ADAGE (Assessment Data Aggregator for Game Environments): A Click-Stream Data Framework for Assessment of Learning in Play

V. Elizabeth Owen, Richard Halverson
Games+Learning+Society Center at the University of Wisconsin-Madison
vowen@wisc.edu, 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 group to facilitate standardized collection of in-game 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 (cf. Loh, 2013; Halverson & Owen, in press). ADAGE assessment structures also inform iterative, data-driven design of GLS games. 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. Well-designed games reward players for mastering content and strategies, scaffold player activities toward greater complexity, engage players in organized social interaction toward shared goals, and provide feedback that allows players to monitor their own progress (Gee, 2005). Rather than ignore the motivating and information-rich features of games in capturing learning, designers need to attend to the ways in which gameplay 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.

Well-designed games are examples of situated learning environments in which learning exists *in situ*, inseparable from environment or context (c.f. Brown et al., 1989; Greeno, 1997). Virtual game worlds have been shown to provide a powerful environment for learning, supporting apprenticeship and collective higher-order thinking skills (Steinkuehler, 2004; Steinkuehler & Duncan, 2008). Videogames afford this environment by providing *designed experiences* in which players explore worlds to understand how knowledge and skills interact in a context (Squire, 2006). From a player perspective, good video games include just-in-time information and cycles of expertise that scaffold play experience. The data channels available to the player act as formative feedback displays which inform play. To maintain this immersive context for learning, good games consist of ongoing assessment balanced with engaging mechanics and narrative (Squire, 2006). Games can thus provide an experience which is distinct from – but relies upon – the core design mechanics of the game. Game design icon Jesse Schell is careful to distinguish early in the design process that “the game is NOT the experience” (2008, p. 10; see Figure 1). Salen and Zimmerman assert that “the careful crafting of player experience through a system of interaction is critical” (2008, p. 61). Additionally, in moments of transgressive play, users often interact with the gamespace in unanticipated ways (Salen & Zimmerman, 2008). How, then, can we further explore the connection between design, interaction, and experience? Applied specifically to educational games, how does it then connect with in-game data collection for assessment of learning?

The GLS approach to bridging these worlds is ADAGE (Assessment Data Aggregator for Game Environments), a click-stream (telemetry) data framework that looks inside the black box of educational games. ADAGE identifies key gameplay verbs as occasions for interaction, providing a click-stream data framework for collecting evidence of learner trajectories. In looking at in-game data, we avoid the “Heisenburg” problem of usertesting – that a user experience “cannot be observed without disturbing the nature of that experience” (Schell, 2008, p. 18). As Val Shute notes, telemetry-based assessment can be a “quiet, yet powerful process” through which we can unobtrusively observe player patterns (2011, p. 504). However, with the affordance of subtlety comes the problem of abundance; log files from digital spaces can produce millions of data points with little to no context (c.f. Baker & Yacef, 2009). ADAGE addresses this core question specifically for educational games: how do we identify, record, and output click-stream data salient to learning analysis?

ADAGE (Assessment Data Aggregator for Game Environments)

ADAGE was designed to transform game-based log file data into evidence of learning. It articulates a bridge between educational game design and player experience, which is then structurally integrated into a framework for an otherwise inchoate mass of log data. ADAGE organizes click-stream data framework that allow developers and researchers to trace trajectories of player experience by tracking interaction with core mechanics in the educational gamespace. It articulates key mechanics for recording (or "tagging") in the game data, and tags concurrent instructional game cues and gameworld context. The ADAGE tagging procedures are developed to create minimal interference with the development process, yet to yield data rich enough to be make inferences about learning. Because it builds on features core to educational game design, ADAGE is flexible enough to use across genres, and is currently implemented in four vastly different GLS games.

Below, we will identify and describe ADAGE assessment mechanics and telemetry features. Together these layers create context-rich raw click-stream data that can be filtered and processed data into sequential blocks or performance indices, facilitating the feature engineering process vital to later analysis.

Assessment Mechanics

Assessment mechanics are structures built into the game that allow for research on play and learning. Understanding game-based learning requires two levels of assessment mechanics: one to trace the paths players take through a game, and the other to access the player experience of game play (Schell, 2008). Squire asserts that games as designed experiences (2006) provide endogenous engagement (Costickyan, 2002) for the player through "roles, goals, and agency" (Squire, 2011, p. 29). Thus, in learning games, there can two core kinds of designed mechanics: one set related to progression through the gameworld (as an engaging learning context [Gee, 2007; Salen & Zimmerman, 2008]); another may be designed as more direct measures of the content the game is trying to teach (e.g. Clarke-Midura et al., 2012). Ideally, these also overlap; good educational games meld learning mechanisms with the core mechanics of the game, where gameplay itself is the only necessary assessment (Gee, 2012; Shute, 2011).

The ADAGE framework identifies underlying game mechanics for which serve as 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). As "Assessment Mechanics", they serve as data-collection (or assessment) anchor points, which yield data informed by core educational game design structures. This terminology also parallels concepts of formative and summative assessment in formal learning environments (Harlen & James, 1997), and formalizes them as powerful elements of game design (c.f. Gee, 2012).

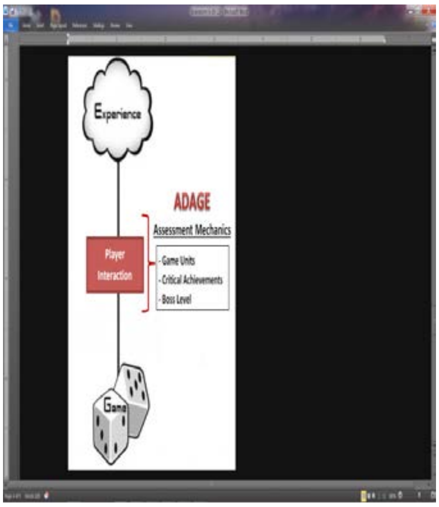

Figure 1: Schell's distinction between player experience and game design (2008, p.23); ADAGE assessment mechanics as bridge between.

Through Assessment Mechanics (AMs), ADAGE operationalizes player interaction (Salen and Zimmerman, 2008) as the vital link between experience and game design (Schell, 2008; Figure 1). These three core AM types can easily overlap within a gameworld; they are not mutually exclusive, though they have distinct categories. Additionally, every game does not have to have all AMs in order to use ADAGE. In this section, we will describe each mechanic, and connect it to ADAGE's underlying telemetry structure.

Game Units. The game Units represent the core progress mechanic of the game. For example, in a game like *World of Warcraft (WoW)*, the core unit is quests. By definition, game units have the property of being a repeating, consistent vehicle for making progress through the gameworld. Units can also be part of a hierarchy – for example, one set of quests may make up a particular map area, and completing all the maps means finishing the game. Thus, from broadest to smallest, game Unit hierarchy might be: game-map-quest. The idea behind Units is that they are flexible enough to work across genres; for example, in Tetris, the core Units are level completion and placement of shapes (different from *WoW*'s quest structure). Currently, ADAGE Unit structure is applied to five different GLS games (*Progenitor X, Fair Play, Anatomy Pro Am, Tenacity,* and *Crystals of Kaydor*) each with different genres and Unit types. The concept of Unit is logistically integrated into ADAGE's telemetry, with the term specifically connected to click-stream tags in ADAGE's API. The Unit AM 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 (CAs) in ADAGE are direct formative assessment slices of the content model (what the game is trying to teach). They are moments of direct content measurement within the context of normal gameplay. Seamlessly woven into the fabric of the game, CAs use naturalistic game mechanics to measure underlying educational content. For example, *Fair Play* is a GLS game which teaches about implicit bias in graduate education settings. In one *Fair Play* CA, the player needs to correctly identify a given bias to another character in order to progress. This is a direct demonstration of bias knowledge (as opposed to indirect movement through the learning context, like in game Units). Evidence Centered Design (ECD) is an analytic framework which focuses entirely on CA-like structures – direct demonstration of content knowledge (Mislevy & Haertel, 2006), recently applied to virtual spaces (e.g. Clarke-Midura et al., 2012; Behrens et al., 2012). For this reason, the CA data structure aligns very well with ECD-specific analyses. CAs (analogous to the "task model" in ECD) are intended to be one kind of direct content assessment embedded in gameplay, looking at selected moments of performance as learning measures. These moments can be compared throughout gameplay to give one snapshot of learning growth; moving beyond a task model, they can also be triangulated with ADAGE mechanisms like broader gameworld interaction data (Units), boss level performance, and pre-post learning measures. Although CAs are a great educational game design feature that lends to robust learning analysis, games don't have to contain CAs to use the ADAGE framework. The concept of CA formative assessment is manifested logistically in ADAGE's click-stream data structure, with CA-specific terminology in the API. Ultimately, 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 is a culmination of 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). Games are an ideal medium for this summative assessment, he asserts, since they can provide just-in-time performance feedback with low cost of failure (Gee, 2007). Thus, summative assessment mechanics in games can give us an unobtrusive measure of performance (c.f. Shute, 2011) in an agency-inspiring context (Squire, 2011) in which players receive instant feedback and appealing opportunity to improve (Gee, 2007). By formalizing the Boss Level as an Assessment Mechanic in ADAGE, we encourage deliberate inclusion of summative assessment in game design, and provide corresponding telemetry API structures for implementation. 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, in press).

Telemetry Framework

The Assessment Mechanics, informed by game design and assessment research, create a conceptual framework for identifying interaction data. The next ADAGE step moves us from concept (AMs) to implementation (telemetry). The telemetry framework hinges on the AMs to create a schema of context-rich data tags for implementation in the game code. Interpretation of student interaction often hinges on the context of the learning environment (in this case, the designed gameworld). The telemetry schema addresses this need by seeding the AM interaction data with vital contextual information.

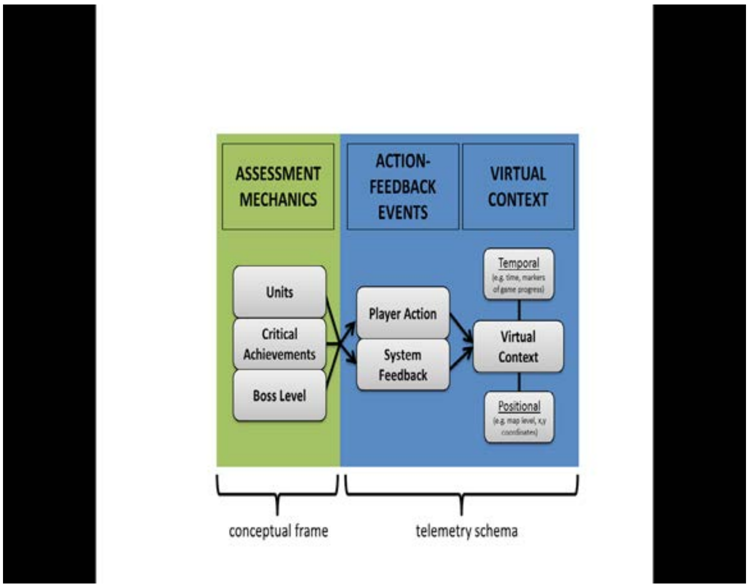

Figure 2: ADAGE Assessment Mechanics and telemetry schema.

The telemetry schema has two layers: an action-feedback layer, and a Virtual Context layer. First, for each Assessment Mechanic, it identifies two sources of direct interaction: user action, and system feedback. It articulates the vital action-feedback loop (c.f. Salen & Zimmerman, 2008) that comprises interaction between the player and the game. The second layer, called the Virtual Context, attaches important contextual information to each action-feedback event. The Virtual Context can include things like timestamp, map level, and screen x,y coordinates. These two layers work in tandem to provide context-rich telemetry data on AM-based gameplay trajectories (Figure 2).

One example of the applied telemetry schema is in the game *Progenitor X*. *Progenitor* is a puzzle-based zombie game about stem cell biology (playable from the footnote link). The core Units of the game are cycles of cell, tissue and organ creation. Table 1 applies the telemetry framework to a single cycle. In column 1, we identify the Assessment Mechanic – a Unit, specifically the first game cycle. Column 2 asks: for the start of that cycle, game cues are going on? To help the player begin, the game makes the start button flash. The feedback event becomes "Start button flashes". Next comes the corresponding player action for Column 2, which is "Player clicks 'start' button". Lastly, for each of the action-feedback events, we define the contextual information we need (column 3). To understand player progress, we attach information about which map the player is on, and elapsed time. Location of click is also recorded, in case heat mapping or place-based performance analysis is desired. The resulting Virtual Context is "Timestamp," "Map Level," and "x,y Coordinates."

Unit	Action-Feedback Events	Virtual Context
1st Cycle	Start button flashes Player clicks "start" button	Timestamp Map Level x,y Coordinates

Table 1: Telemetry schema example: *Progenitor X*

In implementing this framework, this process is completed for every sequential Assessment Mechanic in the game. In other words, each unit, critical achievement, and boss level section is laid out sequentially, then mapped to action-feedback events and Virtual Context. More detailed process information and templates are laid out in ADAGE's DevDoc, a working document for connecting ADAGE with new games. However, ADAGE's core telemetry structure is presented here, centered on the AM sequence, the action-feedback events, and the Virtual Context. Each of these elements has a counterpart in ADAGE code, mapping conceptual AMs to click-stream structures of user actions, system feedback, and the Virtual Context around each.

Raw Data. Essentially, ADAGE identifies core game design features that provide occasion for interaction. It then delineates a framework for tagging this data in the massive influx of click-stream input, and attaches systematic contextual information to each data point. This, in turn, produces an abundant stream of telemetry data informed by the game design structures. Raw ADAGE data contains all action-feedback data of each AM in the game, enriched with the telemetry structure's Virtual Context (Figure 3). The beauty of this rich stream is that it gives contextual data raw enough to be used in almost any analysis.

ADAGE 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. 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).

Feature Engineering & Analysis Lenses

ADAGE's context-rich data make ideal building blocks for feature engineering. Features are essentially variables of interest in the data, which can range from simple click locations to complex measures like accuracy over time. Features of interest across a variety of methods can be generated from ADAGE output, including evidence model performance (ECD), quantitative ethnographic data (c.f. Efferson et al., 2007), or sensor-free affect detectors (Baker et al., 2012).

The features constructed, in turn, can be used across a broad range of analysis techniques. Data lenses can include descriptive statistics, hypothesis-driven applied statistics, and machine learning techniques. For general descriptive stats, ADAGE data can be used for simple aggregation of behaviors in the gamespace, including figures of average elapsed time, number of units completed, time per level, etc. Hypothesis-driven applied statistics (used in methodologies like ECD) can use ADAGE data as dependent variables, independent variables, and covariates for use in associative or predictive modeling. Specific to educational games, this often means testing hypotheses that compare play to content models (cf. Loh, 2013; Halverson & Owen, in press). Lastly, ADAGE data lends itself to learning analytic techniques often used with big data sets. Recent "State-of-the-Art" reports in Educational Data Mining (Baker & Yacef, 2009; Romero & Ventura, 2011) articulate various machine learning analysis techniques used with log file data. These include Social Network Analysis, classification and regression trees, cluster analysis, Markov chain modeling, and Bayesian networks. GLS researchers have also utilized ADAGE data to create heatmaps of most frequently visited in-game areas.

Design Implications 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. 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 assessment structures also serve to inform iterative, data-driven design of GLS games. The articulation of formative and summative Assessment Mechanics inform core educational game design. ADAGE AM data are also utilized as well in the iterative data-driven design process. In the recent GLS Tenacity project, a collaboration with the Center for Investigating Healthy Minds, early usertesting telemetry informed design refinements during game development (Owen et al., 2013). Additionally, ADAGE data output can be used to inform adaptive tutorial help overlays, potentially providing pivotal support for learners in hotspots of game dropout or failure.

ADAGE bridges design and experience, while creating a standard framework for producing salient telemetry data of play and learning. It encourages best practices in iterative game design, specifically around integrated formative and summative assessment mechanisms in gameplay. Overall, it 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. Through integration of content, design, and interaction data, design efforts like ADAGE model technology standards for transforming click-steam data into evidence for learning analysis.

References

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.

Baker, R. S., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A.,Aleven, V., Kusbit, G., Ocumpaugh, J. & Rossi, L. (2012, June). Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. In *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 126-133).

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.

Brown, J. S., Collins, A. & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.

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.

Costikyan, G. (2002). I Have No Words & I Must Design: Toward a Critical Vocabulary for Games. F. Mäyrä (Ed.), *Proceedings of Computer Games and Digital Cultures Conference*. Tampere: Tampere University Press.

Efferson, C., Takezawa, M., & McElreath, R. (2007). New Methods in Quantitative Ethnography. *Current Anthropology*, *48*(6), 912-919.

Greeno, J. G. (1997). On claims that answer the wrong questions. *Educational researcher*, *26*(1), 5-17.

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. (2007). *What video games have to teach us about learning and literacy*. New York, NY: Palgrave Macmillan.

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

Gee, J.P. (2012). Games can drive assessment to a new place. Retrieved from http://gamesandimpact.org/wp-content/uploads/2012/09/Games-Can-Drive-Assessment-to-a-New-Place.pdf

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

Harlen, W. & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education*, *4*(3), 365-379.

Martinez, R., Yacef, K., Kay, J., Kharrufa, A. & Al-Qaraghuli, A. (2011). Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, and J. Stamper (Eds*.), Proceedings of the 4th international conference on Educational Data Mining.* Eindhoven, the Netherlands.

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

Owen, V.E., Slater, S., Solis, E., Smith, A., Farajian, R.,Beall, M., Steinkuehler, C. & Davidson, R. (2013b). Games for Mindfulness and Pro-Social Behavior: the Tenacity Project Collaboration. In A. Ochsner (Ed.), *Proceedings of the 9h Annual Games+Learning+Society Conference.* Pittsburg PA: ETC Press.

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

Schell, J. (2008). *The Art of Game Design: A book of lenses*. Taylor & Francis US.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer games and instruction*, *55*(2), 503-524.

Squire, K. (2006). From content to context: Videogames as designed experience. *Educational researcher*, *35*(8), 19-29.

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 6th international conference on learning sciences* (pp. 521-528). Mahwah, NJ: Erlbaum.

Steinkuehler, C., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology*, *17*(6), 530-543.

Squire, K. (2011). *Video games and learning: teaching and participatory culture in the digital age: Technology, education--connections (the TEC series)*. New York, NY: Teachers College Press.

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.

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 Allison Salmon, Kurt Squire, Craig Kasemodel, Dennis Ramirez, and Becky Torrisi.