Crossing the Bridge: Connecting Game-Based Implicit Science Learning to the Classroom

Elizabeth Rowe, Educational Gaming Environments (EdGE) group @ TERC
Jodi Asbell-Clarke, Educational Gaming Environments (EdGE) group @ TERC
Erin Bardar, Educational Gaming Environments (EdGE) group @ TERC
Emily Kasman, Educational Gaming Environments (EdGE) group @ TERC
Barbara MacEachern, Educational Gaming Environments (EdGE) group @ TERC

Overview

This paper explores implicit science learning through the game, *Impulse,* and how that learning might be leveraged to improve classroom learning. The assumption inherent in the design of *Impulse* is that by building foundational learning through games, students may be better prepared to study related science in class. We are testing that conjecture through an implementation study with hundreds of high school students.

In this paper, we present the results of pre-post assessments of Newton's laws of motion from three groups of high school students. The Game group (209 students in 21 classes) was encouraged by their teachers to play *Impulse* outside of class and some played in class. The "Bridge" group (179 students in 18 classes) was encouraged to play the game outside of class, some played the game in class, and their teachers used examples from the game in their classroom instruction on Newtonian mechanics. One hundred eight (108) students in 11 classes in the Control group neither played the game nor had game examples in class.

Researchers gathered pre and post assessment results as well as daily logs of teacher activities. HLM analyses look at the relationship between study group—Bridge, Game, Control—and students' post-assessment scores. Results show higher post-assessment scores for students in the bridge classes, as compared to the control classes after accounting for the students' pre-assessment scores, student characteristics, teacher characteristics, and school demographics. Whether or not the class was a Honors/AP class, however, moderated the study group effect. This confirms our conjecture that *Impulse* can help prepare some students for improved science learning in class.

Making the Bridge from Games to Classroom

Implicit learning may be foundational to all knowledge (Polanyi, 1966) but is particularly challenging to measure because it is, by definition, largely unexpressed by the learner. Polanyi described implicit knowledge as foundational for explicit knowledge building (1966). McCloskey (1983) and diSessa (1993) studied learners' misconceptions in science in terms of their implicit understandings of the physical world around them, pointing out that many learners have inaccurate fundamental implicit models of forces and motion because of our daily experiences with friction and other complicating factors.

Games offer a unique opportunity to promote and study implicit learning. They are a highly engaging environment that can be designed to use mechanics that mirror authentic science and are rich research environments because of the digital log data they generate (Asbell-Clarke et al., 2012; 2013; NRC,2011). The *data exhaust* from games (Halverson et al., 2012) can be used along with education data mining methods (EDM) to detect patterns of play that learners use and how those patterns change as players advance towards more successful and sophisticated activities (Baker & Yacef, 2009; Martin et al., 2013).

Games that are made for educative purpose, however, often lack the polish desired by typical game players (Isbister, 2010). There are many games on the market that are widely popular and have game mechanics that immerse players in an environment where they need to grapple with accurate physical phenomena to succeed (e.g. *Badlands*, *Where's my Water?,* and even *Angry Birds* for the most part). If a player can successfully navigate game spaces like these, however, and cannot understand and explain how objects move in the real world, it is not of much value in an educative sense. Educational game designers must walk the balance between fun and purpose.

Our group has addressed this challenge by creating games that will be played voluntarily by a wide variety of players in a popular venue, and that are also driven by challenging authentic scientific game mechanics. Through an NSF-funded research grant that looks ahead to how learning environments may be designed later this decade, we are working with professional game developers, teachers, students, scientists, and gamers in the general public to design and study a series of games played in free-choice time that are based on core high-school science content.

Our research attempts to capture the strategies players develop during gameplay that may reveal implicit knowledge and can be leveraged for high school classroom learning and assessment.

Even a good educational game is likely only one part of a complete learning experience. Post-game debriefing and discussions connecting gameplay with classroom learning are critical in helping students apply and transfer learning that takes place in games (Lederman & Fumitoshi, 1995; Ash, 2011; Ke, 2009). To exploit learning that happens in games, teachers may build on the games' "aha" moments and help their students make connections between their actions in games and the content being covered in the classroom.

Description of Research

This paper reports preliminary results from an implementation study of 40 classes and 23 teachers divided into three groups, Bridge, Game, and Control. In Fall 2013, teachers in the Game and Bridge classes were asked to recruit students to play the game *Impulse* in their free time. In Bridge classes, teachers used video clips from the game that were suggested by the research team when teaching about Newton's Laws of Motion. In the Control class, students did not play the game and teachers did not use the game examples. All classes covered their typical content on Newton's Laws of Motion during the implementation period.

The implementation study is being conducted, in part, to examine the conjecture that implicit learning in game play can help prepare students for classroom learning. To test this conjecture, changes in assessment scores before and after classroom instruction were compared for the Game, Bridge, and Control groups using a series of hierarchical linear models using the SPSS MIXED linear models procedure. HLM was chosen to account for the clustering of students within classrooms and the clustering of classrooms by teachers.

Our conjecture would lead us to anticipate that the Bridge group would show the greatest gains in the pre/post assessments. Whether or not the Game group would show gains over the Control group is an open question. Our conjecture is that the game will help prepare them for learning and we hypothesize that having the teacher use examples from the game when teaching the related content will help forge the connections to leverage the implicit learning from the game. While we believe that there may be implicit learning even among the students in the Game group, our framing did not predict whether or not that implicit learning would be apparent without an explicit bridge made by the teacher.

Description of the Intervention: *Impulse*

Impulse immerses players in a particle simulator in which they must predict the Newtonian motion of the particles to successfully avoid collisions and reach the goal (see Figure 1). The motion of all particles obey Newton's laws of motion and gravitation. Players use an impulse (triggered by their click or touch) to apply a force to particles. If the player's particle collides with any ambient particle, the level is over and they must start again. Each level of the game gets more complex, requiring players to grapple with the increasing gravitational forces of an increasing number of particles and also particles of different mass (and thus inertia). *Impulse* has been played by over 10,000 players online and through iOS and Android app stores.

Data Collection

One hundred thirty-five (135) teachers applied to be part of the study. Based on information in their applications and responses to emails, 42 teachers were assigned to the Bridge, Games, or Control group. Teachers were assigned to groups to ensure balance between the groups in terms of: the percentage of students at the school receiving free/reduced price lunch, the percentage of minority students, and the extent to which students had access to the technology needed to participate in and out of school. To a lesser degree, we also sought balance between private vs. public schools, years of science teaching experience, region of the country, and their prior experience using educational games.

Figure 1: A screenshot from *Impulse*. The player is the green particle.

All students were required to return signed parental consent and student assent forms to participate in the study. All Bridge and Game student data were collected through the game portal BrainPlay.com. Each class was given a unique ID that students entered as part of registering in *BrainPlay*. Upon registering, students were immediately taken to the online pre-assessment. Once they completed the pre-assessment, the game was unlocked in *BrainPlay* and all game play activity was logged. When the teacher finished instruction, s/he asked students to complete the online post-assessment. Teachers in the Control group were given URLs outside of *BrainPlay* for the pre- and post-assessments and were asked to assign students unique IDs.

Sample

This paper reports results from the 23 teachers who taught 40 classes to finish participating in the study in the 2013-14 academic year. Eight teachers were in the Bridge group, ten teachers were in the Game group and five teachers were in the Control group. Up to three sections of the same course were included in the study. The schools these 23 teachers work at are mostly public schools (18 or 78%) with 15 (65%) teachers reporting their schools had more than a quarter of their students receiving free/reduced price lunches and 4 (17%) teachers reporting that the majority of their students were from non-white groups. Their schools are located in 16 different states. All except six teachers had more than 5 years of science teaching experience and 18 (78%) reported some prior use of educational games or game-like simulations in their instruction.

Class sizes ranged from 8 to 28. Of the 40 classes in the study, 19 were Honors/AP classes (6 of the 18 Bridge classes, 5 of the 21 Game classes, and 8 of the 11 Control classes). The percentage of students in each class with complete data ranged from 21 percent to 100 percent. 'Complete' data means the students returned the parental consent forms, answered all items on the pre and post assessments, and, in the Bridge and Game groups, played to at least Level 2 of the game (Level 1 is a tutorial mode). Most students were not included in the study because they did not return parental consent forms.

A sample of 556 students had complete enough data to be included in these analyses. Thirty-nine (39) students who answered all pre-assessment items correctly were excluded to avoid a ceiling effect. Eleven students were excluded because they did not provide their gender. Of the remaining 496 students included in these analyses, 270 (54 percent) were females, 108 (22 percent) were in the Control group, 209 (42 percent) were in the Games group, and 179 (36 percent) were in the Bridge group. Students were concentrated at higher grade levels: 15% were 14 years old, 14% were 15 years old, 29% were 16 years old, 35% were 17 years old, and 7% were 18 years old. Almost half (211 or 42%) of the 507 students were in an AP/Honors course—38 percent of those in the Bridge group, 23 percent of those in the Game group, and 85 percent of those in the Control group.

Measures: Pre-Post Assessments

Consenting students took a pre- and post- online assessment on relevant content. As part of the assessment development process, think-aloud interviews were conducted with 10 high school students to confirm the items assessed the intended physics concepts. Each assessment included six items, three dealing with Newton's First Law and three dealing with Newton's Second Law. For each topic there was one question that resembled an animated version of a question from the Force Concept Inventory (Hestenes, Wells & Swackhamer, 1992; Thornton, Sokoloff, 1998), one question using an example from *Impulse*, and one using an excerpt from a NASA astronaut video. All items were written to be answerable with an intuitive understanding of the physics concepts. The post-assessment had the same format with slightly modified questions. Both assessments had a maximum of 10 points possible. To ease interpretation of the HLM results, the pre- and post-assessments were standardized as Z-scores (subtract mean and divide by standard deviation) to have a mean of 0 and a standard deviation of 1. Thus, all coefficients are reported in effect sizes. For the pre-assessment, the mean was 7.44 with a standard deviation of 1.530. The post-assessment had a mean of 8.25 and a standard deviation of 1.586.

Measures: Teacher & Classroom Level Characteristics

All teachers were assigned to one of three groups—Bridge, Game, and Control. In the HLM analyses, the coefficients for Bridge and Game groups are comparisons to the Control group—how many standard deviations higher or lower did students in the gaming groups score on the post-assessment relative to the students in the control groups. In their applications, teachers were asked to classify their schools in terms of the percentage of students receiving free/reduced price lunches and minority status by picking one of these categories: 0-25%, 26-50%, 51-75%, 76-100%.

Classroom level characteristics included the number of students enrolled in the class, whether the class was a Honors/AP course or not, and what percentage of the enrolled students had complete data. Classes were divided into two groups: those with the two-thirds of their students completing the study vs. those with less than two-thirds completing the study. The percentage of students completing the study was included to provide a sense of how representative the data was of the entire classroom.

Measures: Student Level Characteristics

All student level characteristics were collected as part of the registration process into the game-based data collection system called *BrainPlay. BrainPlay* was developed by our team with our game development partners to organize and export pertinent data from each game along with player data from surveys. To register in *BrainPlay*, students were asked to provide their gender and birthdate. From this, their age at the time of registration was calculated.

Data Analyses

Using the SPSS MIXED linear models procedure, HLM analyses began with an unconditional 3-level model with students, classrooms, and teachers using Restricted Maximum Likelihood (REML) and unstructured covariances. In that model, 14 percent of the variance in the post-assessment was attributable to teacher level variation while 9 percent of the variance was attributable to the classroom level. Neither of those variance components was statistically significant at the 0.05 level. Because group membership was assigned at the teacher level, subsequent HLM analyses were two-level models with students nested within teachers. In an unconditional 2-level model, a statistically significant 20 percent of the variance in the post-assessment was attributable to the teacher level.

Sets of covariates were added to the unconditional HLM model in this order:

> **Set 1**. Pre-assessment score (standardized)
> **Set 2**. Study Group (Bridge vs. Control; Game vs. Control)
> **Set 3**. Student Level Characteristics: Student gender & age; whether or not they were enrolled in class where more than half of the students completed the study; whether or not they were enrolled in an AP/Honors science class
> **Set 4**. Teacher Level Characteristics: Teacher gender; whether or not more than 25% of students in their school receive free or reduced price lunch; and whether or not more than 50% of the students in their school are minority

Only statistically significant covariates were retained in the HLM model presented in this paper. Interactions between Study Group and each of the remaining covariates were examined. Only one interaction was retained in the final model. The model with the interaction was a slightly better fit than the model without interactions (X^2 (2 df, N=496), 4.64, $p<0.10$).

Findings

Our original submission from the first 14 teachers to complete the study showed a significant positive effect of the Bridge group compared to the Control group on student's post-assessment scores after accounting for pre-assessment scores. This Group effect, however, was significantly moderated by whether more than half of the students enrolled in the class completed the study (strong vs. weak cohort). For the strong cohort classes, those with high participation rates, the bridge intervention did not have a significant effect, however for weak cohort classes it had a strong effect.

Post-submission, however, we completed data collection, and the best-fitting HLM model, which accounts for 75 percent of the variation at the teacher level, is presented in Table 1. The intercept coefficient represents the estimated outcome for male students who scored at the mean level of the pre-assessment, were in the Control group, were not in a Honors/AP class, and were in classrooms where less than two-thirds of the students completed the study. These students would score 1.24 standard deviations below the mean post-assessment score. The Pre-Score coefficient reflects the change in number of standard deviations of the post-assessment for every increase of 1 standard deviation on the pre-assessment. For every standard deviation increase on the pre-assessment, students would be expected to score 0.31 standard deviations higher on the post-assessment. Even after accounting for study group and pre-assessment scores, female students scored 0.17 standard deviations lower than male students on the post-assessment. Students in classes where two-thirds or more the students participated in the study scored 0.22 standard deviations higher on the post–assessment than students in classes where a smaller proportion of students participated.

Covariate	Coefficient Estimate	Std. Error	df	t	Sig.
Intercept	-1.239835	0.23524	11.898	-5.271	0.000
Pre-Score (Standardized)	0.306283	0.041667	483.389	7.351	0.000
Bridge (vs. Control)	0.677532	0.279544	41.038	2.424	0.020
Games (vs. Control)	0.608235	0.267469	47.321	2.274	0.028
Female	-0.165283	0.079182	493.804	-2.087	0.037
2/3 Students Enrolled Completed Study	0.223509	0.095319	111.527	2.345	0.021
Honors/AP class	0.616081	0.260715	175.479	2.363	0.019
Bridge * Honors	-0.362800	0.32063	77.794	-1.132	0.261
Game * Honors	-0.858206	0.347803	29.1	-2.468	0.020

Table 1. Estimated Fixed Effects on Standardized Post-Assessment Scores

Figure 2 shows the estimated marginal means. Marginal means are calculated by using the mean levels of all covariates to arrive at a predicted post-assessment score. A univariate F test suggests the overall effect of Study Group is statistically significant (F (2, 13.11)=3.9, p=0.047), but pairwise comparisons show the Bridge group to be significantly higher than the Control group (Mean difference=0.527, p=0.016) but not significantly different from the Game group (Mean difference=0.275, p=0.10).

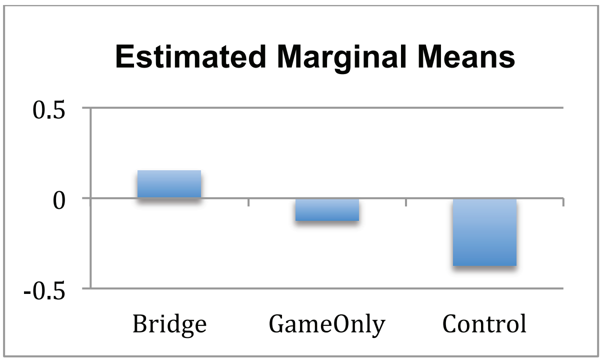

Figure 2: Estimated marginal mean post-assessment scores by study group

Figure 3 shows the predicted post-assessment scores (measured in effect sizes) for students scoring at the mean level of the pre-assessment.

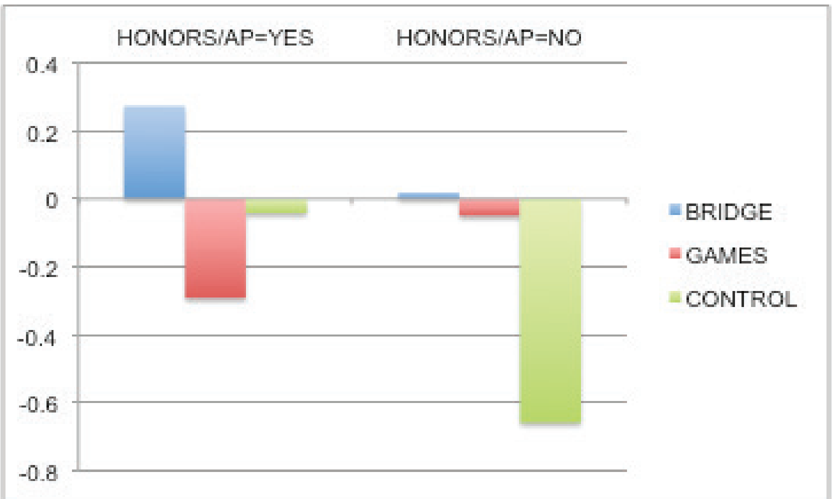

Figure 3: Predicted post-assessment scores across study conditions in Honors/AP classes versus non-Honors/AP classes

Students in Honors/AP classes in the Bridge study group would expect to receive mean level scores on the post-assessment 0.56 SD higher than students in Game classes and 0.31 SD higher than students in Control classes. Among students not in a Honors/AP class, students in Bridge and Game groups had similar mean post-assessments (a difference of 0.07 SD). Both were significantly higher than the mean post-assessment among students in the Control group (0.68 SD higher for Bridge classes, 0.61 SD higher for Games classes).

Discussion

This research was intended overall to study if playing a game designed to foster implicit science learning will better prepare students for classroom learning of related content. To study this question, we recruited high school students to play the game in their free time and gave their teachers game examples that demonstrate the phenomena of interest. We compared the learning gains from those 'bridge' classes to those of classes who only played the game ('game' classes), and 'control' classes who did their typical related curriculum. When looking at the data in aggregate, we find that overall the bridge condition students had results that were significantly different from the control condition, but not better than the game classes.

The difference between Honors/AP students and non-Honors/AP students might be explained if we assume that students in the Honors/AP classes are more academic and are served well by traditional curriculum, and that their counterparts may, on the whole, be more receptive to alternative curricula. This is consistent with our findings in a previous survey of 1500 youth about gaming preferences. Those youth who identified strongly with science preferred games that connected to science in the real world (Sylvan et al., 2013). The game, in this case, serves as an alternative form of scaffolding for the less academic students, but may be a disruption to academic students causing them to falter when given the game without sufficient bridging to the class material.

These results may indicate that an intervention such as Impulse is particularly important in less cohesive classes. Our next research steps include a comparison of the student demographics and instructional practices reported via teacher logs between the Honors/AP and regular classes, to understand the difference in impact that the *Impulse* intervention made in these different audiences.

References

Ash, K. (2011). Gaming Goes Academic. *Education Week, 30*(25), 24-28.

Asbell-Clarke, J., Edwards, T., Larsen, J., Rowe, E., Sylvan, E., & Hewitt, J. (2012). Martian Boneyards: Scientific Inquiry in an MMO Game. *International Journal of Game-Based Learning, 2*(1), 52-76.

Asbell-Clarke, J., Rowe, E., & Sylvan, E. (2013, April). Assessment Design for Emergent Game-Based Learning Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'13). Paris, France.

Baker, R.S.J.d., Yacef, K. . (2009). The State for Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining, 1*(1), 3-17.

diSessa, Andrea A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction, 10*(2/3), 105-225. doi: 10.2307/3233725

Halverson, R., Wills, N., & Owen, E. (2012). CyberSTEM: Game-Based Learning Telemetry Model for Assessment. Paper presented at the 8th annual Games+Learning+Society (GLS) Conference 8.0, Madison, WI, USA.

Hestenes, David, Wells, Malcolm, & Swackhamer, Gregg. (1992). Force concept inventory. *THE PHYSICS TEACHER, 30*, 141.

Isbister, K., Flanagan, M., & Hash, C. (2010, April). *Designing Games for Learning: Insights from Conversations with Designers*. Paper presented at the CHI 2010 Conference, Atlanta, GA.

Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools. In R. E. Furdig (Ed.), *Handbook of Research on Effective Electronic Gaming in Education* (pp. 1–32). New York: IGI Global.

Lederman, L. C., & Fumitoshi, K. (1995). Debriefing the Debriefing Process: A new look. In D. C. K. Arai (Ed.), *Simulation and gaming across disciplines and cultures*. London: Sage Publications.

Martin, T., Aghababyan, A., Smith, C., Olsen, D., Pfaffman, J., Phillips, R., Baker, S., & Janisiewicz, P. (2013). *Nanogenetic learning analytics: Illuminating student learning pathways in an online fraction game*. Paper presented at the Proceedings of the Third Conference on Learning Analytics and Knowledge, Leuven, Belgium.

McCloskey, M. (1983). Intuitive Physics. *Scientific American, 248*(4), 122-130.

National Research Council. (2011). *Learning Science Through Computer Games and Simulations. Committee on Science Learning: Computer Games, Simulations, and Education*. Margaret A. Honey and Margaret L. Hilton (Eds.), Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press. Retrieved from http://www.nap.edu/catalog/13078.html.

Polanyi, M. (1966). *The Tacit Dimension*. London: Routledge. (University of Chicago Press. ISBN 978-0-226-67298-4. 2009 reprint).

Sylvan,E., Asbell-Clarke J., Fraser, J. , Gupta, R., and Rowe, E. (2013). It's not who I am, its what I do: How Youth Gamer Identity and Beliefs about Social Perception Relates to Their Science Understanding Paper presented at DIGRA, August. Atlanta GA.

Thornton, Ronald K., & Sokoloff, David R. (1998). Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics, 66*(4), 338-352.

Acknowledgements

This research was funded as part of a DRK12 grant from NSF (#1119144) to develop and study the *Leveling Up* games. We gratefully acknowledge the rest of the EdGE team: Teon Edwards, Jamie Larsen, and Katie McGrath for their design and outreach efforts, and Game Gurus, our game development partner.