# Modeling User Exploration and Boundary Testing in Digital Learning Games


V. Elizabeth Owen
University of Wisconsin-Madison
225 N Mills St.
Madison, WI 53715
+16082634600
v.elizabeth.owen@gmail.com

Gabriella Anton
Northwestern University
633 Clark St.
Evanston, IL 60208
+18474913741
gabby.anton@gmail.com

Ryan Baker
Teachers College Columbia University
525 W 120th St.
New York, NY 10027
+12126783000
ryanshaunbaker@gmail.com



## SUMMARY

Digital games can be potent problem solving environments which afford discovery learning through thoughtful exploration [1, 2]. As such, game microworlds facilitate self-regulated learning through sandbox elements in which students have agency in individualizing their pathways of interaction [3]. These agency-driven environments can support learning via individual discovery of problem space constraints and solutions, particularly through boundary testing and productive failure [cf. 4]. Thus, modeling of user interaction in digital learning games can provide considerable insight into emergent trajectories of discovery-based progression, in which equally engaged players may interact differently with the system. To this end, this research leverages educational data mining (EDM) [5] to investigate organic player trajectories of thoughtful exploration (around boundary testing and productive failure) in a learning gamespace. We align behavioral coding with log file data to automatically detect sequences of thoughtful exploration (TE) in play. Results include a robust predictive model of event-stream TE, with multiple trajectories of emergent student behavior—offering insight into organic learning pathways through the game-based problem space, and informing iterative design in optimization of user experience and student engagement.



## Keywords

Serious games; game-based learning; microworlds; exploration; productive failure; student model; classification; behavior detection; educational data mining.


## 1. METHODS

This study investigates thoughtful exploration in emergent learning trajectories of students in educational games, specifically in the STEM game *Progenitor X*. *Progenitor* is a biology game developed by the Games+Learning+Society (GLS) center at UW-Madison[1]. As a regenerative biologist, the player must save infected patients from a zombie epidemic by using stem cell science to regenerate healthy tissue and organs. Key virtual lab procedures include *starting* with base cells, *treating* them, and *collecting* the transformed cells. This core loop occurs in increasing difficulty in three systems-level biology layers: cell, tissue and organ phases. 110 middle school students played *Progenitor* at the Wisconsin Institute for Discovery as a part a public summer school program. Students completed a pre- and post- survey after an hour-long game session, which included biology items and demographic questions (e.g. grade level, gender, and gaming habits). Student choices within play were captured in the form of log file data; this was driven by ADAGE [6], an event-stream data framework designed to capture salient learning data in educational games.

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



UMAP '16, July 13-17, 2016, Halifax, NS, Canada
ACM 978-1-4503-4370-1/16/07.
http://dx.doi.org/10.1145/2930238.2930271


## 1.1 Analysis Methods: Building a Detector for Thoughtful Exploration in *Progenitor X*

Prediction modeling in *Progenitor* was used to build a behavior detector, an automated model that can infer from log files whether a student is behaving in a certain way [e.g. 7]. These models can be employed to detect a variety of student behavior aspects, including affect and performance [e.g. 8, 9]. Detectors often leverage human judgment of student behavior, in a process used to train models which can then replicate that judgment. In this study, the analysis process includes: 1) distilling salient data features; 2) identifying instances of the behavior through human evaluation; and 3) predictive modeling with synchronized log file data.

### *1.1.1 Feature Distillation*

Distilling event-stream data into salient features for analysis is vital in data mining for user modeling. In *Progenitor*, game progress markers helped guide feature distillation (including eight broad gameplay objectives, with multiple lab cycles in each). Figure 1 summarizes final features, with objectives and cycles corresponding to event-stream interaction, organized by progression (game navigation) and performance (success/failure).

| | **Progression** | **Performance** |
|---|---|---|
| **Objective** | • objective added<br>• objective starts<br>• objective ends<br>• UI tab: mission screen<br>• mission screen: city selected<br>• UI button use: view objective / add objective | • successes<br>• UI event: complete<br>• cycles in a finished objective<br>• failures<br>• restarts<br>• attempt number |
| **Cycle** | • cycle starts / ends / quits<br>• cycle type: cell, tissue, organ<br>• type of cell started in cycle (type I or II)<br>• type of cell collected in cycle (type I or II)<br>• UI tool select / use: "start" / "move" / "treat" / "collect" | • successes<br>• num + type of cells collected<br>• total turns in a cycle<br>• health remaining at end of cycle<br>• failure (health gone)<br>• failure (wrong cell collect)<br>• % of health used in each cycle |
| **Overall** | • time elapsed<br>• game starts<br>• UI button use: "next" / "back"<br>• UI button use: almanac<br>• UI button use: sound off/on<br>• Grid select: cell, tissue, organ | • game win/loss<br>• mission success events<br>• total objectives complete<br>• total missions complete<br>• mission/game quit<br>• mission/game restart |

**Figure 1. Summary of base data features.**

### *1.1.2 Coding for Thoughtful Exploration*

These data enabled evaluation of the target behavior: thoughtful exploration (TE). In recent research, "exploratory behavior" has proven central to learning growth [10], in natural alignment with game-based learning environments [1, 3]. The construct of productive failure in these contexts emphasizes opportunities to explore constraints of a problem space, and test multiple solution methods [cf. 4]. Since boundary-testing and failing forward can be central to productive learning contexts [e.g. 10, 4]—and games support exploration-based discovery of problem space constraints [2]—the construct of thoughtful exploration emerges as a behavior of focus. Broadly, the target construct here entails investigating system tools, discovering constraints, and testing solutions within the game's learning space.

For TE detection, researchers observed a stream of student actions and identified instances of the behavior. For observing this stream, text replays (a text-based representation of student action during a given span of play) were utilized for their efficiency and accuracy [11]. Text replay clip size for *Progenitor X* was one objective, shown for one player at a time. Clip features displayed *sequences* of key laboratory actions, along with all UI interactions and context (time, position, etc.). Performance data was also displayed, including quit, re-start, and completion (see Figure 1).

Through the use of these text replays, player actions were observed and evaluated for the target behavior with the binary coding schema “thoughtful exploration” (TE) or “not thoughtful exploration” (no_TE). Acceptable inter-rater reliability was achieved between coders, yielding a Cohen's κ [12] of .908.

### *1.1.3 Behavior Prediction: the TE Detection Model*

The final predictive model aligned input variables (1.1.1) with behavioral outcome variables (1.1.2). The TE detector was built using WEKA, a standard tool for data mining. Aligned to the outcome variable (number of TE instances per student), a set of algorithms were selected accordingly: RepTREE, linear regression, K*, and M5'. Models were cross-validated at the student level (the unit of analysis). A single final model was chosen based on the goodness metric of a cross-validated Pearson correlation.

```
Total collects <= 14.5 : Linear Model 1 (63/55.337%)
Total collects > 14.5 :
| Total num of times type II cells collected in Obj 5 <= 0.5 : LM2 (32/51.697%)
| Total num of times type II cells collected in Obj 5 > 0.5 : LM3 (15/48.256%)

Linear Model 1:
total Thoughtful Exploration instances =
0.0011 * duration of Obj 0 (training)
+ 0.0282 * total cells collected in Obj 0
+ 0.7095 * average % health used in Obj 0
+ 0.0256 * total optional UI buttons used in Obj 0
+ 0.0026 * duration of Obj 1
- 0.0178 * total type I cell cycles in Obj 1
- 0.2307 * average % health used in Obj 2 (1st half)
+ 0.0007 * duration of Obj 2 (2nd half)
- 0.196 * average % health used in Obj 2 (2nd half)
+ 0.104 * total times type II cells collected in Obj 5
+ 0.0022 * total cells collected in Obj 5
+ 0.0957 * total type II cell cycles in Obj 8
- 0.0033 * duration of Obj 8 (2nd half)
+ 0.1094 * total successful cycles in Obj 8 (2nd half)
- 0.0076 * total cell or tissue collection instances
- 0.1585

Linear Model 2:
total Thoughtful Exploration instances =
0.8926 * average % health used in Obj 0
+ 0.0322 * total optional UI buttons used in Obj 0
+ 0.0008 * duration of Obj 1
- 0.1119 * total type I cell cycles in Obj 1
+ 0.0033 * duration of Obj 2 (2nd half)
- 0.2903 * average % health used in Obj 2 (1st half)
+ 0.0009 * duration of Obj 2 (2nd half)
- 1.4842 * average % health used in Obj 2 (2nd half)
+ 0.321 * total times type II cells collected in Obj 5
- 0.0143 * total cell collection instances in Obj 5
+ 0.0027 * total cells collected in Obj 5
+ 0.2328 * total type II cell cycles in Obj 8
- 0.0041 * duration of Obj 8 (2nd half)
+ 0.1377 * total successful cycles in Obj 8 (2nd half)
- 0.0095 * total cell or tissue collection instances
+ 2.3512

Linear Model 3:
total Thoughtful Exploration instances =
- 0.196 * total type I cell cycles in Obj 1
+ 0.8926 * average % health used in Obj 0
+ 0.0322 * total optional UI buttons used in Obj 0
+ 0.0008 * duration of Obj 1
- 0.0687 * total type I cell cycles in Obj 1
- 0.0547 * total type II cell cycles in Obj 2
- 0.2903 * average % health used in Obj 2 (1st half)
+ 0.0009 * duration of Obj 2 (2nd half)
- 0.7426 * average % health used in Obj 2 (2nd half)
+ 0.4288 * total times type II cells collected in Obj 5
- 0.0224 * total cell collection instances in Obj 5
+ 0.0027 * total cells collected in Obj 5
+ 0.5849 * total type II cell cycles in Obj 8
- 0.0041 * duration of Obj 8 (2nd half)
+ 0.1377 * total successful cycles in Obj 8 (2nd half)
- 0.0095 * total cell or tissue collection instances
+ 2.7171
```

**Figure 2. The *Progenitor X* detector of thoughtful exploration.**

## 2. RESULTS

Ultimately, M5’ produced the best model performance, achieving a cross-validated correlation of .627, comparable to levels in similar detector models [e.g. 9]. Output is shown below.

## 3. DISCUSSION AND CONCLUSION

This paper presents a predictive student model that serves as a real-time detector of thoughtful exploration in *Progenitor*, using an EDM approach to reveal organic trajectories of student behavior. The M5' detector yielded several branches of play interactions in relationship to thoughtful exploration, revealing multiple emergent user trajectories. Future research entails deeper investigation of students within each branch. Mining these emergent, varied user models also has strong design implications. These data-driven insights fuel the potential for highly effective, game-based adaptive learning systems—which can respond to the exploration and discovery-based learning inherent to games, and optimize personalized, engaging student experiences at scale.


## ACKNOWLEDGMENTS

This work was made possible by a grant from the NSF (DRL-1119383), although the views expressed herein are those of the authors’ and do not necessarily represent the funding agency. We also deeply thank Dr. Steinkuehler and the entire GLS team.


## REFERENCES


[1] R. R. Burton and J. S. Brown, “An investigation of computer coaching for informal learning activities.,” *Int. J. Man-Mach. Stud.*, vol. 11, no. 1, pp. 5–24, 1979.

[2] K. Squire, *Video Games and Learning: Teaching and Participatory Culture in the Digital Age.* Teachers College Press, 2011.

[3] L. P. Rieber, “Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games.,” *Educ. Technol. Res. Dev.*, vol. 44, no. 2, pp. 43–58, 1996.

[4] M. Kapur, “Productive failure,” in *Proceedings of the International Conference on the Learning Sciences*, 2006, vol. 0, pp. 307–313.

[5] R. S. Baker and K. Yacef, “The state of educational data mining in 2009: A review and future visions,” *J. Educ. Data Min.*, vol. 1, no. 1, pp. 3–17, 2009.

[6] R. Halverson and V. E. Owen, “Game Based Assessment: An Integrated Model for Capturing Evidence of Learning in Play,” *Int. J. Learn. Technol. Spec. Issue Game-Based Learn.*, vol. 9, no. 2, pp. 111–138, 2014.

[7] R. S. Baker, A. T. Corbett, and K. R. Koedinger, “Detecting student misuse of intelligent tutoring systems,” in *Intelligent tutoring systems*, 2004, pp. 531–540.

[8] Z. A. Pardos, R. S. Baker, M. O. C. Z. San Pedro, S. M. Gowda, and S. M. Gowda, “Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes,” *J. Learn. Anal.*, vol. 1, no. 1, pp. 107–128, 2014.

[9] R. S. Baker and J. Clarke-Midura, “Predicting Successful Inquiry Learning in a Virtual Performance Assessment for Science,” in *Proceedings of the 21st International Conference on User Modeling, Adaptation, and Personalization*, 2013, pp. 203–214.

[10] D. L. Schwartz and T. Martin, “Inventing to Prepare for Future Learning: The Hidden Efficiency of Encouraging Original Student Production in Statistics Instruction,” *Cogn. Instr.*, vol. 22, no. 2, pp. 129–184, 2004.

[11] R. S. Baker and A. de Carvalho, “Labeling student behavior faster and more precisely with text replays,” in *Proceedings of the 1st International Conference on Educational Data Mining*, 2008, pp. 38–47.

[12] J. Cohen, “A coefficient of agreement for nominal scales.,” *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.