning ($M_{<2h}$ = .04, $M_{>3h}$ = .29, p = .152), compared to students who play video games 3 or more hours per week.

The pre-survey also includes items that measure Math Self-Efficacy, (relevant) Math Knowledge, and Interest in AI. We conducted a series of regression analyses to investigate these three student level factors that may be predictive of observed learning gains. We found that the prior Math Knowledge (as demonstrated on the pre-survey item set) predicted observed AI learning gains (R = .1, p = .006). This suggests that relevant math knowledge is weakly but significantly related to higher AI knowledge gained through ARIN-561.

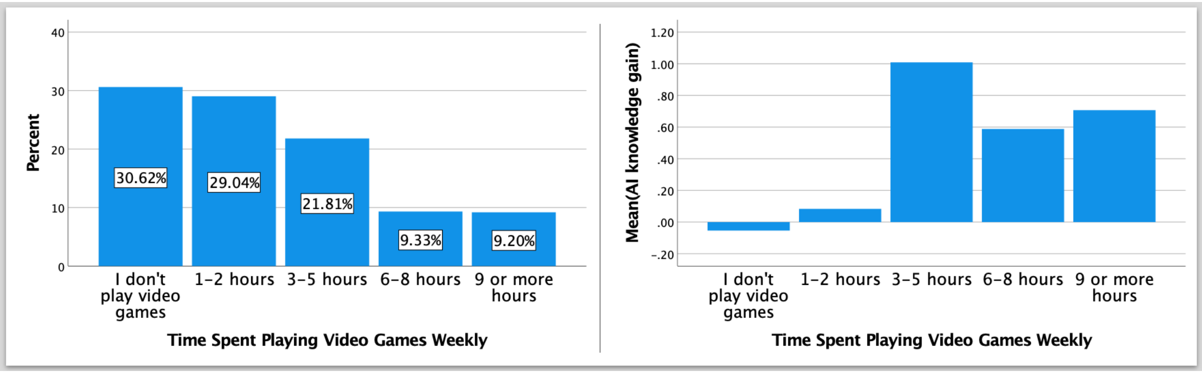

Figure 2. Left: Percentage of students with different weekly gaming experience. Right: AI Knowledge gain broken down by weekly gaming experience.

5.3 In-Game Progress and AI Learning

The pedagogical design of ARIN-561 is based on the hypothesis that AI algorithms build on each other. Algorithms, such as DFS introduced later in the game are discussed in comparison to previously introduced algorithms, such as BFS. While students can jump through different modules by going through the menu selection screen in the game, overall, students took a relatively linear path through the game, by going through BFS, DFS, then Greedy game modules. Thus, as students progress through the game, mastering previously discussed algorithm should help students' learning of the new ones, while learning the new algorithm helps student reinforce the learning of the older ones. We analyzed how reaching milestones in the game, such as completing the DFS module (both tutorial and transfer problem modules), impacts overall AI learning and the learning of individual algorithms. Independent sample t-test shows there is no significant difference in AI knowledge gain between students who completed all modules of the game and students who did not (N_T = 556, N_F = 208, T: completing, F: not completing, p = .642). Students who completed the BFS module did not gain significantly more AI knowledge overall (F = 1.169, p = .28) or BFS knowledge (F = .5, p = .48) than those who started but didn't complete BFS modules. However, completing BFS did help students gain more knowledge on DFS (F = 4.545, p = .033) and greedy search (F = 3.204, p = .074). Completing the DFS or greedy modules did not have a significant impact on overall or individual AI algorithm learning. Given that all students are given the same amount of time to play the game in the classroom, we did not analyze how time in game impacted AI learning. Overall, students spent between 3 seconds and 338 minutes in the game, with a mean/median gameplay time of 89/84 minutes. The outlier of extremely long game-play time is likely due to students forgetting to log out of the game at the end of the class.

6. Discussion

This study demonstrates that a relatively brief in-school exposure to AI learning experiences, via an educational game, can result in learning gains for AI content with pre-college aged youth. Examination of the subscale scores for the AI Knowledge assessment indicated that the learning that took place was concentrated on the BFS algorithm, with smaller gains for DFS, and no gains for Greedy. Drawing on the learning and assessment design concept of a learning progression (Wilson, 2009; Duncan and Hmelo-Silver, 2009), our conjecture is that the design of gameplay, in which students first encountered BFS then compared it to each of the next two algorithms as the game progressed (DFS, then Greedy), led to consolidation of understanding related to BFS and thus a deeper opportunity to learn that content. However, we did not observe a statistically significant impact of completing the BFS, DFS, or Greedy modules on BFS learning gains. Completing the BFS modules however, did contribute to learning DFS and Greedy search. This suggests the efficacy of the progressive roll out of content in our design, where each new content area is explicitly related to prior content. Later design iterations will look to extend opportunities for students to connect and consolidate their emerging understanding of content encountered later in the game, e.g., through additional integrative activities.

The significance of prior mathematical knowledge for predicting observed AI learning gains suggests an educational game that is optimized for youth who already enter with a strong mathematical foundation. This would challenge efforts at using the current iteration of the game for a broad high-school population with a wide range of prior math competencies. Future design iterations will look to support students with varying levels of prior mathematical knowledge, either through focused tutorials for related math content and/or through improved game design that better resonates with students who have not yet taken advanced math courses.

We also see promise for this game-based instructional model in the feasibility of its implementation. First, the youth who engaged with the game did so largely independently of a teacher. The minimal need for outside expertise means that implementation is likely to be less dependent on having educators with AI and computer science expertise, an important consideration given widely reported shortages of high-school teachers with such expertise. Also related to feasibility of implementation, completing the game took roughly 2 class periods on average, which minimizes the time it may draw away from existing scope and sequence. Thus, the educational game is well-positioned to be integrated into a wide range of courses and instructional contexts. On the other hand, the AI knowledge growth was relatively small (effect size =.105), indicating that spending 2 classes playing an AI-themed video game is unlikely to contribute to learning gains that might be expected from a fuller instructional sequence or dedicated course. Interestingly, during post-implementation conversations, the participating teachers expressed strong interests to integrate classroom discussions with game-based learning, and suggested dividing the classroom time into independent gameplay and post-gameplay whole-class discussion, where teachers organize discussions to help students reflect upon what's learned through the game. Such an integrated approach has the potential to further enhance the efficacy of the educational game.

The study was dependent on a researcher-developed measure of AI knowledge, with limited evidence available of its validity with the population sampled. This speaks to the current dearth of AI knowledge measures developed for precollege-aged youth, a challenge that our research team, and others, are working to address through ongoing research and measurement development. In this specific case, we note that the assessment was likely too difficult for the sampled population (for example, the mean score on the post-intervention administration was 14.53 out of a possible 34 points), limiting its potential for demonstrating the learning of high-school-aged youth. Additionally, the measure included different numbers of items for each of these subscales, with fewer items for DFS and Greedy compared with BFS. This constricted the available range for movement on those constructs, potentially impacting observability of changes that may have taken place. Our team is conducting psychometric tests (classical and IRT methods) and triangulating that with in-game opportunities to demonstrate understanding to inform revisions to our instrument for this population.

Acknowledgements

This research was supported by the United States National Science Foundation under Grant #1842385. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

AI+ (2021). Developing an artificial intelligence curriculum adapted to European high schools. https://aiplus.udc.es/.

Alvermann, D. E., Smith, L. C., and Readence, J. E. (1985). Prior knowledge activation and the comprehension of compatible and incompatible text. *Reading Research Quarterly*, 420–436.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18(5)*:382–393.

Duncan, R. G. and Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching, 46(6)*:606–609.

Forsyth, S., Dalton, B., Foster, E. H., Walsh, B., Smilack, J., and Yeh, T. (2021). Imagine a more ethical AI: Using stories to develop teens' awareness and understanding of AI and its societal impacts. *Conference Proceedings of Research in Equitable and Sustained Participation in Engineering, Computing, and Technology*, 1–2.

Friel, S. N., Curcio, F. R., and Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32(2)*:124–158.

Gardner-McCune, C., Touretzky, D., Martin, F., and Seehorn, D. (2019). AI for K–12: Making room for AI in K–12 CS curricula. *ACM Technical Symposium on Computer Science Education*, 1244–1244.

Greenwald, E., Leitner, M., and Wang, N. (2021). Learning artificial intelligence: Insights into how youth encounter and build understanding of AI concepts. *Proceedings of the AAAI Conference on Artificial Intelligence, 35*, 15526–15533.

Ho, J. W., Scadding, M., Kong, S., Andone, D., Biswas, G., Hoppe, H., and Hsu, T. (2019). Classroom activities for teaching artificial intelligence to primary school students. *Conference Proceedings on Computational Thinking Education*, 157–159.

Kandlhofer, M., Steinbauer, G., Laßnig, J. P., Baumann, W., Plomer, S., Ballagi, A., and Alfoldi, I. (2019). Enabling the creation of intelligent things: Bringing artificial intelligence and robotics to schools. *IEEE Frontiers in Education Conference*, 1–5.

Lee, I., Martin, F., Denner, J., Coulter, B., Allan, W., Erickson, J., Malyn-Smith, J., and Werner, L. (2011). Computational thinking for youth in practice. *ACM Inroads, 2(1)*:32–37.

Lee, S., Mott, B., Ottenbreit-Leftwich, A., Scribner, J., Taylor, S., Park, K., Rowe, J., Glazewski, K., Hmelo-Silver, C. E., and Lester, J. (2021). AI-infused collaborative inquiry in upper elementary school: A game-based learning approach. *Symposium on Education Advances in Artificial Intelligence, 35*, 15591–15599.

Lin, P., Van Brummelen, J., Lukin, G., Williams, R., and Breazeal, C. (2020). Zhorai: Designing a conversational agent for children to explore machine learning concepts. *Proceedings of the AAAI Conference on Artificial Intelligence, 34*,13381–13388.

Liu, X. and Koirala, H. (2009). The effect of mathematics self-efficacy on mathematics achievement of high school students. *Proceedings of Northeastern Educational Research Association Conference*.

Long, D. and Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the CHI conference on human factors in computing systems*, 1–16.

MIT AI Education Initiative (2021). What's the DAILy curriculum? https://raise.mit.edu/daily/index.html.

Payne, W. C., Bergner, Y., West, M. E., Charp, C., Shapiro, R. B. B., Szafir, D. A., Taylor, E. V., and DesPortes, K. (2021). danceON: Culturally responsive creative computing. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–16.

Perkins, D. N., Salomon, G., et al. (1992). Transfer of learning. *International Encyclopedia of Education, 2*:6452–6457.

Peterson, D., Goode, K., and Gehlhaus, D. (2021). AI education in China and the United States: A comparative assessment. Technical report, Center for Security and Emerging Technology.

Plass, J. L., Mayer, R. E., and Homer, B. D. (2020). *Handbook of game-based learning*. MIT Press.

Ritter, S., Aglio, J., Stetzer, R., Wilson, G., Navta, N., and Lippl, C. (2019). Teaching AI as computational thinking for middle school students. *Workshop on K-12 AI Education at the 20th International Conference on AI in Education*.

Rodríguez-García, J. D., Moreno-León, J., Román-González, M., and Robles, G. (2021). Evaluation of an online intervention to teach AI with LearningML to 10–16 year old students. *ACM Symposium on Computer Science Education*, 177–183.

Russell, S. J. and Norvig, P. (2016). Artificial intelligence: A modern approach. *Pearson Education Limited*.

Sabourin, J., Rowe, J. P., Mott, B. W., and Lester, J. C. (2011). When off-task is on-task: The affective role of off-task behavior in narrative-centered learning environments. *Proceedings of International Conference on Artificial Intelligence in Education*, 534–536. Springer. Educational

Sakulkueakulsuk, B., Witoon, S., Ngarmkajornwiwat, P., Pataranutaporn, P., Surareungchai, W., Pataranutaporn, P., and Subsoontorn, P. (2018). Kids making AI: Integrating machine learning, gamification, and social context in STEM education. *IEEE international conference on teaching, assessment, and learning for engineering (TALE)*, 1005–1010.

Shamir, G. and Levin, I. (2020). Transformations of computational thinking practices in elementary school on the base of artificial intelligence technologies. *Proceedings of EDULEARN20, volume 6*, 1596–1605.

van Loon, M. H., de Bruin, A. B., van Gog, T., and van Merriënboer, J. J. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. *Learning and Instruction, 24*:15–25.

Wang, N. and Johnson, M. (2019). AI education for K–12: Connecting AI concepts to high school math curriculum. *Proceedings of the IJCAI Workshop on Education in Artificial Intelligence K-12*.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 46(6)*:716–730.

Youjun, X., Jiqing, W., and Jinson, H. (2018). Textbook series on Artificial Intelligence for Elementary and Middle Schools. *East China Normal University Press*.

Yukun, C. and Tang, X. (2018). Fundamentals of Artificial Intelligence for High Schools. *East China Normal University Press*.

Zhou, X., Tang, J., Daley, M., Ahmad, S., and Bai, Z. (2021). "Now, I want to teach it for real!": Introducing machine learning as a scientific discovery tool for K–12 teachers. *Proceedings of International Conference on Artificial Intelligence in Education*, 486–499. Springer.