

# UC Irvine

## UC Irvine Previously Published Works

### Title

Situating Big Data

### Permalink

<https://escholarship.org/uc/item/5vq5065q>

### ISBN

978-1-68123-935-4

### Authors

Dalsen, Jennifer

Anderson, Craig G

Squire, Kurt

et al.

### Publication Date

2017-07-20

### DOI

10.1108/978-1-68123-937-820251014

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

## **Situating Big Data**

Jennifer Dalsen, Craig G Anderson, Kurt Squire & Constance Steinkuehler  
University of Wisconsin-Madison

In Washington DC, contemporary educational reform is increasingly defined in terms of shifting education from a “data poor” activity to a “data rich” activity (T. Kalil, White House Office of Science and Technology Policy, personal communication, September 1, 2013). Over the previous decade, we have seen not only a rise in shared national and state (common core) standards and frameworks that articulate what we, as a country, believe young people and adults should be able to think, know, and do, but also a concomitant rise in large scale, data rich strategies for assessing such knowledge, skills, and dispositions. While inroads into so called “big data” techniques (the capture, curation, storage, and analysis of massive, complex quantitative data sets) have been made on some fronts, particularly in relation to game based technologies for learning, they have not yet caught up to our more sophisticated and inclusive frameworks for learning goals.

Take for example the National Academy of Science’s (Bell & Lewenstein, 2009) “six strands of science learning” framework that unifies goals across both formal and informal science learning environments to include not only content knowledge (strand 2) and inquiry practice (strand 3) but also interest (strand 1), epistemological disposition (strand 4), identity development within the domain (strand 6), and longer term participation in the field (strand 5). Big data techniques applied to learning have made some progress in areas such as content knowledge (Hien & Haddawy, 2007; Pardos, et al., 2007; Nebot, et al., 2006) and, to a lesser extent, inquiry practice (Wasson, et al., 2016; Quellmalz, et al., 2013). Both are areas in which more traditional techniques already fare well, but it has not yet made progress in the more challenging assessment areas that link interest, identity, participation, and epistemology. Putting such constructs in conversation with one another would create a more coherent and convincing data ecology for making strong inferences about learning. If we want to catalyze progress toward more expanded frameworks for learning goals that include tricky variables, such as identity and dispositions, then we must make traction on their empirical measure in ways that are commensurate with contemporary “data rich” corpora and techniques -- and we must do so in ways that are theory-driven and comprehensive, and not simply a list of strands or themes.

Theories of situated cognition restore coherence to the study of learning as knowledge, skill, disposition, interest, participation, and identity by focusing on learning as shifting (inter)action with the social and material world. This body of work includes activity theory (e.g., Engestrom, Miettinen, & Punamaki, 1999), d/Discourse theory (Gee, 2010), distributed cognition (Hutchins, 1995), ecological psychology (Gibson, 1979/1986), ethnomethodology (Garfinkel, 1967), mediated action (Wertsch, 1998), situated learning (Lave & Wenger, 1991), sociocultural theory (Vygotsky, 1978), and situativity theory (Greeno & Moore, 1993). Despite differences among these traditions, each shares a common emphasis on social and material interaction as not just the vehicle for the development of domain knowledge and practices, (from social practices such as argumentation to material practices such as bench lab procedures for growing cells), but also as identity within community (Gee, 2010), interest and engagement, and epistemological understanding (Holland & Quinn, 1987). Human understanding is a process that involves not just individuals but also tools, representations, artifacts, and social interaction.

Educational interventions should be designed to produce deepening participation in complex, authentic social practices (Lave & Wenger, 1991) and increase competence in the use of complex, authentic tools and resources (Pea, 1993). Individuals learn to use tools and resources through regular interactions, which are themselves shaped by membership in communities that give meaning to these tools and resources (Derry & Steinkuehler, 2003). Thus, it is through a series of social and material interactions that individuals develop new identities as capable learners and members within a given community (Gee, 2010; Greeno, 1997).

Nowhere are such situated accounts of cognition more empirically tractable than in online, technology-mediated environments. Foremost, games that offer an interactive media environment, the capacity for joint activity within and beyond the game, a normative following of active online communities of practice, and increasing integration into formal and informal educational environments as a means for learning. Games are both *technologies* and *communities* and their potential for learning can be understood in terms of both. As a technology, games are simulated worlds that offer players “designed experiences” (Squire, 2006) of the world from new points of view or from points of view wholly inaccessible in regular life. In terms of community, games function as new “third places” (Oldenburg, 1999) for informal sociability (Steinkuehler & Williams, 2006). Every game, no matter how obscure, develops its own “affinity space” (Gee, 2004) where mastery is collective (Levy, 1999; Jenkins, 2006; Steinkuehler & Duncan, 2009) and reciprocally apprenticed (Steinkuehler, 2004).

Like any other learning technology, however, games are still only sociotechnical artifacts (Bijker, 1995) whose potential for learning, like that of any instructional tool, is highly influenced by its context of use. Whether it’s a textbook, calculator, or high-end 3-D graphical data display, a tool is only as good as the activities and practices in which it is embedded. The online communities around games are one important part of that ecosystem; the in-room formal and informal environments of implementation are another. Meta-analyses of games for learning confirm that the context of the game and not just the technology in isolation are key factors in their effectiveness (Clark, 2013; Sitzmann, 2011; Vogel et al, 2006; Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013; Young et al, 2012). However, to date we have no models for how we might connect the forms of contextual data that theories of situated cognition emphasize as most important with the forms of “big data” produced from the technology in isolation and increasingly the sole focus of assessment.

In this chapter, we describe an on-going project that seeks to marry theories of situated cognition to the big data movement by connecting clickstream data from technologies in isolation to key forms of multimodal data available from their contexts of use in a theoretically principled and empirically validated manner. Our goal is to connect the broader forms of data considered central from the theory of situated cognition – data including individual and group discourse (online and in-room), artifact analysis, classroom assessments, and school performance data (grades and test scores) – to data exhaust from educational games in order to create a more authentic account of learning. In this chapter, we begin by reviewing the current research on big data analyses of game data exhaust and on game-based social interactions. We then describe the game-based curriculum we used to generate and capture both forms of game-related data (telemetry and talk) in order to create a heterogeneous data corpus for mixed method analysis. We then discuss two key issues that confront us in such an analysis, detailing how abstract theoretical concerns materialize in our methodological and practical work, in the everyday mundane activity of obtaining, organizing, and analyzing data such as these. We conclude with

general reflections on the implications of this effort and concerns it raises about our recent national obsession to convert teaching and learning into a “data rich” enterprise.

### **Big Data Analyses of Game Data Exhaust**

Digital games enable embedded assessments of learning through evaluation of how authentic activity changes, a fruitful contrast to more traditional task-extrinsic measures of learning such as post hoc tests or pre/post measures that can mark endpoints of the learning process or aggregated change over time but cannot trace its full trajectory. Data exhaust from games provide a massive volume of automatically transmitted clickstream or “telemetry” data of trackable actions and choices generated as a byproduct of gameplay by the user. Their analysis holds the promise of allowing us to continuously assess, refine, and even personalize learning.

The majority of successful efforts to use game data exhaust to date lie within the commercial games industry. Within game development companies, such analyses typically focus on maximizing replayability and profits through repeated iteration of A/B tested design features (Kohavi, Longbotham, Sommerfield, & Henne, 2009). For example, the online social game company Zynga, makers of *Farmville* and *Words with Friends*, collects over 15 Terabytes of gameplay exhaust data daily for analysis of play patterns in order to constantly iterate and improve their game content and monetized offerings (Asthana, 2011). Game data exhaust has also been analyzed by game studies scholars in order to examine the nuances of play behaviors within various commercial game titles. Thureau and Bauckhage (2010) examined social interactions in the game *World of Warcraft* to detail the patterns of group formation and interaction. Weber and Mateas (2009) used data mining techniques and machine learning to model and predict opponent strategy use. Drachen and Canossa (2011) used telemetry data to examine spatial patterns of game level exploration, showing patterns in how players navigate a virtual space. Drachen, Canossa, and Yannakakis (2009) used such data to introduce methods of determining qualitative “play styles” or groupings based on different ways that people play. In another study, Bauckhage et al (2012) analyzed telemetry data for traces of interest, predicting how long players will persist in a game. Gagné, El-Nasr, and Shaw (2011) use data mining techniques to investigate how players deviate from the designed game paths.

More moderate progress has been made in the analysis of game data exhaust towards pedagogical ends. Big data efforts in educational research fall under the general rubric of educational data mining (EDM) or learning analytics (LA), where EDM typically describes broad analysis of a large number of learners (>1K) and LA describes a deeper analysis of a smaller set of learners. Using such techniques, Baker, D’Mello, Rodrigo and Graesser (2010) examined how in-game behavior relates to cognitive-affect and performance metrics to find that boredom was more detrimental to learning than frustration. Clark and Martinez-Garza (2012) assessed intuitive and formal understanding of specific physics content and found that playing a physics-based game developed players’ general understanding of physics but had no direct impact on formal physics questions. Berland, Martin, Benton, Petrick Smith, and Davis (2013) demonstrated student learning of programming through strategic game mechanics, highlighting the importance of tinkering in novice programmer skill development. Lewis, Trinh, and Kirsh (2011) studied distributed attention among champion competitive real-time strategy game players to discover a new model for task management training. Such research illustrates the promise of big data

analyses for education, although to date no research has integrated contextual “paradata” from implementation that might account for social interaction.

### **Analyses of Game-Based Social Interactions**

With this rise of quantitative analyses of game-based big data sets over the last several years we see a similar rise of qualitative analyses of game-related social context and interaction, although the two bodies of research have heretofore remained largely disconnected. Much of this latter body of research has focused on game-based affinity spaces, social spaces in which individuals who share a common interest in a given game or gather online asynchronously (or, more rarely, synchronously in person) to share and discuss information about the game (Gee, 2004; Gee & Hayes, 2010). Over time, this loose, distributed network of individuals with a common interest in the game becomes a sort of “community of common interest” with its own structures of membership (Gee, 2007; Gee & Hayes, 2010), discursive and virtual practices (Steinkuehler, 2006), peer mentoring (Squire, 2011), and apprenticeship (Steinkuehler & Oh, 2012). Through such structures, participants share not just knowledge and skills but also dispositions or values – loosely shared orientations toward what “counts” as valuable about a given game or accomplishment (Steinkuehler & Oh, 2012).

The body of research on game-related social interaction and affinity spaces documents the formative role of social context and interaction in learning through games. Studying the game of dominoes, Nasir (2005) observed the intertwining relationship between individual and social processes during play. A player’s prior knowledge of gameplay, balanced by teamwork, conversation and negotiation, afforded the discovery of alternate play goals. Simultaneously, players built competency through seeking help and offering unsolicited advice to those around them. Steinkuehler (2004) evaluated the structure of apprenticeship in the massively multiplayer online game *Lineage I* to find that newcomers gained genuine expertise through apprenticeship and joint activity with more advanced players. A study of peer interaction around *Civilization III* by Squire and Giovanetto (2008) found high engagement in player pairs as each group shared strategies and struggles. More experienced players used sophisticated language as a means to communicate their advanced knowledge to less experienced players. Nardi’s (2010) ethnography of *World of Warcraft* described a learning environment comprised of social organizations, ranging from online communities to friends and family members, in which learning resulted not simply from individual exploration but from social participation and conversation. Here, cognition in games was seen as a socially distributed process. In a later study of the same game, Martin (2012) found WoW players as active producers of their community’s knowledge base through the production of transmedia materials including YouTube videos, blogs, forums and personal websites. Even among players who did not actively seek advice from other players still leveraged the collective intelligence of their community through these materials, thereby engaging in an indirect form of learner-directed mentorship. Similarly, Ducheneaut, Yee, Nickell and Moore (2006) found that such indirect mentorship can also be identified via in-game chat channels: While not every player is obligated to participate in discussion, typed information on general advice or background ‘chatter’ nevertheless creates a shared feeling of social presence and community shared information. A study by Seay, Jerome, Lee, and Kraut (2004) on Project Massive, a website designed to study how online game subscribers played, organized and communicated information with each other, demonstrated significant variation in community talk. The majority of players surveyed (77%) shared advice and offered support to each other. In contrast, few players actively shared information about themselves or their personal

experiences to other people. Educational multiplayer online games such as *Quest Atlantis* (Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005) or *River City* (Dede, Ketelhut, Nelson, Clarke, & Bowman, 2004) offer hybrids of online and offline learning wherein players can adopt new roles and actively participate through the game's educationally oriented narrative and content. Such multimodal approaches to learning, in combination with community engagement, are the basis for the design for the *Virulent* project detailed herein.

### **Context of Investigation**

The goal of this project was to connect qualitative analyses of game-related talk and interaction to quantitative analyses of big data sets from in-game activity. Toward this end, we first needed to create a rich context for investigation that would generate sufficiently rich data of both types. Given our situated, sociocultural perspective, we wanted to create an activity that would engage students in authentic social practices (Lave & Wenger, 1991) using complex tools and resources (Pea, 1993) representative of a given domain while yet remaining playful, in keeping with the overall intention and spirit of a games-based approach. It needed to be structured in such a way as to elicit significant game play (i.e. time on task) as well as substantive conversation around and about gameplay (i.e. paradata) with some element of group affiliation and identity. Toward these ends, we developed a week-long workshop for middle school students over spring break entitled "Game-A-Palooza." Over the course of five days, participants worked in collaborative groups to decipher viral function and human cellular response as a means for stopping the fictional "Raven Virus" from reaching global pandemic and, thus, triggering a worldwide apocalypse.

### **Curriculum Context**

Participants in our week-long Game-A-Palooza (GaP) event roleplayed as scientists recruited by the Center for Disease Control (CDC) to stop the "Raven Virus" from spreading across the world. To enhance roleplay, students received a lab coat, fieldnotes journal, and "digiscope" device, which was an iPad loaded with the game *Virulent* whose game levels were called dynamic "microscopic slides" (Figure 3 below). Throughout the week, students played through the "microscopic slides" of the game to figure out how the Raven virus functioned and how the human immune system reacted. Toward these ends, participants created shared paper models of cell and viral systems, debated within and between small group to decide which model and possible solution was most viable, wrote persuasive letters to the CDC, and recorded science presentation videos for final group solutions to the threatening apocalypse.

Throughout the week, participants were encouraged to play through the science of virus and cell structure and function. Each day began with a (pre-recorded) Skype call from mock scientists at the CDC that included updated information about the Raven Virus and its ongoing spread. A map illustrating the increasing infection rate of the Raven Virus was displayed each day, with more countries infected by this unknown microorganism each day (Figure 1). The curriculum was designed with structured questions (e.g. How do viruses spread?) and activities (e.g. small group model construction) guiding each day's activities, with students divided into small groups of 3-4 students and assigned a facilitator (who also served as a researcher gathering fieldnotes and monitoring other forms of data collection, described below). Each small group worked collaboratively to compete with one another to construct a predictive model for how the virus worked inside the body and to make a persuasive argument for a group

proposal for stopping its spread. Through joint activity, conversation and facilitator mentoring, participants advanced toward their assigned challenge: to stop the Raven Virus in its tracks.



**Figure 1.** Game-A-Palooza daily mock CDC skype call with unfolding information about the Raven Virus and its spread.

In order to accomplish this, each small group or “research team” (Figure 2) met daily over five days to work collaboratively and competitively through the problem. On day one, participants met their teammates, completed pre-assessments, investigated the Raven Virus and wrote a preliminary letter to the CDC. On day two, they conducted initial investigations on “microscopic slides” (game levels) and constructed a preliminary model of how the Raven Virus was infecting the body. On day three, students conducted more “digiscope” research and revised their model based on what they discovered, then produced a short video presentation to explain their model to other scientists (other student groups and the mock CDC scientists). On day four, each small group presented their model to the cohort and deliberated collectively on ways to improve or change it based on the accumulated shared knowledge. At the end of day four, players were informed that the CDC scientists themselves were now infected and thus the race to produce a solution based on an accurate model began. On day five, since the CDC was infected and could no longer assist in researching ways to stop the spread, students were told that they were now on their own to collectively, as a cohort, find a solution to stop the Raven Virus from taking over the world. They were given background materials (fictionalized but scientifically based literature including a journal article, a news article, and a book excerpt) on three choices available – create a vaccine, use an RNA inhibitor, or use a drug to inhibit the cell’s mitochondria – and had to decide, based on their collective model, which avenue to pursue. Thus, on the final day, groups came together to hold a final cohort debate with each group presenting their argument and then individual students placing final votes to decide which option they pursue as the new stand-in scientists for the CDC.

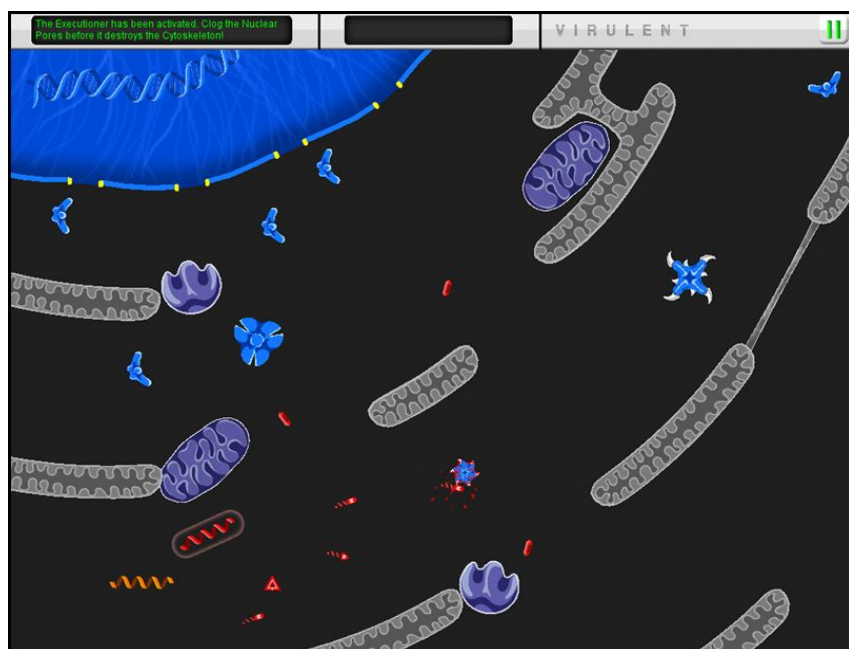


**Figure 2.** A research team works collaboratively to figure out the structure and function of both the virus and the cell.

Throughout the week, all participants were given 24/7 access to their “digiscopes” and were encouraged to take them home and work on the problem on their own time. To drive out-of-session engagement beyond the daily curricula, participants received “coin rewards” for completing optional tasks – for example, finishing a specific challenge level or completing every level of the game, which are both activities outside of the formal curriculum. At the end of the program, students could then exchange these points for various prizes such as pocket microscopes, video game character plushies, and raffle tickets to win an iPad.

### **The Game *Virulent***

*Virulent* (Figure 3) is a real-time strategy game designed to introduce players to viral replication and how the body reacts to infection. Players control the Raven Virus, moving level by level to infect host cells and attack the immune system with viral proteins. Each level takes players deeper into the microscopic world with the support of audio and visual directions. Specific actions that players may accomplish during gameplay include ‘hijacking’ different cellular parts (e.g. mitochondria), evading free antibodies, and ridding the body of slicer enzymes. A graphic almanac serves as reference to different virus and immune system components encountered during gameplay. A series of ‘challenge’ levels also accompanies the game, whereupon players can engage in timed tasks or other “test your mettle” type events.



**Figure 3.** The “digiscope” or GLS tablet game *Virulent* on the Apple iPad.

### **Data Collection**

#### **Participants**

Data was collected across three separate events: As part of a GLS spring break camp called Game-a-Palooza (n=43), within a local private school’s extended day program (n=11), and within a larger summer camp run by the Boys & Girls Club (n=27). Participants were separated into research teams, each of which comprised 3-4 students supported by a Games+Learning+Society facilitator. Thus, there were 21 student research teams across three cohorts. Teams were assembled randomly based on age and served as the primary context for interaction during the 90-minute session.

Recruitment for each cohort varied. For the spring break event, participants were recruited via flyers around the local community and schools. For the extended day program, we worked with administrative staff to fold our program included an existing computer club where students gathered to learn about coding and explored other computational interests. For the Boys & Girls Club event, we worked with directors to advertise our event to children already attending their summer program. Across all three groups, participant time was compensated.

The average participant age was 11.5 years old (range from 9 to 14 years). The majority of participants identified as Caucasian (n=27), African American (n=13) and Hispanic (n=8). One participant identified as multiple race and other students chose not to answer this question. Of the students who listed gender, 35 identified as male and 22 as female.

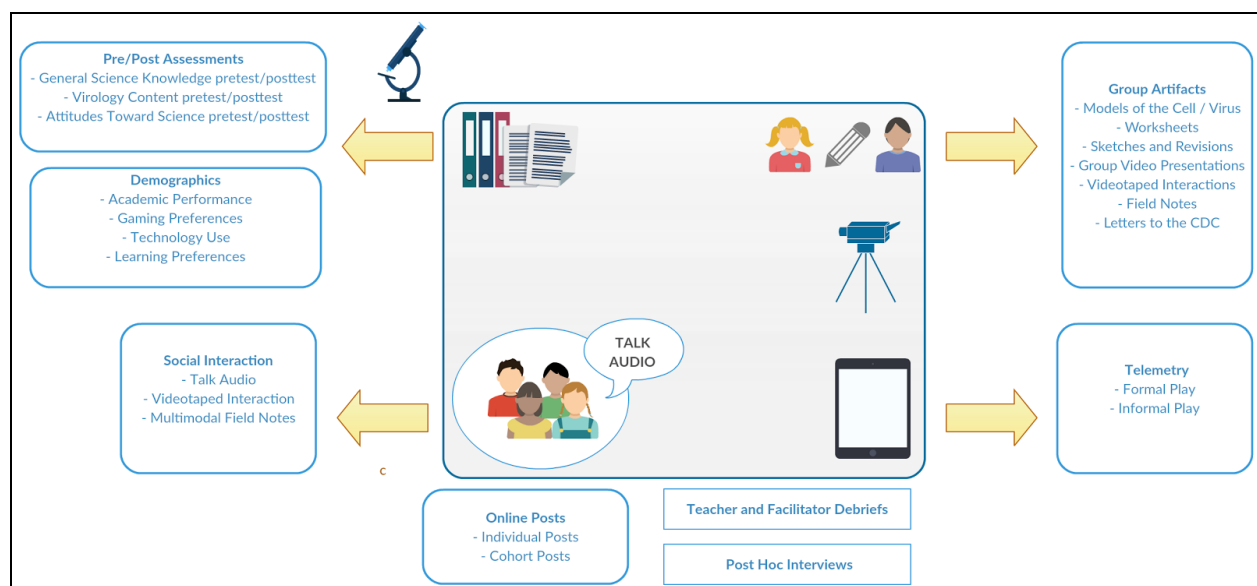
A total of 22 participants attended all five days of the spring break session, with 8 participants four days, 6 attending only three, and 7 attending two or fewer (and thus considered a dropout). No participants dropped from the private school extended day program; two participants dropped from the Boys & Girls Club session. Across all three sites then, a total of our retention rate was 89 percent. Our

final n-size for analysis was 72. To capture group dynamics and effects, participants that dropped out of the event were still included in discourse analyses and any other analysis tied into it.

To ensure player data was tracked across each day (both in and out of formal event sessions), developers at GLS created a QR code login system. QR codes were carried by each participant on paper and scanned using the iPad camera. Each QR code was attached to a particular anonymized, randomly generated ADAGE ID (e.g. “Fluffy Sandals”) in order to track player action. This system ensured that the clickstream data remained organized and tied to a particular player no matter when or where they played. This also simplified the login process, resolving likely problems with students forgetting or losing their unique login details. Gameplay data collected from daily program sessions was structured by the curriculum and therefore considered “formal gameplay”; gameplay occurring outside of sessions was considered “informal play.” Informal play included game activity during break time between sessions, lunchtime, and home activity. Partitioning the data in this way was straightforward using ADAGE generated timestamps.

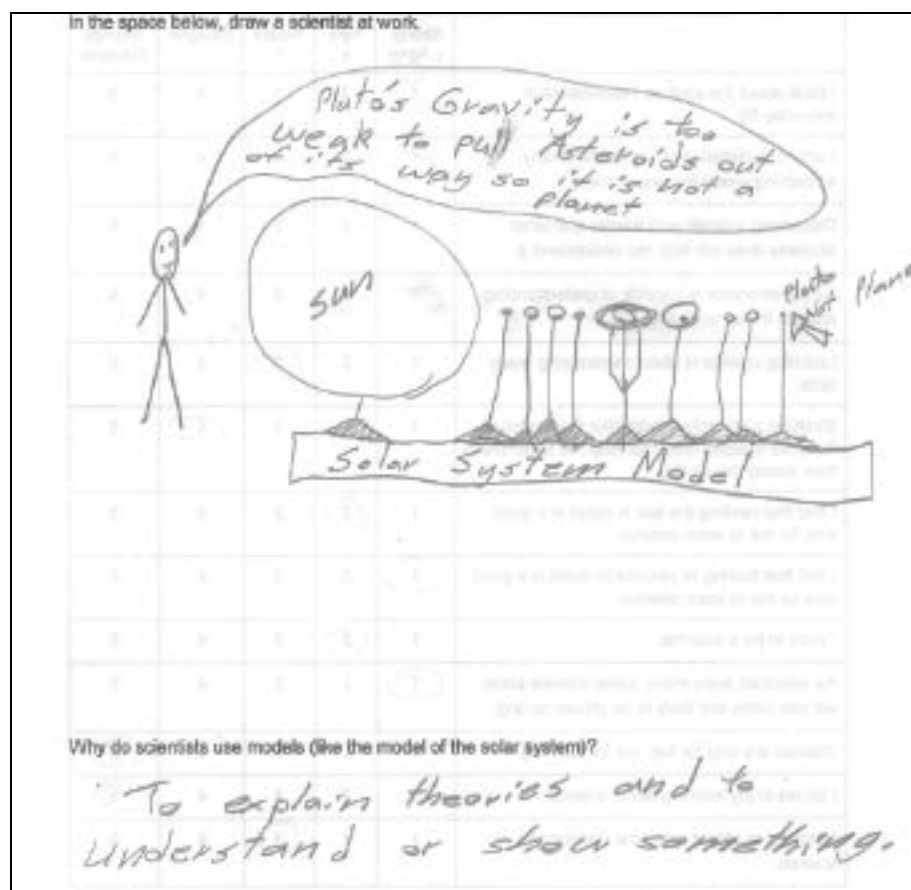
## Data Corpus

A rich, diverse stream of data was collected across these contexts during the five day events as detailed in Figure 4. Here, individual – nested in student research team (small group), which in turn is nested in cohort (each of three locations) – is the primary unit of analysis. Game data exhaust from each participant’s iPad was recorded via ADAGE (described above), with both formal and informal gameplay recorded and stored in the same data corpus. Group artifacts were collected, photographed and uploaded to a secure server, with original materials transferred to a secure office in a university building and stored in locked cabinets. To allow remote teams access to the data for analysis, IRB approved channels were set up to transfer and collaborate on data. Once stored on secure servers, the data was cleaned and anonymized. Anonymized data was shared with collaborating university sites are housed on a secure online account. All procedures were IRB approved.



**Figure 4.** Diagram of the context of investigation, each data stream, and main variables of interest per data stream.

**Pretest/posttest assessments.** Each participant was administered a pre/post assessment on (1) general science content knowledge and attitudes and (2) content knowledge specific to virology. The general science assessment was paper-based, whereas the virology assessment was completed on a tablet. The first was designed to gauge the participants' understanding of and attitudes towards science and included items designed to elicit their understanding of what scientists do and what models are for (Figure 5). The second was designed to assess the participants' content knowledge relate to virology specifically as well as their confidence in their ability to learn similar or more advanced content related to it. Both were framed as assessments of our materials and for feedback on our research agenda and game/material design, not formal assessments of their individual performance.



**Figure 5.** Example data from a posttest general science assessment.

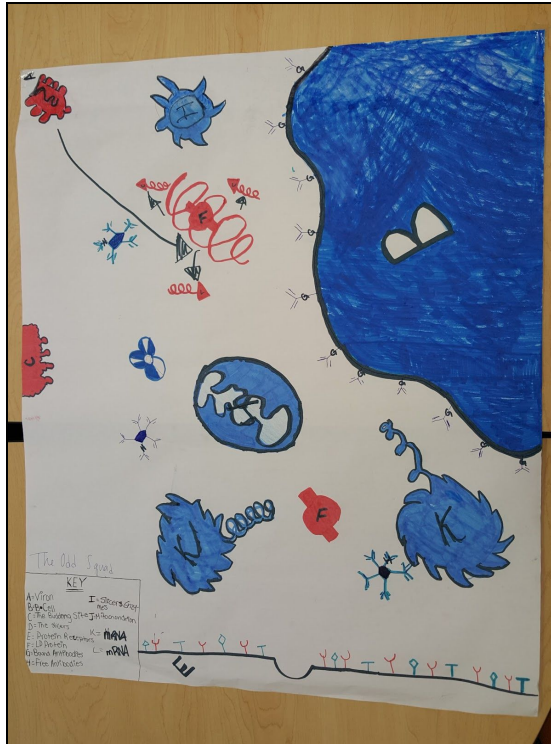
**Demographic data.** A survey of demographic data was administered prior to all activities. Here, we collected standard information related to gender, age and ethnicity, as well as participants' prior experience with technology, game genre preferences, motivations in gameplay (e.g. collecting items, winning the endgame), interest in science, favorite school subject and average grades received. All information was self-reported.

**Social interaction data.** Interactional (verbal) data was captured via small recorders worn around their neck on a lavalier name tag to avoid obstructing movement or conversation. Because some participants are naturally more soft spoken, individual recorders helped ensure a more reliable data stream. Facilitators helped verify that recorders were functioning before small or large group activities, but each participant was in control of their own recorder and its use.

Conversation was recorded for all research team (small group) conversations throughout the curriculum and all cohort (large group) activities and debates. Small group research team conversations included general discussion of the game and game strategies, identification of virus or immune system parts, questions raised, discussion around model making, and social banter. Cohort (large group) conversations included model presentations and cohort debates. Model presentations included small group explanations of virus and immune system functioning, with discussion and feedback about each model. Cohort debates consisted of three stages: (1) small group persuasive presentations of proposals for how to stop the Raven Virus based on their model and collected evidence, (2) collaborative interrogation among fellow cohort members of the proposals, and (3) a final synthetic conversation on what proposal was best. Each cohort was also video recorded, with cameras placed in the back of each classroom and left there throughout the duration of events.

**Group artifacts.** Student artifacts including virus models, structured worksheets, research team video presentations, and letters to the CDC were collected from each participant following each session. All artifacts were digitized, labeled by participant ID and group number, and stored on secure servers at the end of each day less each research team's model, which was photographed but kept on the premises for subsequent daily activity.

**Virus models.** Each research team created and revised a model of the virus and its impact on the immune system (Figure 6). Models were paper-based. No specific template existed for this model making task. Models were based on a collective decision from group members on how they chose to represent the virus and immune system relationship. Available material for model making included: markers, tape, colored pencils, glue, paper, post-it notes, arrow stickers and pipe cleaners.



**Figure 6.** Example group model of the Raven Virus.

**Structured worksheets.** Participants were given structured worksheets on Day 1 to help them identify virus and immune system components and on Day 5 to present three possible ways to stop the Raven Virus from spreading. Participants wrote evidence to support or refute each solution.

**Letters to the CDC.** Each research team wrote a letter to the mock CDC to present their initial proposal (on Day 1) for how to stop the Raven Virus from spreading across the world and to justify that claim based on observations made through their digiscope (ie. from gameplay).

**Research team video presentations.** Each research team produced a 5-minute video presentation of their model and a possible intervention for the Raven Virus (on Day 3) based on it for the mock CDC. All team videos, both practice drafts and final versions, were collected for analysis.

**Online portal postings.** An online portal allowed students to write and post messages to each other within and outside of structured activities each day. Each cohort had password-protected access to a group message board and was encouraged to pose questions, offer advice or challenge other peers in gameplay. All posts were monitored and collected throughout activities.

**Telemetry Data.** Like other games from the Games+Learning+Society Center (GLS), *Virulent* is integrated into a cross-platform and cross-game-title data framework called the Assessment Data Aggregator for Gaming Environments (ADAGE). ADAGE guides game developers to articulate the content model of their games through a metadata tagging process that facilitates data mining of play patterns and key events in game play (Owen & Halverson, 2013). In effect, ADAGE scrapes all game actions from players' gameplay and then stores it digitally on GLS servers. Such data can then be pushed to a JSON or

CSV file for analysis. As Figure 7 illustrates, each line of data constitutes a game action and is tagged with multiple attributes to identify what the action is, who performed it, when it happened, and any other identifying features that are relevant. Such clickstream data can then be manipulated to show patterns in individual or aggregated gameplay – from level attempts made to almanac usage to broader playstyles. Example telemetry variables we culled from *Virulent* data exhaust for this study included starting a session, spawning a unit, unit paths, unit collisions, and unit deaths. Each of the variables were tagged with player ID, timestamp, session token, and other relevant event details. The result is a dense numeric data corpus over two million lines.

| 1  | player      | epoch_time | session_token  | ada_base_types                         | key             | user_id    | group | unit_id | positional_context | virtual_context | button | path | self | self_id | other | other_id | damage | amount | available | type |
|----|-------------|------------|----------------|----------------------------------------|-----------------|------------|-------|---------|--------------------|-----------------|--------|------|------|---------|-------|----------|--------|--------|-----------|------|
| 2  | lazy_cat    | 1.43E+12   | FC021809-DA32- | ["ADAGEData"]                          | ADAGEMouseEvent | 56495      |       |         |                    |                 |        | left |      |         |       |          |        |        |           |      |
| 3  | little_room | 1.43E+12   | B28688F9-CSFB- | ["ADAGEPlayerEventUnitCollisionEvent"] | 56387           |            |       |         |                    |                 |        | left |      |         |       |          |        |        |           |      |
| 4  | lazy_cat    | 1.43E+12   | FC021809-DA32- | ["ADAGEPlayerEventEnergyEvent"]        | 56495           | Genome     |       | -13058  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 5  | little_room | 1.43E+12   | B28688F9-CSFB- | ["ADAGEPlayerEventUnitCollisionEvent"] | 56387           |            |       |         |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 6  | lazy_cat    | 1.43E+12   | FC021809-DA32- | ["ADAGEPlayerEventUnitSpawnEvent"]     | 56495           | N mRNA     |       | -13628  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 7  | little_room | 1.43E+12   | B28688F9-CSFB- | ["ADAGEPlayerEventUnitDeathEvent"]     | 56387           | N mRNA     |       | -16510  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 8  | lazy_cat    | 1.43E+12   | FC021809-DA32- | ["ADAGEData"]                          | ADAGEMouseEvent | 56495      |       |         |                    |                 |        | left |      |         |       |          |        |        |           |      |
| 9  | little_room | 1.43E+12   | B28688F9-CSFB- | ["ADAGEPlayerEventPathCreateEvent"]    | 56387           | N mRNA     |       | -16182  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 10 | lazy_cat    | 1.43E+12   | FC021809-DA32- | ["ADAGEData"]                          | ADAGEMouseEvent | 56495      |       |         |                    |                 |        | left |      |         |       |          |        |        |           |      |
| 11 | little_room | 1.43E+12   | B28688F9-CSFB- | ["ADAGEPlayerEventUnitCollisionEvent"] | 56387           |            |       |         |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 12 | little_room | 1.43E+12   | B28688F9-CSFB- | ["ADAGEPlayerEventUnitCollisionEvent"] | 56387           |            |       |         |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 13 | little_room | 1.43E+12   | B28688F9-CSFB- | ["ADAGEPlayerEventUnitDeathEvent"]     | 56387           | Slicer Enz |       | -16344  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 14 | little_room | 1.43E+12   | B28688F9-CSFB- | ["ADAGEPlayerEventUnitDeathEvent"]     | 56387           | N mRNA     |       | -16218  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 15 | little_room | 1.43E+12   | B28688F9-CSFB- | ["ADAGEPlayerEventUnitDeathEvent"]     | 56387           |            |       |         |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 16 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEPlayerEventPathCreateEvent"]    | 56324           | N mRNA     |       | -25732  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 17 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEPlayerEventPathCreateEvent"]    | 56324           | Slicer Enz |       | -25778  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 18 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEPlayerEventPathCreateEvent"]    | 56324           | Slicer Enz |       | -25816  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 19 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEPlayerEventPathCreateEvent"]    | 56324           | Genome     |       | -25414  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 20 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEData"]                          | ADAGEMouseEvent | 56324      |       |         |                    |                 |        | left |      |         |       |          |        |        |           |      |
| 21 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEPlayerEventPathCreateEvent"]    | 56324           | Slicer Enz |       | -25778  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 22 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEPlayerEventPathCreateEvent"]    | 56324           | Slicer Enz |       | -25816  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 23 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEPlayerEventPathCreateEvent"]    | 56324           | N mRNA     |       | -25862  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 24 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEData"]                          | ADAGEMouseEvent | 56324      |       |         |                    |                 |        | left |      |         |       |          |        |        |           |      |
| 25 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEPlayerEventPathCreateEvent"]    | 56324           |            |       |         |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 26 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEPlayerEventPathCreateEvent"]    | 56324           |            |       |         |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 27 | great_boot  | 1.43E+12   | F8492AAA-8B42- | ["ADAGEPlayerEventPathCreateEvent"]    | 56324           | N mRNA     |       | -25898  |                    |                 |        |      |      |         |       |          |        |        |           |      |
| 28 | bumpy_tow   | 1.43E+12   | 3F506E7-D666-  | ["ADAGEData"]                          | ADAGEMouseEvent | 56243      |       |         |                    |                 |        | left |      |         |       |          |        |        |           |      |

**Figure 7.** Example telemetry data (not all variables included) from *Virulent*.

**Teacher and facilitator debriefs.** Each day, following completion of activities, we conducted a debrief interview with the cohort teacher and led a group debriefing session among all facilitators, providing us an opportunity to identify and address any issues or problems as they arose. Facilitators also recorded their observations about their assigned small groups as fieldnotes immediately following each session. Fieldnote observations focused on issues of individual player engagement, group social dynamics, the model development process, and any idiosyncratic group synonyms or metaphors used for virus and immune system components or other content covered.

**Post hoc interviews.** Finally, a month after all events were over, we conducted follow-up interviews with a subset of participants to test the durability of changes in science understanding and attitudes targeted by the intervention. Interviews were conducted face-to-face on campus and lasted roughly 30 minutes.

## Data Analysis

Obviously, the first question we must ask is, was the intervention useful? The answer is yes. Pre-post assessments show an increase in participants' content knowledge related to virology and biology (a two grade increase or 22.1%,  $t(61)=7.5$ ,  $p < 0.0001$ ), increases in positive orientation and attitudes toward science ( $t(68)=2.5$ ,  $p < 0.05$ ), and confidence in their ability to complete science-related tasks ( $t(49)=3.5$ ,  $p < 0.001$ ). Thus, the program as a whole was demonstrably a success.

For this project, however, demonstrating the merit of the game and its curriculum is merely a preliminary condition that must be met to justify the subsequent in-depth analysis. The main research questions of interest for our project are methodological: Which data streams from the contexts surrounding the game can be meaningfully and practically combined with telemetry data from the game

itself? Specifically, of the contextual data theoretically suggested by a situation cognition approach, which data streams are:

1. Feasible to collect in a standard research or classroom setting?
2. Quantifiable for integration into a single “big data” set?
3. Productive for building algorithms and predictive models of student learning and performance?

In order to answer these questions, we treat the game-based curricular intervention across all three sites as a single instrumental case (Stake, 1995). Within the case, we can then explore what is or is not methodologically feasible, useful, and generative in making sense of the dynamics of individual and group learning. Toward, these ends, we ask questions of the data about the nature of learning within this specific case, use mixed method techniques on the heterogeneous data corpus in order to answer them, and then reflect on the feasibility, usefulness, and generalizability of our methodological approaches for a broader set of cases.

Given the goals of the game and curriculum, we define *learning* in this case context in terms of biology content, argument, and modeling. We also assess student *interest* as a fourth and final construct in keeping with prior work on science learning in informal environments (Bell & Lewenstein, 2009) and given our overarching commitment to pedagogies that increase not only conceptual knowledge and skills but also productive dispositions toward the domain. Table 1 shows the data streams to which each learning construct is applied.

**Biology content.** In order to trace biology content, we developed an inclusive coding scheme based on all terms that appeared in the game, the in-game almanac, and curriculum materials related to viruses (e.g. virion, antigenome, mRNA) and cells (e.g. nucleus, mitochondria, cytoskeleton site). We coded explicit use of each term as well as synonyms, metaphors, and misstatements (such as mispronunciations). For each instance of use, we coded for correct use of the concept used.

**Argumentation.** In order to trace students’ argumentation, we developed a coding scheme based on the work of Berland & Reiser (2011) to capture both individual and group work with claims and their justifications:

- **Construct a Claim.** States a declarative fact (e.g. “We can destroy the virus with a vaccine.”)
- **Defend a Claim.** Provides an example or reason for why the claim is true. (e.g. “No, it's because there are more virions and that's why the body is attacking them more.”)
- **Question a Claim.** Request for clarification or justification for a claim. (e.g. “How could it start at the budding site?”)
- **Evaluate a Claim.** Confirms or denies that a claim is true without providing reason as to why (e.g. “I don’t think so.”)
- **Revise a Claim.** Revises a claim that has already been made due to new information gathered. (e.g. “Well, maybe a virus is alive but it's not like... really intelligent, at least not here.”)

**Modeling.** In order to trace students development and use of models, we constructed a coding scheme based on the work of Russ et. al (2008) that captures relationships among components within the students’ model and connections to outside knowledge and understanding:

- **Relationships**

- **Chaining.** Reasons backwards or forwards in time. (e.g. “If we remove the protein receptors, then the cell won’t be able to get any proteins, and it will die.”)
- **Modification.** Changes, adds, refines, or removes parts of the model. (e.g. “We’re missing the RNA. We need to draw them here.”)
- **Connection to Other Knowledge**
  - **Phenomena.** Compares what the model tells them about viruses to their own experiences in the world in order to evaluate the model. (e.g. “I have to get a flu vaccine every year, so that means vaccines aren’t always effective forever. The virus must change in some way.”)
  - **Analogies.** Compares the virus or cell to some other system. (e.g. “The mRNAs are like the running backs on the virus team. They’re fast and they can avoid the slicers.”)

**Interest.** Finally, in order to trace student interest in science, we constructed a coding scheme based on the work of Hidi and Renninger’s (2008) four-phase model of interest development that captures participants value statements about the ongoing activities and their difficulty, overt displays of participation and curiosity, and identity statements related to their role in the activity or material:

- **Value Statements** (Positive, Negative, Other). Expression of positive or negative feelings towards the game, event, or content matter, including statements about the merit of activities or their appeal.(e.g. “This music is really cool.”)
- **Difficulty Evaluations** (Easy, Hard, Unknown). Statements about the level of difficulty or challenge of the content.
- **Participation Displays** (Volunteering, Persistence, Withdrawing). Overt displays of participation in the form of unprompted volunteering for tasks (e.g. “I want to do it!”), persistence, especially in the face of failure (e.g. “I want to play just a bit longer.”), or withdrawing (e.g. “I’m not doing that.”).
- **Curiosity** (Exploratory, Extension Questions). Extension questions about the content covered in the activity (e.g. “I wonder if...”)
- **Identity Statements.** Statements about about their own attributes or aspirations related to the game, curriculum, or content matter (e.g. “I want to be a...”)

**Table 1.** Analytic constructs and the data streams to which they are applied.

|                 | Pre/Post Assessments | Demographics | Social Interaction | Group Artifacts | Telemetry | Online Posts | Teacher & Facilitator Debriefs | Post Hoc Interviews |
|-----------------|----------------------|--------------|--------------------|-----------------|-----------|--------------|--------------------------------|---------------------|
| Biology Content | X                    |              | X                  | X               | X         | X            |                                | X                   |
| Argument        | X                    |              | X                  |                 |           | X            |                                | X                   |
| Modeling        | X                    |              | X                  | X               |           |              |                                | X                   |
| Interest        | X                    |              | X                  |                 | X         | X            |                                | X                   |

In order to calculate interrater reliability, we randomly sampled 10 minute segments from the data corpus and had 2-3 individual research team members analyze the data using one of the four analytic constructs (biology content, argumentation, modeling, interest). Each coder independently coded roughly 1600 turns of talk. Interrater reliability was calculated using Fleiss's kappa between coders. Constructs with a kappa less than .60 met to discuss and resolve discrepancies before re-coding and recalculating interrater reliability. This process was repeated until sufficient reliability was achieved.

Together, these four constructs for learning (biology content, argumentation, modeling and interest) are applied across the qualitative data corpus using MAXQDA (Gibbs, 2008), quantified, and combined with telemetry data which has been analyzed using the same broad constructs detailed above but here operationalized in terms of behaviors traceable via user clickstream data. Specifically, biology content can be linked to interactions with game units (virion, mitochondria, etc.), along with the player's use of the in-game almanac. The amount of exposure to given game elements can be linked to scientific arguments made with those elements. Information complexity increases with each new level played, incrementally providing additional information on the cell to cell cycle of the virus, which can then be linked back to model development. The amount of time engaging in informal gameplay completed by players can be used to gauge interest. Demographic data collected as part of pretest assessments is used to group participants by age, gender, and gaming experience to investigate differences based on these key variables and to control for science knowledge and interest at time one. Group membership data (small research team group membership, facilitator assignment, cohort membership) is used to examine individual and group dynamics as well as effects of facilitator assignments. Time stamp data is used to explore patterns in the ways that players engage with *Virulent* while in and out of structured activity sessions. Teacher and facilitator daily debriefs are used to triangulate individual and group patterns and to generate potential hypotheses to explain differences between and within student teams (for example, in the basis of varying social dynamics within or across groups).

### **Early Challenges**

The goal of collecting such copious and mixed data and analyzing them using single constructs operationalized in multiple ways is to re-situate "big data" analyses of technology data exhaust with equally rich streams of data from their context of use. In terms of individual analytic strategies, there is nothing new here: Pretest/posttest comparisons, quantified coding of qualitative data (Chi, 1997) from group interaction and interviews, analysis of scientific models generated by a group and revised over time are hardly methodologically risky or innovative. Attempting to reframe such analyses as an important and currently missing component of so-called "big data" approaches is. In truth, this attempt to collect and analyze qualitative data with the goal of not only quantifying it (for inclusion in big data sets) but even (partially) automating the process is a bit audacious. As such, it faces multiple challenges. Here, we describe two.

### **The Lamest Bottleneck is the Greatest Bottleneck: Transcription**

Exploratory talk, or thinking critically about proposed ideas among group members (Mercer, Wegerif & Dawes, 1999), is beneficial to establishing stronger scientific arguments (Rojas-Drummond & Zapata, 2004). As such, we explicitly created a curriculum to promote exploratory talk within small

groups and across cohorts. Such dialogues, especially in the most engaged and fully involved moments throughout the study, were energetic and frequently impassioned, with overlapped speech and more than one student initiating, elaborating, evolving and even contradicting an idea. These conversations in the data corpus represented some of the more substantively dense moments in the corpus – and also the most difficult to disentangle and simply transcribe.

Given our aim of testing the feasibility of automating such data collection, our strategy for recording and representing social interaction was lavalier mics placed on each child and a transcription service that relied on voice recognition software. The mics were durable, portable, and inexpensive (\$8), fitting easily into children's name badges and successfully picking up the audio stream of the wearer quite well. Ambient noise, however, was a continual problem for automated transcription of the data. Transcripts were returned with lexical and syntactic errors and misrepresentations, particularly for the specific terms we were most interested in related to biology and virology content. Even worse, the documents contained few or no speaker designations from the audio given the nature of overlapping speech and an inability to discern one child's voice from another, even within a small group of 3-4. In some transcripts, there were lengthy omissions of entire conversations – especially when multiple individuals were speaking at once in close enough proximity for a mic to capture audio on more than one speaker. In other words, right when exploratory talk emerged and conversations became the most animated and engaged.

To remedy the issue, research team members who had attended the event in person had to review each transcript while listening to the audio files and correct documents line-by-line, clarifying specialized biology and game-related terms, disentangling attributions of the utterances, and filling in large gaps that often represented critical moments of engagement in the curriculum. Such shoring up of the transcripts was expensive, time-consuming, and tedious work that set the project back by several months before any real analysis could begin. Even then, speech pattern and intonation were still not always enough to distinguish one participant from another. As a result, we resigned ourselves to attributing talk to individuals where possible but designating unrecognizable utterances as “speaker unknown.” Of the entire corpus of social interaction data, approximately 5% was lost for analysis of individual learning by the sheer fact that group talk data is onerous to transcribe by hand and simply impossible to transcribe automatically by voice recognition software at this time. Automated services are good enough for single speaker data, especially when calibrated to one person's voice, but are not yet at a point where they can be applied to small group talk, let alone classroom interaction.

### **The Problem of Deixis**

In research, we naturally carve the world into “text” and “context,” foregrounding a portion of the phenomenon (here, the game and talk transcript) as central and backgrounding the remainder as material that can be used to disambiguate or pin down the meaning of the text we are most interested in. There is no sharp dividing line along which text and context are easily separated, however. Context is implicated on all talk through deixis, or “the ways in which languages encode or grammaticalize features of the context of utterance or speech events” (Levinson, 1983, p.54). From the overt use of indexical expressions (e.g. pronouns such as ‘we’ or ‘I’ or pointing words such as ‘here’ or ‘that’) to the covert use of allusions to conceptual and materials worlds presumed or implied by the speaker in their talk, a given utterance always depends on an analysis of the context of that utterance. And thus, we use

context to disambiguate such elements of interaction. And, when contextual information is lacking or falls short, we bemoan the inadequacy of our contextual data, believing that if we simply had more data or the “right” data, we could resolve and disambiguate meaning in interaction itself. Rarely, however, is this task complete. And yet, What context is the right context? How much context is enough? These issues can be addressed but only provisionally; each new utterance holds the promise of a new and shifted context made relevant to ongoing interlocution.

Talking through new concepts and ideas helps solidify understanding and challenge misconceptions (Rennie & Jarvis, 1995; Dove, Everett & Preece, 1999), yet such talk necessarily ranges in expression, precision, and quality. In our intervention, like most, science content was not served up as keywords and vocabulary exercises for participants to use in controlled ways. Rather, new science concepts were situated in their day-to-day conversation and activity: game exploration, small group discussion, model making and cohort debate. Over time, ideas take shape and precision, come to be used accurately and consistently, and stabilize as new tools for “cheating the game” (Squire, 2004). “Squiggly hot dogs” and “ninja stars,” early metaphors that scaffold understanding (Jakobson & Wickman, 2007) at the early stage when its most fragile, evolve into virions and slicer enzymes because the task requirements demand understanding and precision and group discourse demands increasingly precise language in order to make models and claims and defend them. Tracing such evolution is crucial in the analysis, yet disambiguating imprecise terms and their connection to other concepts is tricky and time-consuming work. No amount of contextual data will necessarily get you out of the problem of deixis. No computer code or algorithm readily exists for a given speech act to apply off the shelf in order to catch in its net all terms and synonyms and metaphors and creative monikers and their relation to one another. Such work therefore cannot be easily automated and instead requires old-fashioned reading and interpretation of lengthy transcripts.

### **Summary**

The recent proliferation of big data analytics marks a transformation in the way we collect, interpret, and use information (Mayer-Schonberger & Cukier, 2013) by exponentially increasing scale. Such large corpora provides tremendous analytic power for the detection and comparison of patterns of technology use that, ostensibly, can be used to infer user understanding and learning. But these data are limited by the very fact that contextual data, especially talk and interaction data around the device, are largely absent. From a situated and sociocultural point of view, talk and interaction are an important – perhaps *the* important – vehicle for learning. As such, their inclusion into large scale data analysis is crucial if we are to not fall prey to the streetlight effect (Freedman, 2010) and find ourselves, like the old drunkard in the story, searching for keys under the street light simply because that’s where the light is best.

And yet, including contextual data into larger sets used for research and assessment also raises some real concerns. Is it ethical and wise to record social interaction in classrooms, even if we intend to mine them only for those variables directly related to ongoing learning? Do we, as educators and citizens, really want students to wear to data-recording wearable devices in school settings? More broadly, what role might more inclusive “big data” inadvertently efforts play in our recent national obsession for assessments and accountability? There are significant issues of consequential validity (Messick, 1988) this work necessarily raises; our current project seeks to understand better what’s

possible, but it cannot easily answer what's advisable. For this, broader deliberation and discussion is required.

## References

- Asthana, P. (2011, Oct). Big Data and Little Data. *Forbes*. Retrieved from <http://www.forbes.com/sites/dell/2011/10/31/big-data-and-little-data>
- Baker, R. S. J. d., D'Mello, S. K., Rodrigo, M. M. T., Graesser, A. C. (2010) Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4), 223-241.
- Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development*, 53(1), 86-107.
- Bauckhage, C., Kersting, K., Sifa, R., Thureau, C., Drachen, A., & Canossa, A. (2012). How players lose interest in playing a game: An empirical study based on distributions of total playing times. *Computational Intelligence and Games (CIG 2012)*, (pp. 139-146). IEEE, September.
- Bell, P., & Lewenstein, B. (2009). *Learning Science in Informal Environments*. National Academies Press.
- Berland, M., Martin, T., Benton, T., Petrick Smith, C., & Davis, D. (2013). Using learning analytics to understand the learning pathways of novice programmers. *Journal of the Learning Sciences*, 22(4), 564-599.
- Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. *Science Education*, 95(2), 191-216.
- Bijker, W. E. (1995). *Of bicycles, bakelites, and bulbs: Toward a theory of sociotechnical change*. Cambridge, MA: MIT Press.
- Chi, M.T.H. (1997). Quantifying qualitative analysis of verbal data: A practical guide. *Journal of Learning Sciences*, 6(3), 271--315.
- Clark, D. (2013). *Meta-analysis of digital games and learning in terms of the NRC's Education for Life and Work outcomes*. Paper presented at the Games+Learning+Society Conference 9.0, Madison WI, June 12-14.
- Clark, D. B., & Martinez-Garza, M. (2012). Prediction and explanation as design mechanics in conceptually-integrated digital games to help players articulate the tacit understandings they build through gameplay. In C. Steinkuehler, K. Squire & S. Barab (Eds.), *Games, Learning, and Society: Learning and Meaning in the Digital Age* (pp. 279-305). Cambridge: Cambridge University Press.
- Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2006, April). Alone together?: exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 407-416). ACM.

- Dede, C., Ketelhut, D., Nelson, B., Clarke, J. & Bowman, C. (2004). Design-based research strategies for studying situated learning in a multi-user Virtual environment. In Y.B. Kafai, W.A. Sandoval, & N. Enyedy (Eds), *Proceedings of the International Conference on Learning Sciences*. Mahwah NJ: Lawrence Erlbaum and Associates.
- Derry, S. J., & Steinkuehler, C. A. (2003). Cognitive and situative theories of learning and instruction. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science* (pp. 800-805). England: Nature Publishing Group.
- Dove, J., Everett, L. A., & Preece, P. F. W. (1999). Exploring a hydrological concept through children's drawings. *International Journal of Science Education*, 21(5), 485-497.
- Drachen, A., & Canossa, A. (2011). Evaluating motion: Spatial user behaviour in virtual environments. *International Journal of Arts and Technology*, 4(3), 294-314.
- Drachen, A., Canossa, A., & Yannakakis, G. N. (2009). Player modeling using self-organization in Tomb Raider: Underworld. In *Computational Intelligence and Games (CIG 2009)*, (pp. 1-8). IEEE.
- Engestrom, Y., Miettinen, R., & Punamaki, R.-L. (1999). *Perspectives on activity theory*. Cambridge: Cambridge University Press.
- Freedman, D.H. (2010). Why scientists are so often wrong: The streetlight effect. *Discover Magazine*. Retrieved 20 March 2016 from <http://discovermagazine.com/2010/jul-aug/29-why-scientific-studies-often-wrong-streetlight-effect>
- Gagné, A. R., El-Nasr, M. S., & Shaw, C. D. (2011). A deeper look at the use of telemetry for analysis of player behavior in RTS games. *Entertainment Computing (ICEC 2011)* (pp. 247-257). Springer Berlin Heidelberg.
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs NJ: Prentice-Hall.
- Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. New York: Routledge.
- Gee, J.P. (2007). *Good video games and good learning: Collected essays on video games, learning and literacy*. New York: Peter Lang.
- Gee, J. P. (2010). *An Introduction to Discourse Analysis: Theory and Method (3rd Edition)*. London: Routledge.
- Gee, J. P., & Hayes, E. R. (2010). *Women and gaming: The Sims and 21st century learning*. Palgrave Macmillan.
- Gibbs, G. R. (2008). *Analysing qualitative data*. London: Sage.
- Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale NJ: Erlbaum. (Original work published 1979).
- Greeno, J. G. (1997). On claims that answer the wrong question. *Educational Researcher*, 5-17.
- Greeno, J. G., & Moore, J. L. (1993). *Situativity and symbols: Response to Vera and Simon*. *Cognitive Science*, 17, 49-59.

- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational psychologist*, 41(2), 111-127.
- Hien, N. T. N., & Haddawy, P. (2007, October). A decision support system for evaluating international student applications. In *Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual* (pp. F2A-1). IEEE.
- Holland, D., & Quinn, N. (Eds.). (1987). *Cultural models in language and thought*. Cambridge: Cambridge University Press.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge MA: MIT Press.
- Jakobson, B. & Wickman, P.O. (2007) Transformation through language use: Children's spontaneous metaphors in elementary school science. *Science & Education*, 16, 267-289.
- Jenkins, H., III. (2006) *Convergence culture*. New York: New York University Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge UK: Cambridge University Press.
- Levinson, S.C. (1983) *Pragmatics*. Cambridge: Cambridge University Press.
- Levy, P. (1999). *Collective intelligence: Mankind's emerging world in cyberspace* (R. Bononno, Trans.). Cambridge MA: Perseus Books.
- Lewis, J. M., Trinh, P., & Kirsh, D. (2011). A corpus analysis of strategy video game play in starcraft: Brood war. In *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 687-692).
- Martin, C. (2012). Video games, identity, and the constellation of information. *Bulletin of Science, Technology & Society*, 0270467612463797.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British educational research journal*, 25(1), 95-111.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33-45). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Nardi, B. (2010). *My life as a night elf priest: An anthropological account of World of Warcraft*. University of Michigan Press.
- Nasir, N. I. S. (2005). Individual cognitive structuring and the sociocultural context: Strategy shifts in the game of dominoes. *The Journal of the Learning Sciences*, 14(1), 5-34.
- Nebot, A., Castro, F., Vellido, A., & Mugica, F. (2006, January). Identification of fuzzy models to predict students performance in an e-learning environment. In *The Fifth IASTED international conference on web-based education, WBE* (pp. 74-79).
- Oldenburg, R. (1999). *The great good place: Cafés, coffee shops, community centers, beauty parlors, general stores, bars, hangouts, and how they get you through the day*. New York: Marlowe.

- Owen, V. E., & Halverson, R. (2013). 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).
- Pardos, Z. A., Heffernan, N. T., Anderson, B., & Heffernan, C. L. (2007). The effect of model granularity on student performance prediction using Bayesian networks. In *User modeling 2007* (pp. 435-439). Springer Berlin Heidelberg.
- Pea, R. D. (1993). Practices of distributed intelligences and design for education. In G. Solomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 47-87). Cambridge, UK: Cambridge University Press.
- Quellmalz, E. S., Davenport, J. L., Timms, M. J., DeBoer, G. E., Jordan, K. A., Huang, C. W., & Buckley, B. C. (2013). Next-generation environments for assessing and promoting complex science learning. *Journal of Educational Psychology*, 105(4), 1100.
- Rennie, L. J., & Jarvis, T. (1995). Children's choice of drawings to communicate their ideas about technology. *Research in Science Education*, 25(3), 239-252.
- Rojas-Drummond, S., & Zapata, M. P. (2004). Exploratory talk, argumentation and reasoning in Mexican primary school children. *Language and Education*, 18(6), 539-557.
- Seay, A. F., Jerome, W. J., Lee, K. S., & Kraut, R. E. (2004, April). Project massive: a study of online gaming communities. In *CHI'04 extended abstracts on Human factors in computing systems* (pp. 1421-1424). ACM.
- Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, 64, 489–528. doi:10.1111/j.1744-6570.2011.01190.x
- Squire, K. (2004). *Replaying history: Learning world history through playing Civilization III*. Unpublished dissertation, Indiana University, Indianapolis, IN.
- Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher*, 35(8), 19–29.
- Squire, K. D., & Giovanetto, L. (2008). The higher education of gaming. *E-Learning and Digital Media*, 5(1), 2-28.
- Squire, K. (2011). *Teaching and participatory culture in the digital age*. New York, NY: Teachers College Print.
- Stake, R.E. (1995). *The art of case study research*. New York: Sage Publications.
- 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.
- Steinkuehler, C. A. (2006). Massively multiplayer online videogaming as participation in a Discourse. *Mind, Culture & Activity*, 13(1), 38-52.

- Steinkuehler, C., & Duncan, S. (2009). Scientific habits of mind in virtual worlds. *Journal of Science Education & Technology*. DOI: 10.1007/s10956-008-9120-8.
- Steinkuehler, C. & Oh, Y. (2012). Apprenticeship in massively multiplayer online games. In Steinkuehler, C., Squire, K., & Barab, S. (Eds.). *Games, learning, and society: Learning and meaning in the digital age* (pp. 154-184). Cambridge: Cambridge University Press.
- Steinkuehler, C., & Williams, D. (2006). Where everybody knows your (screen) name: Online games as "third places." *Journal of Computer-Mediated Communication*, 11(4), article 1.
- Thurau, C., & Bauckhage, C. (2010, August). Analyzing the evolution of social groups in World of Warcraft®. In Computational Intelligence and Games (CIG), 2010 IEEE Symposium on (pp. 170-177). IEEE.
- Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34, 229–243. doi:10.2190/FLHV-K4WA-WPVQ-H0YM
- Vygotsky L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge MA: Harvard University Press.
- Wasson, B., Hanson, C., & Mor, Y. (2016). Grand Challenge Problem 11: Empowering Teachers with Student Data. In Grand Challenge Problems in Technology-Enhanced Learning II: MOOCs and Beyond (pp. 55-58). Springer International Publishing.
- Weber, B. G., & Mateas, M. (2009). A data mining approach to strategy prediction. In *Computational Intelligence and Games (CIG) IEEE Symposium* (pp. 140-147).
- Wertsch, J. V. (1998). *Mind as action*. New York: Oxford University Press.
- Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013, February 4). A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology*. Advance online publication. doi: 10.1037/a0031311
- Young, M.F., Slota, S., Cutter, A.B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & and Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*, 82(1), 61–89.

