Working Through Impulse:
Assessment of Emergent Learning in a Physics Game

Jodi Asbell-Clarke, Elizabeth Rowe, Elisabeth Sylvan, TERC
Jodi_Asbell-Clarke@terc.edu, Elizabeth_Rowe@terc.edu, Elisabeth_Sylvan@terc.edu
Ryan S.J.d. Baker, Teachers College, Columbia University, baker2@exchange.tc.columbia.edu

Abstract: Games may offer a unique opportunity to support and observe intuitive learning that could be foundational for further STEM learning. This paper reports on research of players' activity in a game, *Impulse*, that immerses players in the physical laws of Newtonian motion. The aim of this research is to seek evidence of tacit knowledge as it develops through gameplay to see if that tacit knowledge development can eventually be related to players' learning of the underlying science. This paper reports on the identification of a set of strategic moves that players develop in *Impulse* as they advance in the game and discusses how those strategic moves may be indicators of implicit cognitive strategies and knowledge develop over time in the game. Researchers are analyzing video from playtesting and engineering features from associated click data to predict strategic moves hypothesized to reflect tacit understanding of Newtonian motion.

Introduction

There is a growing body of evidence that games foster quality STEM learning experiences (Asbell-Clarke et al., 2012; National Research Council, 2011; Steinkuehler & Duncan, 2008). Well-designed games require active participation, are challenging and designed with clear goals in mind, provide immediate feedback, introduce new concepts in a logical progression, and allow players to take risks and learn from their mistakes with minimal consequence (Shute, Rieber, & Van Eck, 2011; Gee, 2009; Ke, 2009). Furthermore, well-designed games can be used to promote such 21st century skills as creativity and innovation, critical thinking and problem solving, and flexibility and adaptability (Thomas & Brown, 2011). With 97% of US youth ages 12–17 (boys *and* girls) playing video games (Lenhart et al., 2008), and black and Hispanic students accessing the Internet through mobile phones more than their white counterparts (Shuler, 2009), digital games (particularly using wireless devices such as phones and tablets) have real potential as educational tools, especially for reaching learners who are most at risk of becoming "checked out" of school.

Our team is funded by the National Science Foundation to design a series of web and mobile games that focus on high-school science concepts drawn from U.S. standards for science education. Our games use simple game mechanics grounded in the laws of nature and principles of science, thereby allowing players to dwell in scientific phenomena. We are studying how these games can be natural, in situ forms of assessments of learning.

We are focusing on how games may support and enable measurement of intuitive, or tacit, knowledge that may provide an important foundation for more explicit learning that takes place in formal settings. We draw our framework from the notion of *indwelling* (Polanyi, 1966) where people make sense of the system they embody by building a set of subsidiary experiences that become the basis of their intuitive understandings. Thomas and Brown (2011) have more recently emphasized that examination of the indwelling that happens in games and other digital media may be very important to understanding the new culture of learning that is around us.

Framework of Game-Based Assessment and Tacit Knowledge Development

Much of the research in game-based assessment uses an Evidence Centered Design (ECD) framework. ECD models seek to establish a logically coherent, evidence-based argument between the domain being assessed and assessment task design and interpretation [Mislevy & Haertel, 2006]. While this work is important for moving towards formalized assessment in games, our approach looks at emergent behaviors that may be unique to games. Rather than facilitating assessment through constrained interactions, we enable players to dwell, survive, and advance in increasingly complex and difficult physical settings and watch for the strategy development that occurs naturally.

Our research begins with a method analogous to a grounded theory approach to defining our task model, where we make detailed observations of players dwelling in the game and iteratively hone in on themes or strategies that emerge from the observations to describe players' experience in the game. It is from those empirically grounded strategies that we will then build our evidence-based assessments. We use this approach because we are interested in the learning that takes place in games that may be emergent, and not necessarily predicted by the designers and other educators. Our framework places this emergent learning as central to tacit knowledge development that takes place in voluntary games that may be useful for educators to leverage for formal learning.

We are specifically focusing on evidence of strategies that emerge from gameplay that demonstrate tacit understanding of concepts and skills in science, which may not be explicit to the learner. These ideas may be so engrained that they become second nature and thus very difficult to capture. Polanyi (1966) argues however that tacit knowledge is foundational to explicit knowledge. Learners can carry misconceptions (sometimes due to the non-ideal factors in our world such as friction) about fundamental physics that can interfere with their explicit learning about the principles of motion (McCloskey, 1983). Game players do not necessarily perceive their game-based learning as connected to real-world phenomena (Sylvan et al., 2013), In our research, we must look for indicators evident in players game activities and then find ways to validate their inferences that the indicators do indeed represent key cognitive developments.

This paper describes our process of defining emergent strategies within a physics-based game. This is the first step in our longer process of developing game-based science assessments for high school game players. We argue that by identifying cognitive strategies that players develop while dwelling in games that portray and require predictions of accurate scientific phenomenon, we may be able to develop a new, informative, and accessible type of science learning assessments for education. This paper demonstrates the first steps in this process using the game, *Impulse*, developed by EdGE to enable players' indwelling and playing in the context of a physics simulation.

Studying tacit knowledge development in the Game *Impulse*

In *Impulse*, players must predict the motion of the balls around them to survive in a game using very simple game mechanic that is repeated over and over in increasingly difficult contexts. The motion of all of the balls obeys Newton's laws of motion and gravitation. We are examining whether *Impulse* players, while advancing in the game, may be developing and mastering an intuitive sense of Newton's laws that we can systematically measure. Once we can accurately detect what they tacitly understand about Newtonian mechanics from their gameplay, we will use this method to examine how this tacit knowledge develops and how players can build connections between intuitive knowledge they build in games and examples of similar science outside the game.

The process of developing assessments for *Impulse* is a two-phase process, analogous to approaches previously used to develop automated assessments for constructs such as science inquiry skill in simulations (Sao Pedro et al., 2012), and disengaged behaviors in intelligent tutoring systems (Baker et al., 2008). First, human judgment is used to infer a range of learner strategies and approaches within log data. The set of human judgments obtained will then be used as the basis for using data mining (Baker & Yacef, 2009) to create prediction models which can automatically infer the strategies being used by learners, without requiring intensive hand coding of data.

Design of Impulse

Impulse is designed for play on computers, tables, and smartphones Players are immersed in a sea of ambient balls obeying Newton's Laws (see Figure 1). EdGE, and partners GameGurus, developed the game as a proof of concept to understand how a free-choice game can be used to foster and measure intuitive learning about the science. This is different than the large body of research looking at learning and assessment in games within classrooms (Squire & Barab, 2004; Ketelhut & Nelson, 2010; Shute, et al. 2010).

In *Impulse,* players must navigate their own ball (green) to a moving goal, while avoiding collisions with an increasing numbers of ambient balls each new level. The player can impart an impulse to move any of the balls with a touch or click on the screen. Any nearby balls move as they experience the force of the impulse imparted by a player's click. There are also gravitational forces between balls and elastic collisions between ambient balls. A collision between the player's ball and an ambient ball causes an explosion and sends the player back to a previous checkpoint. As the player progresses through Impulse, more balls are added to the screen. At each checkpoint, a new type of ball (different color, size, and mass) is introduced.

Figure 1: The green player ball (near center) in a sea of ambient balls of varying mass in *Impulse*

Players use energy with each impulse. They may use up to 20 impulses and then their energy is replenished if they reach the goal, or they drift without any control. Players may conserve energy, and earn a higher score, by being strategic in when they use the Impulse.

Survival in higher levels of *Impulse* would be virtually impossible without the ability to predict the motion of the balls. While navigating through a sea of up to 22 balls that are colliding with each other, and are attracted to each other through gravity, players need to “study” the behavior of those balls. They need to be able to predict their motion so that they can avoid them as they travel to the goal.

More specifically, the motion that must be predicted can be described by Newton’s First and Second Laws of Motion:

1. Newton’s First Law of Motion (**NFL**) describes how an object will stay at rest or in (straight-line) motion, unperturbed, unless acted upon by a force. This sounds simple, but is not intuitive for much of the public, particularly in cases of circular motion (McCloskey, 1983)

2. Newton’s Second Law of Motion (**NSL**) describes the relationship among force, mass, and acceleration. Simply, heavier objects are harder to move than light objects. The formal expression commonly taught in school is **F**orce = **m**ass x **a**cceleration.

Newtonian mechanics bely the intuition of many learners, as evidenced by research that shows that in general the public’s high school understanding of these concepts are inaccurate and not greatly improved by traditional higher educational instruction (Emarat & Johnston, 2002). Basic misconceptions are typically that constant motion requires constant force (this is not true by Newton’s first law) and that a constant force will cause constant velocity (this is not true, by Newton’s second law the force will cause an acceleration which is a change in velocity).

Identifying Emergent Cognitive Strategies in Impulse

Knowing that the game is designed to have players predict Newton’s first and second laws of motions, we sought to identify cognitive strategies we could observe players using to deal with these underlying behaviors of the balls. We describe here the process we are using to identify the strategies from a series of video captures with think-alouds and interviews from a set of high school students playing the game outside of class time.

Developing System to Code Cognitive Strategies

This paper reports on five playtesting sessions with 15 high school students (3 female) recruited predominantly from urban schools in the U.S. Northeast as well as various other informal settings. Players were observed (and video recorded) while playing the game, and several provided follow-up interviews. Two coders, one a game designer with a physics background and a researcher with no physics background, independently watched segments from the video recordings to develop an initial set of strategic moves. Players were recorded with Silverback software (silverbackapp.com) while playing *Impulse*. This software captures video of the players' onscreen game activities, and audio and video of the player's face. We also recorded a digital log that includes these *raw features*: (a) event type (click, collision, goal reached); (b) timestamp; (c) location of each click, ball, and the goal; (d) mass of each ball in the game space; (e) game level; and (f) player ID. The coding process relies not only on clicks, but also other screen activity such as mouseovers and player commentary.

Researchers used the mouseovers, where players' cursor was visible indicating on the screen their focus of attention, to help code intentions such as purposeful floating--a deliberate act of letting the player ball move under the influence of no forces. We encouraged players to "think aloud", discussing their game play with a friend or the facilitator. In some cases we conducted direct interviews with players to ask them about their moves and strategies, in order to confirm and refine our inferences about the thought processes behind players' moves.

Over 20 randomly selected three-minute segments, two researchers iterated on the coding scheme to introduce all the moves that they observed and provide codes that they could consistently agree upon to describe each move (see Figure 2).

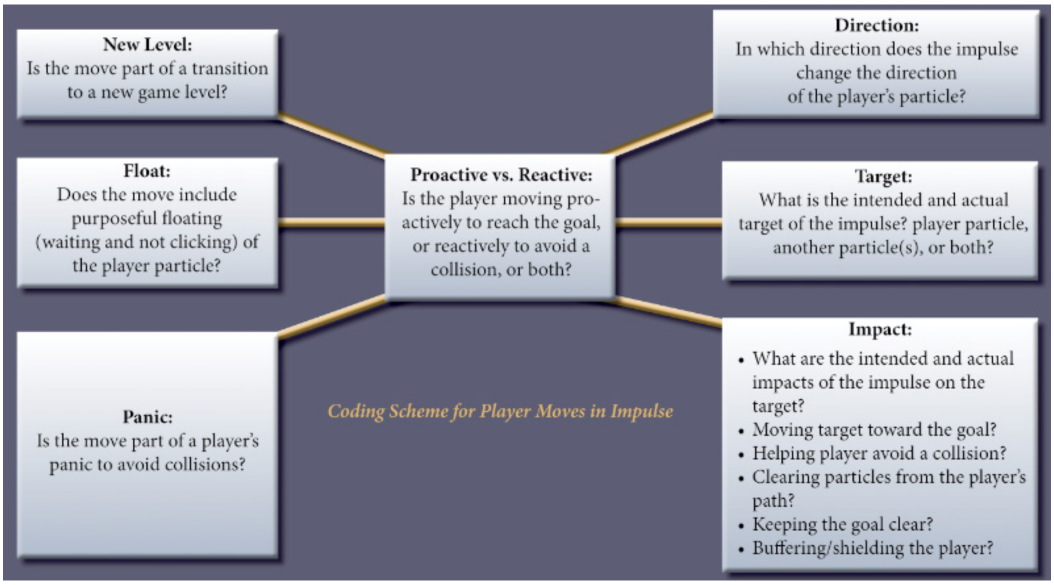

Figure 2: Codes used on players' moves to build towards cognitive strategies

Iterative refinements over several trials have led to improvements in the level of agreement after accounting for chance (Cohen's Kappa). In the approximately 500 clicks coded to date, Cohen's Kappa ranges from 0.49 for Intended Impact to 0.77 for Target. Within the Intended Impact coding, however, Kappas for many individual codes such as Toward Goal, Avoiding Collision, Buffering, and Opposing exceed 0.70. We are using results from this round of coding to further refine the coding system and train to a higher level of agreement.

Based on video observations, think-alouds, and interviews, we have identified the following cognitive strategies that we hypothesize will emerge as distinct sequences of strategic moves in the log data from *Impulse,* which can be automatically identified by models developed through educational data mining. We will also look for other strategies that can be linked with tacit knowledge development about Newton's Laws of Motion.

1. Push to Goal. This is a baseline novice move. Nearly all players use a push-to-goal strategy at times, where they use the impulse along a direct path towards the goal. This indicates at least a baseline concept of the game mechanic and, to some weak (and probably uninteresting) extent, an understanding of straight-line motion under NFL.

2. Oppose Motion. Many players directly oppose straight-line motion with their impulse, particularly at the start of a new level. They clearly demonstrate the need for providing the impulse in the opposite direction of the ball's straight-line motion in order to stop it.

3. Redirect Around Balls. As players become more expert in the game, they are able to predict the motion of the ambient balls and strategically redirect their ball's motion to avoid them.

4. Wait for Balls to Clump and Move. As players become aware of the interactions between the ambient balls, they realize that the sea of balls may change and create a new, unobstructed path towards the goal if they wait.

5. Float and Redirect. Players who can predict the motion of the ambient balls often choose to click less, which uses less energy (yielding more points) and also avoids the possibility of clicking too close to a ball, imparting a greater than intended force and thus accelerating a ball to dangerous speeds.

6. Differentiation of Mass. Players who understand that the white and grey (heavier) balls need more force to move them (and stop them) will click more to get them going and will avoid their path.

These six strategies are common in the videos and appear to grow with the development of the player. These strategies were selected because of their prevalence in the videos and their importance in the explanations given by playtesters in the game. For example, we can observe in the Silverback videos that players exclaim: "I got to wait for them to clump together and a path to clear" or "they are all attracted to each other" as they predict where to safely move their ball. We have also observed other students explain that the blue ones (lower mass) are much harder to handle because they are so reactive to the impulse and that the white ones (higher mass) are harder to move out of the way. We can see in the video that players tap rapidly three times near a white ball when only tapping once near a blue ball. They are experiencing over and over again, and thus beginning to predict, that the white one has more inertia even if they would not use those words.

The videos from our playtesting also show evidence of emergent skills associated with physics learning that we did not anticipate during the design of the game, such as precursors to vector arithmetic. When players encounter a clump of balls between their ball and the goal, they will trace out a detour around the clump. We see through mouseovers that players often trace horizontal and vertical components to their paths, breaking their trajectory up into Cartesian coordinates. They are creating vectors in their mind and tracing them out on the screen for themselves, even though there is no prompt or tool to do this. When we saw exhibited by players who had not taken a formal physics class, we added it to our list of strategies to try to observe, thinking it might serve as the basis of an assessment on vector thinking, or a classroom lesson that a teacher might use.

Distilling Features for Detecting Strategic Moves in Click Data

While the video data allows us to observe and describe the strategic moves that players make during playtesting, these techniques are limited to the samples we can observe directly. To detect these cognitive strategies without video, our first step is to accurately identify their components parts—the strategic moves—from the game log data, discussed in detail below.

Once models are developed that can recognize strategic moves in the data, we can look for sequences of those moves as evidence of cognitive strategies players use to succeed in the game that we hypothesize reflect a tacit understanding of Newtonian motion. We can further mine the data to examine the sets of strategies that apply to players who advance farther or more rapidly in the game (expert players), using regression models to identify the combinations of strategies that are characteristic of greater and lesser degrees of advancement. We can also mine to see the evolution of a player's strategy over time, using Bayesian Knowledge Tracing (Corbett & Anderson, 1995) to track the development and application of new strategies over time, and learnograms (Hershkovitz & Nachmias, 2008) to analyze the emergence and disappearance of strategies over time in a more qualitative fashion.

To obtain data that can be used to develop automated detectors these strategic moves, we will code a larger sample of videos from 70 players using the coding system described above. Those coded clicks will act as 'ground truth' against which the goodness of any predictive models is assessed. To build these models, we will distill sets of features from the raw features available directly from the log files (timestamp, locations of balls, clicks, and player, game level). The feature distillation process will explicitly select features thought by domain experts to be semantically relevant to the strategies observed by the human coders (Sao Pedro et al., 2012), and will be selected with consideration both of construct validity and feasibility for distillation. Distilled features will fall into three main categories. A non-exhaustive list of examples is given in Table 1.

Distilled Feature Type & Example		Rationale
Player Ball		
1	Distance between Player and Goal	Players use different strategies when close to the goal than when farther away
2	Current speed of player ball	If the player is moving faster they need to use different moves and strategies than when slow
3	Distance travelled since last event	This provides an indication of how much the game state will have changed
4	Change in angle between player's path and a straight-line path to goal	Strategies vary depending on whether or not player has a straight-line clear path to the goal
Impulses		
1	Proximity of impulse to player ball	Identifies the likely intended target (player ball or other) of the impulse.
2	Time since last impulse	Very quick actions may indicate panicking or intentional increased force; very slow actions may indicate floating strategies
3	Distance from impulse to closest other ball of each type	Identifies the likely intended target (player ball or other) of the impulse and identifies if players click more near certain color balls.
Other Particles		
1	Number of other balls in play space	Describes the potential complexity of the play space
2	Number of balls in path between player and goal	Describes difficulty of immediate task of getting to goal
3	Number of balls in current path of player ball	Describes immediate danger of collision

Table 1: Distilled feature types, examples, and rationale

Once the features have been distilled from the raw, educational data mining prediction methods (cf. Baker & Yacef, 2009) will be used to build models that automatically infer strategic moves. Classification algorithms that have been successful for similar problems will be tried (such as J48 decision trees, JRip decision rules, logistic and step regression, and K* instance-based classification), within the RapidMiner software (rapid-i.com). The resultant models will be assessed for predictive validity within data from new students, using cross-validation (Efron & Gong, 1983), where a model is repeatedly built using subsets of the data and tested using subsets.

The Challenges of this Emergent Approach to Game-Based Assessment

While not wanting to "break" the gameplay with any form of invasive assessments, and trying to measure knowledge that is tacit and may be unexpressed even within the learners' own minds, we create a challenging assessment task. Even as we gather evidence that illustrates cognitive strategies players use in the game and can show that they are replicated and predictable, we still have the challenge to say that this game progress demonstrates tacit understanding of physical phenomena represented in the game. A related question we return to repeatedly is whether we are measuring the player's learning of science or just becoming expert at the game mechanic. Assessment design has the perennial challenge of ensuring the logical coherence between the behaviors the task (here, the game) elicits the competencies it is designed to assess. In gameplay, the learning of the game may be confounded with learning of the subject matter of the game. This is not a new challenge in assessing learning within gameplay; even in games with known learning objectives, designed to support assessment of specific skills, it can be difficult to distinguish between domain learning and gameplay learning (cf. Baker, Habgood, Ainsworth, & Corbett, 2007).

Acknowledgements

We gratefully acknowledge the DRK12 grant from NSF (#1119144) to develop and study the *Leveling Up* games. We also would like to thank the EdGE team: Erin Bardar, Teon Edwards, Jamie Larsen, and Barbara MacEachern for their design and outreach efforts, and Jasmine Williams for her assistance in the preliminary data analysis.

References

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. doi:10.4018/ijgbl.2012010104.

Baker, R.S.J.d., Corbett, A.T., Roll, I., Koedinger, K.R. (2008) Developing a Generalizable Detector of When Students Game the System. *User Modeling and User-Adapted Interaction*, 18(3).

Baker, R.S.J.d., Habgood, M.P.J., Ainsworth, S.E., Corbett, A.T. (2007) Modeling the Acquisition of Fluent Skill in Educational Action Games. *Proceedings of User Modeling 2007*, 17-26.

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

Corbett, A.T., Anderson, J.R. (1995) Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction*, 4, 253-278.

Efron, B. & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and crossvalidation. *American Statistician*, 37, 36

Emarat, N. and Johnston, I. (2002). The Effectiveness of the Thai Traditional Teaching in the Introductory Physics Course: A comparison with the US and Australian Approaches. *International Journal of Innnovation in Science and Mathematics Education*. 9 (1)

Gee, J. P. (2009). Deep learning properties of good digital games: How far can they go? In U. Ritterfeld, M. Cody & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 67–82). New York, NY: Routledge.

Hershkovitz, A. & Nachmias, R (2008). Developing a Log-Based Motivation Measuring Tool. Proceedings of the International Conference on Educational Data Mining, 226-233

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). NY:IGI Global.

Ketelhut, D.J., & Nelson, B.C. (2010). Designing for real-worldscientific inquiry in virtual environments. Educational Research, 52, 151-167.

Lenhart, A., Kahne, J., Middaugh, E., MacGill, A. R., Evans, C., & Vitak, J. (2008). Teens, Video Games, and Civics: Teens' gaming experiences are diverse and include significant social interaction and civic engagement. Washington, DC: Pew Internet & American Life Project.

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

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

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.

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

Sao Pedro, M., Baker, R.S.J.d., Gobert, J. (2012) Improving Construct Validity Yields Better Models of Systematic Inquiry, Even with Less Information. *Proceedings of the 20th International Conference on User Modeling, Adaptation and Personalization (UMAP 2012)*,249-260.

Shuler, C. (2009). Pockets of Potential: Using Mobile Technologies to Promote Children's Learning. New York: The Joan Gantz Cooney Center at Sesame Workshop.

Shute, V. J., Masduki, I., Donmez, O., Dennen, V., Kim, Y.-J., Jeong, A. C., & Wang, C.-Y. (2010). Modeling, Assessing, and Supporting Key Competencies Within Game Environments In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-Based Diagnostics and Systematic Analysis of Knowledge* (pp. 281-309). New York: Springer-Verlag.

Shute, V. J., Rieber, L., & Van Eck, R. (2011). Games…and . . . learning. In R. Reiser & J.Dempsey (Ed.), *Trends and issues in instructional design and technology, 3rd Edition* (pp. (pp. 321-332)). Upper Saddle River, NJ: Pearson Education Inc.

Squire, K. D., & Barab, S. A. (2004). Replaying history. In Y. Kafai, W. Sandoval, N. Enyedy, A. Dixon, & F. Herrera (Eds.), *Proceedings of the 2004 International Conference of the Learning Sciences* (pp. 505). Mahwah, NJ: Lawrence Erlbaum.

Steinkuehler, C., & Duncan, S. (2008). Scientific Habits of Mind in Virtual Worlds. *Journal of Science Education and Technology, 17*(6), 530-543.

Sylvan, E., Asbell-Clarke, J., Fraser, J. Gupta, R, Rowe, E. (2013). It's not who I am, its what I do: Gaming, Identity and Understanding Science *Proceedings of DiGRA 2013*.

Thomas, D., & Brown, J. S. (2011). *A New Culture of Learning: Cultivating the Imagination for a World of Constant Change*: Printed by CreateSpace.