Challenges to Multimodal Data Set Collection in Games Based Learning Environments

Aybuke G Turker[1], Jennifer Dalsen[1], Matthew Berland[1] and Constance Steinkuehler[2]

[1] University of Wisconsin, Madison, WI 53703, USA
[2] University of California, Irvine, CA 92697, USA
aturker@wisc.edu

Abstract. Our research team examined how middle school students (n=62) used a games-based (GB) activity to learn about virology. This research captured qualitative and quantitative data streams, including talk audio, pre/post assessments, video recordings, and clickstream data. The goal of our study was largely methodological–to combine qualitative and quantitative data channels in new ways to help us make sense of the data for games-based learning (GBL) environments. To this end, our paper describes the design process and the ongoing methodological challenges.

Keywords: Multimodal Dataset, Game Based Learning, Challenges in Data Gathering

1 Introduction

Games provide students with an alternative way to think critically about academic material [2] and connect with each other [3]. Prior research on games-based learning includes history, physics, genetics, and chemistry [2,5]. From games, players can actively explore scientific concepts, make hypotheses, and investigate information within a larger community [4]. Our research team designed a games-based intervention for middle school students that involved gameplay, group collaboration, and the coordination of artifacts. The purpose of this study was develop novel tools to capture and understand connections between qualitative and quantitative data channels in order to create meaningful inferences about STEM learning in games. Our research team's previous reports on situating big data investigated the study design, data collection process [6, 9], the use of heterogeneous data sets [8] and scientific gains [7]. However, we have not critically examined the analysis of multimodal data. Given the numerous data channels captured and million plus clickstream data points obtained, this paper seeks to better outline the analysis landscape of our project, its ongoing challenges during the data analysis phase, and our current strategies for tackling them.

2 Methodology

2.1 Game

Our team used *Virulent*, a tablet-based game where users role-play as the fictional Raven Virus. Each level becomes increasingly complicated as the immune system is alerted to the virus and begins to fight back against it.

Fig. 1. Screenshots of Virulent.

2.2 Curriculum

Participants roleplayed scientists recruited by the Center for Disease Control (CDC) to assist in the elimination of a fictional virus. Each day, participants received additional information about the virus via a Skype video. Participants were divided into groups of 3-4. On Day 1, participants received a tablet or “digiscope” to investigate virus behavior. At the end of the day, each team wrote a letter to the CDC with recommendations on stopping the virus from spreading. On Day 2, participants continued their investigation and models of virus and immune system behavior. Participants continued model construction and gameplay during the following day. This time, each team also produced a video presentation to support their findings. On Day 4, each group presented information to their respective cohorts. Day 5 was a cohort debate, where all team were given fictional articles and asked to determine the best way to stall the Raven Virus. The available options were: vaccine, RNA inhibitor and a mitochondrial inhibitor. At the end of the day, all teams came together and voted on their preferred method.

2.3 Participants

We recruited 62 participants among middle school students from multiple locations across the Midwestern area of the United States. Each session took place outside of the school day and was voluntary. In total, three separate cohorts participated in our data collection: a games camp, an after school club, and a local Boys & Girls club. Participants who chose to participate in our session received monetary compensation for their time.

3 Data Types and Challenges During Data Collection

Multiple data channels were collected over the course of the event. Figure 2 provides a description of these channels along with ongoing challenges.

3.1 Telemetry

Virulent is part of a cross-platform and cross-game-title data framework called Assessment Data Aggregator for Gaming Environments (ADAGE). ADAGE articulates content model of games through a metadata tagging process. This allows our research team to data mine play patterns and other key events in gameplay [1]. Player actions are downloadable onto CSV or JSON files, and tagged with attributes related to player action. Possible attributes include levels completed, unit spawning, time spent and how many times they attempted a level.

Challenges of Telemetry. Although the collection of clickstream data in many different categories (e.g. level completion, almanac use, and player movement during a level) is relatively easy, there are technical and data analysis related challenges. Telemetry data is often time too much to look at, it is messy and hard to analyze. Our research team needed to determine which variables were most important to investigate during our preliminary round of analysis and how these specific categories helped better inform us of participant learning. From technical challenge point of view, every participant had a unique QR code to track and monitor player action. However, QR codes cannot always account for two players sharing a device or technical issues with the login process.

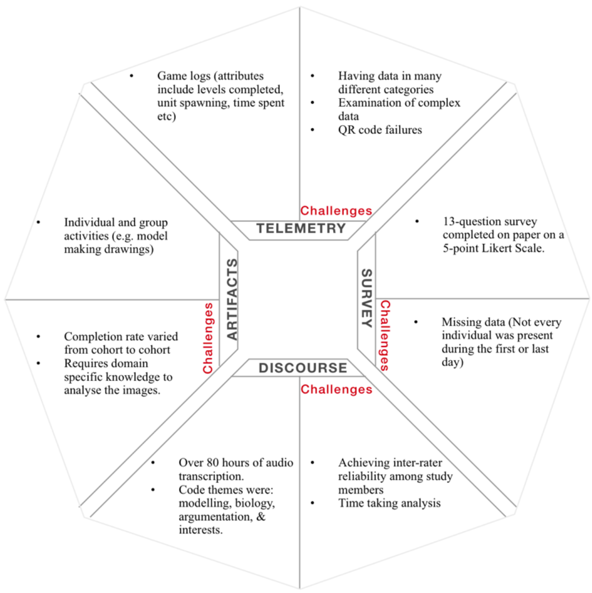

Fig. 2. Data channels and the challenges.

3.2 Discourse

All participants received a USB recorder. This allowed us to track discourse across the room and allow us to decide which audio recording in the group was the clearest. After the study, all audio was sent to a third party transcription service. Transcriptions were reviewed by our research team for errors and imputed timestamps into each transcription. The audio was later uploaded onto MAXQDA software and coded for the following themes: argumentation, modeling, biology and interests. Each code theme had 2 or 3 raters. We randomly pulled data to code until we had consistent inter-rater reliability among raters. The research team coded roughly 1600 turns of talk to reach the threshold agreement. Codes were compared using Fleiss's kappa.

Table 1. Interrater agreement among code theme.

Code Theme	Example	Font size and style
Biology	93%	0.83
Argument	94%	0.93
Model Building	96%	0.71
Interest	98%	0.78

Challenges of Discourse. We examined the following categories during our first round of analysis: biology, interest, argumentation, and modeling. An initial challenge to analyzing discourse was achieving interrater reliability among study members. This varied in difficulty as some discourse categories were harder to interpret. For example, biology group checked correct, incorrect or unsure use of the biology terms. In contrast, argumentation required subjects to gain a mutual understanding of claims and revisions within a complex string of conversations. As a result, each coding team had significantly different challenges in the amount of time and research necessary to create a key that everyone could use. The other challenge currently facing our team is discourse analysis. While our team is interested in using different educational data mining techniques to analyze talk data, we are hesitant about losing the richness of the qualitative data.

3.3 Artifacts

Our research team collected artifacts from each day's session, including: letters to the CDC, scratch paper, and group worksheets. We also photographed and documented model development across each day. All artifacts were scanned and uploaded to a secure university server.

Challenges of Artifacts. While artifacts provide our research team with additional information about individual participants and groups, completion rate varied from cohort to cohort. Not every participant filled out a worksheet during the allotted time available. In some cases, artifacts were missing from participants due to unexpected absences on a particular day. Not every model had the same number of photographs from day to day, and thus our research team's view of models in progress varied from cohort to cohort. A further goal to our participation in this workshop is finding ways to code for information within these artifacts. Further, to find a meaningful way to connect this to both clickstream telemetry and discourse.

4 Challenges of Merging Data Streams

Our team's primary challenge to data analysis is linking discourse, telemetry data and demographic information under a single user interface. Individual data streams are complex and lengthy due to the massive amount of information handled. A description of some challenges explained below.

4.1 Linking Talk Audio to Game Play Telemetry Data

The ability to trace an individual's data stream was surprisingly difficult to navigate for several reasons. First, audio transcriptions were not tagged with User IDs in order to maintain anonymity during the transcription phase. Instead, our team had placeholder names (e.g. Player 1, Player 2). This made analysis difficult to complete on the individual level due to the amount of time that passed. Group facilitators helped us identify speakers. However, this process took additional time and not every facilitator was available to assist. A second challenge to creating a link between gameplay and audio was making sure every player ID matched across all data channels. While a master log existed for pre/post assessment, demographic data, and player profiles, there was no central location to merge talkdata with gameplay. Thus, many of these links were made together through systematically merging separate files based on our research question. A third challenge, while logging each interaction with the system can be very helpful, it had its own data challenges. It took time to get familiar with all the different variables and then come up with ways to do data mining to make sense of the data. Last but not least, ways to quantify discourse data was compelling as we did not want to lose the richness of the qualitative data in expense of quantification.

5 Discussion

The ability to situate big data is a complex process. We recommend that researchers hoping to harness qualitative and quantitative data channels create master logs to track, monitor, and analyze student data. These logs should attempt to capture data channels on the individual, group, and cohort level. In doing this, researchers can easily shift their focus between and across these various levels during analysis. We also recommend the integration of both qualitative and quantitative researchers when examining big data. Our own research team was comprised of curriculum designers, educators, and graduate students majoring in psychology, digital media, engineering, and educational leadership. The interdisciplinary nature of our team not only provided us with a chance to deeply consider changes in the study's design but also contemplate various strategies for data analysis. The multiple perspectives offered by our team members, ranging from discourse analysis to educational data mining techniques, served as a solid foundation for our work.

6 Future Work

We continue to examine clickstream data, discourse, and artifacts to create a more complete picture of the learning environment. Currently, our research team work on various research projects, some of which include visualizing human coded data for exploratory data analysis purposes, understanding productive failure in gameplay, and the co-construction of scientific arguments within a gaming environment.

7 Conclusion

While understanding the gameplay from only telemetry data can be informative, it is not the full picture. It is always hard to answer why questions or do further investigation with one data channel. Analyzing more data channels is needed to answer deeper research questions. Our study is an attempt to capture possible data sources that can happen in a game-based learning environment to explain student behaviors and learning holistically. With this work, we presented the different data channels, design process and the ongoing challenges currently facing our team as we analyze our multimodal dataset.

8 Acknowledgments

The material is based upon work supported by the National Science Foundation under Grant Number (NSF 1418352). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Owen, V. E., & Halverson, R. (2014). ADAGE (Assessment Data Aggregator for Game Environments): A click-stream data framework for assessment of learning in play. In Proceedings of the 9th Annual Games+ Learning+Society Conference, ETC Press (Pittsburgh, PA, USA, 2013
2. Squire, K. (2011). Video Games and Learning: Teaching and Participatory Culture in the Digital Age. Technology, Education--Connections (the TEC Series). Teachers College Press. 1234 Amsterdam Avenue, New York, NY 10027.
3. Steinkuehler, C., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. Journal of Science Education and Technology, 17(6), 530-543.
4. Gee, J.P. (2003). What video games have to teach us about learning and literacy. Computers in Education (CIE), I. 20-20.

5. Squire, K., & Barab, S. (2004, June). Replaying history: Engaging urban underserved students in learning world history through computer simulation games. In *Proceedings of the 6th international conference on Learning sciences* (pp. 505-512). International Society of the Learning Sciences.
6. Anderson, C., Dalsen, J., Stenerson, M., Robinson, J., Binzak, J., Wielgus, L., Ebert, S., Azari, D., Bloker, L., Scaico, P., Bohanen, R., Squire, K., & Steinkuehler, C. (2015). *Game-a-Palooza*. 11th Annual Games + Learning + Society Conference, Madison, WI. GLS 11 Conference Proceedings. MIT Press: Cambridge.
7. Anderson, C. G., Binzak, J. V., Dalsen, J., Saucerman, J., Jordan-Douglass, A., Kumar, V., Turker, A., Scaico, P., Scaico, A., Berland, M., Squire, K., & Steinkuehler, C. (2016). Situating Deep Multimodal Data on Game-Based STEM Learning. Proceedings from ICLS '16: *12th International Conference of the Learning Sciences*. Republic of Singapore.
8. Wielgus, L., Anderson, C., Binzak, J., Dalsen, J., Scalio, P., Azari, D., Berland, M., Squire, K. & Steinkueher, C. (2015). Situating big data across heterogeneous data sets of game data exhaust, class assessment measures and student talk. GLS 11 Conference Proceedings. MIT Press: Cambridge.
9. Dalsen, J., Anderson, C., Squire, K. & Steinkuehler, C. (in press). Situating Big Data. In S. Slota & M. Young (Eds). Exploding the Castle: rethinking how video games & game mechanics can shape the future of education.