Capturing Learner Trajectories in Educational Games through ADAGE (Assessment Data Aggregator for Game Environments): A Click-Stream Data Framework for Assessment of Learning in Play

V. Elizabeth Owen, Dennis Ramirez, Allison Salmon & Richard Halverson
Games+Learning+Society Center @ the University of Wisconsin-Madison
v.elizabeth.owen@gmail.com, dennispr@gmail.com, allisonmariesalmon@gmail.com, halverson@education.wisc.edu

Abstract: A central challenge to educational videogame research is capturing salient in-game data on play and learning. ADAGE (Assessment Data Aggregator for Game Environments) is a click-stream data framework currently being developed by the Games+Learning+Society research center to facilitate standardized collection of embedded assessment data across games. ADAGE integrates core game design structures into a click-stream data (telemetry) schema, which is then seeded with context vital to informing learning analyses. These data can be used to identify patterns in play within and across players (using data mining and learning analytic techniques) as well as statistical methods for testing hypotheses that compare play to content models. Three provided analysis examples show this diversity, combining applied statistics with Markov modeling and classification tree visualization. Overall, ADAGE provides a standardized game telemetry framework with a rich, method-agnostic data yield, efficient enough to have scalability, and flexible enough to use across games.

Introduction and Theoretical Framework

In educational game research, a central challenge is capturing salient in-game data on user experience through the lens of play and learning. A typical approach has been to treat the game as a black box, focusing on data collection via pre- and post- measurements; in relying solely on this, however, we lose the unique characteristics of games as a learning tool. James Gee has suggested that games themselves provide excellent learning assessments (2005). Rather than ignore the motivating and information-rich features of games in capturing learning, designers need to attend to the ways in which game-play itself can provide a powerful new source of assessment data. This requires thinking of games as both intervention *and* assessment; and developing methods for accessing in-game data with a consistent, versatile, context-rich framework for use in learning analysis.

The Games+Learning+Society approach to bridging these worlds is ADAGE (Assessment Data Aggregator for Game Environments), a click-stream (telemetry) data framework that looks inside the data stream of educational games. The GLS Center explores what people learn from engaging with digital worlds, and how the design principles of these worlds can be built into learning environments. Interaction with video games results in a rich data-stream that can potentially provide insight into patterns of player interaction, and ultimately, into learning. ADAGE has helped GLS researchers to addresses a core question in education game assessment: how can we use click-stream data as evidence for learning?

ADAGE (Assessment Data Aggregator for Game Environments)

ADAGE is a protocol for tracking how players interact with game mechanics. It articulates key mechanics for recording (or "tagging") in the game data, and tags concurrent instructional game cues and gameworld context. Logging structures are designed to fit easily into game development practices, yet rich enough to be reliably compare play patterns with learning measures. ADAGE is flexible enough to use across genres, and is currently implemented across GLS games and, in the near future, in games from a variety of developers. ADAGE has three main components: assessment mechanics, a telemetry framework, and data filtering tools.

Assessment Mechanics

The ADAGE Assessment Mechanics (AMs) identify underlying game components that core occasions for player interaction. There are three base types of Assessment Mechanics: *Game Units* (capturing basic play progression), *Critical Achievements* (formative assessment of content), and *Boss Level* (naturalistic summative assessment). These Assessment Mechanics serve as developer specified data-collection anchor points that can be used to construct patterns of play within and across players. AM types can overlap within a gameworld and every game does not have to have all AMs in order to use ADAGE.

Game Units. The game Units represent the core progress mechanic of the game. Game units are repeating, consistent elements of the game that organize play. For example, in a game like *World of Warcraft (WoW)*, the core unit is quests; in a game like Tetris, the core unit is a map. Units can be a nested in a hierarchy– for example, one set of quests may make up a particular map area, and completing all the maps means finishing the game. The definition of a unit (by developers) in ADAGE is flexible enough to work across genres. The concept of Unit is tagged in ADAGE with markers that indicate that players have begun and ended a salient chunk of game play. The Unit informs user experience in setting base interaction with the game environment, a "vital component of design and interaction" (Salen & Zimmerman, 2008, p. 51).

Critical Achievements. Critical Achievements (CA) signal events that developers feel important to learning in the game. Evidence Centered Design (ECD) is an analytic framework which focuses entirely on CA-like structures to document content knowledge (Mislevy & Haertel, 2006) in virtual game spaces (e.g. Clarke-Midura et al., 2012; Behrens et al., 2012). CA events are indicators of the "task model" in ECD, and are intended to embed occasions for direct observation of engagement with content in gameplay. These moments can be compared throughout gameplay to trace a pattern player progress. Embedding CAs in the click stream allows researchers to test whether the anticipated CAs are in fact indicators of learning, and whether other game events are, perhaps, more critical achievements in gameplay than the predicted events. CAs are a unique feature of educational games, and capture both learning AND play dynamics in the user experience.

Boss Level. The Boss Level is a final stage of a game that draws together all the skills learned in gameplay. It is a naturalistic summative assessment, and can include both learning and progress mechanics (like CAs and Units). Gee notes that powerful embedded assessment occurs in "boss battles, which require players to integrate many of the separate skills they have picked up" throughout the game (2008, p. 23). Boss Levels include summative assessment in the data stream, and allow researchers to measure the degree to which the summative experience actually captures player learning. Interaction in the Boss Level shapes user experience as a culminating game encounter, and has also proven significant in ADAGE studies on gameplay progression and learning. For example, in *Progenitor X*, a GLS game about regenerative biology, strong performance in the boss level was predictive of learning gains (Halverson & Owen, 2014).

Telemetry Framework

The telemetry framework transforms concept (AMs) into data (log files). The telemetry framework is defined by the AM schema of data tags that are implemented by developers in the game code. Play data then emerges from the game tagged with the significant events and features specified by AMs. The AMs thus anchor the telemetry schema, which seeds the interaction data with vital contextual information.

The data that result from the telemetry framework must recreate both the experience of the player and the world with which the player interacts. This dual stream model is reflected in two layers of telemetry data - an action-feedback layer that reflects player interaction with the world, and a Virtual Context layer that records the state of the world with which the player interacts. The action-feedback layer captures data on the action-feedback loop (c.f. Salen & Zimmerman, 2008) that comprises interaction between the player and the game. The Virtual Context anchors player interaction in important contextual information such as timestamp, map/unit level, and screen x,y coordinates. These two layers work in tandem to provide context-rich telemetry data on AM-based gameplay trajectories (Figure 1).

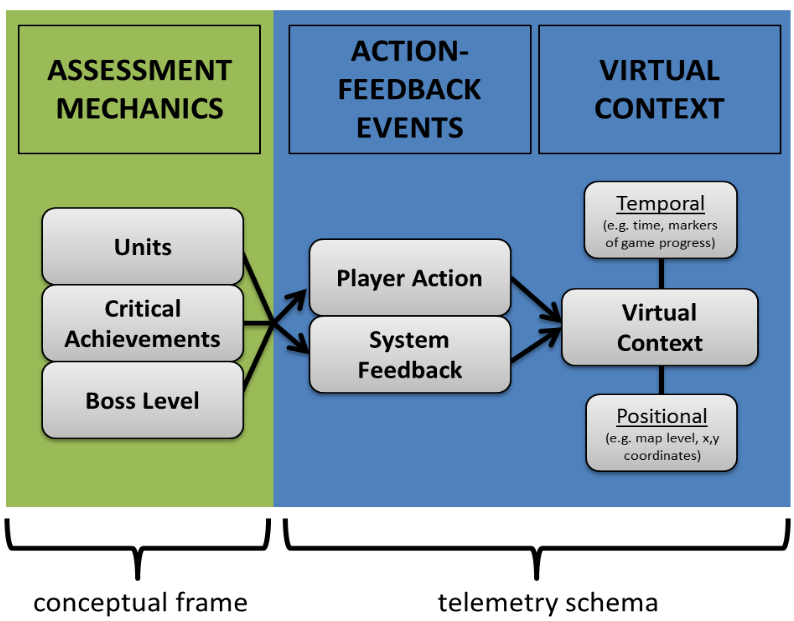

Figure 1: ADAGE Assessment Mechanics and telemetry schema.

Data Filtering

After the raw data from the telemetry schema is tagged, ADAGE features additional processing and filtering affordances. It can build in information about Unit bookends (e.g. the beginning and end of cycles), as well as create performance measures like AM success, failure, and repetition. Data from outside the game, such as pre-post testing, player demographic or academic data can be compared to the patterns of play discerned through ADAGE. Performance measures can be tailored to the research question; for example, one might be interested in Critical Achievement performance (for use with ECD), Unit progression (gamespace trajectory projection), or Boss Level success (in triangulation with a pre-post assessment on learning gains). In the early stages of development, ADAGE data can be used, with usability testing, to improve game design; as the game is played in the wild, ADAGE provides data to determine what players are learning through play.

Application of ADAGE: *Progenitor X*

ADAGE is a data organization strategy designed to fit into any kind of game that generates data about play and learning. To illustrate the range of ways that ADAGE data can provide insight into game play, we turn to a discussion of how ADAGE was implemented in a GLS game, *Progenitor X*[1], and the kinds of analyses and insights that resulted from the data. *Progenitor X* is a puzzle-based zombie game about stem cell biology (playable from the footnote link). Early in the game, players are infected by a zombie strain and must transform pluripotent cells into tissue, then into zombie-free organs.

Utilizing ADAGE data on *Progenitor's* main units of gameplay (cycles and objectives), boss level, and pre-post stem-cell content assessment, three complementary studies were conducted. The data were collected in the summer of 2012 from 110 middle-school students who played *Progenitor X* as part of a summer school curriculum (either in their Dane County classroom, or on-site at the Wisconsin Institute for Discovery). The subsequent investigations leveraged a blend of machine learning, data mining (e.g. Baker & Yacef, 2009; Romero & Ventura, 2010) and statistical methods to explore learning and play patterns. A synopsis of each is given below (to be presented in detail at the 2014 AERA conference).

Investigation One

Using hypothesis-driven statistical methods, we used ADAGE data to draw conclusions about the connections between play patterns and pre-post learning gains. Essentially, a non-parametric analysis revealed that learning gains had a significant positive correlation with gameplay progress and overall game success (Table 1). Shades of failure and success at specific points in the game also had connections to learning, as both aggregate and learner quartile analysis showed. For example, off-task failure overall (Table 2) and especially early in tutorial levels (Table 1 & 2) was negatively related to pre-post gains; in contrast, success in the boss level was positively related to learning gains (Tables 1 & 2).

	Pre-Post Gains
Total Gameplay	19.5% average increase
Objectives Completed	Significant positive correlation (r = +.272, p = .002)
# of Successful Cycles	Significant positive correlation (r = +.216, p = .012)
Tutorial Level Off-Task Failure	Significant negative correlation (r = -.167, p = .04)
Boss Level Success	Significant positive correlation (r = +.272, p = .002)

Table 1: Patterns across all *Progenitor X* players.

[1] http://www.gameslearningsociety.org/project_progenitor_x.php

	Upper Quartile of Pre-Post Learners	Lower Quartile of Pre-Post Learners
# of Off-Task Failures per Objective	1	2
# of Tutorial Level Off-Task Failures	2	4
Total # of Failures	No significant difference.	
# of Successful Boss Level Cycles	3	2
Total # of Successful Cycles	No significant difference.	

Table 2: Patterns across groups of students in the upper and lower quartiles of pre-post learning gains.

Investigation Two

The second study combined machine learning and statistical methods to trace player trajectories through the learning space of *Progenitor X*. The temporal likelihood modeling of Markov chains (Rabiner, 1986) helped us capture naturalistic player navigation through the gamespace. We compared the gameplay trajectories of two groups of students – those who finished the game, and those who did not (given the same amount of time). Non-parametric correlations of completion vs. specific cycle patterns were used along the way to support the Markov analyses. Three interesting themes emerged from the comparison between the two student groups (Figure 2). First, high repetition in early tutorial levels signaled dropout; also, students' persistence in learning the second major skill mechanic (the tissue cycle) was crucial in completion; and lastly, the mid-levels of the game (which involved synthesis of several skills) emerged as critical quitting points.

Figure 2: First-order Markov model of players who did not finish the game (contrasts with finished players marked in yellow).

Investigation Three

The third analysis used predictive analytics to explore the relationship between embedded performance metrics and *Progenitor* game completion. We used machine learning software *WEKA* (Hall et al., 2009) with a variety of ADAGE data on general unit progression & boss level performance, including success, off-task failure, and on-task failure with each unit type. Plugging these data into a J48 classification tree (Figure 3), we found that more successes in the last played objective, lower far failure per objective, and less objective 1 repetition predicted game completion – results which corroborated the findings of the prior two analyses.

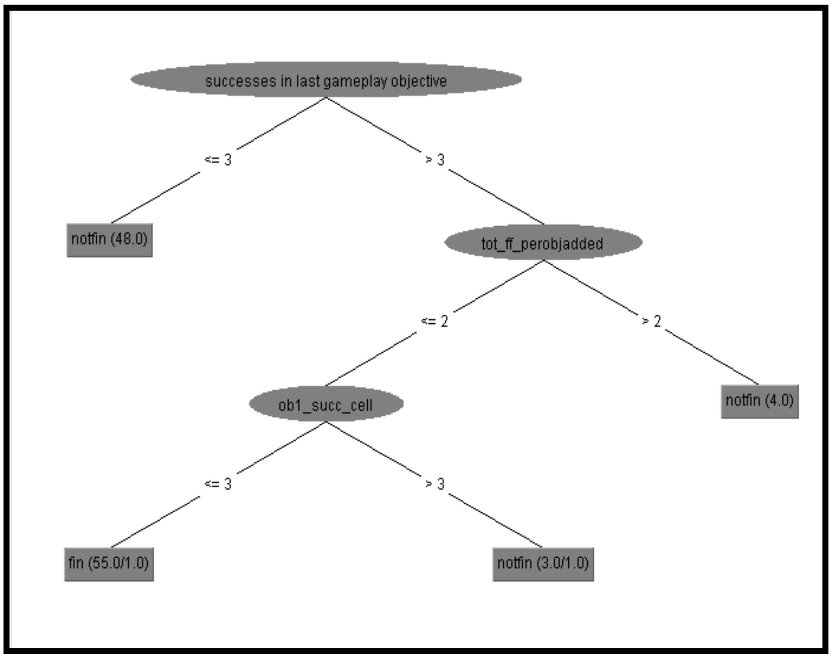

Figure 3: J48 classification tree predicting finished/non-finished game outcomes.

Discussion and Conclusion

By capturing trajectories of player experience via context- rich interaction with core mechanics in the educational gamespace, ADAGE connects design and user experience. It then extends that connection to a standardized framework for collecting salient click-stream data on play and learning. As we can see in the trio of analyses above, these data can be used to identify patterns in play within and across players (using data mining and learning analytic techniques) as well as statistical methods for testing hypotheses that compare play to content models. ADAGE provides a scalable and flexible game telemetry model that provides a standard for transforming click-steam data into evidence for learning analysis.

Acknowledgments

This work was made possible by a grant from the National Science Foundation (DRL-1119383), although the views expressed herein are those of the authors' and do not necessarily represent the funding agency. We would also like to thank the *GLS Center* team, including Kurt Squire, Craig Kasemodel, Becky Torrisi, Nate Wills, Mike Beall, Ted Lauterbach, Jodi Asbell-Clarke, Meagan Rothschild, Shannon Harris, and Ben Shapiro.

References [2]

Baker, R. S. J. D. & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*(1), 3-17.

Behrens, J. T., Mislevy, R. J., DiCerbo, K. E. & Levy, R. (2012). Evidence centered design for learning and assessment in the digital world. In M.C. Mayrath, J. Clarke-Midura, D.H. Robinson, G. Schraw (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 13-53). Charlotte, NC: Information Age Publishing.

Clarke-Midura, J., Code, J., Dede, C., Mayrath, M. & Zap, N. (2012). Thinking outside the bubble: virtual performance assessments for measuring complex learning. In M.C. Mayrath, J. Clarke-Midura, D.H. Robinson, G. Schraw (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 125-148). Charlotte, NC: Information Age Publishing.

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning and Digital Media*, *2*(1), 5-16.

Gee, J. P. (2008). Learning and games. *The ecology of games: Connecting youth, games, and learning*, *3*, 21-40.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18.

Halverson, R. & Owen, V.E. (2014). Game-Based Assessment: An integrated model for capturing evidence of learning in play *[Special Issue: Game-Based Learning]*. *International Journal of Learning Technology.*

Mislevy, R. J. & Haertel, G. D. (2006). Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice*, *25*(4), 6-20.

Salen, K. & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press.

Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. *ASSP Magazine, IEEE*, *3*(1), 4-16.

Romero, C. & Ventura, S. (2010). Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, *40*(6), 601-618.

[2] **<u>This Paper's Citation (APA)</u>:**

Owen, V.E., Ramirez, D., Salmon, A. & Halverson, R. (2014, April). *Capturing Learner Trajectories in Educational Games through ADAGE (Assessment Data Aggregator for Game Environments): A Click-Stream Data Framework for Assessment of Learning in Play.* Presented at the 2014 American Educational Research Association Annual Meeting, Philadelphia, PA.