International Society of the Learning Sciences

Assessing Experimentation: Understanding Implications of Player Choices

Jennifer Scianna, YJ Kim
jscianna@wisc.edu, yj.kim@wisc.edu
University of Wisconsin - Madison

Abstract: Assessing scientific thinking and inquiry skills can be challenging because of the complexity and divergence in student behaviors. Scholars have advocated the use of more open-ended problems and choice for the assessment of scientific inquiry. In this paper, we interrogate an experimentation mechanic in an educational science game to examine challenges that choice introduces to game-based assessment of science inquiry practices. Descriptive analysis of gameplay elucidates the difference between choices to explore and iterating on choices as a sign of struggling to progress.

Introduction

Educational games are popular for engaging students in science inquiry learning. The dynamic engines games are built within allow designers to simulate phenomena for investigation (Kamarainen et al., 2015) while the mechanics allow players to engage in "real" science practices (Boots & Strobel, 2014). Engaging in exploratory, inquiry behaviors allow players to progress within the games, but this may conflict with efforts to break the perception of scientific inquiry as a linear process (e.g. Arlidge et al., 2017). If we look to gameplay as evidence of learner understandings, questions arise around how we might best consider divergent player choices.

In this paper, we consider player practices of experimentation within *Wake, Tales from the Aqualab,* an open-world, science game. In this game, players engage in experimentation to obtain rules, one of the main markers of game progress. We turn to the gameplay interaction log data to explore the nature of experimental differences between players to identify indicators of inquiry skill gaps and begin understanding what additional considerations may be needed when undertaking game-based assessment (GBA) of player inquiry.

Background

Efficiently assessing science inquiry is challenging. Science inquiry performance tasks require students to demonstrate complex reasoning, self-regulation, and planning (Scalise & Clarke-Midura, 2018). These behaviors are only possible when the tasks themselves have sufficiently divergent options for student exploration. Games provide such tasks and by design are constantly assessing players to provide feedback (Shaffer & Gee, 2012). Applying Evidence-Centered Design (ECD) practices to stealth assessment is one mechanism for constructing valid GBA (Kim et al., 2016). ECD has been employed to detect differences in learners playing epistemic games where practices like inquiry are central to the game mechanics (Sweet & Rupp, 2012). Further, analysis of in-game behavior has been used to identify when players are aligned with particular epistemic stances (Martinez-Garza & Clark, 2017).

These analyses are typically supported by evidence from clickstream data which is aggregated into features such as frequencies and time variables and may then be used to correlate with external assessments (Scalise & Clarke-Midura, 2018). Events related to both player and system - and related features – are suggested for prediction of student decision making in game-based systems (Owen & Baker, 2020). Choices, also observed through interaction logs, are also efficient at describing learners. Player choices in educational games have been used to predict mathematics grades (Chi et al., 2014) and identify inquiry strategies (Käser & Schwartz, 2020).

In scientific experimentation, choosing how to experience a phenomenon, with what tools, and how to record such observations is key to demonstrating expertise in scientific inquiry (Duschl & Bybee, 2014). However, choice-based assessments typically are small, independent games designed to feel unlike a traditional assessment and prioritize freedom of choice which is unlike the larger context of educational games (Chin et al., 2016). It is with tension in mind that we seek to explore player choices in an open-world game environment and their connection to assessment with the following research question: *How do divergent player decisions provide evidence for scientific inquiry understanding?*

Method

Context

Wake, Tales from the Aqualab is an educational science game developed by Field Day Lab for players in grades 6-9. Throughout gameplay, players interact with resident scientist NPCs across several ecosystems (kelp forest, arctic, coral reef, bayou, and deep sea), taking on jobs that require them to conduct field observations and experiments, make models, and present evidence in arguments. In this study, we focus on the observation tank mechanic, the first of three experiment mechanics. The Observation Tank is used to observe species:species and species:environment relationships (e.g. "Otters eat urchins," and "Bull kelp produce oxygen").

When players begin an observation experiment, they choose an environment to simulate with the tank's water conditions and species to include in the tank (Figure 1 a, b). Once the experiment begins, the player watches for emoji which signal the presence of an interaction (Figure 1c). The player clicks on the emoji to "document" the presence of the behavior. When the player ends the experiment, the rules they have documented are summarized and added to AQOS, the in-game tablet, for the student to review and use as needed.

Figure 1.
Observation Tank Process: Players Choose the Environment (a) and Species (b) to Observe the Interactions (c).

(a) (b) (c)

Data collection

Interaction log data for *Wake* was obtained through the Open Game Data project (Field Day Lab, 2021). This analysis uses the "Calculated Events" and "Player Feature Data" datasets from September 2023 which are reported to include 4,000 sessions. The "Calculated Events" dataset includes event data from player and game captured within interaction logs as well as several detector features which are embedded alongside the events. The "Player Feature Data" includes aggregate measures such as time played, number of experiments run, etc. for each player over the course of all their sessions. Player data can be tracked across datasets using anonymized pseudonyms which consist of unique adjective-noun pairings.

Data processing and modeling

Player features were used to filter session data for all Wake players from September 2023 (n=1,983) to include players who sufficiently engaged with the game. Included players (n=794) met the criteria either by playing for more than 15 minutes (n=651) or completing more than 3 jobs (n=445); some players met both criteria.

Labeling the Interaction Logs

We used ECD practices to label the interaction data that might indicate player expertise with inquiry in *Wake.* Given the players' decision space in experimentation, we identified the number of species placed in the experiment tank as important for how they were collecting rules. If players add the maximum number of species (4), it may indicate a more exploratory approach to understanding the in-game phenomena whereas conducting an experiment with 2 species may indicate alignment with an interactive inquiry process (See Max Observations and Controlled Experiment in Table 1). We did not include codes for specificity of ecosystem or species.

Table 1
Experimentation Labels and their Definitions.

Label	Definition
Max Observations	The player includes 4 species in the tank at the start of the experiment.
Controlled Experiment	The player includes 2 species in the tank at the start of the experiment.
Missed Something	Player received feedback that they missed a behavior.
Receive Fact	Player obtains a rule that is added to AQOS.
Experiment Task	Player completes one of the tasks related to their current job.

Epistemic network analysis

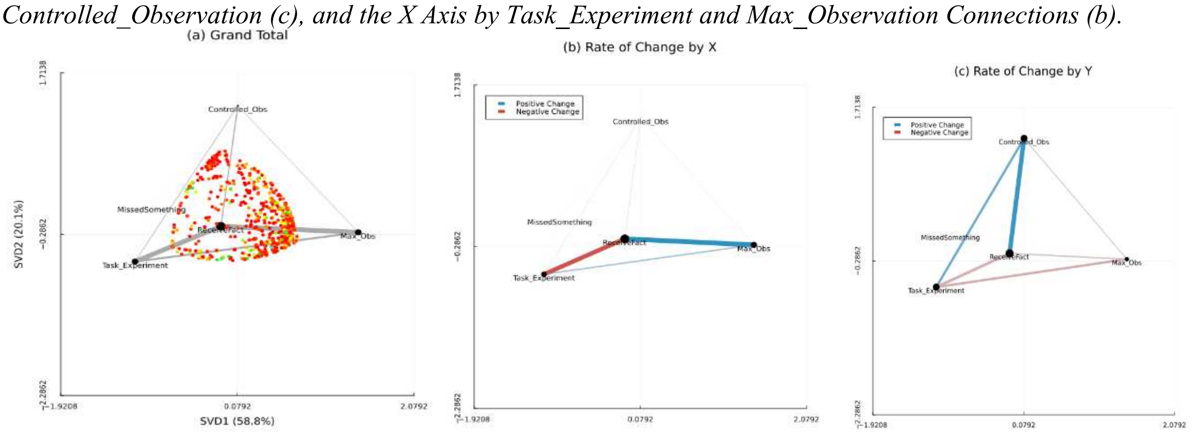
Epistemic Network Analysis (ENA) models connections between codes by consuming a qualitative data table, creating a co-occurrence matrix of codes, and projecting the codes for each unit into space using singular value decomposition to maximize the variance between units. We used jENA, a Julia-based implementation of ENA (Knowles, 2021; Shaffer et al., 2016), to construct our model of player-game interactions. The units for our ENA model segmented players (n=794) into individual units for each job resulting in 2,502 total units. The "conversation" defined in our model included all lines from a particular user allowing crossover between jobs. We used a moving stanza window of 8 to account for the number of intermittent events in the experiment setup. ENA positions units within a multi-dimensional space based on the strength of their connections between codes; a unit with more connections between two codes will gravitate towards the edge that joins them. We used the position of units in our ENA model to identify players whose experiences in the same job were vastly different, and return to the data to better understand the connection differences.

Results

The ENA model for Observation yielded a model with high coregistration values across the first four dimensions (.997, .999, .886, and .918 respectively). We chose to focus on SVD1 and SVD2 as our dimensions as they explained the most variance between players and levels (.86 and .07 respectively).

Figure 2

Results of the ENA Plot: Received_Fact is Centrally Located (a). The y Axis is Defined by Connections to Controlled_Observation (c), and the X Axis by Task_Experiment and Max_Observation Connections (b).



(a) (b) (c)

ReceiveFact is central to the ENA plot, and its connections describe the characteristics of each dimension. Players who lie to the left on SVD1 (x axis in Figure 2) have more connections between *ReceiveFact* and *TaskExperiment*. WobblyTurnip (-0.65, -0.44) begins the job kelp-welcome by asking for help. They run the experiment as suggested by the helper NPC and collect the appropriate rules. This direct approach is characteristic of this location on the plot. In the positive (right) direction, MagneticMink (0.53, -0.41) completes one experiment with all species [*Max_Obs*], but they miss a rule. They try the same experiment again and obtain the missing rule to complete the experiment demonstrating the strong connections between *MaxObservation* and *ReceiveFact*.

Along the y-axis, SVD 2, we see players connecting *ReceiveFact* to *ControlledExperiment* in the positive (upper) direction. RemoteClef is hesitant at the start of the game, asking for help as they start making observations. They begin with *Max_Obs* experiment, but they quickly end it without obtaining a rule and begin to iterate through *Controlled_Experiment.*

Discussion

In this paper, we examine gameplay interaction logs seeking connections between player choices and assessment of inquiry skills. By coding player behavior, we differentiate between when players experience moments of active iteration (MagneticMink) and confusion (RemoteClef). We are able to describe moments where players do not optimize their experiment, leading to repetition. The use of positional data from ENA lays the foundation for potential characterization of playful inquiry and differentiating the influence of the game on player behavior from their own objectives. This study raises questions for formative assessment of inquiry in future implementations. As designers provide digital environments that allow for agency and choice, analytics must be questioned to validate behavior diversity.

There are several limitations to this study which bear noting. This was not a classroom implementation of *Wake*, and thus all descriptions and conclusions are made based on the author's knowledge of the game and the interaction logs themselves. There is the potential for alternative explanations for observations such as lapses in gameplay that may not match our interpretation. Still, this work provides the first steps towards integrating our understandings of choice-based assessment and GBA for the betterment of learner assessment experiences.

References

Arlidge, S. M., Thanukos, A., & Bean, J. R. (2017). Using the Understanding Science Flowchart to Illustrate and Bring Students' Science Stories to Life. *The Bulletin of the Ecological Society of America*, *98*(3), 211–226. https://doi.org/10.1002/bes2.1330

Boots, N. K., & Strobel, J. (2014). Equipping the Designers of the Future: Best Practices of Epistemic Video Game Design. *Games and Culture*, *9*(3), 167–181. https://doi.org/10.1177/1555412014536631

Chi, M., Schwartz, D. L., Blair, K. P., & Chin, D. B. (2014). *Choice-based Assessment: Can Choices Made in Digital Games Predict 6th-Grade Students' Math Test Scores?*

Chin, D. B., Blair, K. P., & Schwartz, D. L. (2016). Got Game? A Choice-Based Learning Assessment of Data Literacy and Visualization Skills. *Technology, Knowledge and Learning*, *21*(2), 195–210. https://doi.org/10.1007/s10758-016-9279-7

Duschl, R. A., & Bybee, R. W. (2014). Planning and carrying out investigations: An entry to learning and to teacher professional development around NGSS science and engineering practices. *International Journal of STEM Education*, *1*(1), 12. https://doi.org/10.1186/s40594-014-0012-6

Field Day Lab. (2021). *Open Game Data* [Python]. https://github.com/opengamedata/opengamedata-core

Kamarainen, A. M., Metcalf, S., Grotzer, T., & Dede, C. (2015). Exploring Ecosystems from the Inside: How Immersive Multi-user Virtual Environments Can Support Development of Epistemologically Grounded Modeling Practices in Ecosystem Science Instruction. *Journal of Science Education and Technology*, *24*(2–3), 148–167. https://doi.org/10.1007/s10956-014-9531-7

Käser, T., & Schwartz, D. L. (2020). Modeling and Analyzing Inquiry Strategies in Open-Ended Learning Environments. *International Journal of Artificial Intelligence in Education*, *30*(3), 504–535. https://doi.org/10.1007/s40593-020-00199-y

Kim, Y. J., Almond, R. G., & Shute, V. J. (2016). Applying Evidence-Centered Design for the Development of Game-Based Assessments in Physics Playground. *International Journal of Testing*, *16*(2), 142–163. https://doi.org/10.1080/15305058.2015.1108322

Klopfer, E., Osterweil, S., & Salen, K. (2009). Moving learning games forward. *Cambridge, MA: The Education Arcade*.

Knowles, M. (2021). *EpistemicNetworkAnalysis.jl* (0.1.0) [Julia].

Martinez-Garza, M. M., & Clark, D. B. (2017). Investigating Epistemic Stances in Game Play with Data Mining: *International Journal of Gaming and Computer-Mediated Simulations*, *9*(3), 1–40. https://doi.org/10.4018/IJGCMS.2017070101

Owen, V. E., & Baker, R. S. (2020). Fueling Prediction of Player Decisions: Foundations of Feature Engineering for Optimized Behavior Modeling in Serious Games. *Technology, Knowledge and Learning*, *25*(2), 225–250. https://doi.org/10.1007/s10758-018-9393-9

Scalise, K., & Clarke-Midura, J. (2018). The many faces of scientific inquiry: Effectively measuring what students do and not only what they say. *Journal of Research in Science Teaching*, *55*(10), 1469–1496. https://doi.org/10.1002/tea.21464

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. *Journal of Learning Analytics*, *3*(3), Article 3. https://doi.org/10.18608/jla.2016.33.3

Shaffer, D. W., & Gee, J. P. (2012). The right kind of GATE: Computer games and the future of assessment. *Technology-Based Assessments for 21st Century Skills: Theoretical and Practical Implications from Modern Research*, 211–228.

Sweet, S. J., & Rupp, A. A. (2012). *Using the ECD Framework to Support Evidentiary Reasoning in the Context of a Simulation Study for Detecting Learner Differences in Epistemic Games*. *4*, 41.

Acknowledgments

This material is based upon work supported by the National Science Foundation under grants DRL-1907384, DRL-1907398, and DRL-1907437. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.