

WOLFRAM in Action: Teaching and Learning (Pseudo)Random Generation with Cellular Automata in Higher Education Settings

Zach Anthis^{1,2,*} and Lefteris Zacharioudakis^{2,3,†}

¹ UCL Knowledge Lab, University College London, Gower Street, London, WC1E 6BT, United Kingdom

² Department of Computer Science, Neapolis University Pafos, Danaes Avenue, Pafos, 8042, Cyprus

³ Igor Sikorsky Kyiv Polytechnic Institute, National Technical University of Ukraine, Kyiv, 03056, Ukraine

Abstract

This article presents ongoing work on WOLFRAM, an interactive EdTech tool designed to teach random generation by visualizing unidimensional Cellular Automata (CA). The web-based prototype integrates a series of gamified tasks with a Learning Analytics (LA) dashboard, to provide students with hands-on experience in elementary CA mechanics whilst delivering detailed insights to instructors in real time. The backend tracks user progress through key performance metrics, including response times, task accuracy, and engagement levels. Preliminary results from a quasi-experimental study demonstrate substantial learning gains across two distinct cohorts: BSc Computer Science (CS) students in a Cybersecurity module and BSc Artificial Intelligence (AI) students in a Machine Learning module. Both cohorts reported high usability and motivation via quantitative Likert scale assessments, with ANOVA showing no significant differences in these areas. Yet, AI students exhibited notably higher improvements in learning clarity, likely due to stronger curricular alignment with CA concepts. In fact, regression analysis confirmed that being in the AI group significantly predicted greater clarity in general, even after controlling for other factors. Next steps involve the integration of adaptive learning features to dynamically adjust content difficulty based on recorded student performance, alongside additional predictive and prescriptive components to provide for automated feedback (in the form of AI-driven hints) on a need-to basis. Future research will focus on expanding the tool's scalability across various (adjoining) academic disciplines and investigating its impact on long-term retention of more advanced concepts such as fractal geometry, entropy estimation, algorithmic complexity, pattern formation, or self-organization.

Keywords

Cellular Automata (CA), Learning Analytics (LA), Computer Science (CS), Artificial Intelligence (AI)

1. Introduction

Recent advances in Artificial Intelligence (AI) and Cybersecurity are jointly reshaping modern technology, with applications ranging from everyday conveniences to complex multi-tier architectures bound to safeguard critical information. As these innovations proliferate, there is a growing demand for individuals who possess a deep understanding of key concepts pertinent to their inherent stochasticity. Afterall, from reactive machines to limited memory and self-awareness ecosystems, predictive modeling increasingly relies on foundational principles of randomness, while navigating statistical and epistemic uncertainties—and for good reason. Controlled randomization has become crucial in several aspects of applied Machine Learning (ML); notably, data shuffling or augmentation, initialization, error bounding, and model training/testing by use of iterative (hyper)parameter optimization [1]. Moreover, as cyberthreats evolve, randomization plays a pivotal role in mitigating risks and maintaining the overall integrity of secure communications, enhancing

1st Workshop on Education for Artificial Intelligence (edu4AI 2024, <https://edu4ai.di.unito.it/>), co-located with the 23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024). 26-28 November 2024, Bolzano, Italy

* Corresponding author.

† These authors contributed equally.

✉ qtnvzan@ucl.ac.uk (Z. Anthis); lzacharioudakis@nup.ac.cy (L. Zacharioudakis)

ORCID [0000-0001-5359-4111](https://orcid.org/0000-0001-5359-4111) (Z. Anthis); [0000-0002-9658-3073](https://orcid.org/0000-0002-9658-3073) (L. Zacharioudakis)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

the robustness of encryption algorithms which prevent malicious actors from easily decoding sensitive data. However, it is crucial to distinguish between pseudo-random number generators (PRNGs) and true random number generators (TRNGs) in this context. While PRNGs use deterministic algorithms to produce sequences that appear random, they can be vulnerable if the initial seed or algorithm becomes known to an attacker [2]. This predictability poses significant risks in hands-on cryptographic applications, but may also introduce selection, confirmation, or algorithmic biases across the MLOps pipeline. In contrast, TRNGs derive randomness from inherently unpredictable physical phenomena, such as electronic noise, radioactive decay, or quantum effects [3]. These sources provide true entropy, resulting in non-deterministic outputs that cannot be predicted or replicated without direct access to the source. Their use is essential in generating cryptographic keys that are truly random, to prevent attackers from being able to guess or calculate them, thus maintaining the integrity and confidentiality of secure communications. Similarly, TRNGs can add to the reliability of ML models (establishing convexity, stability, and generalizability) while minimizing the risks of discernible patterns in scenarios where objectivity/impartiality, transparency, fairness, security, and privacy are of the essence (e.g., adversarial training for fraud detection). On the other hand, the growing accessibility of once-disruptive technologies, such as Internet-of-Things (IoT) [4] and Generative AI (genAI) [5], has fueled offensive strategies that could potentially exploit system vulnerabilities and has driven demand for effective counter-measures grounded in quantum computing [6] or decentralization [7]. Sure enough, this in turn has led to a standalone surge in ML adoption for managing complexity and allocating resources on the receiving end [8]. All in all, in both fields alike, these intricate decisions are furnished by AI one way or another, and that usually means testing out contrastive estimation tactics, iteratively sampling rows (instances) or columns (features), conducting randomized reduction or recovery, and performing random repeats or restarts. In light of these developments, the ability to understand, apply, and critically evaluate random generation has become a highly sought-after skill [9], underscoring the importance of teaching it to students in Computer Science (CS) and its interdisciplinary domains (e.g., AI, data science, or computational engineering).

2. Background

Despite the ubiquity of randomization in CS, teaching this core concept presents unique challenges. If experience has taught us anything, it's that traditional classroom methods often struggle to convey the complexity and importance of random processes, leaving students without the intuitive grasp necessary for effective application. Indeed, even when a strong conceptual understanding is achieved, transferability to diverse AI and Cybersecurity contexts requires adaptive expertise, a challenge often highlighted in both educational and cognitive psychology research [10]. The ontological basis of this study asserts that optimal learning occurs through interactive, contextualized experiences that foster deeper exploration of underlying concepts (e.g., emergent behaviors arising from local interactions, governed by simple, deterministic rules). Epistemologically speaking, this aligns well with constructivist theories [11, 12], highlighting that the knowledge in question is best built through active engagement and meaningful social contexts, which enable learners to integrate new information with prior knowledge for deeper cognitive processing. Thus, the need for educational tools that can transform abstract mathematical theory into tangible, stimulating learning interactions is more pressing than ever.

One promising approach lies in Cellular Automata (CA), a simple yet powerful computational model, particularly suited for teaching complex patterns and behaviors that can arise from simple rules, mirroring the unpredictability of randomization in dynamic systems [13]. Its flexibility, ease-of-use, and suitability for optimizing state-space exploration, outperforms many alternative models in terms of interpretability, scalability, and cognitive alignment. All practicalities in probabilistic model simulation (generating diversity in state transitions) aside, its capacity for interactive experimentation and openness to visualization correspondingly reflect established test preconditions (controllability) and tractable means for empirical verification (observability). This makes CA a

fitting representation framework for illustrating both randomization and complexity in a variety of contexts, including AI and Cybersecurity.

Since Von Neumann's introduction of the “universal constructor mechanism” in the 1940s [14], educational research has persistently explored the benefits of CA in modeling complex natural phenomena such as insect colonies, bird flight paths, and even DNA sequencing [15], as well as in cultivating computational thinking and enhancing problem-solving skills; more so among students in CS [16-18]. Yet, its specific use for teaching principles of complexity theory, particularly in AI and Cybersecurity, is underexplored. The present study fills this gap by examining the role of teaching CA through an interactive, game-based tool, transforming the experience into an integrated Exploratory Learning Environment (ELE) that can empower students from diverse backgrounds to manipulate content in real time. Gamification, recognized for its ability to enhance motivation and engagement, is turning into an evidently powerful feature of modern-day EdTech [19-21]. By integrating game-like elements such as challenges, rewards, and leaderboards, educational platforms can turn passive learning into active participation, shifting from traditional rote learning to a more immersive experience, where students are motivated by the sense of achievement and progress. In doing so, the design itself capitalizes on intrinsic motivators, like curiosity and competition, encouraging students to solve problems, explore concepts, and persevere through challenging tasks in a low-stakes environment.

At the same time, Learning Analytics (LA) is well known to enhance the learning experience (LX) of its own, by collecting and processing usage data on student performance and behavior [22, 23]. Metrics such as task completion time, accuracy, or engagement levels, provide valuable insights into how students interact with the content, which is crucial in Higher Education (HE) settings [24, 25]. For one, LA dashboards allow instructors to summarize relevant data at different granularities with little, or no, programming skill, and thus personalize LXs by adjusting the difficulty or pacing of tasks based on individual performance. Similarly, they enable the early identification of students who may struggle or disengage, fostering contextualized interventions with tailored support, feedback, or resources. This combination of game-based learning and just-in-time analytics creates a responsive environment where data-driven strategies can improve both student outcomes and instructional effectiveness alike.

3. Research Objectives

This study tests the effectiveness of a newly developed EdTech tool as an end-to-end solution for improving learning outcomes across two distinct educational settings: a Cybersecurity module from the BSc in Computer Science course and a Machine Learning module from the BSc in Artificial Intelligence course. This first assessment of the tool aims to address the challenges of teaching (pseudo-)random generation and contribute to a broader understanding of how data-driven learning environments can support AI for Education (AIEd), as much as Education for AI (EdAI). To evaluate the didactic impact of the intervention, it is essential to review how it influences both the user experience and educational outcomes across these diverse learning contexts, in a systematic fashion; specifically, to quantifiably measure usability and motivation subscales, and explore potential differences in reported learning gains.

In response, the Web-based Orchestrated Learning for Random Automata Modeling (WOLFRAM), was developed as a unified platform specifically designed for teaching CA in a structured and accessible way through experiential learning. The tool combines real-time visualizations with gamified tasks and integrates teacher-centered dashboarding, to create an immersive experience that allows students to explore randomization interactively. By tracking metrics such as response times and task accuracy, it enables instructors to monitor student progress and adapt learning pathways accordingly (see Figure 1).

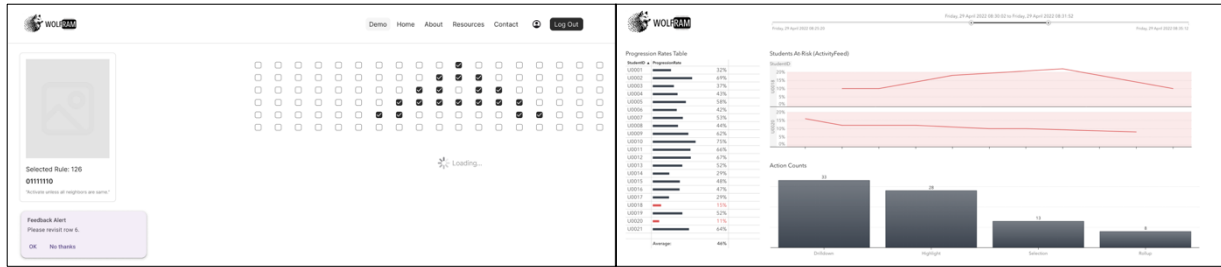


Figure 1: WOLFRAM interface (student view) and the integrated LA dashboard (teacher view).

RQ1: What is the impact of WOLFRAM on students' perceived usability in the context of teaching randomization concepts in Computer Science and Artificial Intelligence courses?

RQ2: How do the effects of WOLFRAM on student motivation and learning outcomes vary between Computer Science and Artificial Intelligence students?

4. Methodology

The broader methodological framework was patterned after the Designed-based Research Collective (DBRC) paradigm [26] which has been used extensively in the past in order to align Technology-enhanced LE (TELEs) with their fundamental epistemological and theoretical assumptions [27, 28]. As a short-term educational program, the intervention in its entirety was based upon the Four-Component Instructional Design (4C/ID) approach for complex learning [29]. To avoid common detachments of EdTech research from policy and practice, the selected approach engages in iterative designs and evaluations (collaborating with research subjects in the process), in the frame of two distinct UG modules, to achieve the right balance between theory-building [30] and practical impact [31].

4.1. Participants

Participants were selected based on their enrollment in courses directly aligned with the research focus. The study involved 60 undergraduate students from our university, divided into two cohorts: a) 36 first-year BSc CS students enrolled in the Cybersecurity module, split evenly into a control group (n=18) and a treatment group (n=18); and b) 24 first-year BSc AI students enrolled in the ML module, similarly divided into a control group (n=12) and a treatment group (n=12). The investigation aimed to gauge the effectiveness of the tool as a means for explicating RNG through CA, while comparing the results of the treatment group (who were granted access to use the platform freely) with the control group who received traditional instruction. The selection of the study population adheres to well-established guidelines in user-centred EdTech research [32] and aligns with design-based research principles, where researchers proceed to empirically test the impact of proposed interventions within real educational settings, while pursuing the generalizability of results to similar academic environments. This methodological approach draws from multiple theoretical perspectives and research paradigms so as to build understandings of the nature and conditions of learning, cognition, and development [33]. Purposeful sampling was employed, to ensure that participants are key stakeholders, directly engaged in learning the complex concepts the tool was designed to address. Undergraduates enrolled in these modules were considered as highly relevant subpopulations, given the CA's twin capacity (as discrete dynamical systems and information-processing systems) and the practical applications of random generators being foundational to both these fields. The cohorts were chosen to assess whether the tool could meet distinct learning requirements against both academic domains. This caters the need for ecological validity (findings being applicable to real-world scenarios), accounting for the experimental circumstances, stimuli under investigation, and behavioral response [34].

4.2. Study Design

A quasi-experimental pre-test/post-test design was employed to critically evaluate the impact of the tool on perceived usability, motivation, and learning outcomes, divided into three distinct phases: (a) Pretest phase: All participants were given 30 minutes to complete a pre-test designed to evaluate their baseline knowledge, including conceptual questions and problem-solving tasks, to provide a comprehensive assessment of their understanding of theoretical/practical aspects of randomness; (b) Intervention phase: Two days later, (randomly assigned) participants in the treatment groups attended a single 3-hour session, featuring a series of in-class interactive tasks via WOLFRAM beta, meant to cover CA discrete evolution and (pseudo-)random generation. At the same time, control group attended a parallel session, receiving only traditional instruction (lectures and textbook-based exercises) instead; (c) Post-test phase: A week after completing the intervention, both groups took a post-test, identical in structure to the pre-test, to evaluate their learning gains. The post-test assessed conceptual understanding, problem-solving skills, as well as the ability to apply randomization principles in new contexts, whereby real-world systems reflect local interactions between individual components leading to emergent global behaviors, or where global order may arise without centralized control.

The selected approach aimed for a model that is robust enough to detect moderate effects in a real-world educational environment and provides sufficient statistical rigor to meet the research objectives of a small-scale study. To ensure statistical validity, a power analysis was conducted, assuming a modest effect size (Cohen's $d = 0.5$), power of 0.80, and a significance level of 0.05 [35], which suggested a sample size of at least 64 participants (32 per group). However, due to practical constraints in class settings, we ended up targeting slightly less participants per cohort, consistent with typical EdTech studies [36, 37]. These numbers still provide adequate power, especially given the inclusion of repeated measures, which tend to enhance statistical efficiency by controlling for within-subject variability [38].

4.3. Data Collection and Analysis

The LA dashboard monitored several real-time metrics during the intervention, which included: (a) Task Completion Times (duration taken to complete tasks involving CA-based randomization); (b) Task Accuracy (the correctness of responses in problem-solving tasks e.g., CA expansions); (c) Engagement Levels (elapsed time spent on tasks and interaction frequency with the platform). In addition to these metrics (and the pre- and post-tests administered to all participants), the study utilized two technology-agnostic (validated) instruments to assess self-reported motivation and usability respectively: a) the Intrinsic Motivation Inventory (IMI), to gauge student engagement, interest, and perceived competence (clarity) [39, 40]; and b) the standardized 10-item System Usability Scale (SUS) to calculate the perceived usability of the platform in the treatment group [41]. Importantly, in this study, learning clarity is taken to be subjective (with task accuracy representing its objective counterpart), allowing for a balanced assessment of both perceived and demonstrated understanding.

In terms of primary statistical methods, an Analysis of Variance (ANOVA) was used to compare usability, motivation, and learning clarity across the two cohorts, providing insight into whether significant differences existed between the groups in their response to traditional instruction versus the intervention. A regression analysis was then conducted to examine whether belonging to the AI cohort predicted higher learning clarity (and/or task accuracy), while controlling for engagement and usability, for a more focused exploration of the specific factors that could drive looked-for learning outcomes. Analyses of data gathered from multimodal usage metrics in combination with pre- and post-tests, SUS scores, and IMI assessments, reveals significant findings related to the effectiveness of the intervention.

4.4. Preliminary Results

Firstly, both cohorts demonstrated notable learning gains. The CS group improved by 42%, while the AI group saw a 46% improvement, overall (using weighted averages). While the post-test scores were significantly higher for both ($p < 0.01$), the difference in improvement between groups was not statistically significant ($p = 0.07$) though the AI cohort did show a trend of higher knowledge retention.

Table 1
Pre-Test and Post-Test Scores (Inter-group Breakdown)

Cohort	Group	Pre-Test	Post-Test	Improvement (%)
AI (n=24)	Control	56.5	81.5	44
	Treatment	58.5	86.6	46
CS (n=36)	Control	52.5	74.5	42
	Treatment	52.5	75.5	43.8

Nevertheless, the breakdown analysis of various LA metrics traced during the intervention phase, resulted in highlighting significant differences in task completion times, accuracy, and engagement levels, which were further corroborated by the submitted feedback on usability and declared student motivation (SUS and IMI respectively) across the treatment groups, as summarized in Table 2.

Table 2
ANOVA Metrics (Treatment Subgroups)

Metric	CS (n=18)	AI (n=12)	p-value
Task Completion Time (mins)	14.2 ± 3.5	12.8 ± 3.1	0.12
Task Accuracy (%)	77.0 ± 7.8	80.4 ± 6.9	0.04*
Engagement (hrs)	1.8 ± 0.3	2.2 ± 0.4	0.05*
Perceived Usability (SUS)	84.3 ± 7.2	87.1 ± 6.9	0.20
Intrinsic Motivation (IMI)	4.2 ± 0.5	4.6 ± 0.4	0.03*

5. Discussion

According to our own in-class observations, integrating RNG principles into CS curriculum through elementary CA strengthens student understanding of the vital role of randomization in handling uncertainty, in general. As expected, by simulating genuinely volatile systems and enabling users to generate and test random numbers, the tool leveraged visual perception to help connect abstract theoretical concepts to practical real-world applications across the two cohorts, aligning with well-established constructivist theories [42, 43]. The flow successfully comprised disparate scenes (interconnecting to form a looping map made up of several branching scenarios) that let learners independently work through complex search subroutines, explore how their choices can lead to different outcomes, or start over and see where another path might take them. This has equipped students with the skills to grasp ML internal mechanics (e.g., overfitting or variance-bias tradeoff), apply common troubleshooting techniques (e.g., early stopping, regularization, and cross-validation), and deepen their understanding of cryptographic security and appreciation of true randomness in defense against evolving cyberthreats. Nonetheless, the AI subgroup demonstrated significantly higher task accuracy ($p = 0.04$) and showed higher engagement (in terms of duration), compared to CS students ($p = 0.05$). Although both subgroups seem to have rated WOLFRAM highly, with no significant difference in perceived usability ($p = 0.20$), AI students reported significantly higher intrinsic motivation ($p = 0.03$). Regression analysis revealed that membership in the AI cohort significantly predicted higher learning clarity ($R^2 = 0.14$, $p = 0.02$) and task accuracy ($R^2 = 0.12$, $p =$

0.04), when controlling for engagement and usability. This suggests that the direct relevance of randomization in ML has likely contributed to the better outcomes in this group, and that the tool can meet our expectations in demystifying analogous stochastic processes in the future (e.g., random walks, Monte Carlo simulations, and noise generation).

In conclusion, WOLFRAM has proven to be an overall effective tool, versatile enough for teaching random generation concepts in both AI and Cybersecurity education. Results show that it enhances teaching, learning, motivation, and engagement, particularly among AI students, where the alignment of CA with the curriculum is perhaps more pronounced. From an LX perspective, the findings highlight the role of interactive, gamified environments in improving both student conceptual understanding and task accuracy, while fostering interest-related motivational constructs such as active participation, reflection, self-regulation, and sustained autonomy. With regards to instructional design implications, the differential impact of WOLFRAM across the two cohorts suggests that content relevance is indeed critical for maximizing learning outcomes [44], whereas tailoring learning tools to the specific domain—such as integrating CA into straightforward scenarios relating to cybersecurity—may further enhance accuracy and engagement in fields where direct applicability is less apparent [45]. This strong relationship between curriculum alignment and outcome underscores the importance of designing EdTech tools that closely integrate with course-specific objectives and individual learning paths.

6. Limitations and Future Work

While yielding promising immediate results, the study presents acknowledgeable limitations. Firstly, the small sample size and quasi-experimental design restrict the generalizability of the findings. The limited participant pool may not fully represent the diverse range of learners and educational contexts, potentially skewing the portability and replicability of the results. Moreover, the focus on short-term learning gains does not address the critical question of long-term retention. Assessing how well participants retain and apply the learned material over extended periods, remains an essential but unexplored dimension. Finally, perceived usefulness and dependability are yet to be tested in broader contexts, beyond the scope of theoretical CS. For instance, it is still unclear how the findings translate to teaching modern combinatorics in other fields, such as pure vs. applied mathematics or electrical vs. mechanical engineering, which may entail different pedagogical requirements and learning outcomes.

To address these gaps, future research should prioritize larger, more diverse samples to enhance the external validity of the findings. Also, longitudinal studies are needed to empirically assess the long-term retention of knowledge and the pertinency of the tool in varied academic disciplines. Specifically, future work should systematically examine the scalability of the WOLFRAM interface across different domains and educational levels. This includes integrating adaptive learning features that adjust content difficulty based on real-time performance data, thereby personalizing the learning experience. Finally, a mixed-methods approach is recommended for future investigations, incorporating qualitative judgements through semi-structured interviews and/or focus group discussions, to capture the richness of human experience in terms of learning needs, preferences, or barriers (e.g., self-efficacy, cognitive load, and affective evaluation aspects). Such an approach will provide a more nuanced understanding of learner perceptions and the (meta-)cognitive processes involved. These steps are essential for assessing WOLFRAM's practicality (its ability to function effectually across diverse scenarios), adaptability (its capacity to adjust to evolving requirements), and overall consistency (its reliability in providing a uniform experience and maintaining high performance standards). Together, these factors determine how effectively the tool can support diverse learning environments (and how successfully it supports various subject-specific learning contexts) and thus will make key areas of future investigation.

References

- [1] Z. Anthis, "The Black-Box Syndrome: Embracing Randomness in Machine Learning Models," in *Artificial Intelligence in Education*, Cham, M. M. Rodrigo, N. Matsuda, A. I. Cristea, and V. Dimitrova, Eds., 2022// 2022: Springer International Publishing, pp. 3-9.
- [2] D. Eastlake 3rd, J. Schiller, and S. Crocker, "Rfc 4086: randomness requirements for security," ed: RFC Editor, 2005.
- [3] A. J. Menezes, P. C. Van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*. CRC press, 2018.
- [4] C. Vijayakumaran, B. Muthusenthil, and B. Manickavasagam, "A reliable next generation cyber security architecture for industrial internet of things environment," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1, p. 387, 2020.
- [5] M. Gupta, C. Akiri, K. Aryal, E. Parker, and L. Praharaj, "From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy," *IEEE Access*, 2023.
- [6] C. Abellan and V. Pruneri, "The future of cybersecurity is quantum," *IEEE Spectrum*, vol. 55, no. 7, pp. 30-35, 2018.
- [7] O. O. Malomo, D. B. Rawat, and M. Garuba, "Next-generation cybersecurity through a blockchain-enabled federated cloud framework," *The Journal of Supercomputing*, vol. 74, no. 10, pp. 5099-5126, 2018.
- [8] M. Mosca, "Cybersecurity in an era with quantum computers: Will we be ready?," *IEEE Security & Privacy*, vol. 16, no. 5, pp. 38-41, 2018.
- [9] L. Crocetti, P. Nannipieri, S. Di Matteo, L. Fanucci, and S. Saponara, "Review of methodologies and metrics for assessing the quality of random number generators," *Electronics*, vol. 12, no. 3, p. 723, 2023.
- [10] S. M. Barnett and S. J. Ceci, "When and where do we apply what we learn?: A taxonomy for far transfer," *Psychological bulletin*, vol. 128, no. 4, p. 612, 2002.
- [11] D. H. Schunk, *Learning theories an educational perspective*. Pearson Education, Inc, 2012.
- [12] A. Pritchard, *Ways of learning: Learning theories for the classroom*. Routledge, 2017.
- [13] A. G. Hoekstra, J. Kroc, and P. M. Slood, *Simulating complex systems by cellular automata*. Springer Science & Business Media, 2010.
- [14] J. Von Neumann and A. W. Burks, "Theory of self-reproducing automata," 1966.
- [15] R. J. Gaylord and K. Nishidate, *Modeling nature: Cellular automata simulations with Mathematica®*. Springer, 2013.
- [16] G. Faraco, P. Pantano, and R. Servidio, "The use of cellular automata in the learning of emergence," *Computers & Education*, vol. 47, no. 3, pp. 280-297, 2006.
- [17] T. Staubitz, R. Teusner, C. Meinel, and N. Prakash, "Cellular Automata as basis for programming exercises in a MOOC on Test Driven Development," in *2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 2016: IEEE, pp. 374-380.
- [18] M. Voskoglou and S. Buckley, "Problem Solving and Computational Thinking in a Learning Environment," 12/02 2012.
- [19] B. Marín, J. Frez, J. Cruz-Lemus, and M. Genero, "An empirical investigation on the benefits of gamification in programming courses," *ACM Transactions on Computing Education (TOCE)*, vol. 19, no. 1, pp. 1-22, 2018.
- [20] M. Sailer and L. Homner, "The gamification of learning: A meta-analysis," *Educational psychology review*, vol. 32, no. 1, pp. 77-112, 2020.
- [21] C. V. de Carvalho and A. Coelho, "Game-based learning, gamification in education and serious games," vol. 11, ed: MDPI, 2022, p. 36.
- [22] M. Bienkowski, M. Feng, and B. Means, "Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief," *Office of Educational Technology*, US Department of Education, 2012.

- [23] B. Rienties and L. Toetenel, "The impact of 151 learning designs on student satisfaction and performance: social learning (analytics) matters," in Proceedings of the sixth international conference on learning analytics & knowledge, 2016, pp. 339-343.
- [24] J. T. Avella, M. Kebritchi, S. G. Nunn, and T. Kanai, "Learning analytics methods, benefits, and challenges in higher education: A systematic literature review," *Online Learning*, vol. 20, no. 2, pp. 13-29, 2016.
- [25] D. Gasevic, Y.-S. Tsai, S. Dawson, and A. Pardo, "How do we start? An approach to learning analytics adoption in higher education," *The International Journal of Information and Learning Technology*, 2019.
- [26] D.-B. R. Collective, "Design-based research: An emerging paradigm for educational inquiry," *Educational researcher*, vol. 32, no. 1, pp. 5-8, 2003.
- [27] [27] F. Wang and M. J. Hannafin, "Design-based research and technology-enhanced learning environments," *Educational technology research and development*, vol. 53, no. 4, pp. 5-23, 2005.
- [28] L. Zheng, "A systematic literature review of design-based research from 2004 to 2013," *Journal of Computers in Education*, vol. 2, no. 4, pp. 399-420, 2015.
- [29] J. J. van Merriënboer and P. A. Kirschner, "4C/ID in the context of instructional design and the learning sciences," in *International handbook of the learning sciences*: Routledge, 2018, pp. 169-179.
- [30] M. Warr, P. Mishra, and B. Scragg, "Designing theory," *Educational Technology Research and Development*, vol. 68, no. 2, pp. 601-632, 2020.
- [31] R. Tormey, C. Hardebolle, F. Pinto, and P. Jermann, "Designing for impact: a conceptual framework for learning analytics as self-assessment tools," *Assessment & Evaluation in Higher Education*, vol. 45, no. 6, pp. 901-911, 2020.
- [32] M. Schmidt, A. A. Tawfik, I. Jahnke, and Y. Earnshaw, "Learner and user experience research: An introduction for the field of learning design & technology," *EdTech Books*, 2020.
- [33] S. Barab and K. Squire, "Design-based research: Putting a stake in the ground," in *Design-based Research*: Psychology Press, 2016, pp. 1-14.
- [34] M. A. Schmuckler, "What is ecological validity? A dimensional analysis," *Infancy*, vol. 2, no. 4, pp. 419-436, 2001.
- [35] M. A. Kraft, "Interpreting effect sizes of education interventions," *Educational researcher*, vol. 49, no. 4, pp. 241-253, 2020.
- [36] J. W. Creswell and J. D. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2017.
- [37] L. Castañeda and B. Williamson, "Assembling New Toolboxes of Methods and Theories for Innovative Critical Research on Educational Technology," *Journal of New Approaches in Educational Research*, vol. 10, no. 1, pp. 1-14, 2021/01/01 2021, doi: 10.7821/naer.2021.1.703.
- [38] A. C. Cheung and R. E. Slavin, "How methodological features affect effect sizes in education," *Educational Researcher*, vol. 45, no. 5, pp. 283-292, 2016.
- [39] M. R. Lepper and T. W. Malone, "Intrinsic motivation and instructional effectiveness in computer-based education," in *Aptitude, learning, and instruction*: Routledge, 2021, pp. 255-286.
- [40] C. Bosch, "Assessing the Psychometric Properties of the Intrinsic Motivation Inventory in Blended Learning Environments," *Journal of Education and e-Learning Research*, vol. 11, no. 2, pp. 263-271, 2024.
- [41] P. Vlachogianni and N. Tselios, "Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review," *Journal of Research on Technology in Education*, vol. 54, no. 3, pp. 392-409, 2022.
- [42] D. C. Phillips, "The good, the bad, and the ugly: The many faces of constructivism," *Educational researcher*, vol. 24, no. 7, pp. 5-12, 1995.
- [43] M. Tam, "Constructivism, instructional design, and technology: Implications for transforming distance learning," *Journal of Educational Technology & Society*, vol. 3, no. 2, pp. 50-60, 2000.

- [44] E. A. Kohler, L. M. Elreda, and K. Tindle, "EdTech Context Inventory: Factor analyses for ten instruments to measure edtech implementation context features," *Computers & Education*, vol. 195, p. 104709, 2023.
- [45] M. R. N. King, S. J. Rothberg, R. J. Dawson, and F. Batmaz, "Bridging the edtech evidence gap," *Información Tecnológica*, vol. 18, no. 1, pp. 18-40, 2016.