S'cape from Formality: Embedded and Automatic Assessments within Simulation Games

Jeffrey A. Olsen
Utah State University
Logan, UT
Jeffrey.A.Olsen@aggiemail.usu.edu

Brett E. Shelton
Utah State University
Logan, UT
Brett.Shelton@usu.edu

Todd Campbell
University of Massachusetts
Dartmouth, MA
todd.campbell@umassd.edu

ABSTRACT

Simulations provide an environment to experiment safely, openly, and repeatedly for learning mastery. However, many simulation environments experienced within a classroom fail to include the assessment components meaningful for instructor interpretations in a way that translate to standardized "scores". Even when a measure of standard assessments is included, often it fails to account for the unpredictable nature of decision-making within the complex, 3D, open-ended simulation environment. Embedding assessments within a virtual simulation environment poses several issues. First, the program must provide assessments that will fulfill educational requirements that will not take the learner cognitively "away" from their activities. Second, the program must provide an engaging game-like experience for educational purposes. Third, it must provide assessments that maximize the unique capability inherent within digital deliveries, that allows for geographically disparate and asynchronous schedules between instructor and learner. This study addresses each of the above concerns through an integration of the classroom requirements and simulation affordances. Through the inclusion of a "replay" function for self-regulated after-action review, students answer questions about their understandings, but also thoughtfully reflect on their process and the applications for those understandings in novel scenarios. Created within an educational curriculum, S'cape is designed to function as a stand-alone module to teach and evaluate understandings about core concepts. This platform was piloted as a first person explorer game addressing various levels of complexity about chemical and physical properties of substances. Refined within gaming and technology best practices, this novel architecture combines educational and gaming principles. Engaging learners through the use of the automated assessment features (i.e., automated embedded assessments as after action review (AAR) and a 'replay' function for metacognitive support), in this case, holds promise for military and corporate scenarios to aid in the valid systemic needs of complex, open-ended assessments. The paper describes how automated embedded assessments may provide a means for safe experimentation while supporting metacognitive practices crucial for 3D training environments.

ABOUT THE AUTHORS

Jeffrey A. Olsen has earned a BA in history with a minor in English from Brigham Young University, an MA in secondary Social Studies Education from Pacific Lutheran University, an MS in Instructional Design and Technology from Western Illinois University, and is currently a PhD student in Instructional Technology and Learning Sciences at Utah State University. He has taught in public schools, universities, and the corporate world in a variety of capacities. In each of these he has worked to design curriculum, create assessments for local and statewide usage, and focused on skill and concept mastery assessment. Currently he is working on a National Science Foundation grant as lead qualitative researcher engaged in studying science teacher orientations and beliefs and the effective integration of technology and inquiry.

Brett E. Shelton is an associate professor in the Department of Instructional Technology and Learning Sciences at Utah State University. He holds a Ph.D. in Educational Technology from the University of Washington, an M.T. in Industrial Management and Supervision from Arizona State University and a B.S. in Computer Engineering from the University of Idaho. Dr. Shelton uses a variety of mixed-method research approaches to study vision, perception, cognition, and the design and assessment of innovative technologies for learning. He is the director of IDIAS—Interactive Design for Instructional Applications and Simulations—whose mission is to build on university strengths in instructional design, interactive simulations, and interface design to inform technology research and develop commercially viable and innovative products. IDIAS research activities center on the design and development of effective teaching and learning strategies—grounded in cognitive studies—using visualization technologies.

Todd Campbell is an associate professor at the University of Massachusetts Dartmouth. His research focuses on factors influencing current reform in science education. This is supported by investigating science teacher professional development, scientific inquiry/modeling instructional practices and science/technology integration. Dr. Campbell is the principal investigator for a National Science Foundation project focused on integrating technology into science instruction and past investigator for a state-level Mathematics Science Partnership Professional development project partnering science teachers and scientists in curriculum development. He has published in the International Journal of Science Education, the Journal of Science Teacher Education, and the Journal of Science Education and Technology, among others.

S'cape from Formality: Embedded and Automatic Assessments within Simulation Games.

Jeffrey A. Olsen
Utah State University
Logan, UT
Jeffrey.A.Olsen@aggiemail.usu.edu

Brett E. Shelton
Utah State University
Logan, UT
Brett.Shelton@usu.edu

Todd Campbell
University of Massachusetts
Dartmouth, MA
todd.campbell@umassd.edu

INTRODUCTION

The creation of interactive simulation systems for education has been increasing in complexity in traditional education, military, and corporate contexts (e.g., Barab, Gresalfi, & Arici, 2009; Freeman, Salter & Hoch, 2004; Kirkley & Kirkley, 2005; Shute, 1993) and the value of these digital games for learning content, gaining skills, creating artifacts, and developing systems thinking for the 21st century is apparent (Thai, Lowenstein, Ching, & Rejeski, 2009). However, assessing student-learning outcomes in open-ended virtual 3D spaces has been of keen interest for instructional technologists and learning scientists especially of late (Ketelhut, Nelson, Clarke, & Dede, 2010; Nelson, Erlanderson, & Denham, 2010). These researchers acknowledge the difficulty of creating embedded assessments, and issues related to functionally applying current assessment mechanisms into a virtual 3D engine.

Within the complex environment of creating simulation games within schools, as well as effective embedded assessments, this research seeks to provide a space for safe experimentation by students while supporting metacognition and refection practices and assisting classroom needs. Utilizing after-action review (AAR) as a method of reflection, together with a replay feature during assessment, students will be assisted in answering the questions and performing meaningful reflection on their learning. After providing some groundwork in the literature and some background on this project, methods are reported, findings are detailed, and preliminary conclusions around this pilot data are provided.

Background

While games and simulations provide students opportunities and methods to potentially improve learning, determining the effectiveness of this medium is contingent on how the learning is assessed within these contexts. Gee (2008) stated that "virtual experiences centered on problem solving, recruit learning and mastery as a form of pleasure" (p. 36). This learning as a form of pleasure creates the need for assessments to seamlessly blend into the structure of game play to avoid altering the learning environment. Designers should consider the purpose behind the design of an instructional environment and select the pedagogical learning theory that connects with that purpose (Ketelhut, et al., 2010).

It stands to reason if games and simulations can provide new methods of learning, they should also provide new methods of assessment as well. Nelson, et al. (2010) are working on three areas of embedded assessments; content understanding, process understanding, and contextual understanding as conceptualized by Mislevy, et al. (2003) as a means of parsing out the complexity of an environment. Current efforts in automated assessment generally fall within one or more of these general constructs. Assessments that offer high adaptability are difficult to produce, but one model toward effective embedded assessments is achieved in science by combining strategies, such as data and task tracking with pedagogical assessment models. Based upon this idea, the National Science Foundation has supported the building of the *S'cape* game designed to teach middle-school science standards while capitalizing on key instructional design features provided by unique instructional design components layered on a game engine. These components allow for automated assessments and player review capabilities of their own activity, thereby fostering key reflective and metacognitive activities, while intertwined within the game narrative. In *S'cape*, players navigate through micro- and macro experiments in an alien landscape, attempting to get away from their alien captors while discovering new elements of the story as they reveal themselves.

Metacognition refers to having knowledge of your own knowledge and regulating that, or put another way, thinking about your thinking and knowing how you regulate that thinking (Hacker, 1998). Research has indicated that students require an increased awareness and means to promote practice of their metacognitive skills (Schraw, 1998; Young & Fry, 2008). While military and academic researchers have implemented and studied tools such as simulations to support metacognitive and reflective practices (e.g. Azevedo, 2005: Salmons, 2008; van Lent et al., 2004), there is still much to be learned about student observational skills within the simulation environments and the way in which effective embedded assessments can monitor and provide effective feedback systems for teachers and students. Some of this understanding may come from engaging in the subcomponents of metacognition; *planning*, *monitoring* and *evaluating* (e.g. Schraw, 1998; Young & Fry, 2008). *Planning* involves planning and selecting the strategies and cognitive resources necessary for a cognitive task. *Monitoring* involves awareness of progress and the ability to decipher our performance through a cognitive task. *Evaluating* includes an examination of outcomes and deciphering whether the learning outcomes and goals matched, as well as evaluating the regulation processes for effectiveness (Schraw and Moshman, 1995). While procedures like after-action review (AAR) allow learners to reflect on their actions, and then allow for a return to the environment to try again from the beginning, this system allows for *individual, independent* visual reflective review at the point of assessment, prior to reevaluation and repeated engagement.

Purpose of the this study

One proven method that helps facilitate metacognition and reflection when utilizing simulations for learning is AAR, and is used extensively in the military in both live and simulated environments (e.g. Salmons, 2008; van Lent et al., 2004). AAR engages a learner in a process to identify mistakes and potential solutions after the simulation has run, and then students may reenter the simulation and attempt to redo or practice actions from the beginning of the simulation.

In consideration of traditional AAR, this study used a first person navigable simulation environment to engage students in chemistry experimentation within a gaming construct. Within this environment, the simulation recorded all the player actions into a playback system that can be paused, played, fast forwarded, and rewound in the same manner as a Digital Video Recorder (DVR) system. These recorded review system features were available to the player within the embedded assessment "interrogation room" to assist the reflective practice on past actions and assist in guiding the instructional process, while still allowing the player to remain immersed in the presence of the game environment. The "interrogation room" is the place where the aliens send the player between successful level escapes to ask them questions to better understand how chemical and physical processes work on planet Earth. When players are placed in the interrogation room they encounter a simple user interface screen (see Figure 1 below) that includes some instructions, questions requiring responses, and a window in the upper right where students could replay their entire scenario using classic scrubber/DVD controls. While players are required to answer questions about core concepts utilized in the room for their escape, they have the ability to review all their actions during their escape to assist in answering the questions. This is the recording "through their own eyes" as they experienced the environment to allow for deeper reflection and thoughtful responses to the questions based on those experiences. Rather than merely being assessed when they escaped the experimentation room, this process allowing for review and reflection is designed to allow for players to focus their attention and draw metacognitive conclusions.

The embedded nature of this form of assessment is directed to the performance of the players. Squire

Figure 1: Initial design of interrogation room interface.

(2003) stated “advances in assessment, such as peer-based assessment or performance-based assessment provide learners multiple sources of feedback based on their performance in authentic contexts” (p. 4). The value of the automated assessment is the instant feedback the player receives during the simulation, and in the ability to meaningfully reflect on the activity. The player is assisted by the opportunity to review their actions in the interrogation room and answer questions to expand their thinking.

As part of a National Science Foundation Discovery Research K-12 Grant (NSF 09-602), this project is focused on addressing DRK-12 Program Challenges 2, 3, and 4, designed to enhance STEM education, even while the actions and topics have relevance for any complicated actions within a 3D environment. Eighth grade science teachers from across the Western U.S. were brought together in a focus group environment to discuss desired aspects to be covered by a simulation environment, as well as the type, style, and functionality of the embedded assessment to meet their educational needs. The creative team used teacher suggestions in designing both the environment of the game play and the embedded assessments. This environment prototype was then presented to two classes of high school students in a rural charter high school for pilot testing of the environment and the assessment feature. This research design engaged elements of observational study (Creswell, 2008) and grounded theory (Corbin & Strauss, 1990). Data was collected through observation, data mining, focus group feedback, and a metacognitive awareness inventory (MAI) survey instrument attached to the gaming environment. The teachers themselves, in connection with the university, were considered the subject matter experts (SMEs) and advisors in creating the assessments and ensuring adequate levels of complexity in the nature of the simulated environments. The use of multiple SMEs ensured both accuracy and necessary complexity within the simulation. Gathering data from teachers across the intermountain west helped to ensure pedagogical applicability and impactful embedded assessment.

The purpose of this research was to better understand how the use of this technology supported metacognition and reflection, and specifically, how reflective processes (individual AAR) embedded in a gaming environment could affect learning with specific regard to metacognition and reflective practice. While there are 4 levels to the complete *S'cape* game, for piloting researchers engaged students only in the simulation environment level 1, embedded assessments to level 1 with reflective replay, repetition of the first two steps if they fail to meet the seventy percent mastery level in assessment, and MAI survey after their experience.

Research Questions

The following questions guided this research:

1. What kinds of metacognitive activities are supported by 3D virtual instructional intervention, and to what extent are these activities correlated with positive educational outcomes?
2. To what extent are metacognitive activities, associated with the review system specifically, correlated with positive educational outcomes?

METHODS

Participants (Sample)

Thirty-nine ninth grade science students in a local rural charter high school were engaged for this pilot study of the gaming simulation environment. While this simulation gaming environment was designed within the NSF grant, these students are not part of that study population, but are merely providing pilot data to further develop and ensure effective general implementation for the research. Data collection took place in two sessions on the same day with twenty-three in the first session and sixteen in the second. The first group completed all necessary components in one hour and fifteen minutes while the second group only took 30 minutes to complete all tasks. Two students that completed the game did not complete the MAI and thus the total number of those utilized in the study reduced to thirty-seven in the sample population.

Procedures

The *S'cape* game was played through a web browser using the Unity platform; all play and collected data were made available and completed using standard Internet practices and protocols. Because no information or programs were required on local machines, the game and assessment can be completed on any computer with Internet connectivity. Each level play within the environment lasted anywhere from five to ten minutes. Few required more time based on their continued return into the environment to complete their understanding and fulfill the assessment at an acceptable level of competency, though their individual level play each time lasted roughly the same amount of time each attempt. The procedures were as follows:

1. Players in the simulation scenario are given the back-story of how they awakened in the room and then each individually engages in the

simulation alone. They are aware that their goal is to determine how to escape the room by utilizing what is available to them in the room.

2. Players enter the environment with this knowledge and complete level 1.
3. When they complete the necessary steps and escape into the next level of the environment, they must first pass through the interrogation room. They must answer the five questions presented per level, but also have the reflective replay feature available to them to organize their thoughts and focus their reflection. Should they complete the assessment with an acceptable level of performance they can move on to the next level (though not in this pilot scenario), but if they do not meet that threshold they are returned to the simulation environment again to repeat actions and fix any missed pieces to their understanding and review until they effectively meet the required level of understanding. In other words, if the student should fail the assessment portion, they must repeat level one again, being sent back into the experiment room to further experiment and explore, and finally redoing the assessment as a repeating cycle until they reach an eighty percent competency.
4. The student completes the MAI survey utilizing an online survey system.

Instruments

Pilot level 1 Game Functionality

The premise of the game is built from the idea that aliens who are studying humans for their problem solving skills, and their understanding of physical and chemical properties, have kidnapped the player. After studying the human race for a short period of time, they have decided that clowns and the circus are two things generally enjoyed by humans, and have thus attempted to theme the environment to attempt to ease the captive and make "it" more comfortable. The player is informed that if they can successfully navigate both the necessary mechanisms to escape the room, and correctly respond to a series of questions related to what they have accomplished, they can proceed to the next experiment. So, what the player experiences is a mousetrap feeling while surrounded by grotesque carnival imagery, but the play itself involves reified molecular phenomena aligned to what they would normally experience within a laboratory setting when learning about chemical and physical properties of substances.

As an example, beginning in the first (most basic) level of the game, the player wakes up in a simple room filled with various elements from which they can work. This first room is designed to teach students the physical properties of water. The student player must determine the methods to create ice, standard water, and steam (gas) from provided elements in the room that will allow them to move to the next level. A screen capture of the environment is provided in Figure 2.

Students wake up in this room and can explore as a first-person shooter (FPS) style game. Within the environment are various hotspots for student

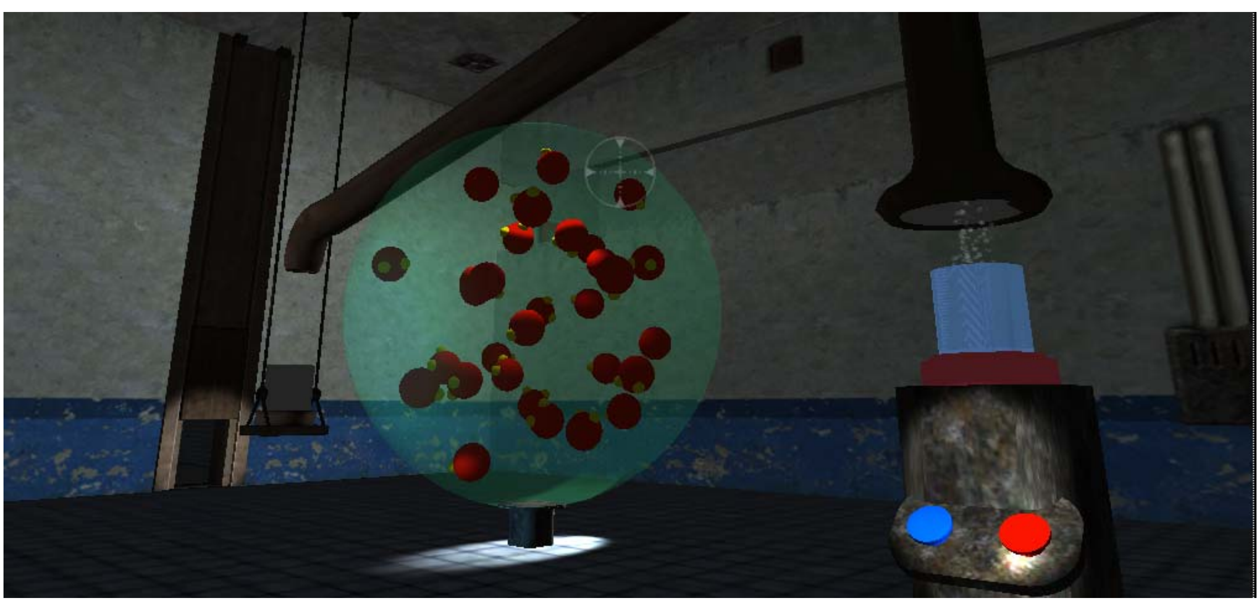

Figure 2: Early level one imagery with molecules shown in motion.

observation and interaction. Should the students click on the red button, they will observe the following two changes: First, the blue basin of water has begun to boil and there are particles rising into the hood. Second, students will be able to observe the droplets (condensed water vapor) falling into the bucket, and that this process is slowly raising the door in the wall. This proves crucial to the student's progression in the game. To help students understand what is occurring at a molecular level, students can click on the hood, the area of where the activity is collecting, and the molecular view appears in the large central orb; shown above with molecules in motion. On a similar score, students may also click the green button causing the steam to stop, and the bucket to stop filling and to turn clear. When students click on the bucket, again the area of molecular activity, they again activate the molecular view in the orb this time seeing a slowly-moving structure of ice. Important to observe within the globe is the density of the molecules between physical states: ice is less dense than water by spatial relationships between molecules, steam is also less dense than water but the energy of the molecules is much different.

The reason students view these molecular alignments relates to how students will be required to apply the information. When students travel through the door into the interrogation chamber they will be required to answer a series of questions related to the molecular nature of the water, demonstrating a level of understanding of the physical and chemical concepts before they can move to level two. Note that one major innovation of S'cape is that during the interrogation/assessment, the student views/replays exactly what they did within the experiment room while they were there, maximizing the opportunity for reflective and metacognitive activity. The player controls the playback of their actions in DVR-like fashion: rewinding, pausing and playing as needed. They may continue to explore and work on this level as long as it takes for them to understand the concepts and develop the necessary understanding applied in the interrogation room described below. The dialogue with their captors during the interrogation is designed to provide hints as to what motivates the alien captors, and provides a vehicle for the overarching storyline.

Pilot Assessment System

As mentioned, within the interrogation room players virtually "discuss" the science underlying the mechanisms that assisted their escape. For example, the first interrogation question asks, "Engage the explanation that describes what happens when you push the red button." Possible responses include: a. the liquid becomes a solid because the molecules are moving more slowly; b. the liquid releases gas as the molecules move quickly and have energy to escape; c. the liquid becomes plasma as the molecules have enough energy to glow; d. the liquid remains unchanged as a liquid. Depending on the selection, responses include: a. Not quite--observe the globe more closely (worth 1 point); b. Observant Human! Consider cooking with grease or water with the loss that boiling brings with it (worth 2 Points); c. Looking closely, is there any observable light emission? (no points); d. Not an observant human. Observe the contents of the globe more closely (worth no points). These questions are designed with a scale for more expert-like understandings to novice-like understandings. A technically accurate response is worth two points, an answer which exposes a level of understanding that may not be entirely accurate but is worth 1 point, and two questions that are scientifically inaccurate and therefore not worth any points (complete level one assessment questions available upon request). This structure, which includes levels of complexity within student answers, also provides a graduated level of assessment to their understanding for the supporting scientific concepts. If they cannot reach a certain threshold for points based on their answers selected, they are sent back into the room to re-investigate what they may have "missed" in the first escape attempt. This new attempt would overwrite their old level and re-records only the latest scenario. The required return to the previous room to re-investigate is designed to eliminate basic attempts at "guessing" at the correct answers, as well as provide a means to further experiment within the room to test hypothesis and gather additional data.

Data Collection

Student performance data is being collected from the simulation environment through an automated program built into the simulation itself. The data is collected as the individual logs-in and is identified by that username throughout their interaction. This data includes: how many times they engaged in the assessment, their responses as well as the correct responses for each time through the assessment, and timestamps for each attempt. This data will aid in understanding if the embedded assessment and the gaming environment are at an appropriate level of difficulty. The data will also provide insight as to if the review features of the interrogation room assisted students in answering the assessment questions. The assessment data is populated real-time into a database and can be viewed by instructors while students are within the 3D environment. The data can be stored for later use and analysis as well. See Figure 3 for a screen-capture of the real-time data collection system accessible through a protected website.

The Metacognitive Awareness Inventory (MAI), originally designed for adult populations, was designed as a method to assess metacognitive awareness. The MAI is composed of 52 questions, originally broken into two subcategories with 17 questions related to their knowledge of cognition, and the remaining 35 questions related to their regulation of cognition. While all questions remain the same on the MAI instrument, the subcategories were broken differently into the three major subcomponents identified in metacognition; *planning*, *monitoring* and *evaluating*. Further, this same instrument was also broken into two alternate categorizations; metacognitive practices directly related to those utilized through the review system in the interrogation room, and other self regulation skills generally acquired. These scores were compared to their student outcome scores in the embedded assessment system and used in answering research question two.

Analysis

To aid in assessing the effects of reflection on metacognitive awareness, MAI scores were gathered after engaging the simulation environment. This five-point Likert scale assessment survey has been found valid (Young & Fry, 2008). Three questions were not included in this analysis, for a new total of 49 questions on the MAI, with the details of their division breakdown below. These divisions operationalized what students thought going into the scenario, during the

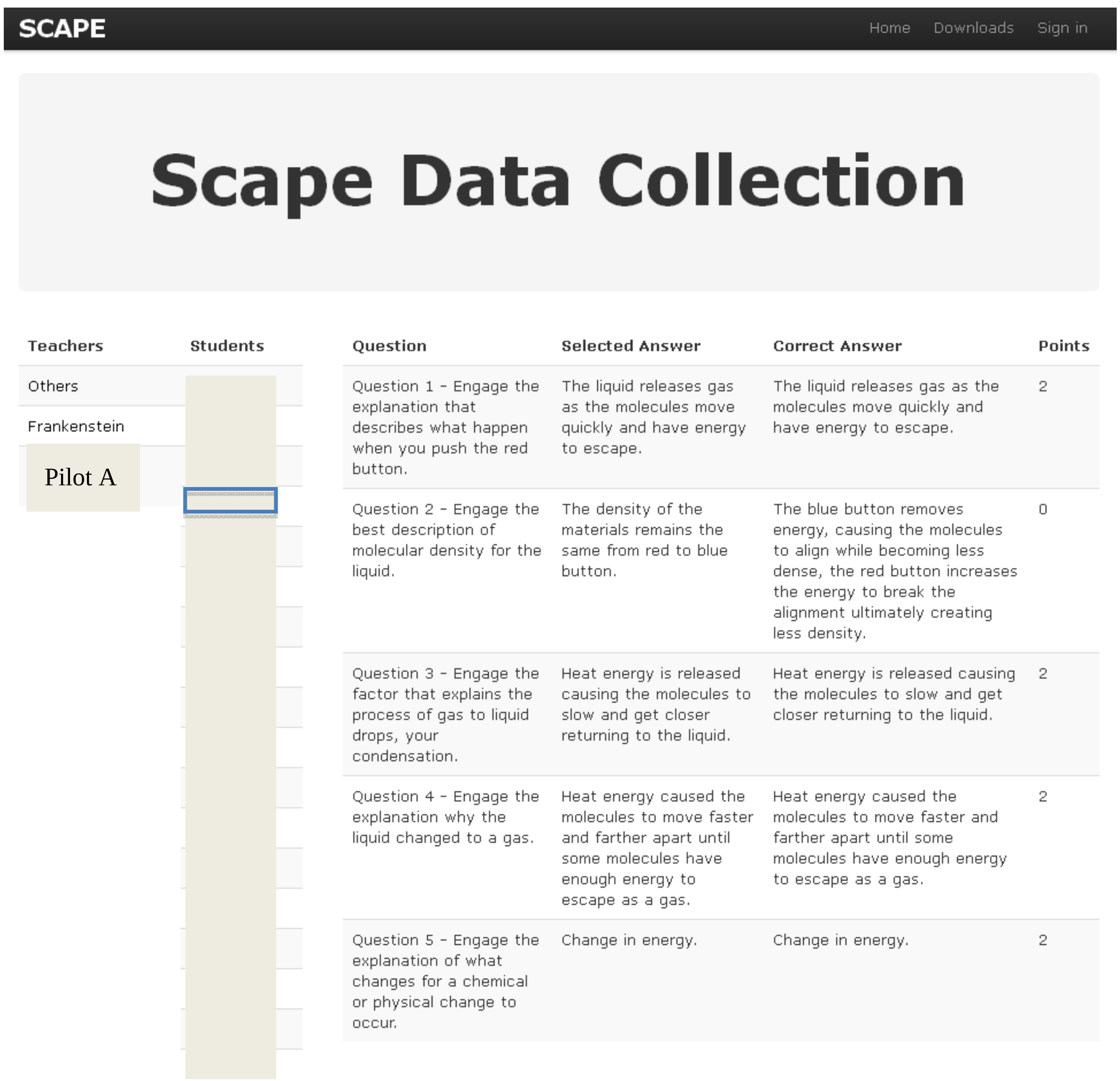

Figure 3: Real time data collection to a secure website of teacher and student organized by teacher, student (pseudonym), school, question, answer, and results. (Student names blocked for confidentiality)

scenario, and once they completed the scenario as they reflected on their activity. Further, it allowed for the division of those self-regulatory skills the students possess as well as those specifically utilized by the build-in independent AAR system within the interrogation room.

Pearson correlation tests are performed showing means and standard deviations for each of the three divisions of MAI questions, sub-grouped into metacognition categories of planning (14 questions), monitoring (17 questions), and evaluating (18 questions) as compared with performance scores. Highly correlated subgroups indicate strong relationships between aspects of metacogitive activity as associated with a strong performance in the educational intervention.

Pearson correlation tests are also run for MAI sub-groups of scores that dealt with the embedded review system (25 questions) or not (24 questions) as compared with performance scores. Similarly, highly correlated metacognitive activity as used with the review system, with that of performance, should indicate if the review system performed as expected. That is, the review system assisted in producing metacognitive awareness during the question/answer portion of the instruction, useful for AAR types of activities.

Performance measures are indicated by the number of tries it takes a student to complete the interrogation with a score of 80% or better. That is, the fewer the number of tries, the better the performance. Therefore, higher MAI scores associated with fewer tries indicates an inverse relationship.

FINDINGS

As seen in Table 1, strong correlations exist between categories of MAI monitoring (r (35) = -.388, p < .05) and MAI evaluating (r (35) = -.474, p < .05) with student performance (statistical significance achieved at r (35) < -.325). These values indicate that students who had high scores on the MAI performed similarly well on the performance measure, that being, completing the interrogation room questions with a score of 8 out of 10, or better. Similarly, those who had poor MAI scores did not receive adequate performance scores, on average.

These correlations indicate that aspects of the *S'cape* game featured elements that assisted in metacognitive strategies of planning and evaluating, but not monitoring. This result could be due to the absence of surrounding instruction to situate this pilot study simulation within or simply the way in which this pilot was administered. However, the nature of the game itself is not designed to support and promote monitoring as a cognitive task. As monitoring involves an awareness of progress and those processes leading to engage our performance, it does make some sense that a game in which the progress toward escape is one of discovery and necessarily unknown to the player at the outset of the game play that this knowledge of progress would be difficult to know and assess. However, as the player has the ability to select strategies necessary to accomplish the task and examine those outcomes to evaluate how well they met the goal of the room, the elements of planning and evaluation would show positive correlations. That is, the nature of the game environment is designed to favor the two positively correlated items of planning and evaluating, while showing a negative correlation for the monitoring aspect of metacognition for which the game structurally is not designed to support. The results within Table 1 would support a view that while the game is effective in assisting students with planning and evaluation, that the idea of not knowing the procedural steps of the game and what they are objectively doing pedantically does not correlate to supporting their monitoring skills.

As seen in Table 2, strong correlations exist between categories of both review system MAI scores, meaning scores on those questions relating specifically to skills supported by the review system (r (35) = -.384, p < .05), and self-regulation MAI scores (r (35) = -.417, p < .05), or those valuable skills developed individually and not directly supported by the review system, with

Table 1: Pearson Correlations with Performance, Means, and Standard Deviations for MAI categories of planning, monitoring and evaluating (*N*=37)

	MAI Planning	MAI Monitoring	MAI Evaluating	MAI Aggregate
Pearson's *r*	-.388	.161	-.474	-.418
M	3.30	3.27	3.56	3.39
SD	.691	.675	.798	.699

student performance. Again, statistical significance is achieved, *r* (35) < -.325. That is, those who had high MAI scores between categories, meaning those who scored well on *planning*, *monitoring*, and *evaluating*, were supported specifically by the review system, and other self-regulated metacognitive strategies, correlated with those completing interrogation room questions of at least 8 out of 10 possible points; the mandatory level of competency to complete the level. Similarly, those who had poor MAI scores did not receive adequate performance scores, on average, and continued to repeat the level until they reached mastery. Those that scored well on the metacognitive inventory appeared to take fewer times through and effectively made it through the level. While the review system may have played a role in that effective processing through the level, it is clear that students with these developed planning and evaluation skills performed more effectively with the assessment and review elements of the environment.

Of those students struggling to adequately meet standard, four of the students never met the threshold of 8 out of 10 possible points, two attempted multiple times but did not meet the threshold and ceased attempting to complete the level (although it must be noted that some of this was related to time demands for this pilot study within the school-day), and two others entered and attempted once, did not reach the threshold of points and did not try again. While a lack of companion instruction likely affected their attempts, as this simulation will be implemented within a structured curriculum, it is difficult to correlate this lack of instruction to the results here. Even of those students completing, during the pilot collection phase students that returned to the environment four or more times began to exhibit exhaustion with the system and several examples of cheating were observed as students fed answers to their neighbors to help them move on. Other students clearly showed improved scores on each attempt up to four-plus times returning through the level, and then markedly dropped off; demonstrating testing exhaustion and decreased internal motivation may have affected their later attempts. So, care must be taken here to recognize the potential role of exhaustion or cheating within this pilot study and to consider ways to mitigate against these concerns into the future. But, while cheating was witnessed during the pilot, this was best described as an isolated event instead of the norm, therefore the impact on the outcomes of this study is limited at best.

Looking back at the research questions, question one asked what metacognitive activities are supported by 3D virtual Intervention and to what extent those activities correlated to positive educational outcomes. At least in this environment, the results indicated that both *planning* and *evaluation* were positively correlated to the system design itself, while *monitoring* as a subset of metacognition did not appear to be supported by the design of S'cape. However, those students that did score well in these three subsets of metacognition did appear to show positive and swift success rates within the embedded assessment and meeting the pedagogical goals of the game over those with poor self reported metacognition scores. Thus, this initial dataset demonstrated significant correlations between those supported metacognitive skills and positive educational outcomes utilizing this form of instructional intervention.

Research question two asked the extent in which the metacognitive activities associated with the review system correlated to positive educational outcomes. While the research did demonstrate that there was a significant positive correlation between the those metacognitive activities for the review system and their positive educational outcomes, as students individual self regulation showed higher statistical significance, it is unknown to what extent students overall metacognitive activity level itself influenced their overall success rate with educational outcomes, rather than merely those associated with the review system. It is safe to state there is a significant correlation between those metacognitive skills associated with the review system and their positive educational outcomes specifically, though it is difficult with the current data

Table 2: Pearson Correlations with Performance, Means, and Standard Deviations for MAI review system and self-regulation (*N*=37)

	MAI Review System	MAI Self-Regulation	MAI Aggregate
Pearson's *r*	-.384	-.417	-.418
M	3.35	3.43	3.39
SD	.803	.616	.691

to tease apart potential contributing skill sets to that rate of success.

CONCLUSIONS

It is significant to note that 21 of the 37 pilot sample population had multiple play times, meaning they did not meet mastery the first time through and were returned to attempt again, with the maximum being eight times returning to the game. Even as 16 passed on a first attempt through the level, the average number of times returning to the game was 2.38 times. As noted, all but 4 students successfully passed the level during the allotted time period for the exercise. Based on these indicators, it would appear the level of difficulty was appropriate for the grade level, and although challenging for most, it was not overly frustrating to the majority of players. As 14 of the 21 students who did not pass on the first time through the level and interrogation room were successful in the second or third attempts, this result could indicate that the review process, utilizing principles of after-action review (AAR) and the replay feature, helped them understand what they were supposed to negotiate in the real-time 3D environment.

While difficult to state definitively whether the assessment and reflective review features of the simulation environment dramatically impacted student metacognitive abilities at this early stage, there are relational connections that can be asserted from the data. Students scoring high in *planning*, and *evaluating* elements of the MAI assessment on average performed higher on the assessment and review. The game design based on the discovery of the process to escape the environment does not support the pedantic *monitoring* of progress as a metacognitive skill, and may explain the lack of correlation in the data. Also, while challenging to complete, the learning objectives proved to be achievable to the majority of the students within the sample pool within three attempts.

Further Research

One limitation of the current study centers on the pilot testing data being restricted to only students at the high school level. It is important to note that while it is being developed to enhance student learning within the classroom, one initial step toward integration and testing the extent of effectiveness is to engage the teachers in developing a product they will integrate, learn how to use, and see the educational merits of before we can engage a set of students in piloting the work and testing their engagement and learning. Thus, teachers engaging in the simulation environment will be the next phase of our pilot testing, both for the simulation itself, and in developing the curriculum that will provide structure around the implementation of this environment. It is anticipated there will be greater levels of correlation between their metacognitive skills and positive educational outcomes when situated within the learning context it is designed to support.

While there are many current benefits and advantages to using this online game in its current permutation, there are limitations to broad based classroom adoption. Additional development is required to develop more advanced back channel archival and retrieval systems permitting the teacher or administrator to access the testing and score results of an entire classroom. Currently the system is designed to allow the instructor to keep track of their progress, score reports from interrogations, and monitor player learning using the embedded features of game play. These embedded features collate multiple attempts through the system by the same player. The entire game is presented and played through a web browser, not requiring additional maintenance or upkeep for IT personnel. Assessment data is collected and archived through the Internet.

The development is at the stage of creating automated functionalities that include saving, storage, record keeping and distribution to the proper administrative personnel, but a more robust back-end system could enhance these features. Further, additional professional development will be provided to assist instructors in using these back channel tools to maximize their usability in the classroom, as well as their own technological knowledge of the tools themselves. Such development will greatly enhance the utility, viability, adaptability and adoption as the game scales. Additional trials that work directly with military and corporate personnel for asynchronous and geographically diverse learning situations must be explored. Continued research into the ways in which this embedded assessment system with reflective replay features can support metacognitive abilities within students will remain critical in advanced improvements within this simulation environment.

ACKNOWLEDGEMENTS

We would like to acknowledge the generous NSF funding for grant # NSF 09-602 , fellow colleagues within the IDIAS group, our grant researcher team at USU, and our “birddog” Kristy Murray.

REFERENCES

Azevedo, R. (2005). Computer environments as metacognitive tools for enhancing learning. *Educational Psychologist, 40*(4), 193-197.

Barab, S., Gresalfi, M., & Arici, A. (2009). Why educators should care about games. *Educational*

Leadership, 67(1), p. 76-80.

Freeman, J., Salter, W. J., & Hoch, S. (2004). *The Users and functions of debriefing in distributed, simulation-based team training.* Paper presented at the Human Factors and Ergonomics Society 48th Annual Meeting, New Orleans, LA.

Gee, J. P. (2008). The ecology of games: Connecting youth, games, and learning. In *Learning and Games*, (Eds. Salen, K.) pp. 21-40. Cambridge, MA: The MIT Press.

Ketelhut, D.J., Nelson, B.C., Clarke, J., & Dede, C. (2010). A multi-user virtual environment for building and assessing higher order inquiry skills in science. *British Journal of Educational Technology, 41*(1), 56-68.

Kirkley, S., & Kirkley, J. (2005). Creating next generation blended learning environments using mixed reality, video games and simulations. *TechTrends, 49*(3).

Mislevy, R., Almond, R., & Lukas, J. (2003). A brief introduction to evidence-centered design. Educational Testing Service.

National Science Education Standards (1996). *National Science Education Standards.* Washington D.C.: National Academy Press.

Nelson, B., Erlanderson, B., & Denham, A. (2010). Global channels of evidence for learning and assessment in complex game environments. *British Journal of Educational Technology.*

Salmons, J. (2008). Five key areas of the 4th sustainment brigade's success. *Army Logistician, 40*(1). 8-11.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*(1), 113-125.

Schraw, G. and Moshman, D. (1995). Metacognitive Theories. *Educational Psychology Review, 7*(4), 351-371.

Shute, V. J. (1993). A comparison of learning environments: All that glitters... In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 47-75). Hillsdale, NJ: Erlbaum.

Squire, K. (2003). Changing the game: What happens when video games enter the classroom? *Innovate, 1*(6).

Thai, A., Lowenstein, D., Ching, D., & Rejeski, D. (2009). *Game changer: Investing in digitial play to advance children's learning and health.* New York: The Joan Ganz Cooney Center at Sesame Workshop.

Van Lent, M., Fisher, W., & Mancuso, M. (2004). *An explainable artificial intelligence system for small-unit tactical behavior (pp. 900-907).* Menlo Park, CA; Cambridge, MA: London: AAAI Press: MIT Press; 1999.

Young, A., & Fry, J. D. (2008). Metacognitive awareness and academic achievement in college students. *Journal of the Scholarship of Teaching and Learning, 8*(2), 1-10.