Why Video Games can be a Good Fit for Formative Assessment

Malcolm Bauer¹*, Caroline Wylie¹, Tanner Jackson¹, Bob Mislevy¹, Erin Hoffman-John², Michael John³ and Seth Corrigan⁴

¹ETS, New Jersey, United States; mbauer@ets.org
²Carnegie Mellon University, Entertainment Technology Center, United States
³University of California, Santa Cruz, United States
⁴LRNG and University of California, Berkeley, United States

Abstract

This paper explores the relation between formative assessment principles and their analogues in video games that game designers have been developing over the past 35 years. We identify important parallels between the two that should enable effective and efficient use of well-designed video games in the classroom as part of an overall learning experience organized and facilitated by teachers. We describe the parallels and then show how game-design elements are used in the service of formative assessment principles in Mars Generation One (MGO): Argubot Academy™, a video game developed by GlassLab that focuses on formative assessment of middle school students’ argumentation skills. MGO was designed and developed together with a broader curricular unit on argumentation within which the game was situated. We discuss how design elements in the game satisfy each core formative assessment principle and how the game is connected to the instructional unit, demonstrating this marriage of game design and formative assessment. Prior work has reported on preliminary evidence on efficacy of the game when used as part of this middle school language arts unit.

Keywords: Formative Assessment, Game Design

1. Introduction

The definition of a good game is therefore “one that teaches everything it has to offer before the player stops playing,” That’s what games are, in the end. Teachers. – Raph Koster (2004).

Formative assessment is one of the most effective approaches to improving student learning in the classroom (Black and Wiliam, 1998; Hattie, 2008). Video games use analogues of formative assessment principles that game designers have been developing and tweaking for the past 30 years. In this paper, we describe some of the parallels between principles of formative assessment and properties of good games. We then discuss the design of a current game-based assessment in use in classrooms to demonstrate how assessment and curriculum designers can leverage the properties of good games to support classroom-based formative assessment and student learning. We address formative assessment within the game itself and through connections made between the game and other teacher-facilitated classroom activities. We also identify some remaining research questions in the field of game-based formative assessment.

Black and Wiliam (1998) define formative assessment broadly as “all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (p. 8). Hattie (2008) conducted an extensive meta-analysis of quantitative measures to better understand the relative effectiveness of different factors on educational outcomes. The meta-analysis focused on studies that manipulated one of the factors of interest and had a control group so that the impact of the manipulated factor could be observed. Of the 138 candidate influences on learning, formative evaluation (i.e., formative assessment) had the third
Why Video Games can be a Good Fit for Formative Assessment

largest effect size. The analysis provides strong evidence that use of formative assessment has valuable payoffs for student learning.

Since well-crafted videogames incorporate many principles of formative assessment, we argue that they are a natural fit for classrooms in which teachers regularly use formative assessment. Videogames can be used in conjunction with other activities in classrooms, for example, to allow teachers to use the formative assessment mechanisms in games as a structure around which to build a curricular unit with increased opportunities to collect evidence of student understanding.

To make this argument, we describe several core principles of formative assessment and its analogous game design element. We then provide an example of a game-based formative assessment—Mars Generation One™ (MGO)—which was designed and developed by a team of professional video game designers, cognitive scientists, assessment and learning designers, and psychometricians in conjunction with a broader curricular unit focused on argumentation within which the game was situated. We discuss how design elements in the game satisfy each core formative assessment principle and how the game is connected to the broader instructional unit in order to demonstrate how games can be used in the classroom as part of a broader formative assessment process. Finally, we discuss the limitations of the work and outstanding research questions.

2. Principles of Formative Assessment

Because we are going to argue that design elements involved in creating an engaging videogame implement principles of formative assessment, we first discuss the principles.

Formative assessment came to many educators’ attention following the extensive Black and Wiliam (1998) literature review. Black, Harrison, Lee, Marshall, and Wiliam (2003) described formative assessment as occurring when “information about learning is evoked and then used to modify the teaching and learning activities in which teachers and students are engaged [emphasis in the original]” (p. 122). Other studies have continued to add support to the Black and Wiliam (1998) review including those from Hattie’s (2008) meta-analyses as well as Brookhart (2005), supporting the strong and positive connection between teachers’ use of formative assessment in everyday teaching and improved student learning.

While the field has described a variety of sets of formative assessment principles, Leahy, Lyon, Thomson, and Wiliam (2005) and Heritage (2010) define principles of formative assessment that cover many components of teacher-student and student-student interactions in the classroom ranging from the setting of learning goals and kind and timing of feedback teachers provide to students to the ways students collaborate and take ownership of their own learning. These principles include:

- Teachers identify and share learning expectations with the students (learning goals and progressions).
- Teachers elicit relevant and quality evidence of student learning on an ongoing basis to inform instruction.
- Teachers structure opportunities for students to take ownership of their own learning.
- Teachers structure opportunities to activate students as instructional resources for one another.
- Teachers provide feedback to move learning forward and create a structure for students to act on it.

While it is often useful to separate out the elements of formative assessment for the purpose of describing and analyzing practice, ultimately all the elements need to be combined into a coherent whole. Teachers can only elicit evidence of student learning when there are clear learning goals or success criteria against which to gauge learning. Similarly, self-assessment requires a clear understanding of the learning goals. The teacher and peers can only provide meaningful feedback when evidence of current learning has been elicited. And all these principles may be enacted in the classroom for the class as a whole, for small groups of students, and for individual students, often with a teacher working on a variety of short- and medium-term goals for students depending on their needs. As noted by Shavelson et al. (2008), formative assessment operates on a continuum that ranges from informal to formal practice: on-the-fly, planned-for-interactions, and embedded-in-the-curriculum formative assessment. The combination of formative assessment within the game and as part of the larger instructional unit spans a range of informal and formal assessment opportunities.
3. Formative Assessment and Video Games

We posit that engaging games contain analogues of these formative assessment practices that video game designers have been using for the past 30 years. Table 1 provides some of the parallels between the formative assessment principles above and design elements in games in the middle column. The final column illustrates how the game-based assessment can then be connected to the broader formative assessment process used for an instructional unit in which a game is embedded.

We describe each practice, its video game analog, and provide an example from an existing commercial video game to demonstrate how game designers have naturally embedded each in entertainment video games. In the section following this one, we use our design of the educational game Mars Generation One™ (MGO) to illustrate both the game analogs of formative assessment and the connections between aspects of formative assessment embedded within the game and those that are part of the larger instructional unit employed in classrooms.

The first formative assessment practice is “Teachers identify and share learning expectations with their students.” In the classroom, this may take the form of teachers providing a clear and specific learning goal for a single lesson such as “learn to add single digit negative numbers” or for a larger curricular unit such as “learn to construct arguments and critique the arguments of others.” Helping students understand learning expectations can also take the form of guided group discussions with students about what the goals should be. Underlying larger learning goals are often learning progressions – descriptions of levels of sophistication of thinking and understanding that students often pass through on their way to mastery (Heritage, 2008). Transparency of learning goals is most critical, and research supports the claim that student learning improves when they understand what the learning goals are (Tell, Bodone, & Addie, 2000; White & Frederiksen, 1998).

Games are similarly transparent because they provide learning expectations through explicit challenges as well as the sequencing of challenges, missions, or quests. There are different ways of providing challenges which are often connected to different game genres. For example, “platform” games follow a fixed sequence of explicit challenges through which a player advances upon successful completion each prior challenge. Some Role Playing Games (RPGs) like the Walking Dead™ embed challenges within a narrative and provide them (and adjust narrative) dynamically based upon player performance. By contrast, sandbox games, such as Minecraft™ and SimCity™, allow more open play where players can create challenges on their own as individuals or as part of the larger game playing community rather than challenges

<table>
<thead>
<tr>
<th>Formative Assessment Practice</th>
<th>Video Game Analog</th>
<th>Connecting the Game to the Broader Learning Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teachers identify and share learning expectations with the students (learning goals and progressions)</td>
<td>Games provide explicit challenges and goals</td>
<td>Game goals and challenges are connected to larger unit learning goals</td>
</tr>
<tr>
<td>Teachers elicit relevant and quality evidence of student learning on an ongoing basis to inform instruction</td>
<td>Games collect telemetry and use analytics to understand player behavior and structure the game experience</td>
<td>Evidence collected and identified in game informs instruction outside the game, and is integrated with other evidence about student learning</td>
</tr>
<tr>
<td>Teachers structure opportunities for students to take ownership of their own learning</td>
<td>Games create a sense of agency by providing choice and flexible pacing</td>
<td>Students have opportunities to reflect on their own learning both within the game and as part of the larger instructional unit</td>
</tr>
<tr>
<td>Teachers structure opportunities to activate students as instructional resources for one another</td>
<td>Games provide structured contexts for collaboration and sharing – multiplayer platforms</td>
<td>Students have opportunities to reflect with peers about their learning within the game, and to provide feedback to peers on their learning</td>
</tr>
<tr>
<td>Teachers provide feedback to move learning forward and create a structure for students to act on it</td>
<td>Games provide just in time feedback linked to rewards that guide players’ actions</td>
<td>The teacher is able to provide feedback that helps students connect learning within the game to the broader instructional unit</td>
</tr>
</tbody>
</table>
being provided by the original game designers. In each of these types of games, players’ goals in the game are made explicit, and, as in formative assessment, there are both immediate goals and larger longer-term goals.

The second practice is “teachers elicit relevant and quality evidence of student learning on an ongoing basis to inform instruction”. In the classroom, this practice means teachers provide activities in which students’ responses, actions, conversations, and work products contain features that teachers can use as evidence of students’ understanding. Teachers can use a variety of approaches to elicit evidence such as 1. Asking students to write short answers to a series of questions on individual white boards to both systematically and quickly get a sense of student understanding. 2. Polling students to see the range of students’ opinions on a topic, and 3. Noticing important features of students’ discourse in classroom conversations. In each case, students are producing something observable that teachers can use to gauge students’ current understanding. The evidence is elicited frequently to inform immediate next steps, but broader patterns in performance may be identified to inform longer term plans for individuals, small groups, or the class as a whole (Graham, Milanowski, & Miller, 2012).

Similarly, games use telemetry and analytics to continually capture and analyze evidence in game play. In many games, all players’ interactions – what players do and when they do it - is collected. This structured data that is collected and acted on during game play is called telemetry. When these data are analyzed and meaningful features are extracted, they become analytics. Analytics are created during game play, for example, in a game that examines the pattern of successes of a player and dynamically increases the difficulty of challenges, as a teacher might do during a class. It can also occur over a much longer time scale with a game community as a whole involved. The release of the game StarCraft™ provides an example of these two ways of using analytics. When StarCraft™ was first released; it collected a variety of telemetry data and used analytics to report on player performance, analogous to teachers collecting evidence of student learning. Players used that individually to reflect on and improve their performance, but also the player community as a whole used it to form the basis for their evidence-based beliefs about what effective performance was and what skilled players did. Eventually, the game developers used on-line discussions and analysis of these data and what they implied about skilled game play to revise and expand the reporting (or game analytics) done in the game. Game analytics like these highlight patterns in players’ actions and their consequences. They are directly useful to players as feedback to analyze and improve their play and to stimulate and structure conversations among them. They can also serve as input to models that relate the data to higher-level, more cognitively structured models cast in terms of learning goals; e.g., students’ understanding of concepts and strategies, or, as in our example game, MGO, described in the next section, levels of performance in a learning progression.

A third formative assessment practice is “Teachers structure opportunities for students to take ownership of their own learning”. In the classroom, this includes providing opportunities for students to engage in self-assessment of their own work, reflect on learning strategies that help them to organize their thoughts, and engage others in thinking about and expanding their explanations through classroom discussion processes like Accountable Talk (Michaels, O’Connor, & Resnick, 2008). Engaging students in this way can have positive effects both on student attitudes and learning outcomes (Cohen, Raudenbush, & Ball, 2003).

Similarly, video games engage students in setting their own goals and reflecting on their progress through choice and flexible pacing. Game designers focus on giving players a sense of “agency” - that their decisions are consequential and they feel empowered by this (Koster, 2014). Agency is achieved at different grain sizes, for example, at larger grain sizes by enabling players to define their characters and make decisions about which challenges to accept, and at smaller grain sizes by providing choice of what things to say to another character, or simply where to place or remove objects in the game environment. While agency is especially clear in “sandbox” games that allow players to create their goals, set their own timing, and choose actions as they wish and still receive immediate feedback, it is also true of other game genres. As in the analogous classroom-based formative assessment principle, the intent in video games is to create a rich, satisfying experience that engages a player more fully.

A fourth formative assessment principle is “Teachers structure opportunities to activate students as instructional resources for one another”. In the classroom, this practice might take the form of paired or group
discussions or activities so that students can bounce ideas off each other. It also can include providing structures and opportunities for students to provide feedback on each other's work leading to improved learning outcomes (King, 1992; Mercer, Dawes, Wegerif, & Sams, 2004). Many forms of collaboration in the classroom can foster these kinds of opportunities, for example, working in teams to develop arguments, providing feedback to each other using characteristics of a high quality argument, and having time to then apply that feedback to the arguments before a debate.

Video games do this by providing structured contexts for collaboration and sharing on multiplayer platforms. One example is the game Little Big Planet™ whose motto is “Play, Create, Share”. Players collaborate to control and move avatars through a world, solving puzzles along the way. They can also create and share their own game levels. As a community, players have created more than 8 million levels. They share videos of their levels and discuss play and creation of levels, learning from each other. Minecraft™ supports this with group and team competitions, and other collaborative experiences.

A fifth formative assessment practice is “Teachers provide feedback to move learning forward and create a structure for students to act on it”. There is a wide range of formative feedback that teachers can provide in the classroom, such as 1. Verbal feedback or written comments to individuals or groups of students to support their learning and 2. questions and follow-ups that a teacher can ask during a classroom discussion to help students elaborate and extend their thinking. Appropriate feedback given at a time when it can still influence learning has a powerful impact (Shute, 2008).

Video games do this through just-in-time feedback linked to rewards and next step actions. Making a choice (e.g., to jump or not) in a game often has an immediate consequence from which a player learns. Succeeding at a challenge can lead to a more complex challenge. By achieving a certain level of performance, a player is rewarded by “leveling up” and gains new game mechanics (i.e., “powers”) that they have to learn. In some cases, the next game activity may be a reward in itself. For instance, it is not uncommon after advancing through a difficult challenge for a video game to provide an easier challenge that was once hard for the player in the past but now easy, fostering a sense of mastery.

Overall, well-designed games incorporate parallels of the five principles of formative assessment. We hypothesize that video games that address academically relevant competencies have the potential to be as important a part of teaching and learning grounded in formative assessment principles as any other classroom learning activity. Using game design concepts, we are exploring formative assessments with content that is central to academic areas. We think that these games, within the context of a larger instructional unit, can support formative assessment practice and therefore, lead to deeper student learning. As noted in Table 1, connections may need to be made between the learning that can take place within the game and the larger goals of the instructional unit to help students transfer the ideas from one situation to another.

4. An Example: Mars Generation One: Argubot Academy (MGO)

MGO is an original game developed by GlassLab in collaboration with the National Aeronautics and Space Administration (NASA) and the National Writing Project to support learning of argumentation in the classroom. The game serves as formative assessment that is embedded within broader classroom experiences and curriculum.

As part of the development process, members of the design team observed classrooms in eight diverse middle schools in California and spoke with teachers about their approaches to teaching argumentation and the challenges they faced. Teachers said they and their students lacked a vocabulary for discussing and teaching argumentation. To address this concern, one purpose of the game is to make argumentation concrete to provide a common vocabulary so that teachers and students can more easily discuss it. The game provides visual representations of terminology such as argument, claims, and rebuttal.

We also took the perspective of argumentation as a dialog, drawing upon the work of Deanna Kuhn (Kuhn, Goh, Iordano, & Shaenfield, 2008; Kuhn & Udell, 2007) who provided input to the project. Additionally, we looked to Walton’s (2002) perspective that different kinds of arguments are represented by different “argumentation schemes”- knowledge structures that provide a template for each kind of argument. For example, an argument that relies upon a trusted authority for its support has the basic template of the claim, what authority supports the claim,
and why the authority is trusted, while an argument that relies on empirical evidence for support has a different template focusing on why the evidence supports the claim and under what conditions. Schemes also include critical questions that are commonly used to challenge the scheme and must be rebutted. Building upon these notions, one insight the game designer brought was that a dialog-based debate is much like a battle, and that the game mechanics of a battle game like Pokémon™ closely parallel mechanics of argumentation. In the game, students collect evidence for an argument, select a position, construct one or more argubots (robot-like creatures) representing their argument, and then battle with other argubots representing an alternative view (see Figure 1 below for example game screens).

Players learn to make argubots based on different argument schemes and construct more complex arguments combining multiple argubots. There are Authoritron, Observatron and Consebot argubots that correspond to arguing by authority, by observation, and by consequence. For example, an Authoritron argubot embodies Walton’s Appeal to Authority scheme (Walton, 2013). It is open to attacks using the critical questions associated with the scheme, and it can be outfitted with shields, representing rebuttals, to protect against those attacks. The game is set in the near future when the first colony on Mars has been established. The player takes the role of Ren, a visitor to the colony from earth and the first person not born on Mars to join the Mars school, Argubot Academy.

While there is a narrative, the primary play in the game involves taking positions on policy issues concerning the students in the academy and arguing their case, for example, that the food at the academy should be primarily plant-based protein vs. bugs or fish.

As shown in Figure 2, there are three basic sections of the game that form a loop: explore evidence, equip an argubot, and battle argubots. In the first section, students identify evidence for a particular argument by engaging in dialogue with other characters or uncovering evidence in the Mars environment. In the second section, they organize that evidence to take a position, construct an argument by equipping their bot(s), and in the third section they apply their argument, and evaluate and critique an opposing argument by battling their bots.

There are 33 issues (missions) in which players choose a position, construct an argument, and battle opposing positions. Players may take the same position multiple times (e.g., after collecting more evidence through further conversations with other characters or exploration of the environment and strengthening their argument) or they may take different positions. As a result, the game provides hundreds of unique argumentation experiences.

Figure 1. Images from MGO: Title screen (top-left), claim-evidence pairing (bottom-left), and argubot battle (right).
5. MGO and the Principles of Formative Assessment

The game was designed to address the five principles of formative assessment using approaches taken by game designers. In this section of the paper, we reprise each formative assessment principle from Table 1 and describe how it was designed into the game, and how the aspect of formative assessment was extended beyond the game by instructional designers to connect to the larger classroom unit.

FA Principle: Teachers identify and share learning expectations with the students (learning goals and progressions).

The first formative assessment principle translates to a game design strategy of creating explicit goals and challenges. In MGO, we incorporated this principle in two ways inside the game itself and also by connecting the game goals to the larger instructional unit. First, the designs of the video game and associated materials are based upon an argumentation learning progression (Deane & Song, 2014) that we deconstructed and simplified to match the structure of the game. There are three dimensions of competency that are addressed: 1. Identify components of an argument (e.g., selecting a claim and identifying evidence to support a claim), 2. Organize the components to construct a coherent argument that supports a position, and 3. Use the argument to debate with another position. In the game loop, these correspond to Explore the world to collect evidence for a position, Equip one or more argubots to construct the argument, and Battle with opposing argubots representing an alternative position.

Second, the game is organized into 33 missions including tutorial missions, missions involving simple arguments, and others which are more complicated involving multiple argumentation schemes. Each mission – and the progression of difficulty - is made explicit and shared with the player. A given mission may involve any combination of the game phases (Explore, Equip, Battle), and particular challenges are issued and expressed as the current mission/quest goal. For example, one mission goal in MGO is expressed as the need to decide the primary source of sustainable protein for the inhabitants of the Mars colony. This goal is stated explicitly by one of the other characters which helps the player know what is ultimately expected by the end of the mission. By explicitly sharing these expectations, it allows the game, teacher, and student to have a common context within which to gauge progress and determine the potential next steps which in this case are exploring the world for viable options (identify claims and data), equipping an argubot with supporting evidence (organizing claims with data within one or argument schemes), and battling against an opposing position to determine the most reasonable and evidence-based decision (evaluating claims and evidence).

Finally, teachers are provided with additional instructional resources to help them situate the gameplay within a larger context of argumentation. These resources highlight the argumentation learning progression which is foundational to the game and critical English Language Arts (ELA) standards. They also provide learning goals that teachers can use during the broader curriculum unit to echo those presented to students within the game and which provide additional support for students to transfer learning from within the game to other contexts involving argumentation.

FA Principle: Teachers elicit relevant and quality evidence of student learning on an ongoing basis to inform instruction.

The second formative assessment principle translates to the use of telemetry and analytics within the game to guide and inform the player’s experience and to provide information to the teacher to guide next steps beyond the game play itself. Game mechanics are the ways in which players can interact with a game. To help ensure that game mechanics provided evidence to inform instruction, we followed an Evidence-Centered
Game Design Process (ECgD) that involves creating a set of models to represent the linking of what is to be measured or learned to the evidence of learning and the game mechanics that can provide this evidence (Mislevy et al., 2014). ECgD enabled us to explicitly create game mechanics to elicit evidence of student learning. We created an evidence model provisionally identifying what interactions in each mission (game telemetry) provided evidence (game analytics) of each level of competency. The idea was to structure the game situations and affordances so that the actions students took would simultaneously advance game play and provide bits of evidence about their understanding of the structures and strategies of argumentation represented in the learning progression. We also linked evidence to more granular skill milestones that could provide a foundation for actionable feedback. We created a probability-based model to aggregate evidence from different types of actions and action sequences to make inferences about students’ understanding of the aspects of building and using arguments. The competency levels in the strands of the argumentation learning progression defined variables in a psychometric model to link these inferences back to instructional goals.

This was an iterative process in which we hypothesized different kinds of evidence for each competency level, designed the game to include these kinds of evidence, and collected increasing amounts of data to evaluate the hypotheses and determine the measurement value of each of the different kinds of evidence in the context of several studies including a mini-tryout, alpha and beta testing, and a medium-sized field test in classrooms. Table 2 provides some examples of hypothesized evidence of competency from student play from the early design and development phase of the work. Figure 3 illustrates the synergy between the game design and

Table 2. A portion of the evidence model for MGO

<table>
<thead>
<tr>
<th>Competency</th>
<th>Level</th>
<th>Description of Skill Milestone</th>
<th>Part of Game</th>
<th>Kind of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify</td>
<td>1</td>
<td>Identify one piece of data that relates to specific argument</td>
<td>Explore Equip</td>
<td>Strength of constructed core claim/data pair for argubots</td>
</tr>
<tr>
<td>Use</td>
<td>3</td>
<td>Use one critical question (or counter claim data) against opposing argument</td>
<td>Battle</td>
<td>Appropriateness of player’s attack(s);</td>
</tr>
<tr>
<td>Organize</td>
<td>2</td>
<td>Organize argument with multiple pieces of data within multiple argumentation schemes</td>
<td>Equip, Battle</td>
<td>Number and type of different argubot models submitted for battle</td>
</tr>
</tbody>
</table>

Figure 3. Mapping of game actions to evidence elements.
assessment design aspects of MGO. The dark rectangles at the bottom of the figure represent phases of activity in the game and specific actions players can take. Combinations and sequences of these actions constitute observable variables input to update beliefs about the variables in the hierarchical psychometric model at the top of the figure. Teachers receive feedback on a daily basis as students work through the missions. Feedback comes in the form of an automatically generated “watch out” list of students who are skipping important aspects of the play (e.g., not collecting any data for claims during the explore phase which would prevent them from making any progress) and a “shout out” list of students who have made significant achievements in the game (e.g., making very rapid progress early on or being especially systematic). The next day, the teacher can use this on-the-fly (Shavelson et al., 2008) information to guide discussions or provide other learning activities to develop students’ understanding of argumentation outside of the game.

While the “watch out” and “shout out” lists are generated from data on-the-fly during the game, there is also an underlying statistical measurement model (in this case, a Bayes net (Almond, Mislevy, Steinberg, Yan, and Williamson, 2015)) that makes inferences from students’ overall patterns of game play over the entirety of play to their probabilities of being at different levels of ability of the argumentation learning progression. This more formal information (more aligned to the Shavelson et al., 2008 embedded assessment) is useful for teachers for longer term planning for the class and also for reflecting on the instructional unit in terms of planning revisions or adjustments for future use.

**FA Principle: Teachers structure opportunities for students to take ownership of their own learning.**

The third formative assessment principle translates in game design to developing ways in which players take ownership of learning. Providing a sense of agency is a critical component of game design. MGO was designed for middle school students. The game designer explored the question “what will help middle school students develop agency and take ownership?” by interviewing middle school students and identifying important commonalities across their responses. She reported that while middle school students said they are beginning to realize the world is larger and more complex than they thought when they were younger, they also felt that embracing that complexity makes them feel powerful – “they hunger for relevancy; for their decisions to matter” (Hoffman-John, 2014).

This theme played out in many aspects of the design: the genre of the game, students’ role in the game, and the openness of the game. The game designer created a role playing game that connects to students’ interests and provides many options for students in terms of how they navigate the environment, who they talk to, and what arguments they choose to employ. Students are placed in the role of a newcomer to Mars, meeting new people, and engaging in policy debates that are connected to life decisions of the students on Mars. Finally, the game was developed to provide a high degree of openness in the play to allow students to take different policy positions and replay different segments.

The surrounding classroom materials that accompany MGO extend opportunities for students to take ownership of their own learning by providing structure to help students reflect explicitly on what they have learned about argumentation from the game play to ensure that the game experience is transferred into broader learning outside of the game, and to extend their skills with argumentation beyond the contexts supported in the game.

**FA Principle: Teachers structure opportunities to activate students as instructional resources for one another.**

The fourth formative assessment principle can be enacted through the use of within-game collaborations and multiplayer platforms, but for MGO this principle is the least directly integrated into the game-playing experience. However, we explored having pairs of students play the game together using a single tablet in early tryouts. We found that it created an engaging experience, leading to a larger activity designed for peer collaboration. Figure 4 provides a photo of two boys playing an early version of the game together.

The primary game designer on the project took part in these sessions and afterwards made the following observations:

*As a game designer, I care about one thing — the look in my player’s eyes…I want to see them leaning over, fighting over the iPad, unable to keep their hands to themselves. This is what engagement looks like [see photo]. The two boys also originated my favorite phrase from this playtest, “wait. wait. wait. we want to hear this!” One boy was advancing quickly through the character dialogue, and the second boy had caught on that evidence was going to be present in that*
dialogue and if they rushed through it too fast without reading they were going to get nailed when they went into battle. Teacher lingo for this is “close reading.” (Hoffman-John, 2014)

While anecdotal, this playtest provides a positive case of two players supporting their own learning through the complementary strategies the boys were taking in playing the game. In this case, the boy rushing through was activating the other boy as an instructional resource to shift to a recognized strategy that would lead to better game play and greater learning.

Figure 4. A pair of students playing MGO together.

Since the game was designed to be used in a classroom as part of an instructional unit on argumentation, our primary approach to addressing the formative assessment principle of peers acting as instructional resources for each other involved activities outside of the game. In this case, colleagues at Glasslab took this idea of students working in pairs with argument ‘battles’, and extended it beyond the game – with the intention of enabling students to transfer their in-game learning to more traditional debate settings.

The classroom activities follow routines originally designed by Deanna Kuhn and colleagues (Kuhn & Udell, 2007). After a teacher-led introduction to argumentation and approximately 30 minutes of play within the Mars Generation One game, students work in pairs outside the game to sort and evaluate pieces of evidence for one of the game missions. Pairs select their strongest evidence and choose a claim to support. In a whole class argument duel, members of each pair find others with opposing views and exchange their arguments. In the argument duels, opponents evaluate each other’s arguments with respect to the relevance and support of their evidence. Later cycles of the same activity require students to pose and defend against a series of critical questions by requesting and offering backing for their arguments. The process plays out as follows:

Step 1: Students work in pairs to associate evidence with claims, evaluate evidence, and build an argument.
Step 2: Students present arguments to other pairs and gather feedback.
Step 3: Students revise their approach based on feedback and present the argument again.

Other classroom structures help students reflect with their peers on what they have learned about argumentation from the game play to ensure that the game experience is translated into broader learning outside of the game.

FA Principle: Teachers provide feedback to move learning forward and create a structure for students to act on it.

The final formative assessment principle was enacted through the game design strategy of using just-in-time feedback linked to rewards within the game and supports for teachers to provide feedback to students outside the game. There are several kinds of in-game feedback provided to players. MGO provides immediate in-game feedback to a student. When a student clicks on objects, the game reports back which objects are actually pieces of data and then describes the data. In addition, when the student connects data to a claim by “fusing a core”, the game reports back the strength of the claim-data pair, and the student has the option of connecting different data to the claim to create a stronger claim-data pair. Also, when a student has constructed an argument in the equip phase of the game and has moved to the battle phase, the game provides feedback on data and claim relations that are “not related” or “irrelevant” when the students’ argubots are attacked by the game AI’s argubots. Finally, students cannot advance to harder quests until they successfully complete prior quests and are given feedback when they do and do not succeed in completing each quest.

In addition, the 33 quests are also structured to support acting on feedback. The game is organized in two major levels. The first level includes single claim arguments with no rebuttals. Students get feedback after each quest (e.g., they either win or lose a battle and must
go back and try again if they lose). Once they have won each of the battles in the first level, they advance to the second level where the arguments are more complex and include critical questions (attacking the backing of an argument) and rebuttals (supporting the backing of an argument). The arguments get more complex and can include multiple pieces of data for a single argument scheme as well as multiple argument schemes which are represented by the different types of argubot (e.g., Observatron for the “arguing by observation” scheme, Consebot for the “arguing by consequence” scheme). In addition to immediate feedback within each game loop, as well as at the end of a quest, feedback is provided at the scheme level across quests. When a student has won a specific number of battles using a particular kind of argubot (i.e., argument scheme), the argubot “evolves” to a new version of itself that includes additional components of the scheme. In this way, students can act on short cycles of feedback (Herman, 2013).

Outside the game, teachers are encouraged to use both the “watch outs” and “shout out” lists to provide feedback to students to support their deeper understanding of argumentation, and to help them connect what they are learning within the game to the broader principles of argumentation. Additional research is needed to understand whether and how students apply and internalize the two sources of feedback to deepen their own understanding of argumentation.

6. Summary, Discussion and Conclusions

In this paper, we presented a case for the value of carefully designed video games that can be part of classroom-based formative assessment practice. We first described five core principles of formative assessment and their analogues in features of commercial video games. We then described a video game designed to be used in ELA classrooms to help students learn aspects of argumentation. We also discussed and demonstrated how each of the principles played out in the design of the video game and how the game was incorporated into the larger instructional unit. This work only constitutes a proof of concept. While we demonstrated that it is possible to embody the principles of effective formative assessment in an educational video game, we could make a stronger case if we had empirical evidence that the formative assessment is carried out effectively using the game as part of an instructional unit in classrooms.

Some preliminary evidence for the effective use of formative assessment in educational video games is presented in Bertling, Jackson, Oranje, and Owen (2015). Here, we do not assert that any video game can be used for formative assessment. To the contrary, “MGO” was carefully designed and developed through a collaboration among video game designers, graphic artists, cognitive scientists, psychometricians, computer scientists, educators, and students following an iterative process, Evidence-Centered Game Design process (ECgD) in order to target critical aspects of developing and critiquing arguments. Additionally, we have not shown that MGO is necessarily preferable to other approaches to formative assessment in teaching argumentation. Making that case would require studies comparing the efficacy of MGO with these other approaches. Rather, our position is that game-based formative assessment deserves to be part of teachers’ palettes of assessment approaches. Additional research is needed to address what features lead to the greatest improvements in learning (cf. Delacruz, 2011), what aspects of validity need to be explored and addressed, and how to reduce the cost of design and development of game-based formative assessment to make it more viable for different educational subjects and settings.

7. Acknowledgements

The authors would like to thank the Bill and Melinda Gates Foundation and the John D. and Catherine T. MacArthur Foundation for funding this work. We also wish to thanks our co-workers who contributed to the development and piloting of Mars Generation One including at Glasslab: Rose Abernathy, Greg Bliss, Brenden Conte, Ben Dapkiewicz, Mat Frenz, Jerry Fu, Liz Kline, Michael Louie, Dren McDonald, Uma Menon, Elizabeth Owen, at ETS: Masha Bertling, Katherine Castellano, Chris Kitchen, Andreas Oranje, and Yi Song, at Pearson: Kristen DiCerbo, Valley View Middle School: Matt Farber, Dunlap School District: Amy Stefanski, and the students, teachers, and Administrators of Central Middle School in San Carlos, CA., Lazear Middle School in Oakland, CA., and the Institute of Play in New York, NY.
8. Conflict of Interest Disclosure

This work described was sponsored by grants from the Bill and Melinda Gates Foundation and the John D. and Catherine T. MacArthur Foundation to GlassLab to which the Educational Testing Service was a subcontractor. Further work and analyses, including the writing of this paper, was sponsored by the Educational Testing Service, Princeton, NJ under which Malcolm Bauer is Principal Investigator of the Extended Games project. I have disclosed those interests fully to the Journal of Applied Testing and Technology.

9. References


