Title: The Moment of Learning: Quantitative Analysis of Exemplar Gameplay Supports CyGaMEs Approach to Embedded Assessment

Author(s):

Debbie Denise Reese—Center for Educational Technologies, Wheeling Jesuit University, WV
Barbara G. Tabachnick—California State University, Northridge
The CyGaMEs Group

Background/Context:

We summarize a quantitative analysis demonstrating that the CyGaMEs toolset for embedded assessment of learning within instructional games measures growth in conceptual knowledge by quantifying player behavior. CyGaMEs stands for Cyberlearning through GaME-based, Metaphor Enhanced Learning Objects. Some scientists of learning claim that all cognition is situated, and the only way to effectively study cognition is within authentic contexts (e.g., Brown, Collins, & Duguid, 1989; Greeno, 1997). CyGaMEs assessment does not violate those assumptions. CyGaMEs assessment is authentic because it is embedded in gameplay. The CyGaMEs assessment toolset keeps track of each player's procedural gameplay activity. That is, CyGaMEs measures learning by tracking player behavior. For the player of a CyGaMEs instructional game, progress toward the procedural game goal requires concurrent discovery and application of the targeted concepts. Thus, the game requires the player to use procedural gameplay to build conceptual knowledge. CyGaMEs assessments are algorithms that quantify player gameplay activity and progress toward the game goal as measures of learning. Within this paper we statistically demonstrate the accuracy and sensitivity of the CyGaMEs assessments. We introduce a moment of learning method, a quantitative methodology conceived by the first author in collaboration with CyGaMEs partner Larry Hedges to quantify the degree to which CyGaMEs tools assess learning. Our findings should help assuage critics who might question our claim that CyGaMEs assessment measures learning. A vision of equity and achievement in 21st century education has motivated federal agencies, national organizations, and private foundations to launch initiatives studying how cyberlearning technologies can enhance learner-centered education through game-based instructional environments and embedded assessments (e.g., Borgman et al., 2008; Laughlin, Roper, & Howell, 2006; The Learning Federation Project, 2003; http://digitallearning.macfound.org/). Those agencies and organizations call for the development of assessment toolsets that can be shared by researchers, developers, and learning environments. This CyGaMEs work supports the development of accurate and authentic instructional game assessment toolsets. Such assessments are essential if education is to enhance its responsiveness to the needs and strengths of each individual learner.

The CyGaMEs method is a theory-based approach to instructional game design, embedded assessment, and research (Reese, 2007b, 2009, in press). The CyGaMEs approach translates targeted abstract concepts into concrete, procedural game worlds. In other words, the CyGaMEs method translates what domain experts *think* into something domain novices *do*. The CyGaMEs approach to instructional game design and assessment derives from a synthesis of theories and methods: cognitive science analogical reasoning, game design, instructional design, learning science, and flow. Briefly, CyGaMEs applies structure mapping theory and pragmatic constraints (Gentner, 1983; Holyoak, Gentner, & Kokinov, 2001; Kurtz, Miao, & Gentner, 2001) to specify the design and development of a game world that is relationally isomorphic (consistent and in one-to-one correspondence) with the targeted conceptual domain. The target domain becomes the base for its game world analog. This is possible because game worlds are virtually concrete relational systems, and gameplay is designed to support game goals (Bogost, 2006; Fullerton, 2008; Schell, 2008; Wright, 2003, 2004, 2006). The relational structure of the targeted domain becomes the relational structure game world, the learning goal becomes the game goal, and this makes player progress toward the game goal a quantifiable, behavioral measure of player attainment of the targeted learning goal. The CyGaMEs assessment tool measuring player progress toward the game goal is the *timed report*. CyGaMEs also captures each player's interaction with the game world at the level with which the player changes the game state. This assessment tool is the *gesture report.* Flow theory (Csikszentmihalyi & Csikszentmihalyi, 1988; Csikszentmihalyi & Larson, 1987; Csikszentmihalyi & Schneider, 2000; Hektner, Schmidt, & Csikszentmihalyi, 2007) is an integral component of game design. Every designer attempts to inspire flow—that state in which the player loses all sense of time and self, immersed in the experience of gaming. The first author designed an assessment tool to measure the degree to which CyGaMEs instructional games place players in flow and how flow and the other dimensions (apathy, boredom, routine expertise, control, arousal, anxiety, and worry) interact with learning. The tool is the flowometer, and it produces a *flow report*.

Playing a CyGaMEs game prepares the player to make viable inferences about targeted learning. These inferences serve as prior knowledge. Apt prior knowledge for a targeted domain makes future learning of that domain more intuitive. Learning scientists call this process "preparation for future learning" (Schwartz & Bransford, 1998; Schwartz, Bransford, & Sears, 2005; Schwartz & Martin, 2004). Instructional designers prescribe this process as an event of instruction that activates and/or develops apt prior knoweldge (Gagné, 1965; Gagné, Briggs, & Wager, 1992; Merrill, 2002). Daniel Schwartz and Taylor Martin have developed an experimental design for use in research when interventions are designed to prepare learners for knowledge acquisition (Schwartz et al., 2005; Schwartz & Martin, 2004). CyGaMEs adapted this double transfer paradigm and applied it to design a research environment for studying how game-based learning assists learners to construct preconceptual mental models for targeted concepts (see Figure 1).

(Please insert figure 1 here.)

Purpose/Objective/Research Question/Focus of Study:

We wanted to identify a prototypical moment of learning using the gesture report and then confirm that the timed report could identify when people had accomplished that moment of learning. We asked:

- Can the gesture report identify a prototypical learning moment, the players who have achieved it, and the time it occurred?
- Can the timed report also identify if players have achieved the learning moment? That is, if we use the time at which gestures indicate the learning moment occurred, will the timed report identify an increase in player performance after the learning moment?

Setting:

The CyGaMEs environment comprises (a) the three embedded assessments, (b) the concrete game analog, and (c) the interstitial research environment, *Selene: A Lunar Construction GaME. Selene* is one complete CyGaMEs environment (Reese, 2007a, 2008, 2009, in press). *Selene* is available online to registered players 24/7. The current version of the game is authored in Java and set within a Flash shell that delivers instructional movies and external assessments. *Selene* players slingshot particles to build the Earth's Moon (accretion), and then change it over time by peppering its surface with impact craters and flooding it with lava. As specified for *Selene*, this domain of lunar science contains 101 interrelated subconcepts.

Population/Participants/Subjects:

The first author triangulated video and gameplay data of one female undergraduate from a

Midwest state university psychology pool who self-selected to participate in a collaborative version of the *Selene* study for research credit. We then used the insights gathered by that triangulation to analyze 22 sets of participant gameplay selected from two phases of data collection (study phase 1 *N*=554 and study phase 2 *N*=119) when player data met inclusion criteria. Phase 1 player ages ranged from 13-17 (*N*=16; female=6, male=10). The typical phase 1 player was 15 years old and attending school grade 9 with a self-reported GPA of B and living in Arkansas (3), Arizona (1), Mississippi (1), Missouri (2), New Jersey (1), New York (2), Ohio (2), Oregon (1), or Pennsylvania (2). The majority of phase 1 players are white=11 (African-American=2, mixed=1, Other=2). About 60 percent of the phase 1 players reported parent's education ended at high school. The other 40 percent reported their parents had earned college or graduate degrees. The six phase 2 undergraduates had self-selected from the same Midwest university psychology pool for research credit (μ_{grade}=sophomore; $\mu_{GPA\ (self\text{-}reported)}$=C/B-; female=4, male=2). Each reported father's level of education as college; two reported mothers had completed college, and four reported mothers had terminated education at high school.

Intervention/Program/Practice:
All participants used an access code to log in and play the *Selene* game. The collaborative study player was supervised by a researcher in a lab setting and videotaped in a computer lab. Phase 1 participants were recruited by *Selene* adult volunteers (e.g., educators, parents, club leaders, etc.) who supervised informed consent and issued access codes. Phase 1 and phase 2 participants could play the game 24/7, independently, at any location. Players have taken from 45 minutes to 3-4 hours to complete the entire *Selene* environment. Data within this analysis are drawn from the first section of accretion module round 1 gameplay (accretion scale 1) and examined before and subsequent to a learning moment. The learning moment occurred at an idiosyncratic time for each player. These players took an average of 9.3 minutes to complete scale 1 ($\mu_{prelearning}$=6.0 and $\mu_{postlearning}$=3.4)

Research Design:
The *Selene* environment is constructed for randomized field trials using an adapted double transfer experimental design (see Figure 1). The design implemented to triangulate video and gameplay data for the single collaborative study participant could be partially characterized as a quantitative case study. This moment of learning analysis uses quantitative repeated measures of around 1 accretion gameplay behavior that occurs before players are differentially routed through the game. Phase 1 players who watch gameplay during round 1 were excluded. Phase 2 players were part of a larger study in which they also completed one of two pregame external assessments. Phase 1 players were not exposed to the external assessments.

Data Collection and Analysis:
Selene timestamps all data and sends it to a database. Two *Selene* embedded assessment tools measure learning:

- Timed report: A timed report is the score of player's progress toward the game goal calculated every 10 seconds of gameplay. We interpret the scoring as continuous data, calculated for interpretation as "-1" (away from goal), "0" (no progress), or "1" (toward goal).
- Gesture report—slingshot: A gesture is a player- or game-initiated event (behavior) that changes the game state. Each gesture has parameters. During accretion scale 1, the player

initiates the slingshot gesture by selecting a particle from a ring around the early Earth and shooting it into a protomoon. The slingshot velocity parameter reports the speed of the launch.

Accretion is the concept that high kinetic energy collisions cause fragmentation and low kinetic energy collisions cause accretion (particles stick together). *Selene* players learn to correctly execute accretion via idiosyncratic learning pathways. Using a moment of learning method, the first author reviewed video footage of a single player's gameplay to identify a prototypical accretion learning moment. This learning moment, accretionLM, is the instant at which a player's *behavior* transitions from initiating very high velocity slingshot gestures to sustained low velocity slingshots. The same author triangulated this player's video corpus with the player's gesture slingshot report (velocity) and timed report data. Both of these embedded assessments bifurcated at the learning moment as expected (see Figures 2 and 3). Triangulation confirms the existence and characteristics of accretionLM. It demonstrates that timed report accurately reflects accretionLM for this player. The next step in the moment of learning method is to identify the prototypical learning moment in other players. If the timed report does, indeed, describe accretion learning, then we should be able to look at people who have accretionLM and see a change in their progress at that learning moment. The first author ran scatterplots of all players' velocity data and identified 22 exemplar players who met inclusion criteria for learning moment analysis (i.e., initial high velocity gameplay followed by sustained, low velocity gameplay). She graphed velocity traces for each exemplar and identified the time (in milliseconds) each exemplar's learning moment occurred. She used each exemplar's time to split that exemplar's timed report data into pre/post learning moment. The authors analyzed these data as repeated measures using multilevel modeling on report (trial) level data and using the general linear model on data aggregated within player by pre/post learning moment.

(Please insert figures 2 and 3 here)

Findings/Results:

Multilevel Modeling of Timed Report

A number of preliminary hierarchical models were analyzed through HLM 6.07 software to determine whether factors such as study phase (group types) and slopes of sequence within sets of trials (trials before learning vs. trials at the point of learning and after) added significantly to prediction of timed report changes averaged within each trial set (labeled learning). Full maximum likelihood estimation permitted comparisons among models with and without these factors and showed that neither factor aided the fit of the model to the responses ($p > .05$). Designating timed report changes averaged within each trial set (learning) as a random factor in a three-level model appeared to provide a better-fitting model over one in which learning was designated a fixed factor; however, no significance test comparing the two models is possible, nor was there any difference in interpretation of the learning effect. Therefore, the simpler, 2-level model is reported, based on analyses using restricted maximum likelihood estimation to provide more accurate results. First-level units of the multilevel model were trials for which velocities were measured, a total of 1,232, with the number of trails varying among participants. Second-level units were the 22 participants. A model based on individual differences alone, without predictors, permitted calculation of the variance associated with individual differences. Although there were significant differences among participants (measured as a random effect), $\chi^2(21) = 51.30$, $p < .001$, the intraclass correlation was found to be quite small, $\rho = .023$. This suggests that the multilevel modeling approach may not be necessary, but it does provide some insights beyond those revealed by repeated-measures ANOVA. Table 1 displays the results of the final two-level model with a single fixed predictor: learning. Table 1 shows that for every one-unit change in learning level (from prelearning trials to trials during and after learning), there is a .42 change in timed report, on a scale of –1 to 1. The random intercepts themselves, i.e., individual differences among participants, have decreased to the point that they are no longer statistically significant ($p > .5$) after accounting for differences due to learning. The statistically significant fixed intercept shows that the grand mean of responses is greater than 0, averaged over all subjects and trials.

(Please insert table 1 here)

Repeated Measures Analysis of Timed Report

A 2 x 2 within – between ANOVA (SPSS 15.0.2) evaluated learning (pre vs. post) across the two study phases found the timed report accurately identified learning. A single outlier was retained because corrected results mirror results from the dataset with the outlier, and the outlier dataset provides a more conservative analysis. Alpha was set to .01 to address a variance heterogeneity issue in the postlearning data. The main effect for learning is statistically significant, $F(1,20) = 358.73$, $p<.001$, partial $\eta^2 = .95$. Learners make little progress before the learning moment (Mean = .054, 99% CI_{lower} = -.05, 99% CI_{upper} = .16). After the learning moment players make strong progress. Their postlearning moment mean timed report value approaches 1 (Mean = .94, 99% CI_{lower} = .88, 99% CI_{upper} = 1.0). This indicates their progress is almost always successful. The main effect for study is not, $F(1,20) = .004$, $p = .95$, partial $\eta^2<.001$. The interaction between learning and study also fails to reach statistical significance, $F(1,20) = 4.15$, $p = .055$, partial $\eta^2 = .17$. Although the interaction between study phase and learning is not significant, it does account for a substantial amount of model variance (see Figure 3) but with little statistical power ($1-\beta = .24$), suggesting a significant interaction might be expected with a larger sample (see Figure 3b). Study phase 2 player mean timed report scores evidence greater dispersion before learning. After learning there is very little variance in the scores of the six phase 2 players. Additionally their aggregate mean postlearning gameplay is almost perfect (Mean = .99, 99% CI_{lower} = .89, 99% CI_{upper} = 1.1). This suggests that one or both of the preassessments may act as a prime that enhances achievement after the learning moment.

(Please insert figure 3 here)

Conclusions:

Selene measures learning as quantified behavior. Different people have learning moments at different times. CyGaMEs identified a moment of learning for the underlying science of accretion, i.e., accretionLM, and used gesture data to identify the time at which each of 22 exemplars players achieved it. The timed report successfully ascertained when people had and had not learned accretionLM. The learning moment, in and of itself, explained 95 percent of the variance in player's timed report progress. Thus, the timed report can be a strong and accurate measure of learning when games are designed according to the CyGaMEs approach. Future research should explore the interaction between the pretests (external assessments) and *Selene* learning. Future development work should generate a rule and algorithm that will support the *Selene* environment's backend reporting system to automate discovery, measurement, and reporting of the accretionLM and, eventually, other moments of learning.

Appendices

Appendix A. References

Bogost, I. (2006). *Unit operations: Approach to videogame criticism*. Cambridge, MA: MIT Press.

Borgman, C. L., Abelson, H., Johnson, R., Koedinger, K. R., Linn, M. C., Lynch, C. A., et al. (2008). *Fostering learning in the networked world: The cyberlearning opportunity and challenge: A 21st century agenda for the National Science Foundation*. Retrieved August 13, 2008, from http://www.nsf.gov/pubs/2008/nsf08204/nsf08204.pdf?govDel=USNSF_124

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

Csikszentmihalyi, M., & Csikszentmihalyi, I. S. (Eds.). (1988). *Optimal experience: Psychological studies of flow in consciousness*. New York: Cambridge University Press.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience sampling method. *The Journal of Nervous and Mental Disease, 175*(9), 526-536.

Csikszentmihalyi, M., & Schneider, B. (2000). *Becoming adult: How teenagers prepare for the world of work*. New York: Basic Books.

Fullerton, T. (2008). *Game design workshop: A Playcentric approach to creating innovative games* (2nd ed.). Burlington. MA: Elsevier.

Gagné, R. M. (1965). *The conditions of learning*. New York: Holt, Rinehart, & Winston.

Gagné, R. M., Briggs, L. J., & Wager, W. W. (1992). *Principles of instructional design* (4th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Gentner, D. (1983). Structure mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

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

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage.

Holyoak, K. J., Gentner, D., & Kokinov, B. N. (2001). Introduction: The place of analogy in cognition. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 1-20). Cambridge, MA: MIT Press.

Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *The Journal of the Learning Sciences, 10*(4), 417-446.

Laughlin, D., Roper, M., & Howell, K. (2006). *NASA eEducation roadmap: Research challenges in the design of massively multiplayer games for education and training*. Washington, DC: National Aeronautics and Space Administration's Office of the Chief Education Officer, Technology and Products Office.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development*, *50*(3), 43-59.

Reese, D. D. (2007a, March) Designing Selene: Theory-based game design and data-mining. *Serious Games Summit*. Podcast retrieved from https://store.cmpgame.com/product.php?cat=19&id=2022

Reese, D. D. (2007b). First steps and beyond: Serious games as preparation for future learning. *Journal of Educational Media and Hypermedia, 16*(3), 283-300.
Reese, D. D. (2008). GaME design for intuitive concept knowledge. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (Vol. 3, pp. 1104-1126). Hershey, PA: Idea Group.
Reese, D. D. (2009). Structure mapping theory as a formalism for instructional game design and assessment. In D. Gentner, K. Holyoak, & B. Kokinov (Eds.), *New frontiers in analogy research: Proceedings of the 2nd International Conference on Analogy (Analogy '09)* (pp. 394-403). Sofia, Bulgaria: New Bulgarian University Press.
Reese, D. D. (in press). Games to evoke and assess readiness to learn conceptual knowledge. In R. V. Eck (Ed.), *Gaming & cognition: Theories and perspectives from the learning sciences*. Hershey, PA: IGI Global.
Schell, J. (2008). *The art of game design: A book of lenses*. NY: Elsevier.
Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475-522.
Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1-51). Greenwich, CT: Information Age.
Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129-184.
The Learning Federation Project (2003). *Learning Federation learning science & technology R & D roadmaps: Instructional design in technology-enabled learning systems: Using simulations and games in learning.* Author.
Wright, W. (2003). *Dynamics for designers*. Keynote address presented at the 2003 Game Developers Conference. Retrieved May 31, 2006, from http://www.gamasutra.com/features/gdcarchive/2003/Wright_Will.ppt
Wright, W. (2004). *Sculpting possibility space*. Keynote address presented at the Accelerating Change 2004: Physical Space, Virtual Space, and Interface Conference. Retrieved June 16, 2006, from http://cdn.itconversations.com/ITC.AC2004-WillWright-2004.11.07.mp3
Wright, W. (2006). *What's next in game design*. Retrieved June, 2006, from http://www.gamasutra.com/features/20060324/sanchez_01.shtml

Appendix B. Tables and Figures

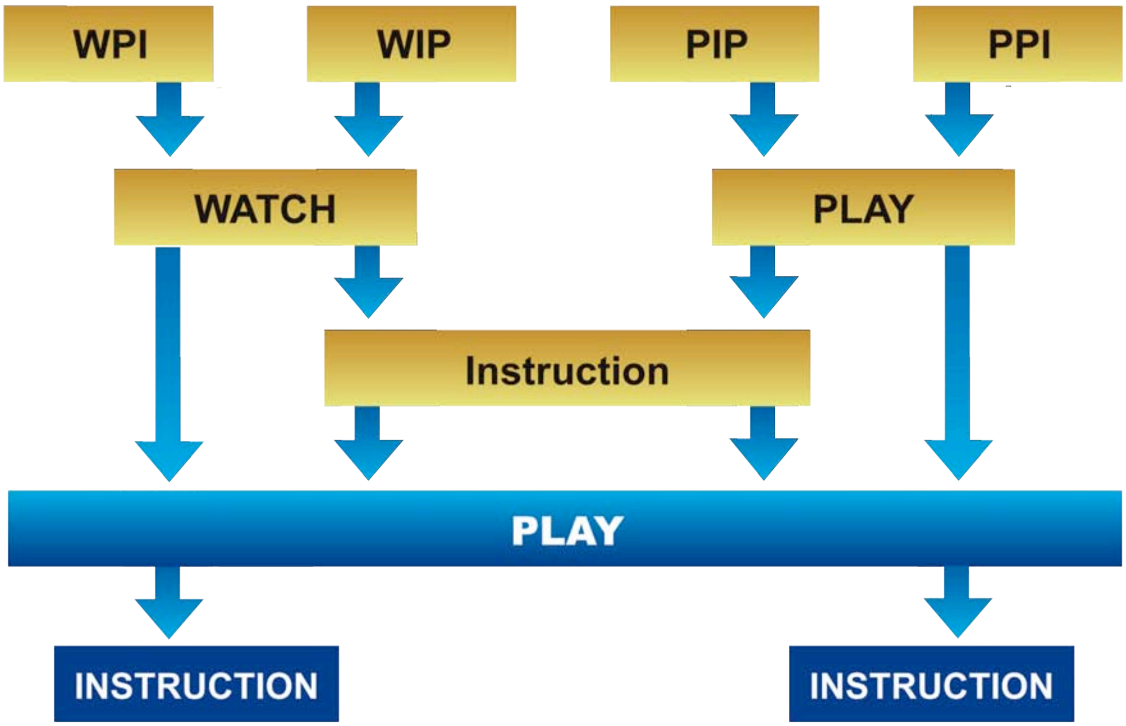

Figure 1. The phase 1 design, an example of one implementation of the double transfer paradigm (Schwartz & Martin, 2004) as adapted for CyGaMEs research. In the phase 1 implementation participants either watched or played the game during round 1. Then all players played round 2. Half the players watched video instruction during round 1. Half watched video instruction after completing the round 2 game. The phase 2 design contained no watcher conditions. Instead, Phase 2 players were assigned to one of two pregame assessments, and then half watched video instruction after round 1 gameplay and half watched video instruction after round 2.

Note: PIP = play-instruction-play, PPI = play-play-instruction, WIP = watch, instruction, play, WPI = watch-play-instruction.

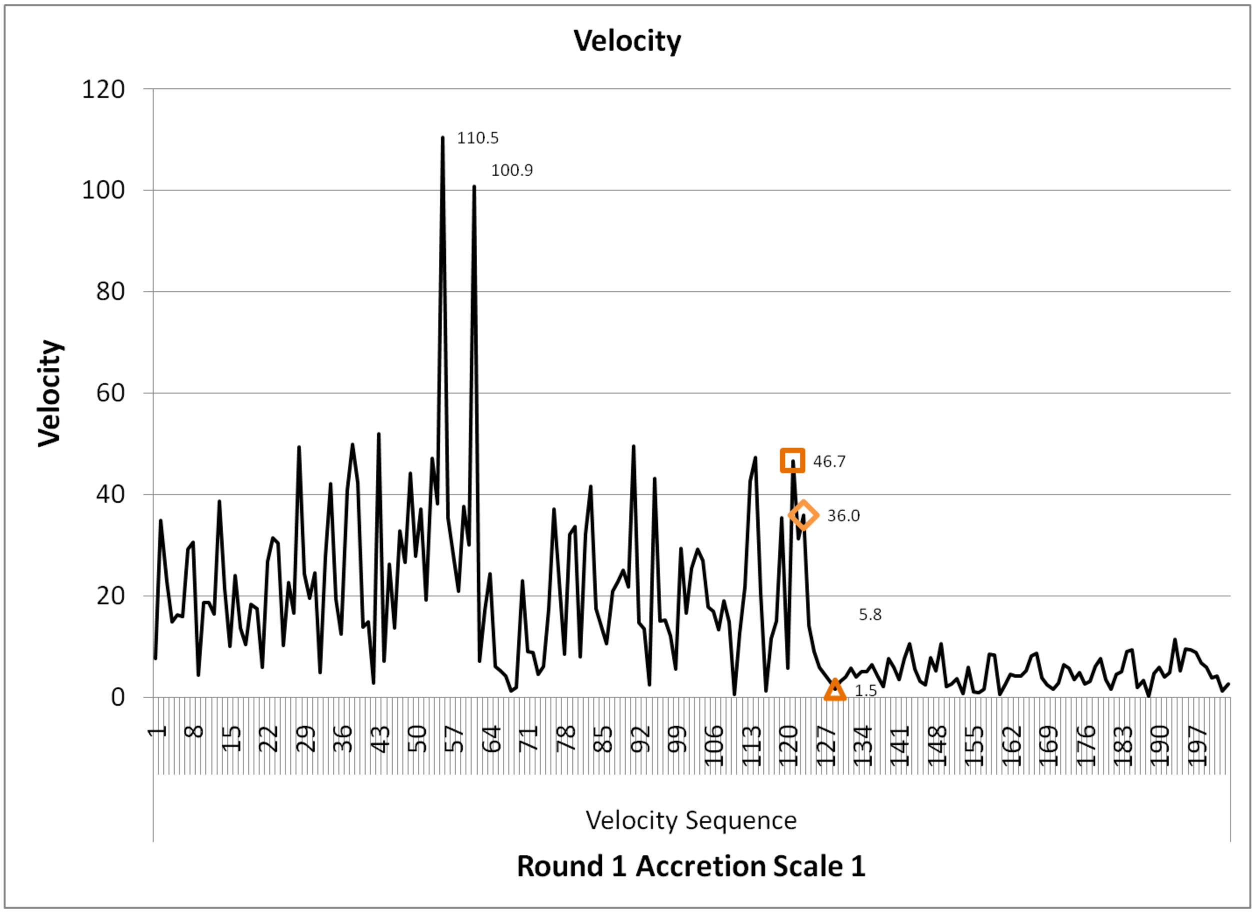

Figure 2. Velocity Bifurcation—Velocity trace for moment of learning, accretionLM, for case study participant. The moment of learning is marked by the orange triangle (velocity=1.5). High velocity collisions cease at the moment of learning, and participant subsequently sustained attenuation of velocity. This graph limits displays to round 1 accretion scale 1 data.

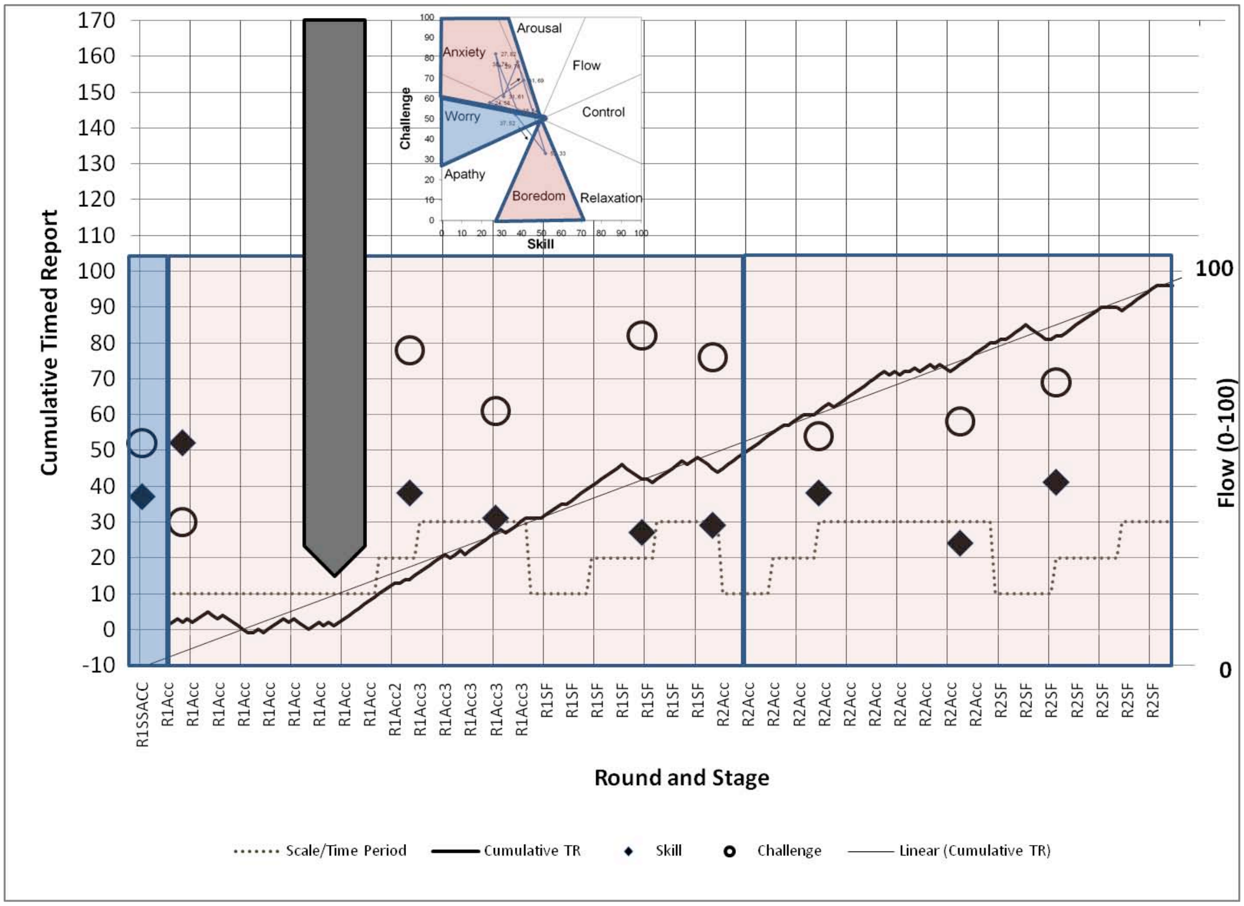

Figure 3. Cumulative timed report trace of moment of learning, accretionLM, for case study participant over two rounds of gameplay, including flowometer reports (skill and challenge) for initial instructional section and subsequent two rounds of gameplay. The dark gray arrow points to the time of the accretionLM, as identified by the velocity data analysis. This player reported a state of worry while watching the solar system accretion gameplay, reported boredom before accretionLM, and reported sustained anxiety for the next half hour of round 1 and round 2 gameplay.

Note: R1SSACC = round 1 solar system accretion, R1Acc = round 1 accretion scale 1, R1Acc2 = round 1 accretion scale 2, R1Acc3 = round 1 accretion scale 3, R1SF = round 1 surfaces features (time periods 1-3), R2Acc = round 2 accretion (scales 1-3), R2SF = round 2 surfaces features (time periods 1-3).

Table 1. Results of Final Two-level Model of Response Timed Report

(a) Random Effect (Individual Differences, Tau)

Effect	Parameter Estimate	Standard Deviation	Chi-square	df	*p* (1-sided)
Intercepts	.00008	0.00869	13.07	21	>.500

(b) Fixed Effects (Averaged over Participants)

Effect	Parameter Estimate	Standard Error	*t*-ratio	Approx. df	*p* (2-sided)
Intercept	0.4944	0.0227	21.73	21	<.001
Learning	0.4222	0.0269	18.62	744	<.001

Table 2. Means and Standard Deviations for Learning by Study

Learning	Study	Mean	Std. Deviation	*N*
Pre				
	Spring 2007 (13-18)	0.101	0.160	16
	Fall 2007 (Undergraduate)	0.008	0.164	6
	Total	0.076	0.162	22
Post				
	Spring 2007 (13-18)	0.894	0.105	16
	Fall 2007 (Undergraduate)	0.992	0.020	6
	Total	0.921	0.100	22

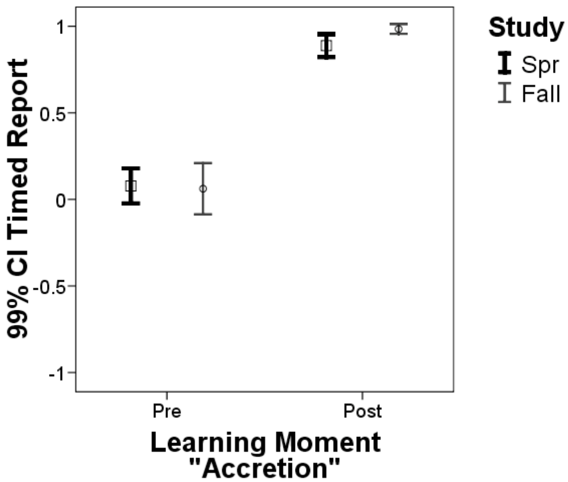

Figure 3(a). Mean scores and error bars by study phase within learning.

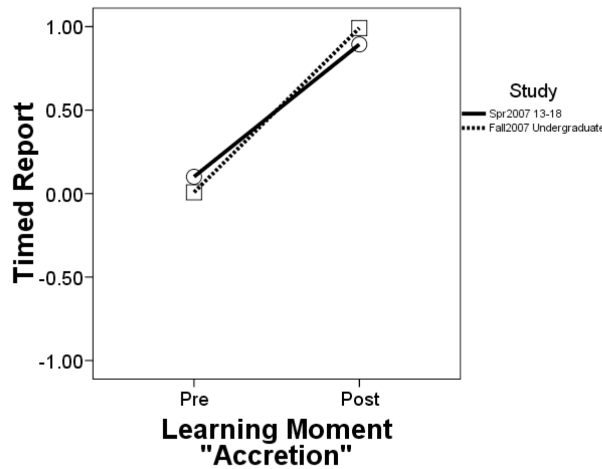

Figure 3(b). The interaction between study phase and learning. Sample size required for 1-β = .80 (α = .01) is 50 players per study phase.