Version of Record: https://www.sciencedirect.com/science/article/pii/S1871187117301682
Manuscript_60c1321c41bedbcdf8b46683ca639e26

Failing up: How Failure in a Game Environment Promotes Learning Through Discourse

Craig G Anderson*[1]
Jen Dalsen[2]
Vishesh Kumar[2]
Matthew Berland[2]
Constance Steinkuehler[1]

[1] Department of Informatics and Computer Science, Donald Brenn School of Information & Computer Science, University of California, Irvine, California, USA
[2] Department of Curriculum & Instruction, Faculty of Education, University of Wisconsin – Madison, Wisconsin, USA
*Corresponding author

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Number 1418352. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Failing up: How Failure in a Game Environment Promotes Learning Through Discourse

1. Introduction

Video games have long been garnering attention from researchers, educators, and policy makers alike for the engagement they create in youth. Countless games have been made for the classroom in hopes that students will enjoy them as much as the ones they race home to play after school, and learn from their play. From *Oregon Trail* to *Math Blasters*, the industry of educational games has seen varied levels of success, and some research has focused on how to make these games more effective. One aspect that may be overlooked in the creation of educational games is player failure. Video games provide an environment in which failure is expected, or even celebrated. However, this may be at odds with how failure is treated in a classroom setting where getting an answer wrong, or failing a test can lead to dire consequences. This leaves open the question of what role failure plays in an educational game.

1.1. The Role of Failure in Developing Understanding

Across educational research, many researchers have argued that failure is an integral part of learning. Piaget argues that these are the moments in which we develop intelligence (1955); we adapt our understanding to account for the gaps in our knowledge when we encounter them. We do so through assimilation of our schemas to new information and accommodation of them to fit new experiences. Thus, learning is a result of encountering experiences that do not fit our conceptual models, forcing us to rethink and modify them. Further, VanLehn (1988) argues that "learning only occurs when an impasse occurs". When a student gets stuck on a problem, they have to lean on resources such as a teacher, textbook, or other students, and these moments provide opportunities for critical reflection. Likewise, Intelligent Fast Failure (IFF) is a strategy in which learners are encouraged to destigmatize failure and try as many ideas as possible in a short period of time in order to learn more about a given topic (Matson, 1991). Building from this, Sitkin (1992) proposes that "failure typically represents an exception that does not conform to expectations and thus requires more active, deeper processing" (p. 237). Furthermore, labeling an outcome as a "minor failure" is beneficial when increased attention and adjustments are desired while labeling an outcome as a success is beneficial when stability is desired (p. 243).

1.2. Productive Failure

More recently, Kapur (2008, 2010, 2011, 2013, 2014) explored the benefits of failure to problem-solving ability by having students struggle through ill-structured problems. Through a series of experiments, Kapur shows that students in small groups with little facilitation before a consolidating lecture show stronger understanding of the underlying concepts. Students have to explore the problem area without direct guidance, encouraging them to think about the problem space more fully. Although they fail more early on, they show a more robust understanding afterwards. Referred to as 'productive failure' (PF), learners come to a deeper understanding by struggling through an ill-structured learning environment. Although students initially struggle in an environment designed to elicit PF, they outperform students in a direct instruction environment in both conceptual understanding and transfer (Kapur, 2015).

Kapur and Bielaczyc (2011) describe environments that promote PF as having 2 phases: generation/exploration and consolidation. Their construction is guided by the following three core design principles:

1. Create problem-solving contexts that involve working on complex problems that challenge but do not frustrate.
2. Provide opportunities for exploration and elaboration.
3. Provide opportunities to compare and contrast the affordances and constraints of failed or suboptimal RSMs (representations and solution methods) and the assembly of conical RSMs.

Further, Kapur and Bielaczyc (2011) argue that PF environments must adhere to four core mechanisms: activation and differentiation of prior knowledge in relation to the targeted concepts; attention to critical conceptual features of the targeted concepts; explanation and elaboration of these features; and organization and assembly of the critical conceptual features into the targeted concepts. Another aspect identified by Kapur and Bielaczyc (2011) as important for a PF environment is an appropriate level of challenge; the environment must keep the learner "challenged yet not frustrated and remain sufficiently engaged in problem solving" (p. 50). Productive failure provides an environment that helps learners explore and fully understand a problem sphere. However, it is unclear if PF contributes to understanding in the same ways in complex interactive environments such as in a video game.

1.3. Failure in Gameplay

Similar to the PF learning environments described above, many commercial video games are designed to allow players to succeed by failing repeatedly (Juul, 2013). Gameplay must be difficult enough to be engaging to players, but not so difficult that players cannot complete them. As Gee (2005) argues, good games must be "pleasantly frustrating." They must ensure the player "feel(s) the game is challenging but doable" and kept within their "regime of competence" (p.6). To ensure players are kept at an appropriate challenge level, many games employ features that adjust the difficulty to fit the player's skill level. For example, Dynamic Difficulty Adjustment (Hunicke, 2005) is a system in which the game monitors how well the player is doing internally and adjusts accordingly. If the player is completing the game tasks quickly and with little failure as defined by the system, a few variable adjustments increase the difficulty until the player is once again being challenged to the height of their ability. Likewise, if the player is struggling too much and making little progress, the system adjusts to lower the difficulty to ensure the player does not get stuck. This system ensures that the player is kept at an appropriate challenge level for their ability. Games use other feedback systems frequently as well (Wark, 2015). For example, racers in popular Nintendo game *Mario Kart* are given advantages or disadvantages that adjust the challenge level based on their current ranking. Players ranked closer to first place are given less powerful items, while players closer to last place are given more powerful items that help them climb the ranks. Through this mechanism, players challenge level is tailored to their abilities, maintaining the challenge level most appropriate to their skill level.

The experimentation of gameplay mechanics affords an individual the opportunity to reconstruct current knowledge of material in new ways (Gee, 2007; Kapur, 2015). Through failure, players can reflect on their current actions (Juul, 2013), self-correct activities, and build mastery through past experiences (Gee, 2007). As is in education, failing in games provides opportunity to adjust conical knowledge to fill gaps that failure revealed. Furthermore, linking the learning goals of the game directly to the player actions through mechanics marries the mechanics to the content. Known as an endogenous game (Squire, 2003), the player goals are virtually the same as the embedded learning goals, allowing the player to explore the concepts through game mechanics directly. Without the impasse that comes with level failure, players are not provided with vital information on how they need to improve to complete the game. Games provide players with the ability to freely navigate, manipulate, experiment, and explore actions; this builds new expertise and models of learning (Popper, 1959). As individuals play games more frequently, they have more insights to gameplay strategy when compared to novice players (Blumberg, 2000), and can pass on this knowledge through mentorship (Steinkuehler, 2004), allowing players to build their knowledge about general game strategy. These strategies have the potential to lead to greater problem-solving techniques and the ability to negotiate unexpected obstacles more easily. In fact, Blumberg et. al (2008) found that after reaching an impasse, players were more likely to comment on gameplay strategies as a way of overcoming it.

As explained by Litts and Ramirez (2014), failure in games is part of the learning process and should not be defined as a stigmatizing end to what learning can become; failure in gameplay should be valued and sought (Gee, 2007; Juul, 2013; Ramirez et. al, 2014). However, when creating a learning environment in games or elsewhere, designers are embedding their own beliefs of knowledge (Bransford, Brown, & Cocking, 2000). As Litts and Ramirez (2014) argue:

> “In games, identifying the difference between activities that fail to achieve the games goals, but serve to further the player’s understanding of the subject, or game world, from actions that do not yield a positive result, is the major challenge facing game based assessment.” (p. 5)

Games also offer students an opportunity to explore and negotiate complex frameworks through collaborative discourse (Osborne, 2010). This discourse is important as individuals learn better when activities are designed for students to collaborate with one another. The social organization of gaming groups, often referred to as an affinity space (Gee, 2005), affords players the opportunity to come together in order to contribute, discuss, and debate information. Joint activity helps to reaffirm and debunk statements while simultaneously amassing player involvement (Squire, 2011). As Stein and Albro (2001) explained, teaching students the value of using argumentation for interpersonal success rather than personal gain over other peers is a critical factor to education. Games are not only a designed object but a shared experience that students can use to better understand complex frameworks together. Glaser (1994) argues, *“Learning environments that involve dialogue with teachers and between peers provide opportunities for learners to share, critique, think with, and add to a common knowledge base.” (p. 19)*

The use of educational games as an intervention tool is expansive. However, few studies are done within the context of an informal game-embedded learning environment (Berland & Resier, 2011), and as such, meaning-making in games and the use of games in the classroom community is, still a relatively new field. The relationship between collaborative or competitive discourse with productive failure in gameplay is just one of many areas to explore.

1.4. Research Questions

In order to investigate the role that failure has in an endogenous educational game-embedded learning environment, we conducted a mixed-methods study of game-based learning in the context of a spring break STEM camp program. Using discourse and clickstream data gathered from 88 participants across a 5-day curriculum about virology, we examined in-game failure, group discourse, and student learning gains to answer the following research questions:

- Research Question 1: What is the relationship between in-game failure and learning gains (as measured by pre-post assessments) in a game-based intervention?
- Research Question 2: What is the relationship between in-game failure and participants’ game-related discourse?

2. Methodology

2.1. Virulent

Virulent (Figure 1) is a game designed and developed by Learning Games Network and the Games+Learning+Society group [redacted for anonymity]. In Virulent, players take the role of a virus invading a host, controlling the virus and its parts within a cell, evading the cell’s defenses, stealing resources, and replicating. Players receive new abilities and controls as they progress through the game’s levels, while also facing new and more challenging content, taking them through the life cycle of a virus. *Virulent* consists of 13 linear “main levels” and 7 optional “challenge levels”.

The core design principles of PF environments detailed by Kapur and Bielaczyc (2011) and outlined above, are core elements in game-based learning, and especially in the design of Virulent. Challenging problem-solving environments, exploration, elaboration, and ability to compare failed and completed solution methods are all inherent to the game’s design. More specifically, *Virulent* was designed to challenge the player’s abilities throughout levels that promote interaction with embedded virology concepts. Players test strategies on how to evade cell defenses such as free and bound antibodies, penetrate the cell wall, steal resources, and propagate before escaping the dying cell.

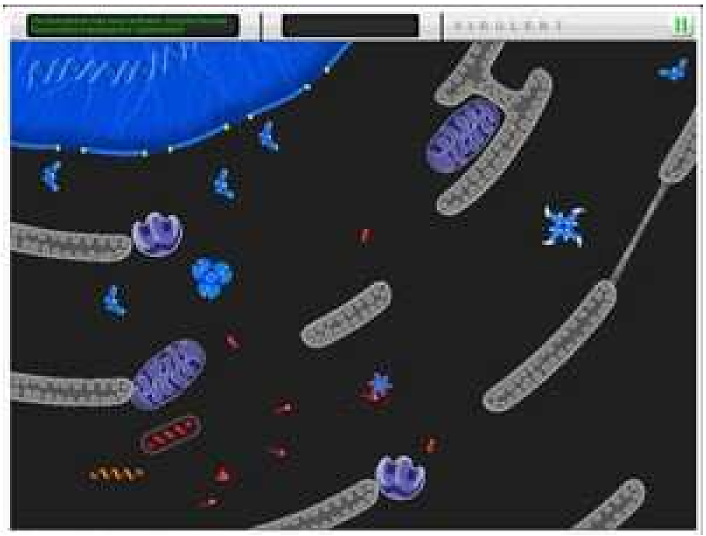

Figure 1: Virulent Gameplay

Further, the four mechanisms required for a PF environment identified by Kapur and Bielaczyc (2011) are core mechanics seen in many games. Activation and differentiation of prior knowledge manifest in games as levels progress; games are often constructed to be well-ordered (Gee, 2005), in that levels build off of the skills that players are developing through play. When a player begins *Virulent*, they have to evade bound antibodies that are attached to a B-cell. This gives them a chance to learn how to maneuver their units through the level with a low level of risk. However, once they master movement and complete the first level, they then face free antibodies that actively seek out their units. They are then challenged to build off of their abilities to move through the level but have to adjust their movements to evade the incoming free antibodies. As the game progresses, players are continually challenged to build off of their previous abilities through “cycles of expertise” (Gee, 2005) that are tied to the learning goals of the game, i.e. how a cell defends against viruses. After a player completes a level, they are asked to spend time reflecting on it with their group members and the facilitator; and are given opportunity to ask questions about the game and the embedded material, elaborating on the concepts to an understanding of how they work outside the game. Within the curriculum, building models of the content also gives the participants opportunity to organize the knowledge they are developing into a full conceptual model. From their gameplay, players reflect on the embedded material, and assemble a complete model of the life cycle of a virus. *Virulent* was designed to be “pleasantly frustrating” (Gee, 2005, p.6) to maintain an appropriate challenge level; the game is designed to build on player abilities from level to level, testing the player’s skill as they progress. It provides moments throughout the game for the player to fail yet develop their abilities, keeping success within sight.

2.2. The Game-Based Curriculum

For this study, we held three after-school summer camps in which we provided participants with tablets to play *Virulent* extensively. Each camp included 5 days of curricular activities around *Virulent*, during which participants roleplayed as scientists situated in a narrative where in the world was becoming infected with the “Raven Virus” and were selected by the Center for Disease Control (CDC) to investigate how to stop the virus from spreading. The tablets they used to play were described as technologically advanced devices called "digiscopes" that allowed them to see inside an infected person's body, control the virus, and see how the immune system responded. Participants were assigned to groups of 3-4, in which they discussed the game, the embedded content, and worked on preparing physical paper models of the virus life-cycle and the immune system. They also recorded and presented their proposed solutions about how to deal with the Raven Virus to the mock CDC and debated among their peers on the best ways to stop the virus. Participants were also encouraged to play the challenge levels on their own time during the event or at home, but they were optional.

It was the facilitator's role to guide players through the camp activities. Players were kept to a schedule and were asked to follow along with the group but were given the freedom to play at their own pace and to contribute in their own ways. The facilitator encouraged players to ask other group members for help when they encountered a difficult part of the game, but only in advice; players were not allowed to take control of each other's' tablets to play the game for them. When asked questions about what something in the game is or does, they were referred to the in-game almanac that detailed each of the game units and their functions. Players were also encouraged to play on their own time and could earn rewards by completing optional challenges outside of the main curricular activities.

2.3. Participants

Eighty-eight participants were recruited through fliers at community centers and schools in a large midwestern city. One participant was asked to leave due to interpersonal conflict with other participants at the end of the first day, and as a result a total of eighty-seven participants' data were used in the subsequent analyses. Twenty-four identified as female, fifty identified at male, and thirteen did not choose to provide any gender information. Thirty-seven identified their race as White, fourteen identified as Black, eight identified as Hispanic/Latino/Latina, two identified as Asian, one identified as White and Asian, one identified as Black and Native American, one identified as White and Hispanic, and twenty-three did not provide any information on racial background. Age of participants ranged from 9-14 and averaged at 11.4. Participants had an average of 6.38 hours per week of tablet usage, 6.97 hours per week of game usage, and average self-reported grades of A's-B's. Three main data streams were collected: learning assessment data, gameplay data, and discourse data. Each data stream is detailed individually below.

2.4. Virology Content Knowledge Assessment

Assessment for virology content knowledge was created with a collaborating middle school teacher and administered before and after each camp. Questions were designed to gauge participants' knowledge of virology concepts, attitudes towards science, and their self-efficacy towards understanding science concepts. The assessment included 7 multiple choice questions, 1 choose all that apply question, 2 open response questions, and 1 diagram labeling question, 3 Likert questions on self-efficacy towards understanding science concepts, and 3 Likert questions on attitudes towards science concepts. Open response questions were scored out of 4 and multiple-choice questions were scored and weighted the same. Participants were encouraged to leave a response blank if they did not know the answer and that the assessment was not a quiz, but a way for us to tailor the event to what they knew.

2.5. Gameplay Data

Players' gameplay data were recorded using the Assessment Data Aggregator for Game Environments (ADAGE) system (Owen & Halverson, 2013; Stenerson et al, 2014). ADAGE allows for recording of each action taken in the game and the context it is taken in. This includes every point and click they make while playing, as well as the level they are currently in and what units are on the screen at the time. This dataset allows us to delve into the ways students engaged with the game, as well as when and how they fail, shedding light on their developing strategies throughout play. In this study, failure is defined as the inability to complete a level in-game, as shown by level failure in telemetry data. Clickstream data were cleaned to remove logging errors caused by technical glitches and were formatted to show individual gameplay patterns across levels. These errors were largely duplicated logs and identified by impossibilities in game behaviors such as completing the same action twice within milliseconds.

2.6. Discourse Data

Each participant received a lavalier audio-recorder in order to capture their discourse around gameplay and curricular activities. Discourse was recorded throughout each 90-minute session, and then cleaned for transcription. Due to the amount of audio recorded, time and cost was a concern. To maintain manageability, three unbroken 10-minute segments per group per day of "most on-topic" audio was selected and sent to a transcription service. These segments were identified by participants talking about the game or the camp, rather than personal matters. Completed transcripts were analyzed using MAXQDA (v. 12) software. For this analysis, our team examined negative value statements and their relation to in-game failure.

2.7. Data Analysis

Previous analyses of pre-post assessments showed overall learning gains, as reported in [redacted for anonymity]. Current analysis focuses on gameplay and discourse factors in this increase. To this end, factorial ANOVA was conducted to compare the main effects of pre-assessment, number of failures before initial success across all main levels, total number of failures across all main levels, total attempts across all main levels, total number of levels completed at least once, and number of minutes spent in main levels on post-assessment scores.

Discourse data were mined for negative value statements around level failures, including "lost, hard, difficult, not easy, tough, impossible, hate, dislike, uncool, sucks, bad, stinks, not good, isn't good, lame, terrible, boring, bored, boredom, failed, failure, fails, failing, and fail". All text was manually reviewed to ensure that captured words were affiliated with moments of failure or discourse surrounding failure, and not a circumstance where a word like "bad" or "hard" was used to describe an activity outside of gameplay or curricular activities. Number of level failures was averaged over groups and groups were split based on high and low failure rates for comparison of discourse, and frequency of utterance was compared between high failure groups and low failure groups. Using negative value statements around level failure as a starting point, we followed discourse to contextualize the discourse on gameplay or in argumentation between group members to look for examples of how failure was influencing group discourse.

3. Results

3.1. Gameplay

Factorial ANOVA showed that number of failed attempts before success across all levels ($F(1,75) = 7.330$, $p = 0.005$, $\eta^2 = 0.087$), total attempts at main levels ($F(1,75) = 4.119$, $p = 0.046$, $\eta^2 = 0.043$), and pre-assessment scores ($F(1,75) = 6.235$, $p = 0.015$, $\eta^2 = 0.065$) were significant predictors of post-assessment scores. Other time on task measures including game progression ($F(1,75) = 1.194$, $p = 0.278$), and total time playing the game ($F(1,75) = 0.631$, $p = 0.430$) were found to be non-significant. Total number of failures at main levels including attempts after initial success were also found to be non-significant ($F(1,75) = 0.023$, $p = 0.880$).

3.2. Discourse

A search of all terms relating to failure resulted in a total of 550 utterances across all groups in the camp. Of this, fifty-four percent (N=300) of failure discourse evaluated difficulty. The most popular negative value statements used by participants were “hard” (N=275) and “bad” (N=82). In the other 25 instances, players described the level being played as impossible. Groups with a higher rate of gameplay failure spoke more frequently about moments of failure (58% in higher failure groups vs. 42% in lower failure groups). In 136 of these 550 instances, statements about gameplay and failure occurred during a scientific argument, although they were not always the subject of the argument itself, but a comment made as team members attempted to resolve a conflicting claim or defense statement. In the following example, participants debate how to evade slicer enzymes and strengthen their forces against the immune system. One of the participants is trying to distract the slicer enzymes by sacrificing a genome (e.g. decoy) in order to bring the remaining ones to safety. The other boys are questioning whether the participant’s decision to sacrifice the genome is successful and if additional genomes need to be reproduced. The participant playing the game nearly loses and is forced to redirect his decoy back toward the slicer enzymes:

Participant 58: He's got one [Genome]. He just got one.
Participant 59: I know, it's coming.
Participant 18: But now you probably over shredded.
Participant 58: You can go now.
Participant 18: Attack you unless it's laying over and they can’t
Participant 58: How many do you want?
Participant 59: As many as I can.

Participant 58: He's using one of his, he's using his antigenome.
Participant 59: Whatever so you can't create MRNA--
Participant 59: Oh crap, crap, crap, no, no, no, no, no, no, no. No, no, no. You shall not pass. I need you as a decoy.
Participant 58: I'll put this away in my pocket.
Participant 59: Then it just falls, dies.

Examination of utterances containing failure tokens showed that negative value statements surrounding in-game failures often led to collaborative discourse and resulted in a focus on mechanics with embedded virology terms. Below, we detail a few of many examples. In the first such event, participants construct knowledge while trying to help a group member complete a level:

Participant 78: I just died. I just died into the brain or whatever that is.
Participant 86: Yeah, me too.
Participant 78: I got sucked into this blue thing… The B cell.
Participant 86: Okay. So what I did was like I pressed the little red thing in the corner when you first start and then you press virion or something like that. And then you start guarding it. Then if it doesn't pop up, you get it right there.
…
Participant 78: So I'm supposed to be sucked in?
Participant 97: Isn't it the brain? A B cell?
Participant 86: Yeah it is.

In another event, a group member explains they do not know the best strategy for completing the level and is attempting to resolve the challenge on their own. Another group member chimes in with a recommendation. Once more, while the conversation is not directly requesting information on virology, students are nevertheless adopting these terms in order to communicate game strategies to each other:

Participant 23: I don’t get where [sic] I’m supposed to do right now but it’s kind of nice because I have a ton of things. I don’t know what I’m supposed to be doing now, though.
Participant 01: I made as many as I could, then I made little N proteins. AN [sic] then the PRRs ate a lot of them, so that made them go away. And then I got to block them all. That’s how I did it.
Participant 81: Is this one hard?
Participant 83: This one? It’s pretty easy for me.
Participant 81: I have to make the Gs? (G-Protein)

Participant 83: Yeah but you have to eat that, and that, and that.

Here, a participant complains about failing a level and is having trouble remembering the name of the unit they are controlling. Another group member chimes in with what they think it's called - an M protein, but is incorrect. The third group member corrects them, and the group comes to agreement that the unit they are using in the level are RNAs:

Participant 63: I died.
Participant 60: Okay.
Participant 60: Okay.
Participant 63: Wait no.
Participant 63: That could be--
Participant 64: That could be an M protein.
Participant 63: Yeah.
Participant 60: RNAs.
Participant 63: Oh yeah, RNAs.

In another example, a participant is struggling to finish a level, which leads them to ask a group member for help. To explain their strategy, the more successful group member uses virology terms, explaining the missing parts the virus needs to create a budding site and escape a dying cell:

Participant 69: And I died again.
Participant 69: Me? I'm just -- for all of them.
Participant 69: Yeah.
Participant 62: And --
Participant 69: Okay, something along those lines?
Participant 62: Sure.
Participant 69: What did you do?
Participant 62: In the budding site, you add M proteins and G proteins.

4. Discussion

4.1. From In-Game Failure to Assessment Success

These results shed light on both of our research questions. First, they suggest that failures before succeeding at a level for the first time supports students reaching the underlying learning goals of the game. Number of failures before initial success and only one of three time on task measures (number of main level attempts) were significant predictors of post-assessment scores in this model, suggesting that these in-game level failures were productive for student learning gains. Further, the only significant time on task measure is likely confounded with number of failed attempts, as this number would increase with failed attempts as well. Students are using their developing understanding to beat the game; when they fail, they are forced to reflect on the strategies they were using and modify them. Students are given opportunities to discover where their misconceptions and inabilities lie and work towards modifying them to succeed.

Second, students show a link between this failure and collaborative discourse focused on the learning-embedded mechanics within the game. As an endogenous game, in which the embedded learning goals are synchronized with the game content, discussion of game content is directly related to the embedded learning goals of the event. Participants not only collaborated on gameplay strategy but simultaneously discussed virology terms embedded in the game as a way to communicate their ideas. As one participant begins by examining how the B Cell contributed to level failure: ("I just died into the brain or whatever that is") and then questions its purpose following gameplay ("So I'm supposed be sucked in?"), it eventually leads the group to understand better what a B cell is ("I got sucked into the blue thing… The B cell") along with its purpose ("Doesn't it like kill it? No, it doesn't kill it. It keeps it"). During this dialogue, group members are following along, helping to cement their own understanding of the underlying material as well ("A B cell?" "Yeah it is.").

The moments of collaborative discourse and strategy development provide participants with an opportunity to become intimately familiar with the game, the underlying material embedded in the game mechanics, and share knowledge with each other. Group members strive to finish difficult levels that their peers have trouble with, and help others when they find a successful strategy. Or, when all members of the group are stuck on a challenging level, they are encouraged to brainstorm ideas together to work towards a solution. In these scenarios, level failures initiate group discourse that revolves around the embedded learning material by virtue of the game mechanics. Participants become more familiar with the terminology and the virology concepts embedded in the game mechanics because they, as a group, encounter moments of failure throughout play.

Traditionally, productive failure (PF) occurs in an environment in which learners are asked to solve a problem with little structure to help them (Kapur, 2008). This study sheds light on a new way that PF can manifest; here, players that struggle through levels before completing them see larger learning gains in the independent virology assessment. This pattern bears resemblance to PF, although it manifests in a slightly different way; instead of failing at a single problem-solving task, players are confronted with multiple points of challenge and failure. In *Virulent*, players struggle through multiple levels with a range of embedded content. Further, the act of failing in *Virulent* is not necessarily due to misunderstanding of the underlying learning objectives, but may be because they lack the game-related skills required to complete the level. This does, however, encourage participants to repeatedly interact and think about the content embedded as mechanics in the game and gives opportunity to collaboratively guide group members, cementing their understanding and furthering others'. As players struggle through the difficult levels of the game, they are often challenged to think about how the virus can most effectively invade the host, and how the body defends against infection. This provides moments of reflection on the content in which players get to manipulate the system they are studying through gameplay to understand the lifecycle of a virus. *Virulent* provides smaller moments of failure in a way designed to encourage deep thought on the embedded concepts through gameplay mechanics. Coupled with a curriculum that encourages collaboration, players have frequent opportunities to struggle through material that is designed to teach them about the embedded learning goals and reflect on the underlying concepts with their group members.

Although all participants came into the camp with little familiarity with virology or immunology, players had varied levels of pre-win failure. Though few, there were some players who completed the game with little trouble. The players who benefited the most from the game were not the ones who completed it easily, but the ones who struggled through it. The number of failures before success more strongly influenced learning outcomes than overall number of failures, or number of levels attempted regardless of success or failure. Even when players complain about the difficulty, similar to PF environments, struggling until you succeed appears to support learning in this environment as well.

4.2. Limitations

Although our analyses show that level failures before success are a strong predictor of pre-post gains, it is important to recognize that it is not the only factor at play. Group discourse played a vital role in helping participants work through and become more familiar with embedded content. Group dynamics play a vital role in fostering these learning gains and need to be explored fully. Likewise, ensuring a game is difficult to all players is non-trivial as some players may have more experience with the selected game genre. This raises questions about the other dynamics involved in level failures that promote content knowledge development. A more direct study on the influences of level failures in other settings such as solo work is needed to tease this out.

Further, the composition of groups is an important dynamic for collaboration. Even if a level failure occurs, it may be meaningless if the group does not use that as an opportunity to talk about the underlying material. This may occur for a number of reasons, including interpersonal conflict. Likewise, there are a myriad of other group dynamics to take into consideration that are outside the scope of this study. During our event we found a good balance by allowing participants to change groups if they wanted to, and having a facilitator available for each 3-4 participant group to help work out any conflict that arose. Many educational settings do not have these luxuries and may not be able to implement a learning environment such as this, in this manner.

4.3. Implications

Leveraging the power that games bring to learning environments has always been an attractive prospect to educators, researchers, and policy-makers alike. Although, an aspect that contributes to the success of games that educational games tend to neglect is the level of challenge expected. One of the largest barriers educators have when using games for educational purposes is insufficient time (Takeuchi & Vaala, 2014), and any impasse a student encounters only takes away more precious time that players could be progressing through content. Creating games that players can easily move through seems like a logical solution to this problem, but this robs the game experience of the challenging moments that exemplify the "edge of your seat" play that make games successful. Too wary are creators of educational games of players missing out on content by getting stuck that they neglect the moments of impasse that encourage players to think deeply about the content in which they are engaged.

It is important to note that number of failures before first success predict post-assessment scores, whereas total number of failures does not. This points to a distinction between pre and post completion play and may be an indication of players approaching levels differently before and after their initial victory. It may be that the players' strategy changes depending on whether they have figured out how to beat a level or not, and post-win failures may utilize strategies that do not support the same learning processes that we see in pre-win failures.

This is a strong instance of how appropriately designed games can invite players to engage with meaningful content, while providing an engaging platform to experience failure, disequilibria, and explicated sensemaking which enables their eventual success. Strategies that commercial game development companies use to ensure players are consistently at an appropriate challenge level could provide educational games with methods to create PF environments in educational games. In the current study, some players progressed through the game with little difficulty, which may be due to varied levels of gameplay experience. Use of development strategies that tailor game difficulty to player ability in an educational game would allow each student to stay at the edge of their competency, remain engaged, and maximize their moments for impasse and reflection of underlying concepts. This provides support for the conscious development of difficult game content in games made for educational purposes. If these games are created with the intention of providing an environment where failure is expected and celebrated, players will spend more meaningful time thinking about and reflecting on the content that they're immersed in, providing moments of productive failure throughout play.

However, this cannot be said without pointing out the tension that lies between educational games and assessment in classrooms. The failure that drives the effects seen here would be meaningless if the negative impact of failing were too high. The classroom environment that learning occurs in is often teeming with high stakes assessments. Introducing failure as a means of progression rather than assessment into these environments may be a productive, albeit difficult to implement, change and deserves more attention.

Productive failure in gameplay shows promise in helping students to not only think critically about strategy, but also to share this information with others. In some instances, especially in endogenous learning games, these strategies also communicate scientific definitions and ideas. The mention of gameplay failure was uttered by multiple participants in varied ways, ranging from "bad" to "impossible," but many times resulting in a further exchange on what made these activities so difficult to complete. In this respect, gameplay failure was something shared and discussed by participants, and was not an activity isolated to a single player. Upon failure, players initiated dialogue that led them to think about the concepts embedded in the game by virtue of the endogenous game mechanics.

4.4. Future work

As a basis, this provides us a good start to think about how some mechanics used to leverage engagement in commercial games can be applied and are integral to ensuring engagement in games developed for educational means. However, more investigation is needed. Contrasting play patterns pre-win and post-win is required to fully understand how failing before and after success are different, and how play patterns in relation to failure illuminate learning gains. Further, failure comes in many shapes and sizes and is not always reacted to in the same ways. A plethora of work is ahead of us to fully understand how failure in games is experienced and how it can be leveraged for productive ends.

5. References

Blumberg, F. C. (2000). The effects of children's goals for learning on video game performance. Journal of Applied Developmental Psychology, 21(6), 641-653.

Blumberg, F. C., Rosenthal, S. F., & Randall, J. D. (2008). Impasse-driven learning in the context of video games. Computers in Human Behavior, 24(4), 1530-1541.

Bielaczyc, K., & Kapur, M. (2010). Playing epistemic games in science and mathematics classrooms.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). How people learn.

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning and Digital Media*, *2*(1), 5-16.

Gee, J. P. (2007). What Video Games Have to Teach Us About Learning and Literacy.

Glaser, R. (1994, July). Application and theory: Learning theory and the design of learning environments. In 23rd International Congress of Applied Psychology, Madrid, Spain.

Halverson, R. (2005). What can K-12 school leaders learn from video games and gaming?. *Innovate: journal of online education*, *1*(6), 3.

Hunicke, R. (2005). The case for dynamic difficulty adjustment in games. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology* (pp. 429-433). ACM.

Juul, J. (2009). Fear of failing? the many meanings of difficulty in video games. *The video game theory reader*, *2*, 237-252.

Juul, J. (2013). The Art of Failure: An Essay on the Pain of Playing Video Games. Mit Press.
Kapur, M., & Bielaczyc, K. (2011). Journal of the Learning Designing for Productive Failure. Journal of the Learning Sciences, 21(August 2012), 37–41.

Kapur, M. (2008). Productive failure. Cognition and Instruction, 26(3), 379–424.

Kapur, M. (2010). Productive failure in mathematical problem solving. Instructional Science, 38(6), 523–550.

Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. Instructional Science, 39(4), 561–579.

Kapur, M. (2012). Productive failure in learning the concept of variance. Instructional Science, 40(4), 651–672.

Kapur, M. (2013). Comparing Learning From Productive Failure and Vicarious Failure. Journal of the Learning Comparing Learning, (October 2013), 37–41. Doi: 10.1080/10508406.2013.819000

Kapur, M. (2014). Productive failure in learning math. Cognitive Science. 38(5), 1008–1022.

Kapur, M. (2015) Learning from productive failure, Learning: Research and Practice, 1:1, 51-65, DOI: 10.1080/23735082.2015.1002195

Kapur, M., & Bielaczyc, K. (2011). Classroom-based Experiments in Productive Failure. In *CogSci*.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. The Journal of the Learning Sciences, 21(1), 45–83.

Litts, B., & Ramirez, D. (2014). Making people fail: Failing to learn through games and making. *Proceedings GLS*, *10*, 160-166.

Matson, J. V. (1991). The art of innovation: Using intelligent fast failure. Pennsylvania State University.

Owen, V. E., & Halverson, R. (2013). ADAGE: assessment data aggregator for game environments. In *Proc. Games, Learning, and Society Conf.*. ETC Press.
Piaget, J. (1955). The child's construction of reality. Routledge & Kegan Paul Limited.

Popper, K. (1959). The logic of scientific discovery. New York, Hagerstown, San Francisco, London: Hutchinson.

Ramirez, D., Seyler, S., Squire, K., & Berland, M. (2014). I'm a Loser, Baby: Gamer Identity & Failure. *Proceedings of DiGRA 2014: Verb that ends in "ing"> the noun> of Game plural noun*.

Sitkin, S. B. (1992). Learning through failure: the strategy ofsmall losses. Research in organizational behavior, 14, 231-266.

Stein, N. L., & Albro, E. R. (2001). The origins and nature of arguments: Studies in conflict understanding, emotion, and negotiation. Discourse processes, 32(2-3), 113-133.

Steinkuehler, C. A. (2004, June). Learning in massively multiplayer online games. In *Proceedings of the 6th international conference on Learning sciences* (pp. 521-528). International Society of the Learning Sciences.

Stenerson, M. E., Salmon, A., Berland, M., & Squire, K. (2014). Adage: an open API for data collection in educational games. In Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play (pp. 437-438). ACM.

Squire, K. (2011). Video games and learning. Teaching and participatory culture in the digital age. New York, NY: Teachers College Print.

Squire, K., & Barab, S. (2003, June). Replaying history: Engaging urban underserved students in learning world history through computer simulation games. In Proceedings of the 6th international conference on Learning sciences (pp. 505-512). International Society of the Learning Sciences.

Takeuchi, L. M., & Vaala, S. (2014). Level up learning: A national survey on teaching with digital games. New York: The Joan Ganz Cooney Center at Sesame Workshop.

VanLehn, K. (1988). Toward a theory of impasse-driven learning. In *Learning issues for intelligent tutoring systems* (pp. 19-41). Springer US.

Wark, M. (2015). Losing Is Fun. The Gameful World: Approaches, Issues, Applications, 163.