Investigating Learner Interest and Observation Patterns in a Minecraft Virtual Astronomy Environment

Xiner Liu[1][0009-0004-3796-2251], Andres Zambrano[1][0000-0003-0692-1209], Amanda Barany[1][0000-0003-2239-2271], Jaclyn Ocumpaugh[1][0000-0002-9667-8523], Jeff Ginger[2][0009-0007-1898-4959], Matt Gadbury[2][0000-0002-8544-8693], H. Chad Lane[2][0000-0002-3046-1751], and Ryan S. Baker[1][0000-0002-3051-3232]

[1] University of Pennsylvania, Philadelphia, PA 19104, USA
xiner@upenn.edu
[2] University of Illinois, Champaign, IL 61820, USA

Abstract. This study explores how student actions in Minecraft-based virtual environments designed to simulate astronomical phenomena shift over time, as their interest in astronomy changes. We analyze observations made by middle school learners participating in the What-if Hypothetical Implementations in Minecraft (WHIMC) project, which adapts the game to immerse learners in scenarios exploring scientific concepts. Combining manual and automated coding techniques, we classify these observations and use epistemic network analysis to investigate how they relate to changes in interest levels as measured by pre- and post-surveys. Our findings show that learners who maintain or increase their astronomy interest produce more complex observational behaviors, such as hypothesis generation and comparisons. Conversely, learners whose interest declines produce more surface-level, factual observations. Results suggest ways to identify and support long-term interest in science education.

Keywords: WHIMC, Automated Coding, Educational Games, STEM Interest

1 Introduction

Ongoing calls for a more STEM-literate workforce [6], coupled with increasing demand for STEM college graduates in astronomy-related fields [25], have prompted educators to seek innovative ways to engage learners and ignite their interest in science. Despite various initiatives, capturing and sustaining student interest remains a significant challenge. One promising approach lies in leveraging popular games and simulated environments [5, 12] like Minecraft [14] to create immersive learning environments that encourage exploration and critical thinking about complex scientific concepts.

Considerable effort has sought to optimize simulations and games to support learner interest [17, 25]. Recently, educators and researchers have investigated ways to leverage the flexibility and immersive nature of simulations in the popular game Minecraft to present learners with STEM-related questions and encourage their overall engagement with scientific content [23, 36, 39]. This study investigates how changes in student interest (from the beginning to the end of a Minecraft learning activity) relate to changes in their behavior, using text-based observational data produced by learners in the What-if Hypothetical Implementations in Minecraft (WHIMC) environment [22]. WHIMC is designed to teach learners about astronomy [23]. Therefore, we investigate the research question of how different patterns of in-game observations (made by learners) correlate with changes in learners' interest in astronomy, as measured by the astronomy subscale of Gadbury & Lane's STEM interest survey [11]. By examining these patterns using Quantitative Ethnographic techniques, we hope to identify key factors that contribute to learners maintaining or increasing interest in STEM.

2 Related Works

This study examines the relationship between student interest and their interactions with the WHIMC system. Considerable research indicates that interest positively influences student perceptions, beliefs, attitudes, and willingness to learn more about that topic [15, 19, 32, 33, 38]. Learners with high interest tend to engage more deeply with material, improving their conceptual understanding [3, 31]. Conversely, learners with low interest are typically less likely to persist with the material [26, 35]. As such, interest is often predictive of future academic choices [18].

Early investigations of interest tended to describe it as an emotion [2, 28], but more recent work notes that interest has a longer duration than most emotions, with Renninger and Hidi [32], noting that a purely affective description does not capture interest's cognitive components [33]. Hidi & Renningers' four-phase model of interest development posits that interest emerges as a product of environmental features in its early phases, referred to as situational interest [16]. As learners' knowledge of a topic increases, individual interest emerges as a self-motivating and enduring state.

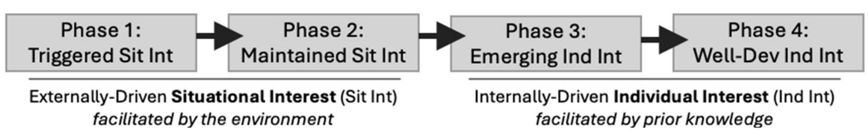


Fig. 1. The Four-Phase Model of Interest [16].

Importantly, research suggests that situational interest is best triggered by novel and attention-getting experiences [13, 30] that lead learners to generate questions [1]. Learners who are in phases 1- 2 are still developing a sense of the value of that content and need support to connect new material to their prior knowledge and skills [29]. As individual interest emerges (phase 3), they become more likely to engage independently, but may see little use in feedback or in the canon of the field; it is not until the interest becomes well-developed that learners effectively manage frustration and actively seek feedback [29].

Eberbach and Crowley emphasize that observational skills develop over time and may require scaffolding [7]. Specific frameworks have been developed to classify and interpret learners' observations in different educational contexts. For example, Yi's framework for coding textual observations made by student includes relevant observations, questions, hypotheses, and texts that might indicate lower engagement with astronomy or the Minecraft platform [42]. Other frameworks consider whether observations made by learners are disengaged [21]. Building on these studies, we expand Yi's system to analyze learners' engagement and learning processes within the WHIMC platform, adding codes relevant to both engagement and off-task behavior [42].

We use Epistemic Network Analysis (ENA) and difference models to compare the connections between different types of observations at the beginning and end of the learning activity. Previous studies have applied ENA specifically in game-based learning environments to analyze player behavior and cognitive skills. For example, Scianna et al. use ENA to visualize player log files and identify variations in how learners responded to game events [37], while Foster et al. explored patterns in learners' identity exploration within play-based environments [9]. Bressler et al. analyzed the conversational discourse of game teams to explore the connections between communication responses, language style, and scientific practice [4].

3 Methods

3.1 Research Context

We conduct this research in the context of WHIMC [23], which exploits the widespread appeal of Minecraft to create scenarios and experiences that stimulate interest in astronomy, Earth science, and other STEM fields related to space exploration. Utilizing Minecraft's Java Edition, WHIMC offers learners simulation environments where they can explore hypothetical astronomy scenarios, addressing *what-if* questions (e.g., *What if Earth had no moon?* or *What if the sun was cooler?*). During informal learning settings (e.g., summer camps), learners are guided by pedagogical agents and human facilitators to make scientific observations that assess each world's potential for human habitation.

In addition to these hypothetical worlds, the WHIMC server features a NASA-inspired launch site, a lunar base, a space station, a Mars map based on real Martian terrain data, several known exoplanets, and phenomena such as black holes and quasars. Learners can access various science tools that support their investigations, enabling them to measure critical habitability factors (i.e., temperature, air pressure, radiation, gravity, and atmospheric composition). Learners are taught to use these measures as evidence by making descriptive, comparative, and inferential observations that underpin testable hypotheses. As part of their game exploration, players record their observations on wooden signs that are displayed to other players and collected with our data.

Previous WHIMC research used association rule mining to examine the ways in which a more limited range of learner behaviors is correlated with learner interest [10], as the open-ended nature of this environment provides many opportunities for supporting student STEM interest development [41]. This study analyzes data collected in 2022 from 76 learners (49 male, 20 female, and 7 who either reported a third option or preferred not to answer) across 5 locations in 3 states. Participants were drawn from populations rural, suburban and urban populations and represented a wide range of racial backgrounds (12 Black/African American, 3 American Indian, 2 Asian/Pacific

Islander, 10 Hispanic/Latino, 22 White/Caucasian, 1 multiple categories, 6 other, and 19 who preferred not to answer). Socio-economic backgrounds varied considerably between locations, with nearly half of the learners coming from two of the wealthiest counties in the US and the remainder coming from mixed or lower income areas. Disability status was not consistently captured, but some learners reported receiving accommodations. Data was collected during 12–15-hour informal learning camps hosted at planetariums, libraries and community centers. Recruitment was run with community partners as part of their regular programming, with sign-ups occurring online. Participation in the research was entirely optional. Written consent was obtained from all parents, and participants assented on the first day.

3.2 Codebook Development

This study expands on Yi's initial efforts to code the observations learners make while exploring WHIMC (Table 1) [42]. Yi's work classified observations into only 4 codes (Noun, Measure/Descriptive, Comparison, Hypothesis), which were designed to capture observations that closely align with WHIMC's learning goals [42]. We augment this framework using an inductive thematic analysis approach to select themes [40]. This process involved identifying and iteratively refining topics that emerged from an open-ended review and discussion of the data. New codes include several constructs that reflect other forms of game-related interactions that were not fully captured by the coding schemes used in [42] and social communication. Finally, codes reflecting non-game interactions (i.e., random characters, unrelated content, or off-topic conversations) were coded separately (to facilitate accurate automatic coding) but then collapsed because they all appear to reflect off-task conversations. As codes are not mutually exclusive, observations may be labeled more than once.

Table 1. Inductive Themes/Constructs Derived from Observations Made by Learners

Code Name	Definition/Example
Noun	**Def**: Stating nouns without any elaborations. (Previously labeled as "factual" in [42].) **Ex**: "I see trees"
Measure/ Descriptive (Meas)	**Def**: Related to measures of physical attributes that learners are encouraged to take in each of the different planets and moons they visit, including color, temperature, quantity, weight or size, radiation, temperature, airflow, pressure, altitude, etc [42].**Ex**: "the temp is -20.6 C, -5.1 F, 252,5 K"
Comparison (Comp)	**Def**: Observations that compare/contrast conditions either (a) among in-game worlds (e.g., 2 planets they've explored) or (b) their real-life experiences on Earth to the in-game worlds. Also inc. examples that suggest their expectations were violated [42]. **Ex**: "the grass is greener in the habitable strip."
Hypothesis (Hypo)	**Def**: Making hypotheses or guesses, showing speculative thinking, forming conjectures, or making predictions or explanations [42]. **Ex**: "this world is probably closer to the sun"
Questioning (Quest)	**Def**: Asking questions about game mechanics or world elements; Seeking to understand the game better, showing curiosity. **Ex**: "Why is there no grass?"
Exclamations (Exc)	**Def**: Pure exclamations without any accompanying explanation of observations, including exclamatory grammatical markers or words. **Ex**: "Wow!"

Continuing Discussion (ContDisc)	**Def**: The same user's observations are not syntactically exact, but they represent the continuation of discussion around a specific topic. **Ex**: "Can't find [NAME]," "[NAME] where are you?".
Repetition (Rep)	**Def**: A student repeats the exact same observations 3 or more times w/o respect to case sensitivity. **Ex**: "temperature," "temperature," "temperature."
Social Comunication (Soc)	**Def**: Interactions with other players through observations: communication or shared exploration. **Ex**: "[NAME] Hello!"
Non-game: True Nonsense	**Def**: A sequence of characters, emojis, or symbols repeated excessively, including random numbers or letters without associated explanations or observations. **Ex**: "AAAAAAAAAAAAAAAAAAAAAAAA"
Non-game: Unrelated Phrases	**Def**: Sentences or phrases unrelated to the purpose of making observations during Minecraft gameplay. **Ex**: "this will expire in a week"
Non-game: Out-of-Context Ref.	**Def**: References to movies, books, celebrities, etc., without relevance to the game. **Ex**: "Subscribe to Mr.Beast Gaming"

Two researchers independently coded 200 lines of data for the presence or absence of each construct using these definitions. After each researcher coded approximately 100 lines, interrater reliability (IRR) was checked; those constructs that did not reach sufficient Kappa were discussed before additional data were coded and interrater reliability was re-checked. All discrepancies between the two human coders were resolved using social moderation before evaluating the performance of the automated coding models.

3.3 Automated Coding of Observation Data

To effectively scale the coding of this data for real-time analysis, we automated our codebook using Python scripts and GPT-4 [27]. Table 2 also provides Kappa between the 2 human coders before resolving disagreements, and the performance metrics for each code, including Kappa, Precision, and Recall scores between GPT and human coding. Kappas between GPT and human coders ranged between 0.72 to 0.95, and Shaffer's rho was sufficient for all codes ($p \leq 0.05$) [8].

Table 2. Performance Metrics for Each Automated Model

Construct	Freq.	Method	Hum-Hum Kappa	Hum-GPT Kappa	Prec.	Recall
Noun	17%	Zero-shot	0.85	0.84	0.88	0.83
Measure/Descriptive (Meas)	36%	Few-shot	0.80	0.78	0.88	0.83
Comparison (Comp)	14%	Few-shot	0.73	0.74	0.73	0.79
Questioning (Quest)	9%	Zero-shot	0.96	0.95	0.9	1.00
Hypotheses (Hypo)	6%	Zero-shot	0.73	0.77	0.81	0.76
Exclamation (Exc)	6%	Zero-shot	0.95	0.86	0.81	0.96
Continuing Discussion (ContDisc)	13%	Embed.	0.88	0.93	0.97	0.95
Repetition (Rep)	9%	Python	1.00	1.00	1.00	1.00
Social (Soc)	3%	Python	0.95	0.94	1.00	0.9
NonGam: True Nonsense	4%	Zero-shot	0.97	0.95	0.94	0.97
NonGam: Unrelated Phrases	7%	Few-shot	0.75	0.77	0.82	0.91
NonGam: Out-of-context Reference	3%	Few-shot	0.88	0.86	0.93	0.81

Coding using GPT-4. We used gpt-4-turbo-2024-04-09, which was the latest available version, through OpenAI's application programming interface (API). We adhered to default hyperparameter settings throughout our analysis, except for setting the temperature hyperparameter to 0 to ensure the consistency of the output. Because the stochastic nature of GPT models can produce variable outputs even with temperature set to 0, we replicated each input configuration three times during model evaluation.

Coding with Zero-Shot Prompting. In the zero-shot prompting approach, we present the GPT-4 model with definitions of each construct without examples and query the model to code across the full 200-line dataset. This approach achieved good IRR for constructs with well-defined definitions that could not be readily associated with keywords.

Coding with Few-Shot Prompting. This method builds on the foundation laid by zero-shot prompting but adds illustrative examples to improve the model. This approach performs better than zero-shot for more nuanced constructs that may require more sophisticated contextual interpretations. To ensure generalizability, we limit training examples to instances that model is not asked to code during the validation process.

Coding with Embedding Models. Embedding is a process that converts words, phrases, or larger texts into numerical vectors that can be compared. We used OpenAI's text embedding model ext-embedding-3-small to code the construct Continuing Discussion, which identifies observations that are not syntactically identical but represent the ongoing discussion of a specific topic. In this method, each observation is first converted into embeddings using OpenAI's text embedding model. Next, we compute the cosine similarity in the spatial domain between the current observation's embedding and the embedding of the previous line. Similarity scores greater than 0.6 (selected based on experimental performance) are coded as '1'; otherwise, it is coded as '0'.

Coding with Python Programs. Two constructs—Social and Repetition—are coded using Python, due to their specific coding criteria. For the Social construct, coding involves identifying keywords such as identity pronouns or other players' username. The Repetition construct requires a systematically defined rule; it is coded when the same student in the same context repeats the same observation more than three times.

3.4 Analysis of Student Interest

This study analyzes existing data on astronomy interest, using a subscale of Gadbury & Lane's STEM interest survey [11]. Gadbury & Lane's [11] subscale contains items that closely reflect the learner characteristics of Hidi & Renninger's [16] stage 3 of interest development—emerging individual interest (as described in [29]). As such, this subscale differentiates those with emerging individual interest from those more likely to need support in developing astronomy interest.

We first validated Gadbury & Lane's STEM survey using a confirmatory factor analysis (CFA) using the data from all 76 learners in the study [11]. Item loadings demonstrated that the items reflect the same latent variable: emerging individual interest in astronomy (pre-survey loadings: 0.57-0.83; post-survey loadings: 0.64-0.84). This survey utilizes a Likert scale (1=*not at all interested*, 5=*extremely interested*). We use a 5-item subscale to assess learners' interest in astronomy (e.g., *I search for information about space in my free time*) with questions that are compatible with emerging individual interests [29]. We then classify learners into 4 groups based on their relative interest in the pre and post surveys (above/below median for each).

3.5 Epistemic Network Analysis

Epistemic networks were generated with the ENA Web Tool [24]. Specifically, we use difference models to compare learners who started with a similar level of astronomy interest but who finished the camp with a different interest gain (HiHi vs HiLo, and LoHi vs LoLo). For these difference models, observations made by each learner were coded temporally, with observations split into thirds. Difference models compare learners with different interests during the initial and final third of each student's observations. Doing so offers a detailed view of how initial and final interest levels interact with game elements and how these interactions might influence the development of astronomical interest.

To analyze the observations for each learner in the order they were produced, data was sorted by world (i.e., the order used in camps), username, and timestamp. Data was segmented by both user and world because the last observation a learner makes in one world is unlikely to influence the first observation made in the next. Several moving stanza window sizes were tested before selecting 4. This decision was made by randomly selecting 20 lines from the observation data to determine how far back connections extended (i.e., 3 was the most common number). ENA models were generated from all 10 codes, with smaller line weights (LW≤0.04) excluded for visual clarity.

4 Results

This study groups leaners by pre and post measures of interest to explores how astronomy interest reflects changes in learner observations within the WHIMC learning system. ENA difference models are then used to compare the networks of these different groups between the first third and last third of the observation period, looking for patterns that might signal the need for interventions.

4.1 Descriptive Statistics

This study explores the differences between groups that began with relatively similar levels of interest but diverged in their final level of interest. Therefore, our difference models (sections 4.2-4.3) compare the LoLo group to the LoHi group, and the HiHi group to the HiLo group. Table 3 shows the descriptive statistics for the interest surveys for each of the four groups of learners: (a) those who started and ended with astronomy interest that was above the median (HiHi), (b) those who started above but ended below the median (HiLo), (c) those who started below but ended above the median (LoHi), and (d) those who started and ended below the median (LoLo). Table 4 shows frequency statistics for each code across the 4 groups of learners for the initial and final third of their observations. As this shows, Measurement is the most common code across all groups and times (23-49%), but Nongame codes are surprisingly high among the early gameplay of the HiHi group (24%). Comparison codes also emerge relatively often across this data (12-15%).

Table 3. Pre- and Post-Interest Scores and Gains for Each Interest Groups

Interest Group	N	Pre, Med. (Std)	Post, Med. (Std)	Gains, Med. (Std)
High-High (HiHi)	15	4.00 (0.45)	4.20 (0.50)	0.00 (0.43)
High-Low (HiLo)	13	3.40 (0.34)	3.00 (0.35)	-0.40 (0.38)
Low-High (LoHi)	12	3.00 (0.23)	3.40 (0.48)	0.50 (0.44)
Low-Low (LoLo)	12	2.30 (0.48)	2.20 (0.38)	0.00 (0.33)

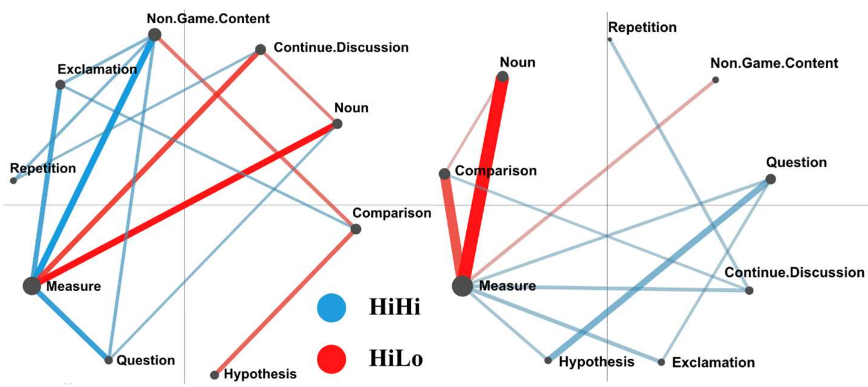
Table 4. Frequencies of Codes Across Learner Groups

	Initial Third of Observation						Final Third of Observation					
Code	**HiHi**	**HiLo**	**Diff**	**LoHi**	**LoLo**	**Diff**	**HiHi**	**HiLo**	**Diff**	**LoHi**	**LoLo**	**Diff**
Total (N)	168	93	**75**	149	123	**26**	152	100	**62**	128	123	**5**
Quest	4%	1%	**3%**	12%	12%	**0%**	23%	5%	**18%**	12%	11%	**1%**
Hypo	4%	6%	**-2%**	2%	6%	**-4%**	9%	2%	**7%**	4%	6%	**-2%**
Soc	5%	3%	**3%**	2%	16%	**-14%**	0%	0%	**0%**	0%	0%	**0%**
Noun	7%	25%	**-18%**	18%	17%	**1%**	22%	21%	**2%**	17%	14%	**3%**
Rep	7%	0%	**7%**	0%	3%	**-3%**	8%	3%	**5%**	3%	2%	**1%**
ContDisc	9%	6%	**3%**	7%	6%	**1%**	11%	3%	**8%**	11%	3%	**7%**
Exc	11%	0%	**11%**	6%	2%	**5%**	5%	3%	**3%**	6%	1%	**5%**
Comp	12%	19%	**-7%**	12%	13%	**0%**	15%	14%	**2%**	9%	9%	**1%**
Meas	23%	28%	**-5%**	32%	28%	**4%**	32%	49%	**-17%**	42%	39%	**3%**
NonGam	23%	8%	**17%**	10%	7%	**3%**	4%	5%	**-1%**	13%	3%	**10%**

Other codes vary more between groups or across time. Exc codes are considerably more common in the initial third of the HiHi group's observations (11%) than they are in the final third (5%) or for any other group/time combination (0-6%). Conversely, the Noun codes are relatively infrequent for the initial third of the HiHi group's observations (7%) compared to all other group/time combinations (18-25%), including their own. Moreover, Meas codes increase sharply for the HiLo group from 28% to 49%, which shows a more pronounced shift towards making observations involving quantitative measures for learners in this group towards the later part of the game. Notably, Quest codes rise 4% to 23% for the HiHi group and 1% to 5% in the HiLo group, but remain stable across time for both the LoHi and LoLo groups (11-12%). Social codes are infrequent overall, but the LoHi group decrease sharply (16% initial to 0% final).

4.2 Temporal Changes in the Difference Models for HiHi and HiLo Groups

Fig. 1 shows the difference models for learners who started with high astronomy interest, and Table 5 shows the corresponding line weights for these models. Both difference models compare learners who start and end with high interest (HiHi) to those who starts high but ends low (HiLo). Fig. 1a (left) shows the observations that were made early in the learners' game play, while Fig. 1b (right) shows the same learners' observations later in the game. For both models, a Mann-Whitney test shows a significant difference between the two groups along the X-axis (initial: U=141.00, p<0.01, effect size of r=-0.68 at α=0.05; final: U=23.50, p<0.01, effect size of r=0.76 at α=0.05). That is, the two groups start and end their gameplay with differences in their observation patterns.



a. Initial third of observations b. Final third of observations

Fig. 1. Difference models for the HiHi group (blue) and the HiLo group (red).

Table 5. Line Weights for Epistemic Visualization in Fig 1.

	Initial Third of Observation			**Final Third of Observation**		
Transition	**HiHi**	**HiLo**	**Diff**	**HiHi**	**HiLo**	**Diff**
Meas ↔ Noun	0.14	**0.25**	**-0.11**	0.10	**0.35**	**-0.25**
Meas ↔ ContDisc	0.09	**0.18**	**-0.09**	**0.07**	-	**0.07**
Comp ↔ Hypo	0.01	**0.08**	**-0.07**	**0.03**	-	**0.03**
Noun ↔ ContDisc	0.02	**0.08**	**-0.06**	0.03	0.03	-
Comp ↔ NonGam	0.02	**0.08**	**-0.06**	**0.04**	0.03	**0.01**
Meas ↔ Comp	0.19	**0.23**	**-0.04**	0.2	**0.38**	**-0.18**
Meas ↔ Hypo	0.08	**0.11**	**-0.03**	**0.07**	0.02	**0.05**
Noun ↔ Comp	0.04	**0.08**	**-0.03**	-	**0.05**	**-0.05**
Comp ↔ ContDisc	0.01	**0.04**	**-0.03**	**0.05**	-	**0.05**
Exc ↔ Quest	-	-	-	**0.05**	-	**0.05**
Hypo ↔ Quest	**0.02**	0.01	**0.01**	**0.12**	-	**0.12**
Exc ↔ NonGam	**0.05**	-	**0.05**	**0.02**	-	**0.02**
Comp ↔ Exc	**0.05**	-	**0.05**	**0.02**	-	**0.02**
Rep ↔ NonGam	**0.05**	-	**0.05**	**0.03**	-	**0.03**

Noun ↔ Quest	**0.05**	-	**0.05**	0.05	**0.08**	**-0.03**
Rep ↔ ContDisc	**0.05**	-	**0.05**	**0.10**	0.03	**0.07**
Quest ↔ NonGam	**0.06**	-	**0.06**	**0.03**	-	**0.03**
Meas ↔ Exc	**0.08**	-	**0.08**	**0.08**	-	**0.08**
Meas ↔ Quest	**0.11**	0.02	**0.09**	**0.12**	0.06	**0.06**
Meas ↔ NonGam	**0.17**	0.06	**0.11**	0.03	**0.10**	**-0.07**

In the initial third of observations, both HiHi and HiLo groups start with connections between a wide, evenly-distributed range of observational categories. However, HiLo learners increasingly produce 2 specific connections (Meas↔Comp and Meas↔Noun), which rise in the final third of observations. This shift suggests that HiLo learners' engagement pattern becomes more focused and potentially less exploratory.

HiLo learners also show more connections between continuous discussions and other types of content-related observations in the initial third of their observations, as indicated by higher line weights associated with this construct (i.e., Meas↔ContDisc, Noun↔ContDisc, and Comp↔ContDisc). However, these connections decreased in the final third and became more prevalent for the HiHi group. This decline suggests that sustaining these discussions might be an important factor for maintaining a high level of interest among learners started with high astronomy interest.

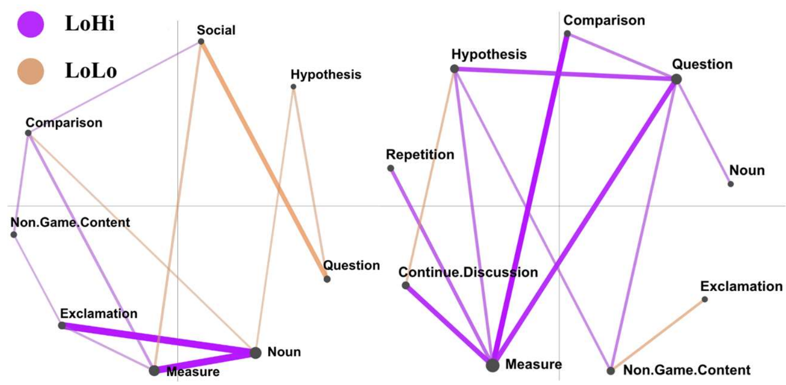
Interestingly, instead of predominantly making meaningful observations, the HiHi group shows higher levels of excitement (i.e., Meas↔Exc, Exc↔NonGam) and curiosity (i.e., Meas↔Quest, Quest↔NonGam) in their observations about the game throughout the study. They also occasionally go off-task to discuss topics unrelated to the game (i.e., Meas↔NonGam, Quest↔NonGam). These findings highlight that sustained interest in astronomy may be linked to a more enthusiastic and inquisitive approach to observations, rather than just making the most meaningful observations.

4.3 Temporal Changes in the Difference Models for LoHi and LoLo Groups

As the difference models in Fig. 2 show, the observation patterns of players in the LoHi and LoLo groups are noticeably different in both the initial and final thirds of observations. For both models, a Mann-Whitney test shows significant differences along the X-axis (initial third: $U=24$, $p=0.01$, effect size of $r=0.64$ at $\alpha=0.05$; final third: $U=109$, $p=0.03$, effect size of $r=-0.51$ at $\alpha=0.05$). Like the HiHi and HiLo groups, the LoHi and LoLo groups start and end their gameplay with different patterns of observations.

During the initial third of the observations, the LoLo group shows more connections involving social communication (i.e., Quest↔Soc; Meas↔Soc), which may indicate a tendency to focus more on interactions with others than on the game content itself (though note exception, Comp↔Soc). Conversely, the LoHi group shows more connections involving exclamations (i.e., Exc↔NonGam, Noun↔Exc) than the LoLo group, implying higher initial levels of excitement and perhaps greater interest.

In the initial set of observations, the LoLo group also shows stronger connections between Noun↔Hypo and Hypo↔Quest than the LoHI, but this difference diminishes in the final third of their observations. This early link between questions and hypothesis generation might suggest that situational interest is not being properly supported.



a. Initial third of observations   b. Final third of observations

Fig. 2. Difference networks for the LoHi group (purple) and the LoLo group (orange).

Table 6. Line Weights for Epistemic Visualization in Fig 2.

	Initial Third of Observation			**Final Third of Observation**		
Transition	**LoHi**	**LoLo**	**Diff**	**LoHi**	**LoLo**	**Diff**
Quest ↔ Soc	-	0.11	**-0.11**	-	-	-
Meas ↔ Soc	0.04	0.10	**-0.06**	-	-	-
Hypo ↔ Quest	-	**0.06**	**-0.06**	**0.08**	-	**0.08**
Noun ↔ Comp	0.04	**0.09**	**-0.05**	0.03	0.03	-
Noun ↔ Hypo	0.02	**0.07**	**-0.05**	-	**0.01**	**0.01**
Meas ↔ Rep	-	**0.04**	**-0.04**	**0.07**	-	**0.07**
Meas ↔ Hypo	0.03	**0.06**	**-0.03**	**0.07**	0.01	**0.06**
Hypo ↔ NonGam	0.01	**0.03**	**-0.02**	**0.05**	-	**0.05**
Noun ↔ Quest	0.08	**0.09**	**-0.01**	**0.07**	0.02	**0.05**
Hypo ↔ ContDisc	-	**0.01**	**-0.01**	**0.04**	-	**0.04**
Quest ↔ NonGam	0.07	0.07	**0**	**0.05**	-	**0.05**
Meas ↔ ContDisc	**0.15**	0.14	**0.01**	**0.13**	0.04	**0.09**
Comp ↔ Quest	**0.02**	0.01	**0.01**	**0.06**	-	**0.06**
Meas ↔ Quest	**0.07**	0.05	**0.02**	**0.18**	0.09	**0.09**
Comp ↔ Soc	**0.04**	-	**0.04**	-	-	-
Exc ↔ NonGam	**0.05**	-	**0.05**	-	**0.05**	**-0.05**
Comp ↔ NonGam	**0.05**	-	**0.05**	-	-	-
Meas ↔ Comp	**0.2**	0.13	**0.07**	**0.23**	0.13	**0.1**
Meas ↔ Noun	**0.31**	0.14	**0.17**	**0.35**	0.33	**0.02**
Noun ↔ Exc	**0.17**	-	**0.17**	0.03	-	**0.03**

We also observed that, in the final third of observations, the LoLo group shows a markedly less diverse interaction network, characterized by fewer connections and lower interaction intensity compared to the LoHi group. Conversely, the LoHi group demonstrates substantial growth in network density and interaction diversity from the initial to the final third of observations. As illustrated in Table 6, only two connections in the LoLo group are stronger than those in the LoHi group during the final third of observations. The observation data also shows that in the final third of observations made by players in the LoLo group, more than 8% of these observations were non-game-related exclamations (e.g., "TECHNOBLADE NEVER DIES!").

These findings indicate that players in the LoHi group not only increase their frequency of interactions but also diversify the types of interactions they engage in. This diversification may be contributing to the increase in their interest levels. In contrast, the LoLo group shows minimal change in observation patterns, and may benefit from interventions that encourage more diverse and interactive observations.

5 Discussion & Conclusion

This study used ENA difference models to explore how learners with varying levels of interest interact with an online learning system early and late in their usage of a Minecraft learning activity. The goal was to identify how qualitative characteristics related to students' text-based observations--may be used to identify learners who are in need of additional support. By identifying the initial observation patterns of learners with low interest, we hope to provide actionable insights for educators and facilitators to design targeted interventions and effective education programs.

Interest was measured using a scale that approximated stage three in Hidi & Reninger's four-stage model. To analyze the observational data, a group of researchers reviewed the data to distill themes inductively, and we then used automated methods with GPT-4 and Python to qualitatively code the large amount of data. To study how these patterns evolved, we divided the data into temporal thirds and compared the initial and final thirds.

ENA results reveal distinct patterns of observation connections among different interest groups. More specifically, learners who maintained or increased their interest (HiHi and LoHi) tended to make a wider range of types of observations and more frequently posed questions, formed hypotheses, and expressed excitement. These learners showed higher levels of exploratory behavior and curiosity-driven interactions. These findings suggest that fostering an environment that encourages curiosity and diverse observations might help in maintaining or even increasing interest among players.

Conversely, learners whose interest declined or stayed low (the HiLo and LoLo groups) showed a narrower range of observations in their later activity, increasingly transitioning between specific observations that involve measurement and comparisons (i.e., Meas↔Comp, Meas↔Noun). This shift may indicate that they are still working to obtain the basic knowledge required to move into more advanced stages of interest, and may require additional scaffolding to facilitate this process. Alternatively, perhaps they were compliant but not genuinely interested, and needed activities better tailored to their interests to capture their attention. Interestingly, the HiHi group showed not only meaningful observations but also moments of excitement (i.e., Meas↔Exc) and curiosity (i.e., Meas↔Quest), with occasional off-task behavior throughout the game. These trends were also observed in the LoHi group. Future research should investigate the specific triggers of these learners' exclamations to better understand what captured their interest. Since this excitement seems to reflect a more developed individual interest, it may indicate that learners are connecting to prior knowledge. Future research should consider the degree to which these explanations are doing so, and should explore the extent to which helping learners with less-developed knowledge and interest might increase the apparent excitement of those learners who start with lower survey scores.

Meanwhile, these findings also suggest that sustained interest in astronomy might be linked to balancing structured scientific observations with spontaneous or off-task moments (e.g., Meas↔NonGam, Exc↔NonGam). It is possible that learners may be regulating their emotions using off-task behavior (e.g., [34]) or that these non-game observations are showing some sort of social function [20]. Future research could interview learners who are showing these observation patterns; fostering their enthusiasm and curiosity could play a significant role in maintaining or improving interest.

That said, observations that appear to be more social in nature do not always correspond with higher interest. For example, the LoLo group had more social interactions during the initial third of their observations (16% vs. 3-5%). This suggests that for learners with consistently low interest, social interactions may be more distraction than a support. However, it is possible that the learners regulate this tendency, as these occurrences disappear uniformly in the final third of the observations. Facilitators might consider interventions that refocus these learners on individual exploration or otherwise encourage them to record and analyze their own observations and insights, but be alert to ensure that eliminating these experiences does not reduce situational interest.

Some limitations impact what conclusions can be drawn from this study. With fewer than 15 participants per group, the generalizability of our findings is limited. The small sample size also limits our ability to account for demographic differences (and the cultural factors that may align with them), which can sometimes be associated with limited opportunities for learners to interact with these concepts. Because these opportunity gaps may affect the knowledge development that is critical to the transition between situational interest and maintained individual interest [16], future research should examine the degree to which these results may replicate in other populations.

Another limitation is the use of measurements related to individual interest—a more advanced level of interest found among learners who have preexisting domain knowledge. Augmenting this analysis with surveys of situational interest would highlight how learner observations change as they acquire such knowledge, which would likely help us to better understand our LoLo and LoHi.

This study uses ENA provides insights into the ways different interaction patterns within a digital learning environment like WHIMC may relate to shifts in learner interest. We identified key patterns that can guide targeted interventions, though future research is still needed to improve our understanding of some of these patterns. Overall, these findings identify indicators that could help us to develop strategies for interventions that can support the development and maintenance of student interest in STEM.

References

1. Ainley, M., Ainley, J.: Early science learning experiences: Triggered and maintained interest. In: Renninger, K. A., Nieswandt, M., Hidi, S. (eds.) Interest in Mathematics and Science Learning, pp. 17–31. American Educational Research Association (2015)
2. Ainley, M.: Being and feeling interested. Emotion in Ed., pp. 147–163. Elsevier (2007)
3. Andre, T., Windschitl, M.: Interest, epistemological belief, and intentional conceptual change. In: Sinatra, G., Pintrich, P. (eds.) Intentional Conceptual Change, pp. 179–203. Routledge (2003)
4. Bressler, D., Bodzin, A., Eagan, B., Tabatabai, S.: Using epistemic network analysis to examine discourse and scientific practice during a collaborative game. J. Sci. Ed. & Tech. 28, pp. 553-566 (2019)
5. Crider, A.: Astronomy education in virtual worlds and virtual reality. In: Astronomy Education, V2: Best Practices for Online Learning Environments. IOP Pub. Ltd. (2020)
6. Deitz, S., Freyman, C.: The state of U.S. science & engineering 2024. Nat. Sci. Foundation, Nat. Sci. Board, Nat. Center for Sci. & Eng. Statistics, Alexandria, VA (2024)
7. Eberbach, C., Crowley, K.: From everyday to scientific observation: How children learn to observe the biologist's world. Review of Educational Research 79(1), pp. 39-68 (2009)
8. Eagan, B., Brohinsky, J., Wang, J., Shaffer, D.: Testing the reliability of inter-rater reliability. In: Proc. 10th Int. Conf. on Learning Analytics & Knowledge, pp. 454-461 (2020)
9. Foster, A., Shah, M., Barany, A., Talafian, H.: Tracing identity exploration trajectories with quantitative ethnographic techniques: A case study. In: Inter. Conf. Quant. Ethnography, pp. 77-88 (2019)
10. Gadbury, M., Lane, H.: Mining for STEM interest behaviors in Minecraft. In: 23rd Artificial Intell. in Ed. Conf., Durham, UK, vol. 13356, pp. 236–239. Springer Int. Pub. (2022)
11. Gadbury, M., Lane, H.: A Bayesian analysis of adolescent STEM interest using Minecraft. In: Int. Conf. on Art. Intell. in Ed., pp. 384-389 (2023)
12. Gee, J.: What video games have to teach us about learning and literacy. Computational Entertain. 1(1), pp. 20-20 (2003)
13. Harackiewicz, J., Smith, J., Priniski, S.: Interest matters: The importance of promoting interest in education. Policy Insights fr. the Behav. & Brain Sciences 3(2), pp. 220–227 (2016)
14. Hewett, K., Zeng, G., Pletcher, B.: The acquisition of 21st-century skills through video games: Minecraft design process models and their web of class roles. Simulation & Gaming 51(3), pp. 336-364 (2020)
15. Hidi, S., Harackiewicz, J. M.: Motivating the academically unmotivated: A critical issue for the 21st century. Rev. of Educational Research 70(2), pp. 151–179 (2000)
16. Hidi, S., Renninger, K. A.: The four-phase model of interest development. Educational Psychologist 41(2), pp. 111–127 (2006)
17. Ishak, S., Din, R., Hasran, U.: Beyond play: Conceptualising the capability of a good digital game to stimulate interest in STEM. Int. J. of Learning, Teaching & Ed. Res. 20(10) (2021)
18. Knogler, M., Harackiewicz, J., Gegenfurtner, A., Lewalter, D.: How situational is situational interest? Investigating the longitudinal structure of situational interest. Contemporary Educational Psychology 43, pp. 39–50 (2015)
19. Krapp, A.: Interest, motivation and learning: An educational-psychological perspective. European Journal of Psychology of Education 14(1), pp. 23–40 (1999)
20. Kreijns, K.: Sociable CSCL environments: Social affordances, sociability, and social presence. Unpublished MA Thesis, Open U., Heerlen, The Netherlands (2004)
21. Langer-Osuna, J. M., Gargroetzi, E., Munson, J., Chavez, R.: Exploring the role of off-task activity on students' collaborative dynamics. J. of Ed. Psych. 112(3), 514 (2020)

22. Lane, H. C., Gadbury, M., Ginger, J., Yi, S., Comins, N., Henhapl, J., Rivera-Rogers, A.: Triggering STEM interest with Minecraft in a hybrid summer camp. Technology, Mind, & Behavior, 3(4) (2022)
23. Lane, H. C., Yi, S., Guerrero, B., Comins, N. F.: Minecraft as a sandbox for STEM interest development: Preliminary results. In: Proc. 25th Int. Conf. on Computers in Ed. (2017)
24. Marquart, C. L., Hinojosa, C., Swiecki, Z., Eagan, B., Shaffer, D. W.: Epistemic network analysis (Ver. 1.7.0) [Software]. Available from http://app.epistemicnetwork.org (2021)
25. National Science & Tech. Council, Office of Science & Tech. Policy: Interagency roadmap to support space-related STEM education & workforce. (2022)
26. Nieswandt, M.: Student affect and conceptual understanding in learning chemistry. J. of Res. in Science Teaching 44(7), pp. 908–937 (2007)
27. OpenAI.: GPT-4-turbo (Version GPT-4-turbo-2024-04-09) [Large language model]. OpenAI. https://www.openai.com (2024)
28. Reeve, J., Jang, H., Hardre, P., Omura, M.: Providing a rationale in an autonomy-supportive way as a strategy to motivate others during an uninteresting activity. Motivation & Emotion 26, pp. 183-207 (2002)
29. Renninger, K. A.: Interest and identity development in instruction: An inductive model. Educational Psychologist 44(2), pp. 105-118 (2009)
30. Renninger, K., Bachrach, J.: Studying triggers for interest and engagement using observational methods. Educational Psychologist 50(1), pp. 58–69 (2015)
31. Renninger, K., Hidi, S., Krapp, A.: The role of interest in learning & development. Routledge. https://doi.org/10.4324/9781315807430 (1992)
32. Renninger, K., Hidi, S.: The power of interest for motivation and engagement. Routledge. https://doi.org/10.4324/9781315771045 (2015)
33. Renninger, K., Nieswandt, M., Hidi, S.: Interest in mathematics and science learning. American Educational Research Association (2015)
34. Sabourin, J., Rowe, J., Mott, B., Lester, J.: When off-task is on-task: The affective role of off-task behavior in narrative-centered learning environments. In: 15th Int. Conf. Art. Intell. in Ed., Auckland, NZ, vol. 15, pp. 534-536 (2011)
35. Sansone, C., Fraughton, T., Zachary, J., Butner, J., Heiner, C.: Self-regulation of motivation when learning online: The importance of who, why and how. Ed. Tech. Res. & Dev. 59, pp. 199-212 (2011)
36. Saricam, U., Yildirim, M.: The effects of digital game-based STEM activities on students' interests in STEM fields and scientific creativity: Minecraft case. Int. J. of Technology in Education & Science 5(2), pp. 166-192 (2021)
37. Scianna, J., Gagnon, D., Knowles, B.: Counting the game: Visualizing changes in play by incorporating game events. In: Inter. Conf. Quant. Ethnography, pp 218–231 (2021)
38. Silvia, P. J.: Exploring the psychology of interest. Oxford University Press (2006)
39. Tablatin, C. L. S., Casano, J. D., Rodrigo, M. M. T.: Using Minecraft to cultivate student interest in STEM. In: Frontiers in Education, vol. 8, p. 1127984. Frontiers Media SA (2023)
40. Thornberg, R., Charmaz, K.: Grounded theory and theoretical coding. The SAGE handbook of qualitative data analysis 5, pp. 153-169 (2014)
41. Yi, S., Gadbury, M., Lane, H.: Identifying and coding STEM interest triggers in a summer camp. In: Proc. 15th Int. Conf. Learning Sciences, pp. 915-916 (2021)
42. Yi, S., Gadbury, M., Lane, H.: Coding and analyzing scientific observations from middle school students in Minecraft. In: 14th Int. Conf. Learning Sciences (2020)
43. Zambrano, A., Barany, A., Ocumpaugh, J., Nasiar, N., Hutt, S., Goslen, A., Mott, B.: Cracking the code of learning gains: Using ordered network analysis to understand the influence of prior knowledge. In: Inter. Conf. Quant. Ethnography, pp. 18-33 (2023)