International Society of the Learning Sciences

Exploring Students' Interest-Driven Patterns of Scientific Observation in Minecraft

Zhanlan Wei, Nidhi Nasiar, Andres Felipe Zambrano, Xiner Liu, Jaclyn Ocumpaugh, Amanda Barany, Ryan S. Baker, Camille Giordano
zhanlanw@upenn.edu, nasiar@upenn.edu, azamb13@upenn.edu, xiner@upenn.edu, jlocumpaugh@gmail.com, amanda.barany@gmail.com, ryanshaunbaker@gmail.com, camille.giordano.upenn@gmail.com
University of Pennsylvania

Abstract: Students bring different levels of interest to learning experiences, which impacts how they engage with learning materials. This study aims to understand the relationship between student's interest levels and their scientific observation behaviors within a Minecraft-based learning system. Motivated by the growing interest in integrating human-AI collaboration within educational research, we combine the capabilities of Large Language Models (LLMs) with the expertise of human researchers to capture the emerging themes within students' observations. Using epistemic network analysis, we then visualized and compared the observational patterns of students with high and low situational interest. Our findings indicate that students with higher situational interest tend to make observations across a broader range of topics, with a particular emphasis on scientific content. These results highlight the potential for developing timely interventions to support students with low situational interest.

Introduction

Supporting early interest in Science, Technology, Engineering, and Mathematics (STEM), which is critical to future career trajectories in these fields, has long been a focus in education research (Hidi & Harackiewicz, 2000; Lee et al., 2022). Prior research has explored creating more authentic science experiences, such as simulations (Makransky et al., 2020). Other work has designed and implemented digital games aiming to promote deeper engagement (Ishak et al., 2021). Both approaches present opportunities to model active learning experiences that might otherwise be difficult to provide for younger students due to costs and safety concerns that sometimes limit laboratory access.

Recently, developers have leveraged the creative and engaging properties of a popular commercial game, Minecraft, to enable younger students to explore astronomy facts and experiment with scientific hypotheses. This system, called What-If Hypothetical Implementations in Minecraft (WHIMC; Lane et al., 2017) allows students to virtually visit different planets (e.g., Moon, Mars, Kepler, and Earth), including different versions of the same planet (e.g., Earth with two moons, Earth with no moon), where they can make scientific observations and ask questions about their findings. Beyond encouraging active participation through creative exploration, the open-ended game system also provides an opportunity for researchers to understand how interest emerges, develops, and impacts learning experiences. For instance, although autonomy has been shown to enhance student interest (Deen, 2015), previous studies suggests that students with low prior knowledge—often linked to low interest (Hidi & Renninger, 2006)—may struggle in open-ended contexts (Dong et al., 2020). These findings underscore the importance of understanding how interest shapes students' interactions with open-ended, game-based learning environments, such as WHIMC.

Prior studies have examined the role of players' situational interest in game-based learning, including its relationship with learning and self-efficacy (Koskinen et al., 2022) and how interest in a game relates to students' pre-existing individual interest in math (Rodríguez-Aflecht et al., 2018). However, limited research has explored how situational interest shapes student behaviors in open-ended exploratory games like WHIMC, where students have more control over their learning process. Thus, to address this gap, in this study, we examine how differences in situational interest—a temporary, spontaneous interest triggered by specific external stimulus features (Renninger et al., 2014)—are associated with differences in the scientific observations students make as they learn from the game. Specifically, we leverage a Large Language Model (LLM) to develop a codebook (in a hybrid approach involving humans and an LLM working together) and automate the coding of different categories of observations, towards the goal of facilitating real-time support for students with low interest. We further explore the patterns of these observations using a combination of statistical analyses and Epistemic Network Analysis (ENA; Shaffer et al., 2016), which allows us to visualize how a student transitions from one observation to the next across their experience in these worlds. In doing so, we seek to understand how students' observational behaviors, especially scientific observations, are related to their level of situational interest.

Theoretical background

In educational research, interest has long been recognized as a critical factor that impacts students' learning and academic success (Hidi & Harackiewicz, 2000). As a psychological state, interest can not only shape individual behaviors but can also reflect the prior knowledge students bring to an educational task (Renninger, 1992). Interest is typically categorized into two forms: individual interest, a longer-term inclination toward a pre-experienced topic, or situational interest, a short-term response with charged attention and emotions triggered by an external stimulus (Linnenbrink-Garcia, 2010). The latter, situational interest, while transient in nature, holds the potential to evolve into the more enduring and stable form of individual interest (Hidi and Renninger, 2006). As situational interest often emerges in response to materials and features presented within an environment (Linnenbrink-Garcia, 2010), researchers actively explore more interactive and engaging learning environments, such as game-based environments, to cultivate students' interest.

The idea that *"play is our brain's favorite way of learning"* (Prensky, 2005) continues to resonate today, as many students find games both engaging and motivating, leading them to naturally gravitate toward game-based settings as their preferred way of learning (Prensky, 2005). In its simplest form, game-based learning can be defined as students participating in a designed gameplay to achieve pre-determined learning objectives (Plass et al., 2015). Today, game-based learning is often (but not always) implemented within digital environments (Plass et al., 2015). Previous studies suggest that game-based learning can effectively support knowledge and skill acquisition and further improve educational outcomes (Qian et al., 2016). Seeking to better understand how students could benefit from game-based learning, prior researchers have increasingly employed tools, such as Epistemic Network Analysis (ENA; Shaffer et al. 2016), to visualize and investigate complex patterns emerging in gameplay. For instance, Bressler et al. (2019) have utilized ENA to discuss how game-based learning assists students in constructing scientific knowledge through collaboration, discourse, and learner-driven activities. Similarly, Liu et al. (2023) employed ENA to uncover the possible factors that contribute to students' decisions to quit playing video games. Building on these studies, this paper draws on the potential of the ENA to explore the relationships between students' situational interest levels and their in-game observational behaviors.

Inspired by the growing interest in using Large Language Models (LLMs) for qualitative analysis in educational research (Barany et al., 2023), this study leverages ChatGPT, a prominent LLM tool, to support our research process. Recent studies have examined ChatGPT's utility for codebook development and automated qualitative coding (Lopez-Fierro & Nguyen, 2024; Xiao et al., 2023). For instance, Zambrano et al. (2023) used ChatGPT to automate the coding of a human-developed codebook. In addition, Barany et al. (2023) evaluated ChatGPT's performances in codebook development across four distinct approaches: human-only, ChatGPT-only, ChatGPT followed by human iterations, and human input followed by ChatGPT. Their results demonstrated that two hybrid approaches yielded the most effective outcomes. However, despite the extensive efforts in using LLMs for qualitative research, few studies have applied codes that developed with joint efforts of human-ChatGPT to conduct data analysis. Thus, in this study, we employ a hybrid human-ChatGPT approach throughout the research cycle, including theme identification, codebook development, automated coding, and applying the developed codebook to explore our research question: "*When engaged in a game-based learning environment (WHIMC), how do students' observational behaviors vary based on their interest levels?*"

Methodology

Data context and participants

The study was conducted within the What-if Hypothetical Implementations in Minecraft (WHIMC) digital learning game (Lane et al., 2017), in which Minecraft (Java Edition) is used as a platform for astronomy simulations designed to promote students' learning and interest in STEM (Lane et al., 2022). WHIMC is typically implemented in 5-day summer camps or during after-school programs. Students explore open-ended worlds (e.g., the moon, a space station, multiple exoplanets), and engage with hypothetical astronomical scenarios (i.e., *What if Earth had no moon? What if the sun was cooler?).* For the first three days of each camp, students explore and collect environmental measurements (e.g., temperature, humidity, gravity, oxygen) using 18 virtual scientific tools, then post location-specific observations that are visible in real-time to other players (Lane et al., 2022). In the final two days, student teams design and build habitats on Mars.

For the codebook development process, we utilized existing data from prior summer camps in 2022, reported in Liu et al. (2024a), which included 76 students (49 male, 20 female, and 7 who either identified with a third gender option or chose not to disclose) from three different locations within the United States. The students group represented rural, suburban, and urban areas and a range of racial backgrounds: 12 Black/African American, 3 American Indian, 2 Asian/Pacific Islander, 4 Hispanic/Latino, 22 White/Caucasian, 1 identifying with multiple categories, 6 categorized as other, and 19 who preferred not to answer. Participation in the study was entirely voluntary, with written consent obtained from parents and students at the beginning of each camp.

We then applied the developed codebook to newly collected data in 2024, including 322 in-game scientific observations collected from 31 middle schoolers from camps in two regions of the United States, with one camp in urban and the other in suburban areas. Students representing diverse racial/ethnic backgrounds (10 White, 14 African American, 2 Native American or Pacific Islander, and 3 who reported another option) included 19 male students, 11 female students, and 2 students who either reported a third option or chose not to respond. Both datasets used in this study were collected from the same format of workshops across comparable settings at summer camps, with the WHIMC Minecraft game directed at students' learning of astronomy via gameplay. These two datasets were consisted of textual observations made by students during their exploration of various virtual worlds in the game. These observations reflect what captured students' attention or stood out to them. For example, students might post questions prompted by their observations, such as "*What does the rocket provide?*" or describe what they have noticed, like "*The Earth is green and blue.*" Individual students independently posted each observation, and these student-generated observations were visible to other players within the game and collected as part of our data for future analysis.

Codebook development

This study builds on prior efforts to use large language models (LLMs) for thematic analysis of qualitative data to develop codes (e.g., Barany et al., 2024). Following the approach by (Barany et al., 2024), we conducted an initial review of the observation data to understand its context and structure. We then engaged in rounds of prompt engineering using ChatGPT-4 to select a prompt that would identify relevant and high-quality themes in WHIMC student observation data, through an inductive approach. Due to the probabilistic nature of ChatGPT's responses, which can limit the LLM's consistency, each prompt evaluated 100 lines of data that were then re-evaluated across different sessions with ChatGPT, switching accounts to avoid OpenAI's history feature. After identifying a prompt that produced consistent results (i.e., where repeated tests of the same prompt generated no more than two codes that were different across runs), the final prompt was applied to the entire dataset:

> Hi ChatGPT, you are a great qualitative researcher. I want you to analyze the following comments and observations made by students while playing an educational version of Minecraft with different worlds focusing on astronomy. Please develop a qualitative codebook by conducting a thematic analysis to identify the main themes that emerge from the "observations" across different worlds pertaining to students' science learning in the game environment. The structure of the codebook should include names, definitions, example codes, and explanations for the codes. I will divide the data into chunks of 100 lines and give you one chunk at a time. You should use all chunks of 100 lines to do the thematic analysis. With every chunk, you can update and edit the codebook if needed to make sure it still captures all the themes across the previous chunks. Here is the data:

To stay within ChatGPT's 4096-token limit, the data was divided into 18 batches of 100 lines each. Each batch was processed individually, resulting approximately eight themes per batch. However, only themes that recurred at least two times across batches were retained. Authors 1 and 2 then reviewed and discussed themes proposed by ChatGPT to identify and consolidate repeated codes. Across the overlapping definitions identified for each theme, the clearest or most comprehensive definition was selected, or multiple definitions were merged through human efforts to create the codebook. Because codes were not always mutually exclusive, an observation could be labeled with more than one code. Researchers then re-reviewed the original dataset to confirm that code names and definitions were clear, comprehensive, and relevant to the context, and manually selected examples that best represented each code (past work found that ChatGPT-4 sometimes hallucinated examples; Barany et al., 2024). The resulting codebook is presented in Table 1. Four of the identified codes are related to the learning goals (*Astrophysical*, *Environmental*, *Resource*, *Scientific*), while two reflect social needs that do not map to any specific learning task (*Socio-emotional* and *Cultural*).

Table 1
Inductive Codebook developed by ChatGPT with Human Refinement

Codes	Definition	Examples
Astrophysical & Planetary Observation (Astrophysical)	Observations detailing physical astrophysical implications and characteristics of planetary bodies, incl. visual and environmental descriptions.	"lots of ice i guess you need it to be on the moon" "the planet is 10x bigger than earth."

Environnemental Adaptation & Survival Strategies (Environnemental)	Students developed strategies, took actions, or manipulated environmental conditions to adapt and survive.	"NO OTHER SINGS OF LIFE OTHER THAN RABITS" "in human life its to cold to live here"
Resource Identification & Utilization (Resource)	Observations and interactions with game resources, their identification, and discussions on their use.	"what does the lighthouse to since there no moon?" "they use solar panels for energy so there are probally no fossil fuels or anything else"
Scientific Inquiry & Reasoning (Scientific)	Instances where students made logical reasoning, proposed hypotheses or engaged in questioning, and attempted to comprehend scientific concepts.	"theory: the biodome roof opens to regulate sunlight" "how does it get power with no sun"
Scientific Inquiry & Reasoning (Scientific)	Instances where students made logical reasoning, proposed hypotheses or engaged in questioning, and attempted to comprehend scientific concepts.	"theory: the biodome roof opens to regulate sunlight" "how does it get power with no sun"
Socio-emotional Responses (Socio-emotional)	Comments reflect emotional states and social reactions to the game environment or events within the game.	"i can FLY!!!!" "PARY TO BEAD R I P :("
Cultural & Playful References (Cultural)	References to popular culture or playful elements that indicate engagement beyond the educational objective.	"Joe MAMA" "sussy sussy amogus"

Automated coding

Once the inductively developed codebook was complete, researchers then worked with GPT-4o to automate and validate the coding process. To ensure the reliability of GPT-4o's automated coding, the first and second authors first hand-coded a subset (N=200) of student observations separately, achieving acceptable interrater reliability (IRR; >0.70) on 3 of the 6 codes. Social moderation (Herrenkohl & Cornelius, 2013) was employed for codes that did not achieve IRR in the first round to resolve any disagreements and improve coding accuracy; clarifications to the codebook were made as a result. Human coding and review were then repeated with additional sets of 100 lines until the $\kappa > 0.70$ threshold was reached for each construct.

We then automated our coding process using few-shot prompting (Liu et al., 2024b), in which a large language model (GPT-4o) is given a small number of examples in the input prompt to guide it in generating a response. In this case, the code names, definitions, and selected examples were provided. Human-human agreement was established through coding and discussion between two human researchers. This human coded data was used to validate the GPT coding scheme. Following Zambrano et al., (2023), a binary classifier (1 for presence/0 for absence) was implemented to simplify coding tasks. GPT's temperature hyperparameter was set to 0 in order to ensure consistent outputs, but default parameter settings were used otherwise. In this case, the code names, definitions, and examples were provided. Additionally, each batch of data was run three times under identical configuration settings to further minimize any stochastic variation during model evaluation, and a majority vote was used if the runs disagreed. The prompt we used is as follows:

> Please review the provided text and code it based on the construct: {construct}. The definition of this construct is {definition and examples}. After reviewing the text, assign a code of '1' if you believe the text exemplifies {construct}, or a '0' if it does not. Your response should only be '1' or '0'.

Table 2
Performance Metrics of GPT in Coding Each Construct

Codes	Human-Human Initial κ	Human-Human Final κ	Human-GPT κ	Human-GPT Precision	Human-GPT Recall
Astrophysical	0.88	0.88	0.74	0.88	0.71
Scientific	0.55	0.84	0.78	0.91	0.71
Socio-emotional	0.56	0.77	0.79	0.95	0.75
Environmental	0.64	0.71	0.78	0.75	0.86
Resource	0.70	0.71	0.70	0.82	0.70
Cultural	0.78	0.78	0.76	0.95	0.71

Table 2 provides both the initial agreement between two human researchers and the subsequent agreement after discussing the initial disagreements and coding additional data (κ>0.70 for all codes). It also compares the IRR between humans and GPT, including Kappa (κ), precision and recall scores. Precision captures how often the model is correct when it selects a code, and recall captures how often the model selects a code when it is present according to humans.

Measuring situational interest

Students' situational interest (SI) was measured using Linnenbrink-Garcia et al. (2010) validated survey, which was administered on the fourth day of the camp, after students completed all observations. Of the 31 students in the study, 28 completed this survey. Each student's SI score was calculated as the average of their responses to all questions on the scale. Students were then grouped into low (n=10), medium (n=9), and high (n=9) interest categories, based on the lower, middle, and upper thirds of the interest score distribution. To explore how situational interest influences students' scientific observations, we compare the high and low-interest groups.

Epistemic network analysis

Epistemic Network Analysis (ENA) was applied to compare how students with high and low situational interest in astronomy differ in terms of their observation patterns. As a quantitative ethnographic technique, ENA identifies, quantifies, and visualizes connections across themes in complex qualitative data (Shaffer et al., 2016). It has been widely employed in studying factors influencing learning in game-based environments (e.g., Bressler et al., 2019; Liu et al., 2024a). The structure of epistemic networks also supports visual and statistical comparisons of network patterns across groups (e.g., high versus low achieving students). ENA typically uses a moving window to calculate associations between the codes found both within and across lines of data, which we used to visualize the development of patterns as students posted consecutive observations over time.

Epistemic networks were generated using Marquart et al.'s (2021) ENA Web Tool. Unit variables—which organize the data into meaningful groups for comparison—were set as interest groups (low and high), student, camp, and day. The conversation variable—which bounds connections made across lines using the moving window—was set as student, so that development could be assessed within each learner's data. To set the length of the moving stanza based on the structure of the dataset, we randomly sampled 20 lines and qualitatively assessed how many prior chronological observations students self-referenced. After applying the average (window size=2), we tested window sizes of 3 to 6 and found only minimal differences in outcomes. The ENA models we used included all six codes from the codebook (see Table 1). When network models were generated for the low and high interest student groups, a means rotation was used to maximize variance along the x-axis.

Results

In this study, we grouped students based on their self-reported situational interest levels and categorized their observations using predefined codes. The following section presents both descriptive results and ENA difference models that visualize the differences in patterns of observation-making for each group.

Descriptive results

Table 3 displays the frequency distribution of each type of observation among students with high situational interest (High-SI) and low situational interest (Low-SI). A Mann-Whitney U test was conducted to assess whether students' situational interest (SI) levels were associated with making different total numbers of observations when engaging with the gameplay. The Mann-Whitney U results suggest that, overall, High-SI students tend to make statistically significantly more observations than their Low-SI peers across all categories (U=72.5, p=0.027). In terms of specific categories of observations, High-SI students made considerably more *Scientific* (U=75.0, p=0.011) and *Astrophysical* (U=72.5, p=0.023) observations than students in the Low-SI group. Similarly, *Resource* (U=81.0, p=0.002) also appeared more frequently among the High- than the Low-SI group. All statistical tests remained significant when a Benjamini & Hochberg (1995) post-hoc test was applied. It is worth noting that, while no statistically significant differences were observed in the remaining observation categories between the two SI groups, *Socio-emotional* (1.6 observations per student, 24%) and *Cultural* (1.8 observations per student, 27%) observations were, on average, the most prevalent codes within the Low-SI group. In contrast, the two dominant codes among the High-SI group are: *Scientific* (25% with an average of 7.1 observations per student*),* and *Astrophysical* (with the average observation per student reaching 10.6 and a frequency of 37%).

Table 3
Code Frequencies for High and Low SI Groups

Codes	High-SI (n=9)		Low-SI (n=10)		Diff. %	Mann-Whitney U
	Avg. obs./student	%	Avg. obs./student	%		
Scientific	7.1	25%	0.6	9%	16%	75.0 (p =0.011)
Astrophysical	10.6	37%	1.5	23%	14%	72.5 (p=0.023)
Environmental	1.4	5%	0.9	14%	9%	51.0 (p=0.637)
Resource	3.1	11%	0.2	3%	8%	81.0 (p=0.002)
Socio-Emotional	3.8	13%	1.6	24%	-11%	66.5 (p=0.081)
Cultural	2.6	9%	1.8	27%	-19%	61.0 (p=0.191)
Total # of observations (N)	257		66		191	72.5 (p=0.027)

ENA model

Figure 1 shows the epistemic network for the High-SI group (*up left*, red), the Low-SI group (*up right*, blue), and Figure 1 *bottom*) displays a subtracted model highlighting the differences between each group. The corresponding line weights (lw) for each network and the average number of occurrences of each connection per student are displayed in Table 4. A Mann-Whitney U test shows the observation patterns of the two groups are statistically significant along the X-axis (U= 10.00, p<0.01, effect size of r=0.78 at α=0.05). This statistical result points to the distinct difference in how students from the two SI groups engage with WHIMC. Specifically, High-SI students demonstrated more robust and interconnected patterns than those in the Low-SI group, with all sixteen possible links presented in the High-SI network (Figure 1, *up left*), in contrast to only eight connections among the nodes in the Low-SI students (Figure 1, *up right*).

Figure 1

ENA models for the High-SI group (up left), the Low-SI group (up right), and the difference model (bottom).

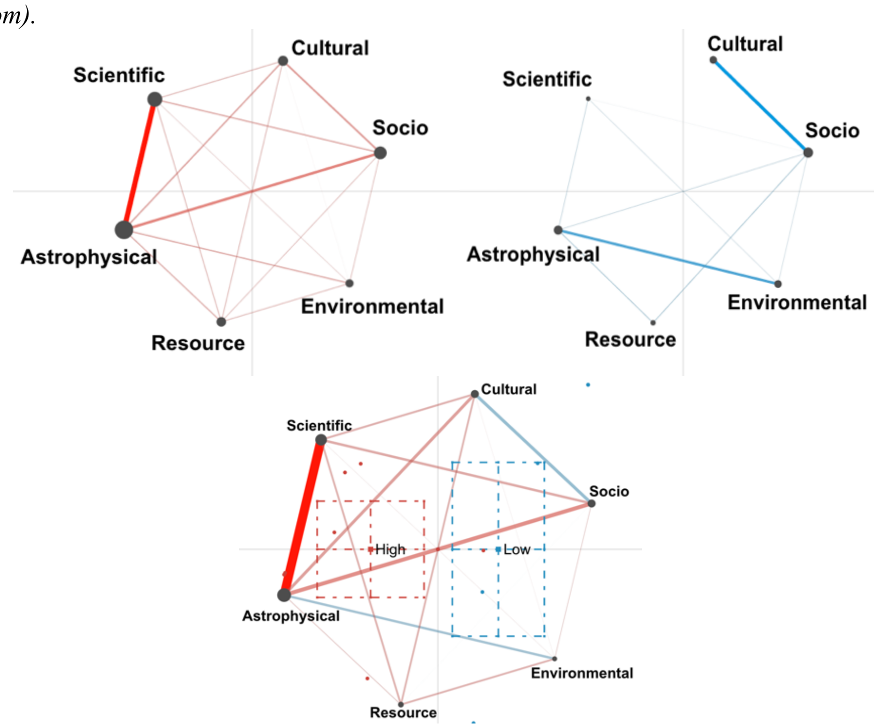

Table 4 presents the line weights (lw) of the ENA models and the average frequency of each connection per student. As shown above, the High-SI group tends to have frequent co-occurrences between *Astrophysical* and *Scientific* (7.7 co-occurrences on average per student), as indicated by the thicker red edge (lw=0.44) connecting these two nodes in the ENA model. This stronger association suggests that High-SI students often integrate scientific thinking with practical observations about planetary phenomena. Examples of these associations from High-SI students may start with an observation about the moon, *"the moon is to aggresive to the water to make it over flow,"* followed by a question of this phenomenon, *"why is the tides so low?"*. Similarly, High-SI students might also pose a scientific question about an astrophysical observation, curious about: *"why is* *the sky orange?"* This inquiry then precedes active obversions and possible explanations, such as: *"there is lava all around the map!"* These observational patterns indicate that High-SI students are constantly putting effort to interpret and reflect the phenomenon they are observing within the game. Correspondingly, the thicker line weight between these two codes is supported by frequency data in Table 3, where *Astrophysical* and *Scientific* represent the two most prominent codes among the High-SI group (37% and 25%).

Table 4
Line Weights for Epistemic Network Difference Model in Figure 1 (bottom)

Connection	High-SI Group		Low-SI Group		Differences
	LW	Avg #/stu	LW	Avg #/stu	
Astrophysical - Scientific	0.44	7.7	0.05	0.5	0.39
Astrophysical - Socio-Emotional	0.22	5.0	0.05	0.8	0.17
Astrophysical - Culture	0.13	2.7	-	0.2	0.13
Scientific - Socio-Emotional	0.12	2.3	0.01	0.3	0.11
Scientific - Cultural	0.09	1.3	-	0.1	0.09
Scientific - Resource	0.08	1.4	-	0.0	0.08
Resource - Environmental	0.08	0.8	-	0.1	0.08
Resource - Cultural	0.08	1.2	-	0.1	0.08
Astrophysical - Resource	0.11	2.2	0.07	0.1	0.04
Environmental - Socio-Emotional	0.07	0.7	0.04	0.4	0.03
Scientific - Environmental	0.05	0.9	0.04	0.4	0.01
Resource - Socio-Emotional	0.07	1.4	0.07	0.1	0
Astrophysical - Environmental	0.11	1.8	0.22	1.0	-0.11
Cultural - Socio-Emotional	0.15	2.1	0.29	1.4	-0.14

Additional associations from the High-SI group can be observed between *Scientific* and *Socio-emotional*, as well as *Resource (2.3 and 1.4 co-occurrences per student, respectively*) though these connections show lower edge weights (lw=0.12, 0.08, respectively) compared to the most pronounced connection. The diverse observational patterns among High-SI students may reveals that they incorporate not only scientific components during gameplay but also draw their emotional reactions and resource-related considerations into their narratives. For instance, High-SI students may first share the resource they have discovered: *"i found an igloo"*, then proceed with a logical inference based on their observation: *"i infer that he polar bears live here"* (as both igloos and polar bears are associate with snow, cold weather and often found around the Arctic regions.). This ability to connect various concepts and bridge different pieces of information to develop a logical hypothesis, suggests that High-SI students are not only more actively engaged but are also capable of identifying and synthesizing the relations between different elements. This, in turn, leads them to a deeper and more holistic understanding of the presented material. Although *Cultural* is also associated with other codes within the network, their edge weights are relatively low compared to associations between *Astrophysical* and *Scientific*, indicating that High-SI students prioritize astrophysical and scientific elements over more playful references when engaged in WHIMC.

In contrast, the strongest connection among Low-SI students is between the two codes that are least related to on-task game activities: *Socio-emotional* and *Cultural* (1.4 co-occurrences per student, lw=0.29). Though *Astrophysical* and *Environmental* are also present in Low-SI group's network (1.0 co-occurrences), the connection strength is weaker than the connection between *Cultural* and *Socio-emotional*, indicating a reduced focus on planetary investigations among the Low-SI students, who seem to engage more often in off-task observations. Instead of articulating their scientific reasoning, Low-SI students may make more playful and non-task-related comments such as, *"WE LIKE FORTNITE."* Only two visible edges (in blue) are defined in the ENA difference model, showing that the Low-SI students' observational behaviors are more narrowly distributed across categories, lacking the multi-dimensional integration observed among High-SI students. This limited interconnectivity may also represent that Low-SI students are not fully engaged in science content offered by the WHIMC environment to connect observations across various domains.

Discussion and conclusion

The goal of this study is to explore how middle-school students with different levels of situational interest (Linnenbrink-Garcia et al., 2010) made observations while learning from WHIMC. We found that High-SI students made observations significantly more frequently during gameplay than Low-SI students. Moreover, High-SI students are also more active in generating observations that reflect scientific knowledge (codes such as

Astrophysical or *Scientific*) than the students in Low-SI group. The epistemic networks provide further insights into distinctions between the two groups: High-SI students exhibit a particularly strong connection between *Scientific* and *Astrophysical* observations, displaying a generally denser network (more connections) than Low-SI students, whose connections primarily related to off-task topics (i.e., *Socio-emotional* and *Cultural*). Further, no connections involving *Scientific* observations are present in the Low-SI group's ENA difference model. The absence of this category implies that students with lower interest levels may not fully engage in the scientific aspects of the learning content. These results underscore the critical role of interest in shaping the breadth and depth of their participation within WHIMC, highlighting the importance of fostering interest and the potential benefits of targeted interventions to build self-regulation skills, helping middle-school students redirect their attention and engage more deeply with educational content and objectives, even when their initial interest is low.

The distinct observation patterns produced by the two groups also show evidence that they are engaged differently by the game, based on their situational interest. The Low-SI group's limited observations, centered on *Socio-emotional* and *Cultural* references, align with the earliest phase of interest development, which Hidi & Renninger's (2010) refer to as *triggered situational interest*. Within this form of interest, students' interest is mainly expressed through affective responses, where their attention gravitates toward gameplay's immediate and entertaining features. On the other hand, High-SI students had more frequent and diverse links that showed strong connections with *Astrophysical* and *Scientific* content. These patterns are consistent with *maintained situational interest*, where students start to recognize the intrinsic value of the content, and engage more deeply with the domain content (Linnenbrink-Garcia et al., 2010). Thus, while both groups show interest during gameplay, their engagement mechanisms differ: High-SI students exhibit a sustained, knowledge-driven interest, actively integrating new insights to build deeper connections throughout their learning. In contrast, Low-SI students engage at a surface level, attending to emotionally appealing but less scientific aspects of the learning experience.

While our analysis offers insights about the relationship between students' interests and their participation during gameplay, we recognize some limitations to this work. While the amount of data we have is sufficient for the type of analyses we are running, the relatively focused sample may limit the generalizability of our findings. Future work could address this by running the study with a broader sample. In addition, future studies could explore automated systems capable of tracking student observations in real-time. Such systems could help camp facilitators support student interest development and optimize their learning outcomes. In general, our findings suggest the potential for future work around interventions. Although the relationships we found cannot yet be treated as causal, encouraging or otherwise supporting the use of more science-related observations among low interest students could improve their knowledge of the space, which is known to be an important component required for interest development (Schraw & Lehman, 2001; Hidi & Renninger, 2006).

Overall, this study contributes to the ongoing conversation on identifying middle-school students in need of support within digital game-based learning environments. Prior research indicates that game-based learning does not consistently stimulate interest across all students (Rodríguez-Aflecht et al., 2018). Our study builds on previous work by investigating how students with different levels of situational interest engage with in-game observations in distinct ways. These findings aim to provide practical insights for future researchers, educators, and facilitators to select interventions that help students develop an enduring interest in science, while also supporting effective engagement and learning for less interested students. Given WHIMC’s open-ended nature, these insights may extend beyond game-based learning and apply to other self-directed learning environments where students have greater autonomy to manage their own exploration.

References

Barany, A., Nasiar, N., Porter, C., Zambrano, A. F., Andres, A., Bright, D., ... & Baker, R. S. (2024). ChatGPT for education research: Exploring the potential of large language models for qualitative codebook development. *Intl. Conf. on Artificial Intelligence in Education*, 134-149.

Bressler, D. M., Bodzin, A. M., Eagan, B., & Tabatabai, S. (2019). Using epistemic network analysis to examine discourse and scientific practice during a collaborative game. *J. of Science Education & Technology*, 28, 553-566.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. of the Royal statistical society: series B (Methodological), 57*(1), 289-300.

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, 591203.

Deen, M. (2015). G.A.M.E., Games autonomy motivation & education: how autonomy-supportive game design may improve motivation to learn (Doctoral dissertation), Technische Universiteit Eindhoven.

Herrenkohl, L. R., & Cornelius, L. (2013). Investigating elementary students' scientific and historical argumentation. *J. of the Learning Sciences, 22*(3), 413-461.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research,* 70(2), 151-179.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational psychologist, 41*(2), 111-127.

Ishak, S. A., Din, R., & Hasran, U. A. (2021). Beyond play: Conceptualising the capability of a good digital game to stimulate interest in STEM. *Intl. J. of Learning, Teaching and Educational Research, 20*(10).

Koskinen, A., McMullen, J., Halme, H., Hannula-Sormunen, M., Ninaus, M., & Kiili, K. (2022). The role of situational interest in game-based learning. *Intl. GamiFIN Conf.*, 54-63.

Lane, H. C., Yi, S., Guerrero, B., & Comins, N. F. (2017). Minecraft as a sandbox for STEM interest development: Preliminary results. *25th Intl. Conf. on Computers in Education*, 387-397.

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

Lee, S. B., Charles, R., Jokodola, A., Lux, K., Molloy, J., Hurford, A., & Hamilton, E. (2022). Promoting STEM careers: Preliminary analysis of the Young Scholars Program. *16th Intl. Conf. of the Learning Sciences-ICLS 2022*, 1673-1676.

Linnenbrink-Garcia, L., Durik, A., Conley, A., Barron, K., Tauer, J., Karabenick, S., & Harackiewicz, J. (2010). Measuring situational interest in academic domains. *Educational & Psychological Measurement, 70*(4), 647-671.

Liu, X., Hussein, B., Barany, A., Baker, R. S., & Chen, B. (2023). Decoding player behavior: Analyzing reasons for player quitting using log data from puzzle game Baba Is You. *Intl. Conf. on Quantitative Ethnography*, 34-48.

Liu, X., Zambrano, A., Barany, A., Ocumpaugh, J., Ginger, J., Gadbury, M., Lane, H., & Baker, R. (2024a). Investigating learner interest and observation patterns in a Minecraft virtual astronomy environment. *Intl. Conf. on Quantitative Ethnography*, 19-34.

Liu, X., Zhang, J., Barany, A., Pankiewicz, M., & Baker, R. S. (2024b). Assessing the potential and limits of large language models in qualitative coding. *Intl. Conf. on Quantitative Ethnography*, 89-103.

Lopez-Fierro, S., & Nguyen, H. (2024). Making human-AI contributions transparent in qualitative coding. *17th Intl. Conf. on Computer-Supported Collaborative Learning-CSCL 2024*, 3-10.

Makransky, G., Petersen, G. B., & Klingenberg, S. (2020). Can an immersive virtual reality simulation increase students' interest and career aspirations in science? *British J. of Educational Technology, 51*(6), 2079-2097.

Marquart, C. L., Hinojosa, C., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2021). Epistemic network analysis (Version 1.7.0). http://app.epistemicnetwork.org.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*(4), 258-283.

Prensky, M. (2005). Computer games and learning: Digital game-based learning. *Handbook of Computer Game Studies*, *18*(2005), 97-122.

Qian, M., & Clark, K. R. (2016). Game-based learning and 21st-century skills: A review of recent research. *Computers in Human Behavior*, 63, 50-58.

Rodríguez-Aflecht, G., Jaakkola, T., Pongsakdi, N., Hannula-Sormunen, M., Brezovszky, B., & Lehtinen, E. (2018). The development of situational interest during a digital mathematics game. *J. of Computer Assisted Learning*, *34*(3), 259-268.

Renninger, K. A. (1992). Individual interest and development: Implications for theory and practice. *The Role of Interest in Learning and Development, 26*(3-4), 361-395.

Renninger, K. A., Hidi, S., Krapp, A., & Renninger, A. (2014). The role of interest in learning and development. *Psychology Press*.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *J. of Learning Analytics, 3*(3), 9-45.

Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational psychology review*, 13, 23-52.

Zambrano, A. F., Liu, X., Barany, A., Baker, R. S., Kim, J., & Nasiar, N. (2023). From nCoder to ChatGPT: From automated coding to refining human coding. *Intl. Conf. on Quantitative Ethnography,* 470-485.

Acknowledgments

This research was supported by National Science Foundation Grants # DRL2301173.