Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Behavior-Aware Argumentation: Integrating Toulmin's Model for Adaptive Learning in Digital Environments

Banni Satria ANDOKO[a*], Vivin Ayu LESTARI[a], Usman NURHASAN[a], Rosa Andrie ASMARA[a], Galih Putra RIATMA[b], Ananda Galih PRATIWI[a], Tsukasa HIRASHIMA[c]
[a]*Information Technology Dept, State Polytechnic of Malang, Indonesia*
[b]*Electrical Engineering Dept, State Polytechnic of Malang, Indonesia*
[c] *Graduate School of Advanced Science and Engineering, Hiroshima University, Japan*
*ando@polinema.ac.id

Abstract: This study investigates an adaptive learning mechanism within Viat-map, a learning tool based on Toulmin's Argument Model, to address engagement challenges among struggling and gaming students. In addition to scoring correctness, the system monitors two key behavioral indicators: the number of steps taken and the time spent on tasks. These indicators are used to tailor the number of multiple-choice options, reducing them for overloaded learners and increasing them for those who may be guessing strategically. In a controlled experiment with sophomore EFL students, pre- and post-tests, ANCOVA, and effect size analyses showed significant gains for struggling learners, while high performers benefited less. Findings highlight the value of behavior-aware analytics in optimizing educational technologies, and offer insight into refining personalized learning algorithms.

Keywords: Adaptive Learning, Toulmin Argument, Viat-map, EFL, Constructivism

1. Introduction

Understanding how students interact with digital learning environments is essential to creating effective interventions. Viat-map implements Toulmin's Argument Model, a framework built on claim, ground, and warrant, to help learners develop logical reasoning and critical thinking (Andoko et al., 2023; Andoko, Mubarok, et al., 2022). The Toulmin model, which emphasizes the construction of arguments through Claims, Grounds, and Warrants, offers a structured approach to developing logical thinking skills, a core component of critical reasoning (Al-Ajm & Ambusaidi, 2022; Majidi et al., 2021). Toulmin's structure is especially relevant here because it provides clear scaffolding for reasoning. This makes it easier for learners to focus their working memory on processing ideas rather than holding unstated connections (Le Cunff et al., 2024; Tzafilkou et al., 2021).

Previous evaluations of Viat-map have focused mostly on whether students reached correct answers. This approach risks missing the underlying learning process, particularly for two types of learners (Andoko et al., 2024).:

- Struggling learners, who spend a long time working but make little progress, often because of cognitive overload
- Gaming learners, who exploit system rules to score highly without fully understanding the material

Additionally, phenomena such as "gaming the system"—in which students attempt to exploit platform mechanics to achieve high scores without genuine understanding—highlight the limitations of performance-only assessment methods and underscore the need for behavior-aware analytics (R. Baker et al., 2007; R. S. J. D. Baker et al., 2006).

To address this, Viat-map now incorporates two indicators of engagement: the time spent and the sequence of steps taken. Together, these indicators capture both pacing and exploration patterns. They are widely supported in learning analytics research as useful measures of cognitive engagement and strategic behavior.

A prior clustering study identified both gaming and struggling learner profiles. These informed new adaptive rules: choice complexity is decreased for overloaded learners and increased for rapid guessers. This study tests the effect of those rules and examines how prior knowledge, material type, and learner profile interact with the adaptive mechanisms. The research questions for this study are:

1. How does the adaptive learning intervention impact student learning outcomes after controlling prior knowledge, and does its effect vary across different types of instructional materials?
2. To what extent does prior knowledge influence the effectiveness of adaptive learning, and how do learning gains differ between low-performing and high-performing students?

2. Literature Review

2.1 Cognitive Load

Cognitive Load Theory distinguishes between intrinsic load, which relates to the complexity of the task and the learner's prior knowledge; extraneous load, which comes from irrelevant demands; and germane load, which is the mental effort devoted to learning (Leppink et al., 2013). In EFL reading and argumentation, structured scaffolds such as Toulmin's clear claim-evidence-warrant layout can reduce extraneous load and free mental resources for deeper processing. Research shows that such structures help mitigate working-memory strain, which improves retention and transfer (Wang et al., 2020; Le Cunff et al., 2024).

2.2 Gaming The System

Gaming occurs when learners bypass the intended cognitive process to achieve success, for example by cycling through options or taking advantage of predictable patterns in multiple-choice questions. This behavior can hide the true level of understanding and distort assessment data. Adaptive strategies such as varying complexity or giving targeted feedback have been shown to redirect engagement back to meaningful learning (Petre et al., 2019; R. S. J. D. Baker, 2006).

2.3 Toulmin Argument

Toulmin's model breaks argumentation into components that can be taught explicitly. For adaptive systems, this is valuable for two reasons: it has proven benefits for reasoning skills (Majidi et al., 2021), and its component-based design fits naturally with step-by-step interaction data (Rismanto et al., 2021), making it possible to analyse behaviour in detail (Magalhães, n.d.).

2.4 Viat Map

Viat-map is designed to structure learning using Toulmin's basic argument model, where In Viat-map, claims are set by the teacher and learners choose the grounds and warrants. This keeps the activity structured while giving students a chance to evaluate the quality and relevance of different arguments. The system records each selection step and the time taken, which makes it possible to distinguish deliberate reasoning from random trial-and-error. This information feeds directly into the adaptive mechanisms.

3. Method

3.1 Adaptive Learning Function in Viat-map

Viat-map adjusted the number of answer options in real time, responding to how students interacted with each question. When a student took more than seven seconds to choose, the system treated this as a sign of difficulty and reduced the options from three to two, easing cognitive load. If a student changed their answer more than twice within ten seconds, the system assumed they might be guessing and increased the options to four, prompting more careful consideration. Most students remained at the default of three options unless their behavior matched one of these patterns. In the control group, the number of options never changed.

3.2 Experimental Setting

The study involved 41 second-year students from two intact Information Technology classes at the State Polytechnic of Malang. The classes followed their regular timetable, with one class assigned to the control condition and the other to the adaptive condition. Two reading passages from the standard English syllabus were used: Computer and iPod Nano. Each was paired with a ten-item multiple-choice pre-test and a post-test, and the answer keys were only revealed once the study ended. The experiment was designed following the flow shown in Figure 1

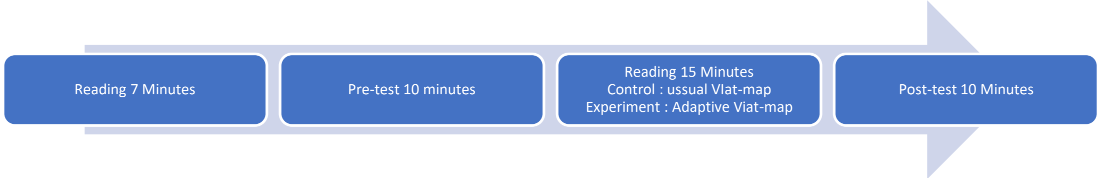

Figure 1. Experimental setting flow

4. Result

Post-test scores were compared between groups using ANCOVA, with pre-test scores as the covariate to control for prior knowledge. Hedges's g was calculated to gauge the practical significance of the results. The first analysis should focus on establishing baseline comparability between the experimental and control groups before measuring the impact of the intervention. Descriptive Statistics for Pre-Test Scores using Welch two-sample t-test was conducted for both materials and here is the result:

To examine differences in prior knowledge before the intervention, we analyzed pre-test scores using Welch's two-sample t-test. As shown in Table 1, the control group had a slightly higher average score (5.375) than the experimental group (4.235), but the difference was not statistically significant ($p = 0.0769$). A second analysis for the second material (Table 2) showed similar results, with the control group scoring 5.416 and the experimental group 4.764 ($p = 0.2408$). In both cases, the p-values exceeded 0.05 and the confidence intervals included zero, indicating no significant baseline differences between the groups.

Table 1. Summary of Pre-Test Score Comparison Between Control and Experimental Groups for iPod Nano Material

Statistic	Control Group	Experimental Group	Test Value
Mean Pre-Test Score	5.375	4.235	—
t-test Value (t)	—	—	1.8172
Degrees of Freedom (df)	—	—	38.859
p-value	—	—	0.0769
95% Confidence Interval	—	—	-0.129 to 2.408

Table 2. Summary of Pre-Test Score Comparison Between Control and Experimental Groups for Computer Material

Statistic	Control Group	Experimental Group	Test Value
Mean Pre-Test Score	5.416667	4.764706	—
t-test Value (t)	—	—	1.1919
Degrees of Freedom (df)	—	—	37.619
p-value	—	—	0.2408
95% Confidence Interval	—	—	-0.455 to 1.760

After ensuring that the baseline of each group are comparable, the analysis now is divided into two sections to show a detailed analysis of each material within groups.

4.1 ANCOVA Analysis for Computer Material

Students' prior knowledge clearly shaped their post-test outcomes, as shown in Table 3. Still, the adaptive learning intervention made a meaningful difference, boosting performance with a statistically significant result ($p = 0.0229$). The model's low residual variance adds confidence that these findings genuinely reflect the intervention's impact.

Table 3. ANCOVA Results for Adaptive Learning Intervention

Factor	Degreesof Freedom (Df)	Sum of Squares (SS)	Mean Square (MS)	F-Value	p-Value
Pre-Test	1	54.35	54.35	64.486	1.04e-09*
Group (Intervention Impact)	1	4.74	4.74	5.627	0.0229*
Residuals	38	32.03	0.84	—	—

*Significance codes: *** (p < 0.001), ** (p < 0.01), * (p < 0.05)*

The adaptive learning program clearly improved student performance, with a strong effect size shown in Table 4. The negative indicator suggests a shift in how students learned. The confidence interval confirms that the improvement was meaningful and not just by chance..

Table 4. Effect Size Calculation (Hedges's g)

Effect Size Measure	Estimate	95% Confidence Interval
Hedges's g	-0.836 (large)	Lower: -1.491, Upper: -0.181

4.2 ANCOVA Analysis for iPod Nano Material

Table 5 shows that pre-test scores had a strong influence on post-test performance, underlining the key role of prior knowledge in shaping learning outcomes. Still, the adaptive learning intervention led to a clear improvement (p = 0.00673), confirming its effectiveness. The moderate residual variance suggests the model was well-fitted and dependable, meaning the results reliably reflect the intervention's impact.

Table 5. ANCOVA Results for Adaptive Learning Intervention

Factor	Degrees of Freedom (Df)	Sum of Squares (SS)	Mean Square (MS)	F-Value	p-Value
Pre-Test	1	134.04	134.04	104.090	1.95e-12*
Group (Intervention Impact)	1	10.58	10.58	8.218	0.00673
Residuals	38	48.94	1.29	—	—

*Significance codes: *** (p < 0.001), ** (p < 0.01), * (p < 0.05)*

Table 6 confirms that the adaptive learning function had a strong impact on student performance, supported by a large effect size. The negative value reflects a shift in performance patterns, indicating potential underlying changes in how students engaged with and benefited from the intervention. The confidence interval reinforces the reliability of this effect, showing that the program produced meaningful improvements in outcomes. Complementing this, Table 7 presents ANCOVA and effect size results for low- and high-performing students across materials, revealing that the strongest and most significant gains occurred among lower performers, particularly for certain content areas. This convergence of findings underscores both the overall effectiveness of the adaptive function and its differential benefits across learner profiles.

Table 6. Effect Size Calculation (Hedges's g)

Effect Size Measure	Estimate	95% Confidence Interval
Hedges's g	-1.001 (large)	Lower: -1.667, Upper: -0.335

Table 7. ANCOVA and Effect Size Results for Low vs. High Performers Across Materials

Material	Performance Group	p-value (ANCOVA, Group Effect)	Hedges's g	95% CI for g
Computer	Low performers	0.0239 *	–1.415 (large)	–2.644 to –0.186

Computer	High performers	0.4003	–0.611 (medium)	–1.411 to 0.188
iPod Nano	Low performers	0.01010 *	–1.276 (large)	–2.291 to –0.260
iPod Nano	High performers	0.3420	–0.600 (medium)	–1.508 to 0.307

5. Discussion and Conclusion

Research Question 1 examined how adaptive learning influenced student outcomes and whether its effect varied with different materials. After accounting for prior knowledge, post-test scores rose significantly ($p = 0.0229$; $p = 0.00673$) with large effect sizes (–0.836 to –1.001). Prior knowledge strongly predicted performance ($p = 0.00281$; $p = 0.000203$). High achievers relied more on existing knowledge, while lower-performing students benefited more from the program. The effect also varied by topic, with both Computer and iPod Nano materials improving results but to different extents, underscoring the value of matching adaptive strategies to the subject.

Research Question 2 showed that low-performing students gained the most from the adaptive learning intervention, with significant results ($p = 0.0239$; $p = 0.01010$) and large effect sizes ($g = -1.415$) compared to smaller, non-significant effects for high performers ($p > 0.34$; $g = -0.600$; CI included zero), whose outcomes were largely driven by prior knowledge. Effectiveness also varied by material, suggesting that subject-specific tailoring can further optimize impact. These findings align with prior work linking domain-specific prior knowledge (Simonsmeier et al., 2022) and cognitive load considerations (Dong et al., 2020) to adaptive learning benefits, and with evidence of domain-dependent effects (Mirari, 2022). They highlight the importance of refining adaptive algorithms to meet diverse learner needs and expanding research to address long-term retention and engagement (Martin et al., 2020).

While these findings support adaptive learning, especially for lower-performing students, they come with limitations. The study involved a small sample from one institution, limiting generalizability. Learning gains were measured only immediately after the intervention, so long-term effects are unknown. The adaptation relied on a narrow set of behavioral indicators and fixed thresholds, which may not fully reflect individual differences. Future studies should use larger, more diverse samples, track long-term outcomes, and draw on richer learning analytics for more flexible, learner-specific adaptations. Including motivation and engagement measures could also reveal how adaptive systems shape the learning experience beyond grades.

6. Acknowledgement

We would like to thank the State Polytechnic of Malang for funding this research under Research grant number: SP DIPA-139.03.2.693474/2025.

References

Al-Ajm, B., & Ambusaidi, A. (2022). The Level of Scientific Argumentation Skills in Chemistry Subject among Grade 11th Students: The Role of Logical Thinking. *Science Education International*, *33*(1), 66–74. https://doi.org/10.33828/sei.v33.i1.7

Andoko, B. S., Asmara, R. A., Lestari, V. A., Ikawati, D. S. E., Prasetyo, A., Hirashima, T., & Hayashi, Y. (2023). Experimental Comparison of Promotion Effect for EFL Reading Comprehension between Conventional Summarization and Toulmin Argument Reconstruction. *Proceedings of the 31st International Conference on Computers in Education, Volume 1*, 732–741.

Andoko, B. S., Lestari, V. A., Pramudhita, N., Nuraini, A., Nisak, K., & Hirashima, T. (2024). Analyzing Student Behavior in Viat-map: Steps and Time as Performance Indicators. In Kashihara (Ed.), *Proceedings of the 32nd International Conference on Computers in Education.* Asia-Pacific Society for Computers in Education.

Andoko, B. S., Mubarok, F. U., Hirashima, T., Arhandi, P. P., Astiningrum, M., & Najib, M. F. (2022). Constructing Toulmin's Logical Structure Through Viat-map Application for

Reading Comprehension of EFL Students. *2022 Fifth International Conference on Vocational Education and Electrical Engineering (ICVEE)*, 196–200. https://doi.org/10.1109/ICVEE57061.2022.9930471

Baker, R., Mercedes, M., Rodrigo, T., Baker, R. S. J. D., Lagud, M. C. V, Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., Santillano, J. Q., Sevilla, L. R. S., Sugay, J. O., Tep, S., & Viehland, N. J. B. (2007). *Affect and Usage Choices in Simulation Problem-Solving Environments. Affect and Usage Choices in Simulation Problem-Solving Environments*. https://www.researchgate.net/publication/221297694

Baker, R. S. J. D. (2006). *Detecting When Students Game the System 1 Developing a Generalizable Detector of When Students Game the System*.

Baker, R. S. J. D., Corbett, A. T., Koedinger, K. R., & Roll, I. (2006). Generalizing detection of gaming the system across a tutoring curriculum. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *4053 LNCS*, 402–411. https://doi.org/10.1007/11774303_40

Dong, A., Jong, M. S.-Y., & King, R. B. (2020). How Does Prior Knowledge Influence Learning Engagement? The Mediating Roles of Cognitive Load and Help-Seeking. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.591203

Le Cunff, A. L., Giampietro, V., & Dommett, E. (2024). Neurodiversity and cognitive load in online learning: A systematic review with narrative synthesis. *Educational Research Review*, *43*. https://doi.org/10.1016/j.edurev.2024.100604

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Magalhães, A. L. (n.d.). Teaching How to Develop an Argument Using the Toulmin Model. *International Journal of Multidisciplinary and Current Educational Research*. www.ijmcer.com

Majidi, A. el, Janssen, D., & de Graaff, R. (2021). The effects of in-class debates on argumentation skills in second language education. *System*, *101*. https://doi.org/10.1016/j.system.2021.102576

Martin, F., Chen, Y., Moore, R. L., & Westine, C. D. (2020). Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. *Educational Technology Research and Development*, *68*(4), 1903–1929. https://doi.org/10.1007/s11423-020-09793-2

Mirari, K. (2022). The Effectiveness of Adaptive Learning Systems in Personalized Education. *Journal of Education Review Provision*, *2*(3), 70–76. https://doi.org/10.55885/jerp.v2i3.194

Petre, C., Duffy, B. E., & Hund, E. (2019). "Gaming the System": Platform Paternalism and the Politics of Algorithmic Visibility. *Social Media + Society*, *5*(4). https://doi.org/10.1177/2056305119879995

Rismanto, R., Nurul Asri, A., Satria Andoko, B., Hirashima, T., & Arifiandi Leonanta, A. (2021). A Preliminary Study: Toulmin Arguments in English Reading Comprehension for English as Foreign Language Students. *Proceedings - IEIT 2021: 1st International Conference on Electrical and Information Technology*, 6–10. https://doi.org/10.1109/IEIT53149.2021.9587398

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2022). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist*, *57*(1), 31–54. https://doi.org/10.1080/00461520.2021.1939700

Tzafilkou, K., Perifanou, M., & Economides, A. A. (2021). Negative emotions, cognitive load, acceptance, and self-perceived learning outcome in emergency remote education during COVID-19. *Education and Information Technologies*, *26*(6), 7497–7521. https://doi.org/10.1007/s10639-021-10604-1

Wang, C., Fang, T., & Gu, Y. (2020). Learning performance and behavioral patterns of online collaborative learning: Impact of cognitive load and affordances of different multimedia. *Computers and Education*, *143*. https://doi.org/10.1016/j.compedu.2019.103683