Investigating Student Interest in a Minecraft Game-Based Learning Environment: A Changepoint Detection Analysis

Yiqiu Zhou
University of Illinois at Urbana-Champaign
yiqiuz3@illinois.edu

Luc Paquette
University of Illinois at Urbana-Champaign
lpaq@illinois.edu

ABSTRACT

Extensive research underscores the importance of stimulating students' interest in learning, as it can improve key educational outcomes such as self-regulation, collaboration, problem-solving, and overall enjoyment. Yet, the mechanisms through which interest manifests and impacts learning remain less explored, particularly in open-ended game-based learning environments like Minecraft. The unstructured nature of gameplay data in such settings poses analytical challenges. This study employed advanced data mining techniques, including changepoint detection and clustering, to extract meaningful patterns from students' movement data. Changepoint detection allows us to pinpoint significant shifts in behavior and segment unstructured gameplay data into distinct phases characterized by unique movement patterns. This research goes beyond traditional session-level analysis, offering a dynamic view of the learning process as it captures changes in student behaviors while they navigate challenges and interact with the environment. Three distinct exploration patterns emerged: surface-level exploration, in-depth exploration, and dynamic exploration. Notably, we found a negative correlation between surface-level exploration and interest development, whereas dynamic exploration positively correlated with interest development, regardless of initial interest levels. In addition to providing insights into how interest can manifest in Minecraft gameplay behavior, this paper makes significant methodological contributions by showcasing innovative approaches for extracting meaningful patterns from unstructured behavioral data within game-based learning environments. The implications of our research extend beyond Minecraft, offering valuable insights into the applications of changepoint detection in educational research to investigate student behavior in open-ended and complex learning settings.

Keywords

Changepoint Detection,Time-series Analysis, STEM Interest, Unstructured Data, Game-based Learning Environments

1. INTRODUCTION

Educational games hold significant potential for facilitating active learning and the development of STEM interests [9,47]. Among these, Minecraft gained particular prominence not only as a game but also as a platform for creativity, problem-solving, and skill development [4,6,21,35]. With its open-ended nature and potential for customization, Minecraft has been widely adopted as a powerful tool for learning [31]. There is thus increasing need to understand student behavior within these environments. However, quantitatively analyzing the complex behavioral data, especially from these open-ended environments, poses analytical challenges.

One of the primary challenges in analyzing data from open-ended environments is the difficulty in translating unstructured behavioral data into meaningful insights. Traditional structured learning environments often involve clear assessment criteria and easily identifiable outcomes [28,41]. In contrast, open-ended environments present ill-structured tasks that encourage exploration, information synthesis, and creative problem-solving [28]. These environments allow for a variety of behaviors and multiple attempts over extended periods [8], resulting in complex datasets that reflect dynamic learning processes. For example, Minecraft players may engage in diverse activities such as construction, exploration, and collaboration, generating datasets that lack predefined answers that can be used as evidence of outcome measures [22]. The relationship between granular action, the development of higher-order skills, and overall student performance is not immediately clear [34]. Without explicit scoring rubrics and criteria of assessment, it becomes challenging to apply standard analytical approaches. This necessitates the development of innovative analytical methods to extract meaningful information from unstructured gameplay data [37,44].

Another challenge is related to the complexity of the skills being measured or developed in educational games [22]. While research has established that educational games contributed to the development of STEM interest and enhance aspects like self-regulation, collaboration, problem-solving, and enjoyment in learning [42], the process of how interest evolves from initial curiosity to a sustained engagement remains under-explored [17, 24, 30]. Interest appears to emerge as a desire to fill an information gap, then evolves into a flow-like experience in later phases [16]. However, the internalization process underlying this transition is not docu-

mented [1, 17, 24], presenting a challenge for designing game-based learning environments that engage and sustain student interest. Traditional data analysis often falls short in capturing this complexity, highlighting the gap in uncovering the relationship between in-game behaviors and STEM interest.

This study aims to address these challenges by leveraging data mining to analyze students' Minecraft movement data, with the goal of uncovering exploration patterns and their relationships to interest development. By employing methods including change point detection and clustering, this work identifies behavioral shifts and distinct exploration patterns based on movement trajectories. Triangulating these patterns with survey data, this study sheds light on how students navigate open-ended environments and how such explorations manifest interest. The research questions guiding this study are:

RQ1: How can advanced data mining techniques be applied to unstructured movement data to identify exploration patterns within Minecraft?

RQ2: How do these identified exploration patterns relate to interest development?

2. RELATED WORK

2.1 Minecraft as an educational platform

The integration of digital gaming in educational settings, particularly in the realm of STEM, has gained increasing interest [26]. Minecraft, a sandbox-style video game, emerged as an effective platform for engaging students and fostering STEM interest (see the review [31]). Several studies have demonstrated its positive impacts, including the development of spatial skills critical to STEM learning [6], enhanced collaboration skills [4], and increased computer science competencies and motivation [21].

Empirical studies have highlighted the potential of Minecraft in fostering STEM interest [25, 35, 46]. Sandbox environments like Minecraft can spark STEM interest by allowing students to interact with natural phenomena in a digital setting [46]. Studies have also investigated the relationship between specific in-game behaviors and levels of STEM interest, such as the usage of scientific tools and the frequency of scientific observations [13]. These findings collectively underscore the potential of Minecraft as a platform for fostering STEM interest and learning, while providing insights into student exploration behaviors.

2.2 Data mining in educational games

Traditional data collection approaches, such as surveys, interviews, and pre/post-tests, primarily focus on capturing data before and after gameplay [38]. While these approaches are effective for assessing performance, they provide limited insights into the learning process during gameplay. Advanced data mining, particularly log data analysis, can track users' in-game actions at a fine-grained level and to investigate the intricate behavioral patterns that emerge [7, 12, 19, 27, 33]. Previous studies have examined students' in-game behaviors such as use of in-game tools [20, 27], gameplay duration and gaming performances [7]. While these metrics

provide insights into engagement, they do not take full advantage of the rich interactions within open-ended games [2]. In environments like Minecraft, where students explore the environment in a natural manner without necessarily making observable interactions, common approaches and metrics suited for structured environments prove inadequate.

Trajectory analysis has emerged as a promising approach to address the limitations of traditional metrics in capturing the intricacies of learning processes [10, 19, 36, 39, 40]. For example, [19] investigated students' daily gameplay actions and identified three distinct trajectories of group action similarity to examine the processes of collaborative problem-solving. [33] employed process mining and created a global graph of the quest pathway to visualize how students moved around, made decisions, and completed quests. [40] utilized random walk analysis to visualize students' trajectories and slopes within the game-based environment iSTART-ME and compared individual differences in terms of reading ability and self-explanation quality. Additionally, [36] used time series analysis to compare students' problem-solving approaches with expert solutions, linking these comparisons to learning outcomes.

While these trajectory-focused studies provide valuable insights, they assume consistent behaviors or strategies throughout the session. This assumption may not always hold. As [14] and [39] suggest, learning strategies and behaviors are dynamic, varying across contexts and evolving within the task as students' understanding and gaming skills develop. Therefore, summarizing with an overall pattern or learner profile may overlook these nuances. To bridge this gap, our study introduces a novel approach able to capture the dynamic nature of learning processes in an open-ended environment. By identifying natural segments in students' exploration trajectories as indicators of exploration strategies and examining their association with STEM interests, our approach offers a more nuanced understanding of segment-specific behaviors.

2.3 WHIMC as a learning context

What-If Hypothetical Implementations in Minecraft (WHIMC; https://whimcproject.web.illinois.edu/) presents a diverse set of Minecraft worlds designed to support STEM education. Its core mission is to develop engaging computer simulations that deepen interest in STEM. Utilizing Minecraft Java Edition, this project creates 13 distinct WHIMC worlds, including two orientation worlds, six 'What-if' scenarios, and five exoplanets. A conceptual diagram of the user experience and learning pathways is shown in Fig. 1(a). These virtual environments provide interactive and immersive learning experiences where students can engage with nonplayer characters (NPCs) for knowledge or task guidance (as illustrated in Fig. 1(b)), or use different science tools to conduct experiments and collect data such as temperature and oxygen level (as illustrated in Fig. 1(c)). Additionally, students can enter written observations at any location; these observations are then visible to other players exploring the same world.

Previous studies associated with the WHIMC project have demonstrated its effectiveness in fostering STEM interest [13, 25, 46]. These studies, however, primarily used qualitative coding or interviews to identify interest-triggering mo-

(a) Overall User Experience and Learning Pathways

(b) Screenshot of User Interacting with NPC

(c) Screenshot of User Exploring the World

Figure 1: WHIMC

ments. While these approaches provide valuable insights, they may not fully capture the behavioral patterns evident in the log data, emphasizing the need for alternative methods to understand the intricacies of student behaviors within game-based learning environments.

3. DATA SOURCE
3.1 Participant
The study involved 82 middle-school students from seven different summer camps located in the Northeast and Midwest regions of the United States. The participant demographic comprised 32% female students, with a racial composition of 37% White, 28% Pacific Islander or Native American, 24% Black, 13% Hispanic, 2% Asian, and 1% American Indian. On the first day of the camp, students completed self-report surveys and received a brief introduction to the camp and the "What-if" astronomy scenarios. They then explored the WHIMC worlds. After excluding data from students who dropped out, a total of 55 participants were included in the final analysis. Data was collected with consent from the participants and their guardians and approved by the Institutional Review Board for research purpose.

3.2 Data
This study utilized log and survey data to investigate students' exploration patterns and their relationship with STEM interest levels. The Minecraft Server recorded log data for each participant every three seconds, as well as upon the usage of scientific measurement tools or observation actions. The dataset included timestamps, x, y, and z coordinate locations, utilization of science tools, and observations made by students. Since our primary focus was spatial navigation, we focused on x (east-west) and z (north-south) coordinates, excluding the y coordinate which tracks vertical change. Our analysis focused on worlds that offer a rich potential in uncovering student interactions and engagement, excluding those with limited movement. Selected worlds included No Moon, Colder Sun, Tilted Earth, Mynoa, and Two Moons.In addition to the log data, a validated interest survey was provided as a pre- and post-test to evaluate the impact of WHIMC on students' interest levels [13]. This survey consists of 25 items using a 1–5 Likert scale (1 = "Strongly Disagree", 5 = "Strongly Agree"). For analytical clarity, we organized the survey questions into STEM, Minecraft, and astronomy interest as detailed in Table 1.

4. METHODOLOGY
4.1 Feature Generation and Selection
The log data was processed to extract four features related to their movement: speed $v(t)$, direction $\theta(t)$, and their corresponding rates of change $a(t)$ and $\Delta\theta(t)$ over time. Speed and acceleration reveal aspects of pace in activity levels, while direction and direction change offer perspectives on spatial navigation. We tracked students' positions, represented as $\vec{p}(t) = \begin{bmatrix} x(t) \\ z(t) \end{bmatrix}$ at time t, where $x(t)$ and $z(t)$ are the coordinates of students in the Minecraft world. The speed $v(t)$ and direction $\theta(t)$ at time t can be thus calculated as:

$$v(t) = \|\vec{v}(t)\|, \quad \theta(t) = \arctan 2\left(\frac{dz}{dx}\right) \qquad (1)$$

where dx and dz are the changes in x and z coordinates compared to the previous time point $t-1$. In our context, these two features represent how fast the students are moving within Minecraft and their heading or direction at the given moment.

Additionally, we looked at how these values change over time to capture their dynamics of movement. The change in speed $a(t)$ and the change in direction $\Delta\theta(t)$ at time t are calculated as:

$$a(t) = \frac{dv(t)}{dt}, \quad \Delta\theta(t) = \frac{d\theta(t)}{dt} \qquad (2)$$

Table 1: Survey Categories and Exemplary Questions

| Category	Exemplary Survey Questions
STEM Interest	Please tell us how interested you are in these general activities. - Learning how to make slime in science class
MC Interest	Please tell us how interested you are in these Minecraft activities. - Exploring to find specific biomes
Astronomy Interest	Please indicate how much you agree or disagree with the following statements - I think Astronomy is interesting.
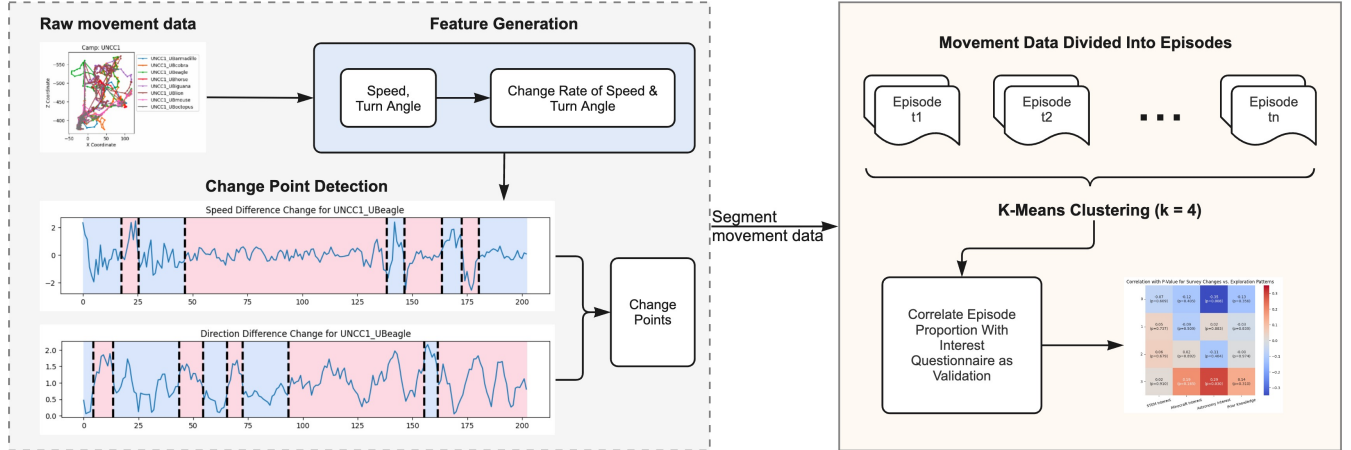
Figure 2: Analysis Pipeline

where $a(t)$ is the acceleration (rate of change of speed) and $\Delta\theta(t)$ is the rate of change of direction. These features enable us to identify significant behavioral patterns in movement, such as sudden increases in acceleration, or periods marked by frequent or abrupt turns.

We experimented with all four features and decided to focus on those quantifying rates of change (i.e., rate of change in speed or direction). While measures of speed and direction provide snapshots of movement, such as their current speed and trajectory at a given moment, they fail to reflect the dynamics and evolution of these movements. In contrast, rates of change provide a more comprehensive view. These features show how movement changes over time rather than in isolated moments, providing a more granular understanding of students' exploratory behaviors.

4.2 Changepoint Detection

Changepoints refer to points within a data sequence where a significant change in statistical properties occurs. Given a multivariate process $\mathbf{y} = \{y_1, \ldots, y_T\}$ that consists of T data points, this process involves determining the optimal segmentation $\hat{\tau}$ that minimizes a criterion function $V(\tau, y)$. This criterion is typically a sum of costs over all segments defined by the segmentation:

$$V(\tau, y) := \sum_{k=0}^{K} c(y_{t_k:t_{k+1}}) \qquad (3)$$

Here, $c(\cdot)$ is a cost function quantifying the goodness-of-fit of the sub-signal $y_{t_k:t_{k+1}} = \{y_t\}_{t=t_k}^{t_{k+1}}$ to a specified model. The sub-signal represents the data points between two consecutive changepoints t_k and t_{k+1}. The best segmentation $\hat{\tau}$

is then the one that minimizes the criterion function $V(\tau, y)$.

Changepoint detection has been studied over the last several decades in data mining, statistics, and computer science for scenarios such as speech recognition, image analysis, and human activity analysis [3, 18]. [43] classifies changepoint detection into two types: (1) online; and (2) offline. Online approaches allow for real-time detection of changes. Offline approaches detect changes in the entire dataset and retrospectively identify the location of abrupt changes (for a detailed overview, see [43]). Offline changepoint detection can be further divided into two paradigms [15, 43]: the constrained problem and the penalized problem (also known as unconstrained problem). The constrained approach operates under the assumption that the number of changepoints is known or predefined. Conversely, the penalized problem handles situations where the number of changepoints is unknown, necessitating more advanced detection techniques. In our context, the number of potential behavioral changes in students' Minecraft movement is unknown, thus we chose the unconstrained changepoint detection approach to identify the number and position of those changes.

4.2.1 PELT Algorithm

We used the penalized changepoint detection approach as an unsupervised technique to identify shifts in the statistical properties of two data sequences, namely the change of speed, and change of movement direction. Changepoint Analysis algorithms vary in terms of which statistics they monitor changes in. Our study chose to employ the pruned exact linear time (PELT) algorithm [15, 23]. PELT is a pruning-based dynamic programming algorithm that efficiently identifies both the quantity and locations of change-

points with a linear computational cost. Its efficiency is achieved through a pruning process that discards unnecessary candidates in the dynamic programming computation, lowering the computational complexity. This method also improves accuracy when compared to other popular choices such as binary segmentation algorithm [23]. We implemented PELT using the ruptures Python package [43].

4.2.2 Cost Functions
The cost function involves the computation of the negative log-likelihood or other statistical measures that are minimized to detect changes in the signal's properties (see [43] for an overview and comparison of common cost functions). Two cost functions $c(\cdot)$ were employed for the speed change and direction change sequence for each student. Specifically, we utilized the L2 norm for analyzing speed changes and the L1 norm for direction changes, a decision informed by their inherent statistical characteristics.

Cost Function $L1$. This algorithm sums the absolute differences between data points and their segment means, which makes it more robust to outliers.

$$c_{L1}(y_{t_i} \ldots y_{t_{i+1}}) = \sum_{t=t_i}^{t_{i+1}} |y_t - \mu_i| \qquad (4)$$

Cost Function $L2$. Conversely, this algorithm sums the squared differences between the data points and the mean within each segment. The L2 norm is more sensitive to outliers as it tends to amplify deviations via squaring. In the context of speed change detection, it ensures that even subtle yet meaningful variations are captured.

$$c_{L2}(y_{t_i} \ldots y_{t_{i+1}}) = \sum_{t=t_i}^{t_{i+1}} (y_t - \mu_i)^2 \qquad (5)$$

4.2.3 Penalty
Another important parameter of the PELT algorithm is the penalty β, which serves to control the model's complexity and to avoid under/over-fitting. This parameter influences the algorithm's sensitivity: a higher penalty value imposes a stricter criterion for detecting changepoints, thereby reducing their total number. Conversely, a lower penalty value encourages the identification of more changepoints. The penalty term is represented as a term proportional to the number of changepoints:

$$V(\tau, y) = \sum_{k=0}^{K} c(y_{t_k} \ldots y_{t_{k+1}}) + \beta K, \qquad (6)$$

where β denotes the penalty parameter and K represents the number of segments. By minimizing the overall cost, the algorithm seeks to find the most likely points where the statistical properties change in the data sequence.

$$\hat{\tau} = \arg \min_{\tau} \left\{ \sum_{k=0}^{K} c(y_{t_k:t_{k+1}}) + \beta K \right\} \qquad (7)$$

In our analysis, we carefully selected penalty values to achieve a balance between sensitivity and specificity. This process involved iterative refinement based on visual inspection of changepoint plots, ensuring that our algorithms distinguished meaningful statistical shifts from random noises. A penalty

value of 1 was decided for speed change and 2.2 for direction change.

To reduce the computational cost, the ruptures package also allows us to consider only a subset of possible changepoint indexes, by modifying the `min_size` and `jump` arguments. The `min_size` parameter defines the minimum gap between consecutive changepoints to avoid recognizing minor fluctuations. After experimentation, we set `min_size=3` for both the speed and direction changes. This choice was informed after a thorough evaluation of the data's temporal resolution and the time scales of changes in speed and direction. This setup ensures at lease three data points between any two detected changes, limiting over-segmentation due to closely spaced or isolated outliers.

The `jump` parameter defines the granularity at which the algorithm searches for potential changepoints. By setting `jump=1`, we ensured a thorough examination over the entire data, without overlooking any possible changepoint locations. This decision ensures an exhaustive search but can be adjusted in future studies to increase computational efficiency, particularly for larger datasets.

4.3 Episode Segmentation and Clustering
After identifying changepoints, we segmented the data sequence into distinct episodes, each representing a period of homogeneity in the students' movement characteristics. Cluster analysis was then performed to group these episodes into broader exploration patterns. Specifically, we focused on five key variables that characterized the episode: average speed, the standard deviation of speed, range of speed, average directional difference, and standard deviation of directional difference during the corresponding episode. These movement features were selected for their potential to highlight distinct aspects of movement, aligning with our overarching goal of identifying episodes indicative of varied exploration patterns.

4.4 Correlation with Questionnaire Data
The final step involved linking the identified exploration patterns with interest measures from pre- and post-gameplay surveys, as detailed in section 3.1.3. This relationship analysis helps validate the meaningfulness of the episodes identified through data mining. Furthermore, it examines the connection between these movement patterns and interest development during gameplay. This step not only validates our analytical methods but also deepens our understanding of of how an abstract concept like interest manifests in Minecraft.

5. RESULT
5.1 RQ 1 Unveiling Patterns via Segmentation
To address RQ1, we took a two-step process: detecting change points in students' movement trajectories, followed by a clustering analysis to categorize these segmented episodes into exploration patterns. Fig 3 presents examples of detected change points, marked by dashed lines that represent statistically significant shifts in movement metrics. It is important to note that these visualizations show the relative sequences in the time series data rather than precise moment. Therefore, the x-axis in Fig 3 represents the sequen-

tial indices of data points. By mapping the data points to raw logs with timestamps, we were able to identify changing moments and segmented entire sessions into exploration episodes with consistent movement characteristics.
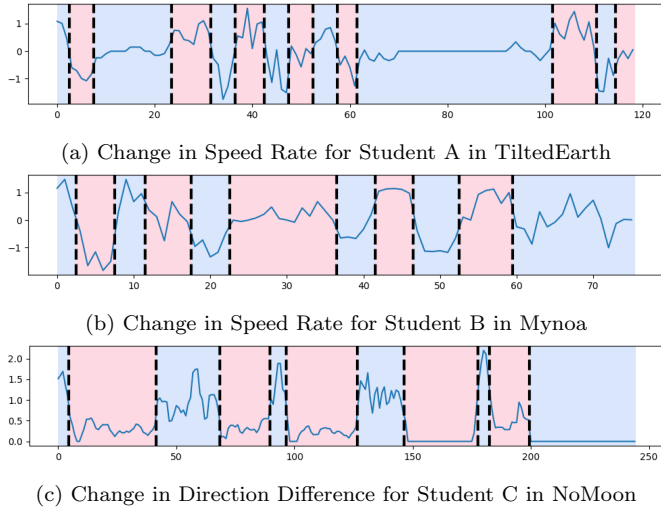

(a) Change in Speed Rate for Student A in TiltedEarth

(b) Change in Speed Rate for Student B in Mynoa

(c) Change in Direction Difference for Student C in NoMoon

Figure 3: Changepoint Detection Examples

To obtain a holistic understanding of these segmented episodes, K-means clustering was employed to group these episodes into higher-level exploration patterns among all students. The elbow method determined the optimal number of clusters by plotting the sum of squared distances (inertia) of samples to their closest cluster center for a range of cluster numbers. This approach identified the point where the rate of decrease sharply changes as shown in Fig 4, leading to the identification of four distinct exploration patterns. We also experimented with three clusters which resulted in cluster counts of 6888, 6311, and 5 respectively. Comparing to the four-cluster, this three-cluster solution might simplify the complexity of student behaviors, potentially merging distinct patterns into broader categories and obscuring nuanced differences in engagement.
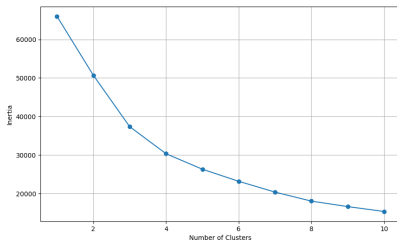
Figure 4: Elbow Method For Optimal Number of Clusters

5.2 RQ 2 Patterns and STEM Interest

Our investigation into RQ2 explored the relationship between specific exploration patterns derived from Minecraft movement data and the development of student interest. This involved correlating the time spent within each pattern with their responses in pre- and post-gameplay surveys. Results from pre-surveys indicated no significant correlations between exploration patterns and STEM or Astronomy interest.

However, significant correlations emerged when examining interest gains comparing pre- and post-surveys. Exploration Pattern 0 showed a negative correlation with changes in Astronomy interest scores ($r(53) = -.35$, $p = .008$), while Exploration Pattern 3 positively correlated with changes in Astronomy interest scores ($r(53) = .29$, $p = .030$). These findings suggest a potential relationship between in-game exploration behaviors and the development of Astronomy interest, although interpretations should remain cautious due to the interdependence of the exploration patterns.

6. DISCUSSION

This research leverages data mining techniques to uncover exploratory patterns in an open-ended game-based learning environment. Changepoint detection and clustering were utilized to analyze the unstructured movement data. Results revealed that students' movement trajectories can inform us of their exploration strategies, which can further be associated with their interest levels. This study built on previous work which established the positive connection between STEM interest and gameplay [13, 25], and extended the research by demonstrating how interest manifest in Minecraft.

6.1 RQ1 Segmenting Unstructured Data for Nuanced Analysis

Traditionally, research aggregated process data over the whole session to capture behaviors or identify learner profiles in game-based learning environments [10, 19, 33, 36, 40]. This approach, however, assume stable behavior and consistent learning strategies, oversimplifying the temporal variation in how student learn and engage [48]. Learning behaviors are not static; they vary across different contexts and evolve as students' understanding deepens and their skills develop [14, 39]. This research takes a different perspective, leveraging changepoint detection to segment movement data into distinct phases. Our approach recognizes the non-linear, dynamic nature of learning in open-ended game-based environments. Rather than a structured linear progression, it involves a series of shifts and fluctuating phases as students encounter challenges, pursue goals, and develop skills. As shown in Table 2, this study successfully identified three distinct movement patterns characteristic of how students advanced their exploration: **surface-level exploration** (Cluster 0), **in-depth exploration** (Cluster 1), and **dynamic exploration** (Cluster 3). Notably, Cluster 2 was excluded from detailed discussions due to its rare occurrence, more likely representing outliers.

Surface-level exploration is characterized by the highest average speed and the greatest standard deviation in speed, indicating a rapid, cursory exploration . Students displaying this behavior likely move quickly between locations, showing a preference for breadth over depth. In contrast, **in-depth exploration** features minimal variations in speed and direction change, suggesting a more consistent and steady pace of movement. The lower average speed also suggests a focus on detailed investigation. **Dynamic exploration** exhibits a medium level of speed and direction change rate. The relatively large standard deviation in speed and direction change point to a changing movement pace where students actively seek out new experiences while also taking the time to engage with the educationally relevant contents.

Table 2: Description Information of Exploration Patterns in WHIMC

Cluster	Count	Duration (s)	Avg Speed	Std Dev of Speed	Min Speed	Range of Speed	Avg Dir Diff	Std Dir Diff
0	4202	12.61	4.33	1.86	2.28	4.09	0.57	0.50
1	3223	16.83	0.85	0.66	0.33	1.40	0.44	0.52
2	5	8.00	30.76	58.31	0.24	115.07	1.53	1.02
3	5774	16.03	1.66	1.18	0.59	2.39	1.50	1.10

By identifying changepoints and segmented movement trajectories, this work provides a more granular view of students' engagement. Pinpointing these changepoints may signify critical moments such as losing focus or sparking interest. These moments can further inform targeted scaffolding and warrant qualitative analysis like detector-driven interviews [5] to understand the context underlying these changes. Furthermore, each segmented episode represents a distinct phase characterized by movement features. This approach identifies meaningful patterns beyond whole-session categorizations, revealing the dynamics of student learning processes as they navigate challenges, pursue various goals, and develop skills over time.

6.2 RQ2 Correlation with STEM Interest

The correlation analysis revealed a relationship between students' exploratory patterns and the development of their interest. Notably, this association was found to be independent of the students' initial interest levels, as measured by a pre-gameplay survey. This finding aligns with previous research [25], which demonstrated that learner interest can be triggered by in-game experiences and interactions, such as the novelty of seeing a giant planet or engaging in conversations with NPCs.

In specific, the **dynamic exploration** pattern characterized by frequent changes in speed and direction exhibited the strongest positive correlation with interest gain. This finding supports the efficacy of active, exploratory learning in STEM education [11, 45]. Such experience of "seeing the science in the world" help students deepen understanding through engagement in scaffolded activities and support the construction of explanations on observable phenomena [29]. Our findings provide preliminary evidence that hands-on exploration and investigation within Minecraft may not only strengthen conceptual understanding, but also relate to interest gains. This corroborates prior studies showing undergraduates who actively observe and explain phenomena are more likely to persist in STEM majors [32]. While promising, it is important to note this work was an initial proof-of-concept analysis examining the potential of using changepoints in movement data to identify patterns associated with STEM interest. Establishing robust links between exploration patterns and interest gains requires further confirmatory research.

6.3 Limitations and Future Study

This study has several limitations. Firstly, the research uses self-reported data to evaluate students' interest in STEM subjects, which may be subject to biases. Additionally, the current optimization process for changepoint detection relies on subjective visual inspection , which may limit the generalizability of our findings. Future studies should develop more objective and automated methods for parameter selection. Lastly, despite the use of surveys to corroborate the relationship between identified patterns and interest, the personal motivations underlying these moments remain unclear. Future study will triangulate log data interpretations by combining automatic detection of key moments with detector-driven interviews to collect personal reflections [5].

7. CONCLUSION

This research leverages novel data mining approaches to examine student exploration patterns in Minecraft and their relationship to interest development. Utilizing changepoint detection to segment the unstructured movement data, this study identifies three distinct patterns: surface-level exploration, in-depth exploration, and dynamic exploration. Notably, dynamic exploration, which is characterized by diverse movement metrics, exhibited the strongest association with increased STEM interest. This approach provides a granular view of student engagement through finer-grained behavioral patterns, capturing the dynamic nature of learning as students navigate challenges and pursue diverse goals during gameplay. Overall, this study demonstrates the potential of changepoint detection for analyzing unstructured data in open-ended learning environments. Future research should refine changepoint detection for wider applicability and integrate qualitative methods to understand underlying motivations and engagement triggers that explain the relationship between interest and exploration patterns. Ultimately, this work paves the way for a more nuanced understanding of student learning and engagement, informing the design of personalized and effective learning experiences that foster deeper STEM interest.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] M. Ainley. Interest: Knowns, unknowns, and basic processes. *The science of interest*, pages 3–24, 2017.

[2] C. Alonso-Fernández, A. Calvo-Morata, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón. Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, 141:103612, 2019.

[3] S. Aminikhanghahi and D. J. Cook. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2):339–367, 2017.

[4] R. Andersen and M. Rustad. Using minecraft as an educational tool for supporting collaboration as a 21st century skill. *Computers and Education Open*, 3:100094, 2022.

[5] R. S. Baker, S. Hutt, N. Bosch, J. Ocumpaugh, G. Biswas, L. Paquette, J. M. A. Andres, N. Nasiar, and A. Munshi. Detector-driven classroom interviewing: focusing qualitative researcher time by selecting cases in situ. *Educational technology research and development*, 2023.

[6] C. Carbonell-Carrera, A. J. Jaeger, J. L. Saorín, D. Melián, and J. de la Torre-Cantero. Minecraft as a block building approach for developing spatial skills. *Entertainment Computing*, 38:100427, 2021.

[7] M.-T. Cheng, Y.-W. Lin, and H.-C. She. Learning through playing virtual age: Exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters. *Computers & Education*, 86:18–29, 2015.

[8] T. Davey, S. Ferrara, Paul W. Holland, Rich Shavelson, Noreen M. Webb, and Lauress L. Wise. Psychometric considerations for the next generation of performance assessment, 2015.

[9] S. de Freitas. Are games effective learning tools? a review of educational games. *Journal of Educational Technology & Society*, 21(2):74–84, 2018. Publisher: International Forum of Educational Technology & Society.

[10] J. de Haan and D. Richards. Navigation paths and performance in educational virtual worlds. In *Proceedings of the 21th Pacific Asia Conference on Information Systems (PACIS 2017)*, pages 1–7, 2017.

[11] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23):8410–8415, 2014.

[12] A. E. J. V. Gaalen, J. Schönrock-Adema, R. J. Renken, A. D. C. Jaarsma, and J. R. Georgiadis. Identifying player types to tailor game-based learning design to learners: Cross-sectional survey using q methodology. *JMIR Serious Games*, 10(2):e30464, 2022. Company: JMIR Serious Games Distributor: JMIR Serious Games Institution: JMIR Serious Games Label: JMIR Serious Games Publisher: JMIR Publications Inc., Toronto, Canada.

[13] M. Gadbury and H. C. Lane. Mining for STEM interest behaviors in minecraft. In M. M. Rodrigo, N. Matsuda, A. I. Cristea, and V. Dimitrova, editors, *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium*, Lecture Notes in Computer Science, pages 236–239. Springer International Publishing, 2022.

[14] A. F. Hadwin, J. C. Nesbit, D. Jamieson-Noel, J. Code, and P. H. Winne. Examining trace data to explore self-regulated learning. *Metacognition and Learning*, 2(2):107–124, 2007.

[15] K. Haynes, I. A. Eckley, and P. Fearnhead. Computationally efficient changepoint detection for a range of penalties. *Journal of Computational and Graphical Statistics*, 26(1):134–143, 2017.

[16] S. Hidi and K. A. Renninger. The four-phase model of interest development. *Educational Psychologist*, 41(2):111–127, 2006.

[17] S. E. Hidi and K. A. Renninger. Interest development and its relation to curiosity: Needed neuroscientific research. *Educational Psychology Review*, 31(4):833–852, 2019.

[18] T. D. Hocking, G. Rigaill, P. Fearnhead, and G. Bourque. Constrained dynamic programming and supervised penalty learning algorithms for peak detection in genomic data. *The Journal of Machine Learning Research*, 21(1):3388–3427, 2020.

[19] J. Kang, D. An, and L. Yan. Collaborative problem-solving process in a science serious game: Exploring group action similarity trajectory. In *Proceedings of the 12th international conference on educational data mining*, pages 336–341, 2019.

[20] J. Kang, M. Liu, and W. Qu. Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior*, 72:757–770, 2017.

[21] T. Karsenti and J. Bugmann. Exploring the educational potential of minecraft: The case of 118 elementary-school students. pages 175–179, 2017.

[22] D. Kerr, J. J. Andrews, and R. J. Mislevy. The in-task assessment framework for behavioral data. *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications*, pages 472–507, 2016.

[23] R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012.

[24] A. Krapp. An educational–psychological conceptualisation of interest. *International Journal for Educational and Vocational Guidance*, 7(1):5–21, 2007.

[25] H. C. Lane, M. Gadbury, J. Ginger, S. Yi, N. Comins, J. Henhapl, and A. Rivera-Rogers. Triggering stem interest with minecraft in a hybrid summer camp. *Technology, Mind, and Behavior*, 3(4), 2022.

[26] Y. Li, K. Wang, Y. Xiao, and J. E. Froyd. Research and trends in STEM education: a systematic review of journal publications. *International Journal of STEM Education*, 7(1):11, 2020.

[27] M. Liu, Y. Cai, S. Han, and P. Shao. Understanding student navigation patterns in game-based learning. *Journal of Learning Analytics*, 9(3):50–74, 2022.

[28] K. R. Lodewyk, P. H. Winne, and D. L. Jamieson-Noel. Implications of task structure on self-regulated learning and achievement. *Educational Psychology*, 29(1):1–25, 2009.

[29] D. Lombardi, T. F. Shipley, J. M. Bailey, P. S. Bretones, E. E. Prather, C. J. Ballen, J. K. Knight, M. K. Smith, R. L. Stowe, M. M. Cooper, M. Prince, K. Atit, D. H. Uttal, N. D. LaDue, P. M. McNeal, K. Ryker, K. St. John, K. J. van der Hoeven Kraft, and J. L. Docktor. The curious construct of active learning. *Psychological Science in the Public Interest*, 22(1):8–43, 2021. Publisher: SAGE Publications Inc.

[30] K. Murayama, L. FitzGibbon, and M. Sakaki. Process account of curiosity and interest: A reward-learning perspective. *Educational Psychology Review*,

31(4):875–895, 2019.

[31] S. Nebel, S. Schneider, and G. D. Rey. Mining learning and crafting scientific experiments: A literature review on the use of minecraft in education and research. *Journal of Educational Technology & Society*, 19(2):355–366, 2016. Publisher: International Forum of Educational Technology & Society.

[32] K. J. Pugh, M. M. Phillips, J. M. Sexton, C. M. Bergstrom, and E. M. Riggs. A quantitative investigation of geoscience departmental factors associated with the recruitment and retention of female students. *Journal of Geoscience Education*, 67(3):266–284, 2019.

[33] J. A. Ruipérez-Valiente, L. Rosenheck, and Y. J. Kim. What does exploration look like? painting a picture of learning pathways using learning analytics. *Game-Based Assessment Revisited*, pages 281–300, 2019.

[34] A. A. Rupp, M. Gushta, R. J. Mislevy, and D. W. Shaffer. Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *The Journal of Technology, Learning and Assessment,*, 8(4), 2010.

[35] U. Saricam and M. Yildirim. The effects of digital game-based STEM activities on students' interests in STEM fields and scientific creativity: Minecraft case. *International Journal of Technology in Education and Science*, 5(2):166–192, 2021.

[36] R. Sawyer, J. Rowe, R. Azevedo, and J. Lester. Filtered time series analyses of student problem-solving behaviors in game-based learning. *International Educational Data Mining Society*, 2018.

[37] V. J. Shute. Stealth assessment in computer-based games to support learning. *Computer games and instruction*, 55(2):503–524, 2011.

[38] S. P. Smith, K. Blackmore, and K. Nesbitt. A meta-analysis of data collection in serious games research. *Serious games analytics: Methodologies for performance measurement, assessment, and improvement*, pages 31–55, 2015.

[39] E. L. Snow, A. D. Likens, L. K. Allen, and D. S. McNamara. Taking control: Stealth assessment of deterministic behaviors within a game-based system. *International Journal of Artificial Intelligence in Education*, 26(4):1011–1032, 2016.

[40] E. L. Snow, A. D. Likens, T. Jackson, and D. S. McNamara. Students' walk through tutoring: Using a random walk analysis to profile students. In *Educational Data Mining 2013*, 2013.

[41] R. Spiro, R. Coulson, P. Feltovich, and D. Anderson. Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. In *Tenth annual conference of the Cognitive Science Society*, pages 375–383. Hillsdale, 1988.

[42] L. P. Sun, P. Siklander, and H. Ruokamo. How to trigger students' interest in digital learning environments: A systematic literature review. 2018.

[43] C. Truong, L. Oudre, and N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

[44] J.-Y. Wu. Learning analytics on structured and unstructured heterogeneous data sources: Perspectives from procrastination, help-seeking, and machine-learning defined cognitive engagement. *Computers & Education*, 163:104066, 2021.

[45] N. Yannier, S. E. Hudson, and K. R. Koedinger. Active learning is about more than hands-on: A mixed-reality AI system to support STEM education. *International Journal of Artificial Intelligence in Education*, 30(1):74–96, 2020.

[46] S. Yi, M. Gadbury, and H. C. Lane. Coding and analyzing scientific observations from middle school students in minecraft. In *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS)*, volume 3, pages 1787–1788, 2020.

[47] J. Zeng, S. Parks, and J. Shang. To learn scientifically, effectively, and enjoyably: A review of educational games. *Human Behavior and Emerging Technologies*, 2(2):186–195, 2020.

[48] Y. Zhang, Y. Ye, L. Paquette, Y. Wang, and X. Hu. Investigating the reliability of aggregate measurements of learning process data: From theory to practice. *Journal of Computer Assisted Learning*, 2024.