Mixed Methods to Study Games and Learning

Constance Steinkuehler, Elizabeth King, Crystle Martin, Yoonsin Oh, Sarah Chu, Caroline Williams, Amanda Ochsner, Shannon Harris, V. Elizabeth Owen, Gabriella Anton, & Jonathan Elmergreen
University of Wisconsin – Madison, 225 N. Mills, Madison, WI
Email: constancies@gmail.com, emking29@gmail.com, crystle.martin@gmail.com, yoonsinoh@gmail.com, sarahnchu@gmail.com, caro.williams@gmail.com, amanda.ochsner@gmail.com, shannon.linehanharris@gmail.com, v.elizabeth.owen@gmail.com, gabby.anton@gmail.com, jelmergreen@gmail.com

Abstract

This paper outlines a mixed methods workshop describing how the methods of quantifying qualitative codes, repertory grid analysis, matched sample comparison, triangulation, and discourse analysis can be used in various combinations to create a more detailed analysis. These methods are presented in the context of researching online games.

Introduction

For the last six years, the Pop.Cosmo research team lead by Constance Steinkuehler at the University of Wisconsin-Madison has been conducting mixed methods research to examine learning in the context of online gameplay. To date, this work has included data collection across multiple *contexts* (from naturalistic studies of gameplay by anonymous fans within the virtual game world to controlled studies conducted within institutionally-affiliated lab environments), drawing on a variety of *modal data types* (video data, multimodal fieldnotes, in-game chatlogs, structured and unstructured interviews, forum data, web pages, structured lab activities, and even surveys), and has used a range of *analytic means* (content coding, discourse analysis, pre-post comparisons, longitudinal analyses, quantitative data mining, and comparisons of means).

In this workshop, we explored the use of mixed methods research in game-related studies of learning as a means for reaching broad audiences. Using our own data corpus as fodder for discussion, we explored a range of methods for studying learning and how those methods can be used in combination to build a persuasive case for (or against) learning related to games. The workshop began with an introduction to mixed methods and a description of the dataset from which the examples presented in the workshop were drawn. Mixed methods are a combination of qualitative and quantitative research techniques. It is a "third wave" research movement building on the idea of pragmatism (Johnson & Onwuegbuzie, 2004). The logic of inquiry that underlies mixed methods research includes the use of induction (discovering patterns), deduction (hypothesis testing), and abduction (uncovering best explanations for results). Using mixed methods offers a variety of benefits as an approach to research (Greene & Caracelli, 1997; Johnson & Onwuegbuzie, 2004). Triangulation of data through the corroboration of results increases the validity of findings. Mixed methods data gives completeness to an analysis, resulting in a more comprehensive account of phenomena. It can offer development of a research trajectory illuminating the next steps in the line of inquiry or offset the weaknesses of a single method through the use of several analytic strategies. While one method specifies the outcomes of the study, a second can make clear the process behind those outcomes. A mixed methods approach can answer related questions as well as the ones asked of the study. Or, while one method provides context for findings, another might enable generalizability. One method might

allow illustration of the data while another provides depth. In sum, by combining methods, you can enhance analysis of your data by augmenting that analysis with other approaches.

The context for the data used as examples in this workshop originated from the Games+Learning+Society (GLS) Casual Learning Lab. The lab ran for two years; 2008 was the pilot year with 9 participants, and in 2009, the formal program ran, which had 22 participants. The participants of the lab were males age 13-18 from nearby rural areas. The lab had one PI, 8 doctoral students, and 6 undergraduates in terms of staff, allowing for a very high ratio of participants to researchers. The lab met monthly in a face-to-face setting in a game lab on campus, as well as online during regular collaborative gaming within *World of Warcraft.* The participants and researchers all joined one guild so that they could easily interact and "lifeguard," that is, take participant observations and function as a resource in the game. Communication also took place asynchronously on forums created on their guild website. In total, the lab collected 8 months of ethnographic data as well as data from 4 studies that focused on games vs. school targeting: reading, information literacy, social reasoning, and epistemological beliefs. The main objective of the lab was to resource and trace individual learning trajectories and interests throughout the duration of participation. The dataset consisted of 454 photos, 66 forum posts, 100+ hours of video, and 2,506 pages of in-game chatlogs. This corpus was then coded by 8 analysts with an a priori content coding scheme consisting of 11 themes and 48 codes. Interrater agreement was 98%.

The workshop was structured so that participants divided into groups. Each group was given a card with a game context (see Figure 1) and a phenomenon; for example: "apprenticeship." With the context and phenomenon in mind, the group created a research question. Then each group member chose to go to two of five hands-on mini-workshops focusing on one method per workshop. The methods that were available to choose from included: quantifying qualitative codes, repertory grid analysis, triangulation, discourse analysis, and matched samples comparison (each is described below). After the mini-workshops, the groups reconvened and discussed which combination of methods they felt were best suited to help them answer their research question, what data sources that combination of methods would require, and assessed its feasibility and drawbacks. The workshop ended with the groups sharing their research question, mixed methods, data collection strategy, and anticipated challenges with the group at large.

Portal 2
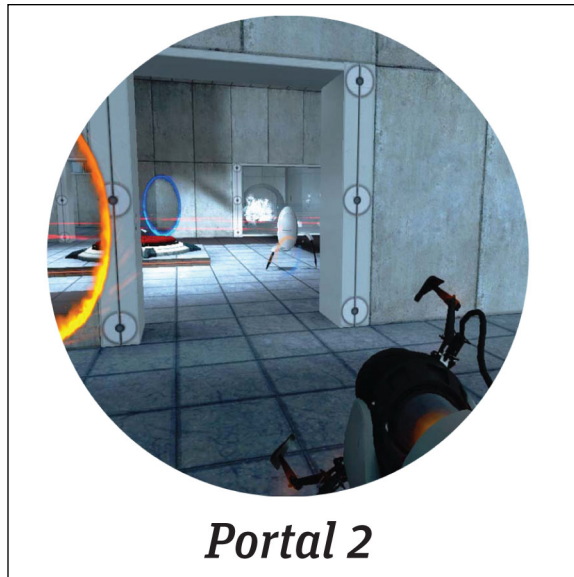
Figure 1: **Example of a game card**

Quantifying Qualitative Codes

Quantifying qualitative data is a method used to analyze coded data (Chi, 1997). During the coding process, an a priori scheme is applied to a set of data to illuminate patterns of themes. Schemes may be generated from the ground up to capture unique occurrences within the data. Codes are counted and examined to find patterns. Graphical representations of the data can be created from these findings. While this method allows for the examination of aggregated patterns of qualitative data, events or items that occur only once can be missed even though they might be significant.

Researchers in the Pop.Cosmo lab frequently utilize this method to understand general patterns occurring within the ethnography as well as specific patterns within a smaller study. Researchers begin by looking for general codes that apply to the data. Quantifying qualitative codes was utilized in the lab data to understand participants' "information literacy" or how adolescents use the web to find information in one of the smaller studies within the lab. The data set, containing audio and video of the guys' progress, is transcribed with actions and speech. Researchers then collaboratively coded one small excerpt of the video corpus with the scheme in order to understand the nuances of the application of codes. During this initial process, generated codes can be easily added to the scheme to more fully represent the events and behaviors occurring within the data. While missing uncommon but important events remain an issue for this method, it can be checked by a general "emergent" code used to highlight any seemingly significant occurrence for post hoc review. After collaborative coding for "calibration" is completed, each researcher individually codes a portion of the data set (our norm is 10%) in order to obtain an inter-rater reliability score of (typically 90% or better is required). If reached, the researchers may then carve up the remaining corpus and code individually with the assumption that they are each coding consistently. The patterns obtained from counting codes allow the understanding and quantification of the general processes occurring within qualitative data. These patterns can then be compared across variables of interest within the study to determine differences of events and behaviors. While there are pitfalls to using this method, it provides a vital picture into the aggregated patterns within data.

Repertory Grid Analysis

The repertory grid method is a form of structured interview that helps elicit the constructs an individual uses to make sense of their world. Generally, the interview focuses on the interviewee's constructs or views around a particular topic. To begin, the researcher sets the topic of the interview, which is some general category. Next, she asks the interviewee to think of as many "elements" as possible within that category that seem relevant to the conversation. Once the interviewee is finished listing or simply repeating elements, the researcher then presents those elements back to the interviewee in random sets of triads asking which of the three elements is the least like the other two and why. The descriptive words used to explain what makes one element distinct from the remaining two are recorded and later become the "constructs" of meaning that represent the participant's world of meaning. Since these are provided in terms of one thing being unlike another, they are always expressed as contrasting concepts or terms. After the researcher is satisfied with the number of constructs provided by the participant (either all triads are used or the participant reaches the point of repetition), s/he has the participant rate each of the initial elements in terms of the entire set of elicited constructs. This rating is typically done on a 5 or 7-point scale, from which the researcher can create a matrix to which can be applied various statistical analyses that cluster the elements and constructs (e.g., RepGrid).

The primary strength of the repertory grid analysis method is that it allows the researcher to elicit the participant's constructs of meaning without supplying them terms or priming certain kinds of responses or language. Additionally, though the data collected by this method are qualitative in nature, the fact that participants rate the elements on a numerical scale makes it possible for the researcher to run various statistical tests to examine clustering patterns. However, since each interviewee provides their own unique set of constructs, this method is less readily used to examine aggregated patterns across entire groups of research subjects. Thus, the strength of this method is to enable the researcher to discern subtle differences between individual meaning making rather than to generalize across populations.

As part of the Pop.Cosmo research team, we used the repertory grid analysis method to elicit the constructs that *World of Warcraft* players form about other players and their interests. Toward this end, we selected six participants and staff from our casual learning lab and had them recall other participants in the lab ("elements") and discern what their play styles (Bartle, 1996) and interests were ("constructs"). In the coming months, we will be using these "interest" constructs to mine the ethnographic data to see if specific play styles and interest led to differentiated learning outcomes.

Matched Sample Comparisons

Matched pair comparisons such as *t*-tests on pre-/posttest performance tasks or attitudes surveys allow comparison of the means (averages) of the two sets of related scores (Trochim, 2006). When analyzing data from isomorphic tasks situated across two points in time or two contexts, *t*-test results gave us a sense of direction for how to differences in the two measures. Such differences can then be compared to related qualitative data through a process called "triangulation" (see discussion below).

The guiding research question for our matched sample comparison study was related to a social reasoning task: How do our participants reason about social/ethical dilemmas within versus outside of online games like *World of Warcraft*? First we borrowed an out-of-game

instrument called the Defining Issues Test, or DIT-2 (Rest, Narvaez, Bebeau & Thoma, 1999) to measure reasoning in one context. Next, we adapted the DIT-2 scenario to create an in-game version with the same core ethical and social issues at stake (yet set in the online game's virtual world). Participants completed both measures with order of instrument counterbalanced to mitigate any ordering effects (i.e. one random half of the participants took the games measure first, the other random half took the out-of-game measure first) so we could compare the two matched samples of performance scores to examine similarities and differences based on contrasting contexts. Based on the statistical results, participants were more willing to abdicate to an authority and suspend personal rights in the context of the virtual world than the real one. We then triangulated these findings with our observational data (see next section).

Triangulation

Another mixed methods technique is the triangulation of quantitative (e.g., survey or performance test) and qualitative (e.g., observation) data (Jick, 1979). One strength of this approach is that it helps ensure that the aggregate quantitative patterns you find are understood in sufficient depth (qualitative data). One potential complication of this strategy is that surveys and their analyses typically assume that people have stable attitudes or dispositions or beliefs (factors) that endure across time and context, which may not always be true.

For example, in the Pop.Cosmo lab, quantitative data to measure individuals' attitudes, dispositions, and beliefs is first analyzed using statistical procedures (e.g., comparison of means, cluster analysis) to group together items in order to identify any latent constructs that might explain patterns in the survey item responses. We then use qualitative data taken from everyday activities involving the same participants (e.g. game transcripts) to triangulate, explain, or otherwise augment the quantitative data analysis to build a more complete and valid picture of what's going on. Data used for the workshop is an excellent case on point: Here, we gave participants two isomorphic surveys (63 items each) designed to measure their epistemological beliefs specific to game versus school contexts. Based on this comparison, we found that participants were significantly more likely to hold naïve beliefs about the nature of knowledge and learning in the context of school versus games. Specifically, they were more likely to believe that "success is unrelated to hard work" and "you cannot learn how to learn" in relation to school than to games. We then searched our coded ethnographic data related to these two themes for both confirming and disconfirming evidence. Once these significant differences were found, we then searched for confirming and disconfirming evidence within our longitudinal ethnographic data where we indeed found the same pattern, corroborating and strengthening our final claims.

Discourse Analysis

Discourse analysis (Gee, 1996) is a method of closely examining language in order to connect the micro-dynamics of language-in-use with the macro-dynamics of culture and society. Discourse analysis approaches language as action and affiliation, attending to both the content of what is said as well as its form (i.e. the way in which it's said, which is also part of its content). Thus, by unveiling the work that is done tacitly through language in social interaction, discourse analysis functions as a particularly powerful means for examining issues the actual activities that participants are involved in, the value structures in play, and the identities that are being performed – all work that social interaction accomplishes covertly more often than overtly. The primary drawback to discourse analysis is the complexity of the analysis (it can sometimes take quite a bit of study to become adept at analyzing language in this way) and its limitations to only

smaller excerpts of language-in-use given its practical constraints. As we say in the lab, analyzing long tracts of data with discourse analysis is like attempting to paint Texas with a fine, camel's hair brush: Not advisable.

Discourse analysis techniques are used in multiple ways within the Pop.Cosmo lab to better understand the nature of social engagement among participants, the forms of collaboration (and competition) that emerged among the adolescents involved, and ways in which the "culture" of gaming is taken up and handed down. This latter issue, cultural apprenticeship, was the example used for the purposes of this workshop. Using transcripts excerpted from the ethnographic data culled over the eight month lab, we had participants in the workshop observe and then apply discourse analytic strategies to a partially analyzed one-page transcript of in-game talk between a master and an apprentice in order to answer two related questions: (a) What moves does the expert make to apprentice the novice into the activity captured in the transcript? and (b) What values are in play throughout their interaction? Responses to these two questions were then used as fodder for reflection on a third and final question: What evidence do we have for the inferences drawn?

Final Reflections on Mixed Methods Research

Mixed methods as a research approach offers multiple avenues for strengthening one's empirical argument and speaking to a range of diverse research communities at once. The right combination of appropriate methods can create a more complete understanding of one's data corpus, increasing the reliability of one's findings and painting a more complete portrait of the phenomenon at hand. Some hold an "incompatibility thesis" and argue that methods with competing or conflicting premises about the nature of the world and of our truth claims about it cannot, in good faith, be used in combination. After all, if your t-test assumes that aggregated averages are useful representations of groups of people while your "repertory grid analysis" assumes that meaning is individual and not usefully aggregated across groups, then there's a way in which your two selected methods are in conflict about not just the best way to make sense of humans but, indeed, the very nature of the world itself. Such conflict can and do arise. We argue, however, that many of these seeming philosophical quandaries can be easily avoided altogether by understanding the different scales at which various methods operate. In the above example, t-tests assume that populations are of central interest while repertory grid analyses assume that individuals are. But surely groups are composed of individuals whose individual meaning-making shapes and is shaped by the groups of which they are a part. Understanding how individuals become group members and how group characteristics reflect their individual membership is surely part of the enterprise of social sciences and educational research. In thinking carefully through the scale or "unit of analysis" on which various methodologies operate, the researcher can thoughtfully combine methods to create robust, analytic, descriptive and predictive analyses of human beings in all their myriad contexts. Resolving conflicts and incompatibilities between methods is a central charge of the mixed methodologist, a process to be explored and explained, not avoided, in one's research proposal, presentations, and publications. We hope this workshop description illustrates, at least in part, some of that heavy work.

Endnotes
(1) For materials from the workshop go to: http://therealca.ro/GLSmixedmethods.html

References

Bartle, R. (1996). *Hearts, clubs, diamonds, spades: Players who suit MUDs.* Retrieved from: http://www.mud.co.uk/richard/hcds.htm

Chi, M. T. H. (1997). Quantifying qualitative analysis of verbal data: A practical guide. *Journal of Learning Sciences, 6*(3), 271-315.

Gee, J. P. (1999). An introduction to discourse analysis: Theory and method. New York: Routledge.

Steinkuehler, C. A. (2004). Learning in massively multiplayer online games. In Y. B. Kafai, W. A. Sandoval, N. Enyedy, A. S. Nixon, & F. Herrera (Eds.), *Proceedings of the sixth International Conference of the Learning Sciences* (pp. 521–528). Mahwah, NJ: Erlbaum.

Greene, J. C. & Caracelli, V. J. (Eds.). (1997). *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms.* New Directions for Program Evaluation, No. 74, San Francisco: Jossey-Bass.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, *24*(4), 602-611.

Johnson, R. B. & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26.

Rest, J., Narvaez, D., Bebeau, M., & Thoma, S. (1999). DIT-2: Devising and testing a new instrument of moral judgment. *Journal of Educational Psychology*, *91*(4), 644-659.

Trochim, W. M. K. (2006). The t-test. *Research Methods Knowledge Base*. Available from http://www.socialresearchmethods.net/kb/stat_t.php.

Acknowledgments