
Combining Learning Analytics and Qualitative Analysis for the Exploration of Open-Ended Learning Environments

Yipu Zheng, yipu.zheng@tc.columbia.edu
Teachers College, Columbia University, New York, USA

Paulo Blikstein, paulob@tc.columbia.edu
Teachers College, Columbia University, New York, USA

Nathan Holbert, holbert@tc.columbia.edu
Teachers College, Columbia University, New York, USA

Constructionist learning environments require learners to build projects, inhabit microworlds, create objects and code. Assessing and tracking the construction of these artifacts has always been a challenge for designers and educators. This work provides methodological input on how qualitative analysis and learning analytics based on log data can build upon each other to explore students' behaviors in an open-ended learning environment. Findings from both the data and the structures of the analysis process are reported. The work could suggest procedures and workflows for constructionist researchers to investigate new ways to examine learning in complex, open-ended environments.

Qualitative research methodologies such as interviews and field observations provide deep and thorough descriptions of complex learning phenomena. However, qualitative data collection and analysis methods are time consuming (Atieno, 2009). Recent research has suggested the potential of augmenting qualitative methods with learning analytics (Berland, Baker & Blikstein, 2014; Fields et al., 2016; Sherin et al., 2018; Worsley, 2018).

In this work we examine how qualitative data analysis and learning analytics can be used to make sense of complex process data in a complementary and iterative manner, where insights from qualitative analysis and learning analytics are put into conversation. In particular, we explore ways to gain deeper insights on students' gameplay behaviors in an online constructionist game (*Beats Empire*) through iteratively analyzing transcripts of interviews and think-aloud protocols, field notes of the social context, and log data of students' gameplay.

Beats Empire (Holbert et al., 2019) is an open-ended constructionist music management role-play game where players work to create a successful music studio by making decisions about what artists to sign, what songs to record, and where to promote their music by engaging with data about listener's musical preferences in a fictional US city. Thirty-five 7th graders in an urban middle school in the Northeastern US were engaged in one hour of gameplay, which was captured by a logging system. Concurrently with their classmates, seven students were interviewed by a researcher and played the game using a think-aloud protocol.

Through this exploratory analysis, we considered the complementarity of qualitative and log data analytics to be highly problem-driven. Depending on the research question, the analysis went back and forth spontaneously between qualitative analysis and log data analytics to understand the behavior of interest.

We started by taking an expansive view of the qualitative data, looking through field notes and interview transcripts, highlighting interesting gameplay behaviors or generating questions about potential game play patterns. For example, one anecdotal observation we generated was that some students seemed to spend more time than others cycling through auto-generated titles for soon-to-be-recorded songs, even though the choice of song title does not lead to any in-game rewards. We were interested in verifying the prevalence of this observation across subjects and

exploring the reasons behind the actions. Thus, we processed and analyzed log data to help answer the question. To our surprise, out of 35 students, 24 students generated more than 50 song titles per song release. Two students even generated more than 175 song titles for some songs. (Noted that a few students initially thought that the song titles they had previously seen would cycle back around, so they kept going through the list to find the titles they liked.) This phenomenon led us to question why the song title generation feature was so prominent in the students' gameplay experience. We also wondered if the feature facilitated or hindered the students from exploring other meaningful features in the game that would better cultivate students' understanding and applications of data. Thus, we dived into the qualitative data again and compared interview transcripts of think-aloud players who generated about 15 titles per songs with those of who generated more than 50 titles per songs. We found students that generated few song titles engaged with song creation using mechanics that are explicitly defined by the game, while the others related their real-life music preference to the game. The qualitative analysis also suggested that some students who spent a lot of time scrolling through the song titles were actually looking for song titles that were suitable to the most popular moods indicated by the bar graphs in the data screen, so they were engaged with data in a different way in their decision making. In other words, an otherwise unnoticed aspect of gameplay (choosing song titles) ended up revealing a very meaningful pattern relating to students' experiences.

Furthermore, our work suggests five major roles log data analytics might complement qualitative data analysis: 1) capture transient actions, 2) identify possible trends in the sample population, 3) quantify a qualitative phenomenon using combinations of log data attributes and verify a qualitative phenomenon across a larger population, 4) identify particular chunks of qualitative data for further analysis, and 5) look for relationships among different phenomena. We believe that this work can also be applied to other types of microworlds, computational environments and constructionist games, enabling researchers to gain a deeper understanding into patterns of learning and interaction.

References

- Atieno, O. P. (2009). An Analysis of the Strengths and Limitations of Qualitative and Quantitative Research Paradigms. *Problems of Education in the 21st Century*, 13(1), 13-38.
- Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational Data Mining and Learning Analytics: Applications to Constructionist Research. *Technology, Knowledge and Learning*, 19(1-2), 205-220.
- Fields, D. A., Quirke, L., Amely, J., Maughan, J. (2016). Combining Big Data and Thick Data Analyses for Understanding Youth Learning Trajectories in a Summer Coding Camp. *Proceedings of the 47th ACM technical symposium on computing science education*. ACM, 2016.
- Holbert, N., Berland, M., Rutstein, D., Basu, S., Disalvo, E., Rochelle, J., Villeroy, M., Kumar, V. (2019). Designing Constructionist Formative Assessment Games. The Annual Meeting of the American Educational Research (AERA) 2019, Toronto, Canada.
- Sherin, B., Kersting, N., & Berland, M. (2018). Learning Analytics in Support of Qualitative Analysis. In Kay, J. and Luckin, R. (Eds.) *Rethinking Learning in the Digital Age: Making the Learning Sciences Count, 13th International Conference of the Learning Sciences (ICLS) 2018, Volume 1*. London, UK: International Society of the Learning Sciences.
- Worsley, M. (2018). Multimodal Learning Analytics for the Qualitative Researcher. In Kay, J. and Luckin, R. (Eds.) *Rethinking Learning in the Digital Age: Making the Learning Sciences Count, 13th International Conference of the Learning Sciences (ICLS) 2018, Volume 1*. London, UK: International Society of the Learning Sciences.