Level Up: A Frame work for the Design and Evaluation of Educational Games

Michael Eagle
University of North Carolina at Charlotte
Computer Science Department
9201 University City Blvd.
1-704-687-8577
mjeagle@uncc.edu

ABSTRACT
Games have the potential to *transform the educational system* in the United States. However, the study of games has lacked a coherent research paradigm. In order for games to research their full positional, there is a need to have a scientific way to evaluate the effectiveness of games and interactive environments for learning. In this paper, we discuss current research into the empirical evaluation of games; we use methods from the Intelligent Tutoring System literature to evaluate an educational computer science game. We do this by mapping empirical learning curves on student game-log data. We believe the results of this type of analysis can help evaluate and improve educational video game research.

Categories and Subject Descriptors
K.3.1 [**Computers and Education**]: Computer Uses in Education
K.3.2 [**Computers and Education**]: Computer and Information Science Education - *computer science education.*
K.8.0 [**Personal Computing**]: General – games.

General Terms
Design, Experimentation, Human Factors

Keywords
Empirical Evaluation, Games, Intelligent Tutoring Systems, Educational Data Mining

1. INTRODUCTION
Games have the potential to *transform the educational system* in the United States. Games have historically been on the cutting edge of technology, yet little has been done to study how educators can take advantage of them [13]. While studies have defined qualities that might be leveraged to improve education [13] [10], and the social context of games in informal learning [8] [12], the study of games has lacked a coherent research paradigm [8]. Too often, games have been used as a reward for learning, rather than a powerful tool for making abstract concepts into concrete interactive visualization and exploration. While game creation has been used as a way to tap into student creativity, the potential for games to provide realistic experiences and accessible metaphors for learning has not been realized.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ICFDG 2009, April 26–30, 2009, Orlando, FL, USA.
Copyright 2009 ACM 978-1-60558-437-9…$5.00.

Enrollments in computing majors are on the decline [14] [4] and leveraging games for computing education could have dramatic impacts on perception, interest, and depth of learning in computing. Video games are already proving themselves powerful motivators for computing assignments and undergraduate research [3] [7]. In order to realize their potential, there is a great need to have a scientific way to evaluate the effectiveness of games and interactive environments for learning.

2. RESEARCH GOALS
The **goal** of this proposed research, *Level Up,* is to develop new ways to *design and evaluate the next generation of future game-based educational learning environments*. UNC Charlotte provides a unique opportunity to research the combination of effective educational game design with innovative educational data mining techniques and interactive visualization. We can also disseminate our games in college classes and through community-based outreach programs that are part of the NSF-funded STARS Alliance program to broaden participation in computing.

We hypothesize; that the *Level Up* framework and methodologies will increase the speed in which educational games can be produced; increase the quality of the games produced; and provide a scientific evaluation of the games educational content. Students developing games with the *Level Up* framework will be exposed to scientific research, user studies, and outreach activities; we hypothesize that this will have an effect on the perception of computing and scientific research.

3. RESEARCH PLAN
Level Up will be realized through an agile process with three main parts. First, we will leverage the existing *Game2Learn* project to engage and guide undergraduate students in making computing educational games that include our logging system. Second, we will evaluate these games with college freshmen, and local high school students in introductory computing courses through the STARS Leadership Corps outreach program. In parallel, we will work to develop a theoretical framework that incorporates lessons learned from intelligent tutoring systems, educational data mining, human-computer interaction, and principles for creating flow in games. We will use this framework along with user studies to evaluate the educational games for learning effectiveness and engagement. These results will be iteratively fed back into the game designs to improve their efficacy.

Although these goals seem particularly broad, we will focus our efforts in two main areas. We will engage students in our game design program, and a new proposed educational games course in developing new games; the scope of the created games will be small enough for undergraduate students to build and test in a single semester. The games will focus on specific learning goals; with the core game-mechanic based on a single concept. Students will learn to test their games with both traditional pre and posttest evaluations and newly designed techniques.

Our work will emphasize data-driven analysis of learning experiences through visualizations, educational data mining, and statistical techniques applied to game logs. The *Level Up* framework will include methods to evaluate educational games by applying techniques from the intelligent tutoring systems literature. There are some important parallels between ITS and video games [2] such as instant feedback and scaffolding techniques; there is a call for finding how we can best leverage the ITS and game communities for future educational learning environments [11]. We believe that this evaluation can strengthen the research in educational games, and help researchers and developers provide better educational systems to students.

Level Up will use a suite of empirical experiments and log-data driven analyses such as empirical learning curves to understand how learning occurs in educational video games; empirical learning curves are often used to evaluate intelligent tutoring systems. [1] [9] Our goal is to use educational data mining on student game-log data to model student learning and identify places where we can improve development of our games. These models can be used to both assess the quality of the learning in game, and find the correct places to apply in game feedback such as hints on difficult problems.

4. ANTICIPATED RESULTS

Level Up will result in the development of many new innovative educational games, each one teaching a small computer science concept. We will study the differences in the successful and unsuccessful games and improve the educational game design methodologies. Students who play the games will benefit from increased learning, while the game designers will gain valuable research and technical skills.

In [6] we evaluated the game *Wu's Castle,* developed by an undergraduate student, using pre and posttest data. The results revealed large pre to posttest learning gains for students who played the game as a homework assignment. While pre and posttest are great ways to evaluate effective learning gains, they are not ideal. Here we include an alternative way of evaluating learning; we analyze the game-log data collected in [6]. Our goal is to use these educational data mining techniques to identify places where we can improve the development of our games, and to better understand how learning occurs in educational video games.

4.1 Preliminary Results

We analyzed the game-log data of 61 students for two levels of the *Wu's Castle* video game described in [6]. The player must complete 14 problems to finish the game, six are in the first level, and the remaining eight are part of the second level. In *Wu's Castle,* the player completes problems by filling in the parameters of for-loops. Figure 1 shows an example of the player answering an array based question. After selecting the parameters, the game will execute the player code. If the player has successfully completed the problem, the next mission will load. If the player makes a mistake, the level will reset and they are able to try again. We keep detailed logs of all the player interactions, including how much time is spent on each problem and the number of overall attempts.

Figure 1: Player solving a problem in Wu's Castle

We mapped learning curves for speed and accuracy. Figure 1 shows the learning curve for speed; as the student progressed through the level they became faster at solving problems despite the problems increasing the complexity. Figure 2 shows the average number of attempts decrease as the level progresses. Both of these figures indicate an increase in learning and accuracy over time as the students move through the six challenges.

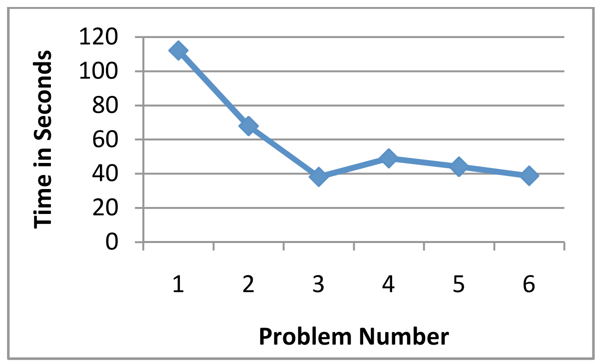

Figure 2: Learning curve for speed

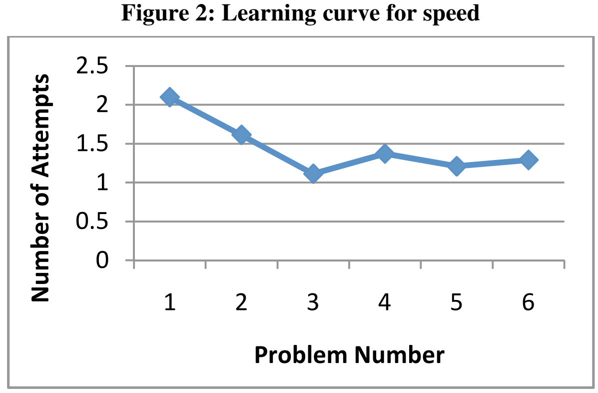

Figure 3: Learning curve for Accuracy

We see a nice learning curve start at problem one and end at problem three. The second curve, starting at problem four and ending at problem six, was not as drastic and could indicate that students have already mastered the learning content of this level. Not only do these curves offer evidence of learning, they allow us to evaluate changes over time.

If we make a change to the problem order, add new problems, or change the difficulty of the problems, we can view the differences in the student game-log files. These techniques combined with pre and posttest control studies will allow for better evaluation of educational games. For the second level of the game, we expected to see two learning curves. Figure 3 shows the speed learning curve for problems 7—14, the second level. Our goal is that each new knowledge component we add to the game will be reflected with another visible learning curve. By collecting more game-logs, we can evaluate game changes over time. We also hope to find ways to use this data to provide intelligent feedback from within the games.

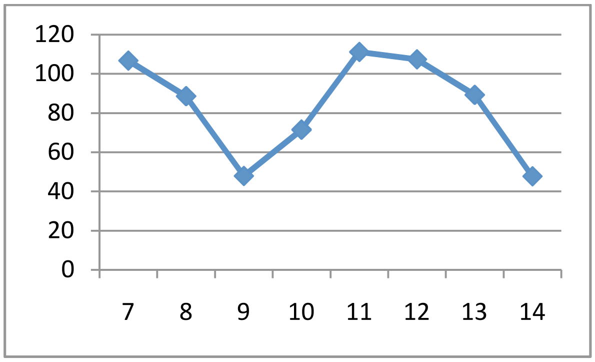

Figure 4: Speed curve for the second level

5. DISCUSION

The *Level Up* framework will provide scientifically sound methods of evaluating educational games. Many of these concepts can be carried over to regular games, simulations, and other interactive systems. The *Level Up* framework will result in a suite of games developed using consistent guidelines and refined using a grounded and data-driven agile feedback process, along with a continuously refined logging system. Together, evaluations of these games and the logging system will provide the basis for developing principles for effective educational game design.

Students can learn to create and evaluate their own educational games. The successful games can be delivered to local underprivileged high schools, which have little access to technology instructors, through STARS Alliance outreach programs. Successful games will be placed online and ranked by their effect size on learning gains; anyone can download and benefit from the games created. Our logging standard will be made compatible with the PSLC's Datashop standard [5]; this will provide other researchers will valuable data on learning in interactive educational systems.

The implications of using learning curves to evaluate educational games are not yet fully known. We hope to expand the practice to non-educational games as well. Learning to mix the right amounts of fun and play will prove critical to the development of effective educational games.

We hypothesize that games often have the correct learning curve frequencies, and that educational games modeled after non-educational game's flow could prove to be more enjoyable and effective learning tools.

6. REFERENCES

[1] Anderson, J.R., Conrad, F.G., Corbett, A.T. Skill Acquisition and the LISP Tutor. *Cognitive Science,* 13 (1989), 467-505.

[2] Baker, R. S. J. D. Habgood, M. P. J. Ainsworth, S. E. Corbett, A. T. Modeling the Acquisition of Fluent Skill in Educational Action Games. *Lecture Notes in Computer Science*, NUMB 4511, 17-26

[3] Barnes, T., E. Powell, A. Chaffin, H. Lipford. Game2Learn: Improving the engagement and motivation of CS1 students. In *ACM GDCSE'08.*

[4] Beauboeuf, T & J. Mason. Why the high attrition rate for computer science students: some thoughts and observations. *SIGCSE Bull.* 37, 2 (Jun. 2005), 103-106.

[5] Pittsburgh Science of Learning Center (PSLC) Data Shop http://www.learnlab.org/technologies/datashop/ (Dec 2008)

[6] Eagle, M. & T. Barnes. Experimental Evaluation of an Educational Game for Improved Learning in Introductory Computing. To appear in *SIGCSE 2009.* Knoxville, TN.

[7] Gee, J. P. What video games have to teach us about learning and literacy. Comput. Entertain. 1, 1 (Oct. 2003), 20.

[8] Hunicke, R., Robison, A., Squire, K., and Steinkuehler, C. Games, learning and literacy. *Sandbox 2006.* ACM Press, New York, NY, 19-19.

[9] Martin, B., Koedinger, K., Mitrovic, A., Mathan, S. On Using Learning Curves to Evaluate ITS. *Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED-2005),* 419-426.

[10] Prensky, M. 2003. Digital game-based learning. *Comput. Entertain.* 1, 1 (Oct. 2003), 21-21.

[11] Rodrigo, M.M.T., Baker, R.S.J.d., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sevilla, L.R.S., Sugay, J.O., Tep, S., Viehland, N.J.B.: Affect and usage choices in simulation problem solving environments. *Artificial Intelligence in Education*, pp. 145–152. Marina del Rey, CA (2007)

[12] Steinkuehler, C. Learning in massively multiplayer online games. In Proc. Intl. Conf. Learning Sciences, Santa Monica, CA, June 22 - 26, 2004, p. 521-528.

[13] Squire, K. (2003). Video games in education. *International Journal of Intelligent Simulations and Gaming*, vol. 2, 49-62.

[14] Zweben, S. 2003-2004 Taulbee Survey. *Computing Research Association Taulbee Survey*, May 2005.