Computers & Education 122 (2018) 179–193

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

# An exploratory study of blending the virtual world and the laboratory experience in secondary chemistry classrooms

Georgia Wood Hodges*, Lu Wang, Juyeon Lee, Allan Cohen, Yoonsun Jang

*The University of Georgia, 110 Carlton Street, Athens, GA 30602, United States*

ARTICLE INFO

*Keywords:*
Serious educational games
Telemetry data
Blended reality
Scientific reasoning

ABSTRACT

Research suggests that serious educational games may support learning of microscopic phenomena by making the *invisible* visible. The current study presents a novel instructional approach that blends a traditional lab experience with a serious educational game (SEG) to engage students with the hands-on laboratory experience while examining the molecular level mechanisms that cause the phenomena. We conducted a mixed methods study to examine teacher use of the blended reality environment (BRE) that utilizes a real-time data capture system and the learning gains associated with use of the BRE with chemistry students (n = 578) in a public high school in the southeast. The results demonstrate that students who experienced the blended reality environment (a) experienced significantly higher learning gains than students who experienced the business as usual (BAU) condition and (b) improved on specific scientific practices outlined in the Next Generation Science Standards (NGSS) by practicing science skills in an authentic environment as scientists do. Analysis also reveals the explanatory power of examining student interaction within the BRE. Thematic analysis of teacher interviews identified three strengths of the BRE environment: (a) a coherent explanation of scientific phenomena, (2) an authentic science inquiry, and (3) a connection between the student and teacher within the learning environment.

## 1. Introduction

### 1.1. Teaching science with serious educational games

As computer technologies become an integral part of modern life, school administrators, teachers, and researchers strive to incorporate the technologies into classroom instruction to improve student learning outcomes. Science, technology, engineering, and mathematics (STEM) classrooms are often environments where new technologies are implemented due to the nature of the content addressed. In science classrooms specifically, the implementation of computer-based science teaching methods has been regarded as an important strategy because "it adapts to today's students who grew up in an increasingly digital world and are more accustomed to visual learning" (Xie & Pallant, 2011, p. 122). As access to new technology becomes widely available, practicing teachers and educational researchers must examine and test the new technologies in school settings to determine if the new technology enriches student learning.

Interaction with immersive games has increased across the world as access to computers, tablets, and smartphones increases. According to the Lenhart, AKahne, Middaugh, Macgill, Evans, and Vitak (2008), 99% of the boys and 94% of the girls were reported

* Corresponding author.
*E-mail addresses:* georgiahodges@uga.edu (G.W. Hodges), Lu.Wang15@uga.edu (L. Wang), juyeon.lee25@uga.edu (J. Lee), acohen@uga.edu (A. Cohen), todekfr@gmail.com (Y. Jang).

https://doi.org/10.1016/j.compedu.2018.03.003
Received 5 January 2017; Received in revised form 27 February 2018; Accepted 1 March 2018
Available online 27 March 2018
0360-1315/ 

to play digital games, and the time they spent on playing video games was approximately 7–10 h per week. Given the popularity and the high level of engagement, researchers have argued that incorporating games into education would be beneficial to students learning (Plass, Homer, & Kinzer, 2015). A recent increase in the adoption of serious educational games (SEG) by teachers, policymakers, parents, and schools (U.S. Department of Education (DOE), 2010) has led to more students engaging in SEGs during instructional time. Many schools have implemented gameplay due to understanding of modern learning theories that characterize effective learning experiences as active experiential, situated, problem-based, and capable of providing immediate feedback (Boyle, Connolly, & Hainey, 2011). SEGs address many of these characteristics. While scholars continually struggle to agree on a definition for "games", many scholars accept the definition provided by Clark C. Abt in 1968, "any contest (play) among adversaries (players) operating under constrains (rules) for an objective (winning, victory pay-off)" (as cited in Ellington, Addinall, & Percival, 1982, p.9). In Connolly, Boyle, MacArthur, Hainey, and Boyle's, (2012) meta-analysis, which reviewed 129 empirical studies of SEGs, increased perceptual, cognitive, behavioral, affective, and motivational outcomes were quantified. The SEGs are most effective at promoting students' knowledge acquisition, affective, and motivational outcomes (Connolly et al., 2012). In this paper, we adopted Marsh's (2011) definition of serious educational games that states, "Serious games are digital games, simulations, virtual environments and mixed reality/media that provide opportunities to engage in activities through responsive narrative/story, gameplay or encounters to inform, influence, for well-being, and/or experience to convey meaning" (p.63).

Within the field of science education specifically, learning outcomes associated with gameplay have measured content knowledge gains (Clark et al., 2011) and scientific habits of mind (Steinkuehler & Duncan, 2008), which supports the use of SEGs in science classrooms (Honey & Hilton, 2011; Squire, 2011). Although learning gains associated with SEGs have been detected, they are explained primarily by differences in scores on a pre-test and a post-test administered *before and after* instruction rather than during instruction. Researchers (Hodges et al., 2015) have found that what students do *within* gameplay provides powerful, compelling insight into student learning that pre- and post-test assessments fail to measure. Some scholars have applied educational data mining techniques to extract useful information from large sets of data for further examination (Berland, Baker, & Blikstein, 2014). Although valuable insights are found in these data, the data are not collected in real-time and, thus, cannot inform instruction in the classroom in real-time. Learning analytics (LA) are designed to provide real-time information for educational decision-making (Ifenthaler & Widanapathirana, 2014). As stated by the Society for Learning Analytics Research (SoLAR), "games are quickly gaining momentum as a tool for learning and assessment. However, the method by which to harness this data to understand learners and improve our games are less than clear (Loh, Sheng, & Ifenthaler, 2015). Theoretically, scholars have described serious game analytics as "actionable metrics developed through problem definition in training/learning scenarios and the application of statistical models, metrics, and analysis for skills and human performance, improvement, and assessment, using serious games as the primary tools for training (Loh et al., 2015, p. 23)." Through years of rigorous research, our team has developed a new framework for SEGs, which we describe below, that provides relatable and engaging authentic inquiry experiences for students and also provides real-time data analytics to teachers to utilize within the classroom setting.

### *1.2. Leveraging telemetry data within a SEG*

Telemetry data is defined as the systematic collection of specific actions within a game environment (Chung, 2015). Scholars have identified the struggles faced by researchers who seek to utilize SEGs within traditional educational environments (e.g. Bousbia & Belamri, 2014), namely the lack of standardization of telemetry data, which limits the potential use of a standardized tool to process data. This lack of standardization leads to huge time sinks that require data analysts to analyze the data set retrieved from gameplay and transform them into a useful output style, often requiring 70%–80% of the data analyst's time (U.S. DOE, 2010). Thus, analysis of the data occurs after gameplay, rendering these data useless for teachers' and students' formative use within the moment at which learning actually takes place. As such, our team created a novel telemetry capture and analysis tool that distills the specific moves taken by students within gameplay and reports findings to the teacher in real-time. The telemetry data are aggregated by discrete skills as the system grades the majority of assessment items embedded within gameplay. The discrete skills align with the scientific practices of analyzing and interpreting data, forming hypothesis, predicting, communicating findings, and supporting claims with evidence to construct an explanation. Each of the learning environments addresses a disciplinary core idea identified by teacher partners and the literature as a concept that students struggle to master. The research team has studied the development of the data capture system throughout a four-year timeframe, and the development team has created multiple learning environments that utilize the same system. Due to the focus on transferable skills, the research and development team named the learning environment SABLE (skills and assessment-based learning environment), as the system focuses on specific scientific reasoning skills and disciplinary core ideas (Hodges et al., 2015). With this tool, student performance at each stage can be sent to their teachers as they progress, which allows teachers to monitor student progress in real-time throughout the learning experience and intervene when needed. Teachers may also examine, grade, and leave useful comment for students' written responses. Students can access their scores and comments with their unique usernames and passwords. This feature enables students to get feedback immediately. What is novel within this research is the tethering of a SABLE SEG with a laboratory experience. As such, our team garnered insight into student movement and response within the Blended Reality Environment (Robertson & Ducrest, 2016) during a wet lab.

### *1.3. Chemistry teaching at the pre-college level*

#### *1.3.1. Teaching molecular chemistry concepts*

The domain of chemistry has been described as consisting of three interacting levels of understanding: macroscopic (observable), microscopic (molecular), and symbolic forms of representation (Bowen, 1998; Johnstone, 1993). The macroscopic level refers to phenomena that can be observed through naked eyes, such as a combustion reaction, sodium chloride dissolving, and gas formation during chemical reactions. The microscopic level refers to phenomena that cannot be observed directly, such as the structure of an atom, while the symbolic forms of representation refers to the expression of the observed physical phenomenon. A typical symbolic concept is the Periodic Table, a model that characterizes the properties of different elements.

In high school chemistry curricula, students are tasked with learning far more microscopic or molecular level concepts than macroscopic or symbolic level concepts. For example, traditional high school curricula require students to learn concepts such as the mole, chemical bonding, and chemical equilibria. These concepts require students to apply the symbolic representation of phenomena to a microscopic phenomenon, such as the model of an atom or molecule. Chemistry students throughout the world struggle to grasp molecular concepts, and research has identified two primary reasons for the difficulty. First, students fail to connect concepts such as the models of atoms to their daily life, thus they fail to engage deeply with the concepts (Barke, Hazari, & Yitbarek, 2009). For example, the concept of chemical bonding, the formation and breaking of chemical bonds, undergirds all chemical reactions. But the processes that form and break these bonds cannot be observed through naked eyes in students' everyday life. To grasp this concept, students must understand the structure of atoms and activities of electrons, which are complex, molecular level concepts. To address this, teachers implement varying strategies to relate chemistry to students' daily lives. Second, the flexibility of some atoms confuses students as some models change when applied to different situations. For example, when covalent bonds are formed, the three electrons in the outer spheres of an aluminum atom are shared with oxygen atoms, yet when an ionic bond is formed with chloride ions, the same three outer sphere electrons transfer to the outer sphere of chloride atoms. Thus, students have difficulty understanding what factors influence the behavior of electrons of the same element in different situations. To address these known limitations of student understanding, teachers implement laboratory experiences that address these microscopic concepts to support knowledge construction.

#### *1.3.2. Chemistry as a laboratory science*

Laboratory work is an integral part of chemistry instruction. Laboratory work provides students opportunities to interact with varying phenomena, which engages students and improves their understanding of matter, conversion, and chemical changes (Leite & Afons, 2002; Tezcan & Bilgin, 2004). The laboratory environment plays an important role in students understanding of basic concepts in a particular scientific topic (Jagodziński & Wolski, 2015). More importantly, student work in the laboratory effectively engages students with scientific practices, one of the three dimensions of high quality learning experiences in the sciences (National Research Council, 2012). Designing experiments, manipulating equipment, and collecting data are each experiences in which students enrich their scientific reasoning and critical thinking skills through the application of specific scientific practices. With these skills acquired through their laboratory experiences, students can learn chemistry by acting as scientists do. The NGSS explain that laboratory teaching is required for students to develop the type of thinking and reasoning skills needed to understand science, especially in chemistry.

Although laboratory experiences are viewed as essential to chemistry instruction, research regarding the actual effectiveness of laboratory teaching and learning is contradictory. Hawkes (2004) insisted that traditional laboratory classes did not "help students to understand how chemical principles affect their universe" (p. 1257). Similarly, McKeachie's (1999) review paper found that laboratories are not effective in teaching the practices associated with the scientific method. However, the researchers concluded that the specific teaching contexts and laboratory experiences should be studied in more depth. Based on these results we could say that laboratory instruction is important for chemistry teaching and learning as students practice science skills, yet the traditional teacher-centered model of laboratory instruction lacks compelling evidence that conceptual learning has taken place. As Hawkes (2004) discussed, the laboratory experiment can help students to develop an understanding of designing an experiment, as well as many other scientific practices if the labs are carefully arranged and taught. He also argued that the objective of promoting students' problem-solving, which is highly praised by advocates of laboratory instruction, can be achieved by computer simulations.

### *1.4. Redox reaction*

The redox reaction is one of the most important concepts in chemistry as it underpins virtually every biological and chemical reaction. Conceptual understanding of the redox reaction requires mastery of four major concepts which include: naming oxidation numbers and values, defining of reduction and oxidation, balancing complex redox equations, and identifying the relative strength of reducing and oxidizing agents (De Jong, Acampo, & Verdonk, 1995). To understand these concepts, students must construct an abstract, mental model of the molecular interaction. Therefore, a laboratory experience is usually implemented to demonstrate the mechanism of the reaction; one of the common experiments used is "*gold rush*" lab. When students conduct an experiment that involves redox reactions, students see the macroscopic phenomena, or the effect of the reaction, but they do not see the microscopic reaction mechanisms, where the changes actually occur. Thus, teachers traditionally end up taking students out of the lab to teach related concepts, with the hope that students will be able to connect the experiential lab experience with the notes they take that focus on chemical reactions or the problems they work. Yet, this approach removes students from the lab, from the place where the reaction actually happens, disconnecting the conceptual learning from the actual phenomena under study.

Based on this disconnect, we created a new instructional approach to teach the molecular concepts related to redox reaction, which we call the Blended Reality Environment (BRE). This new learning environment blends a serious educational game with a hands-on laboratory, which immerses students in an inquiry learning experience, provides real-time feedback to the students, and equips teachers to examine student performance in real-time. Thus, the BRE intersects the macroscopic, microscopic, and symbolic

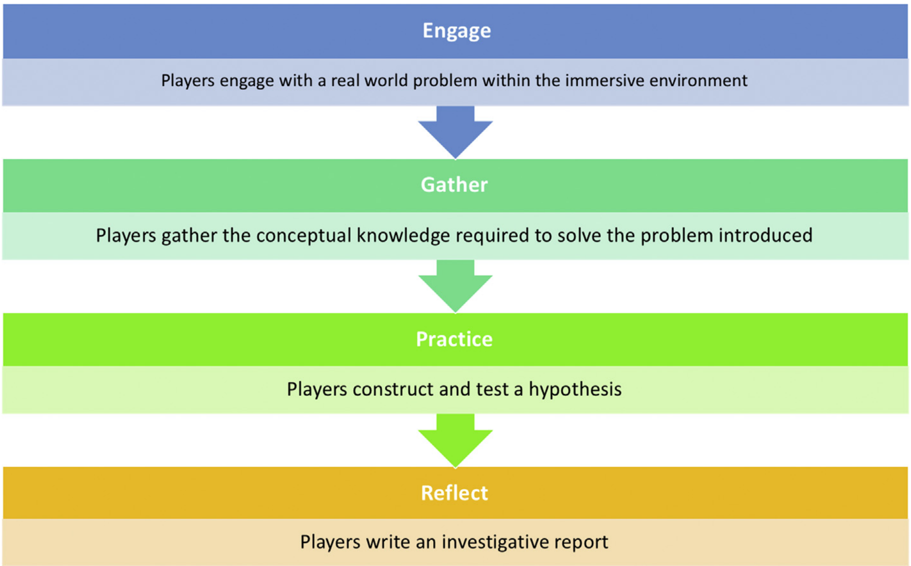



**Fig. 1.** Flow of the Blended Reality Environment.

representations that the research has characterized as essential to acquiring chemistry conceptual understanding.

### *1.5. The Blended Reality Environment (BRE)*

#### *1.5.1. The flow of the Blended Reality Environment*

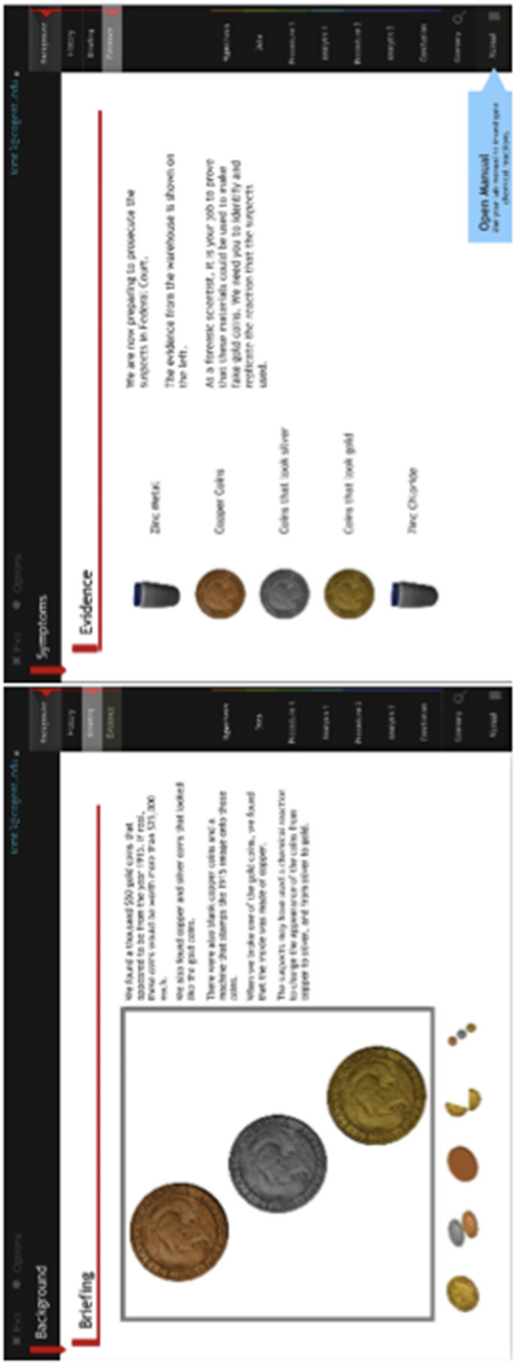
The flow of the Blended Reality Environment (BRE) is demonstrated in Fig. 1. Within the Blended Reality Environment (BRE), students first log into the SEG where students are given the role of a forensic scientist working for the Secret Service Students are "hooked" in to the learning experience through high quality visualizations that outline the challenge of determining how suspects used a chemical reaction to make fake coins. By conducting a "*gold rush*" experiment, students provide evidence about how the suspects transform the copper coins to appear silver and gold so that the suspects can be prosecuted in Federal court (Pics-1). A navigation panel details the steps students follow during the experience that align with the scientific method.

In order to explain how the fake coins are made, students first complete the "laboratory manual" which provides the background knowledge required to solve the problem. Prior to using the BRE, teacher addressed this information by providing worksheets that explain the procedures of the "*gold rush*" lab. Compared to the traditional didactic instruction, the "laboratory manual" is an interactive guide that addresses the key learning objectives required to complete the experience using powerful visualizations and simulations. These core concepts scaffold the learning for the student and are further supported by an embedded glossary that allows students to review key terms by clicking on any text that is orange in color. Throughout the laboratory manual, embedded items monitor and assess student learning (pic-2). Students are provided real-time feedback as to whether their answer was correct or incorrect as well as elaborated feedback to reinforce the concepts addressed (pic-3). Meanwhile, the teachers are provided with a heat-map that provides feedback student performance on the assessment items in real time, which enables teachers to immediately identify students that may be struggling with a particular concept or skill (pics-4). The manual was designed to equip students to learn more complex concepts in a step-wise manner by providing students opportunities to reflect on their new knowledge, review concepts they found difficult, and refine their understanding through answering questions that require use of the knowledge. In other words, rather than just "telling" the students the concepts, the student builds their own knowledge by interacting with the new concepts, in this case, atoms and ions.

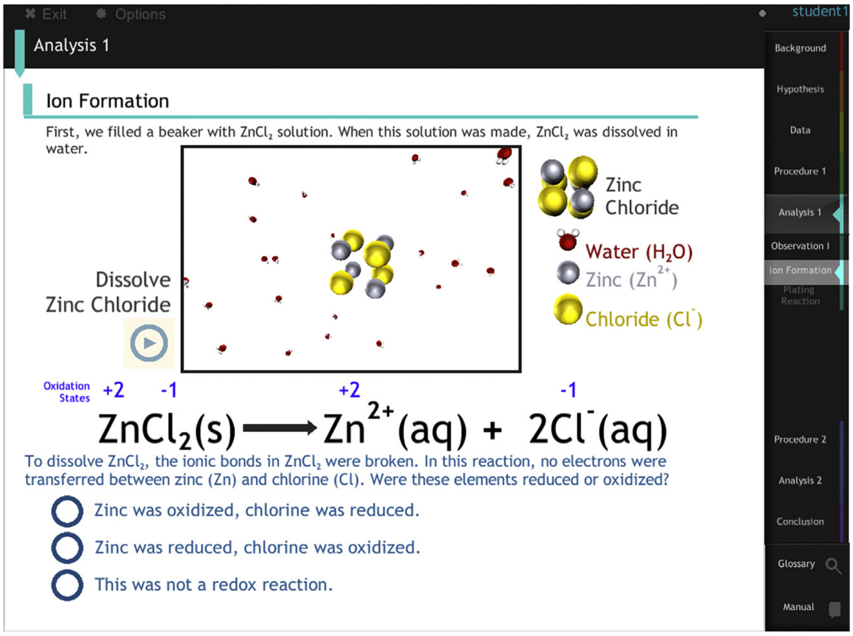
After students successfully complete the manual, which requires by exhibiting mastery of the requisite knowledge, students are asked to make a hypothesis regarding the type of chemical reaction used by the suspects to make the coins appear silver and gold. To test their hypothesis, students use the "gold rush" lab to make a copper penny appear silver, then gold. Since safety in the chemistry classroom is a vital consideration, students are provided with a virtual checklist (pics-5) that they must complete before they perform the hands-on experiment.

In the first stage of the experiment, students make the copper appear silver by using a redox reaction to plate a layer of zinc onto the surface of the penny. Students respond to the embedded assessment question whether the reaction was successful and, if not, they are able to play a video of a successful reaction. Students then review the reaction by viewing an animation of the reaction in 3D; students are then assessed to determine whether they understand the key concepts of the reaction. Unlike a traditional lab, students are able to see an atomic view of the plating reaction generated by the redox reaction simulator.

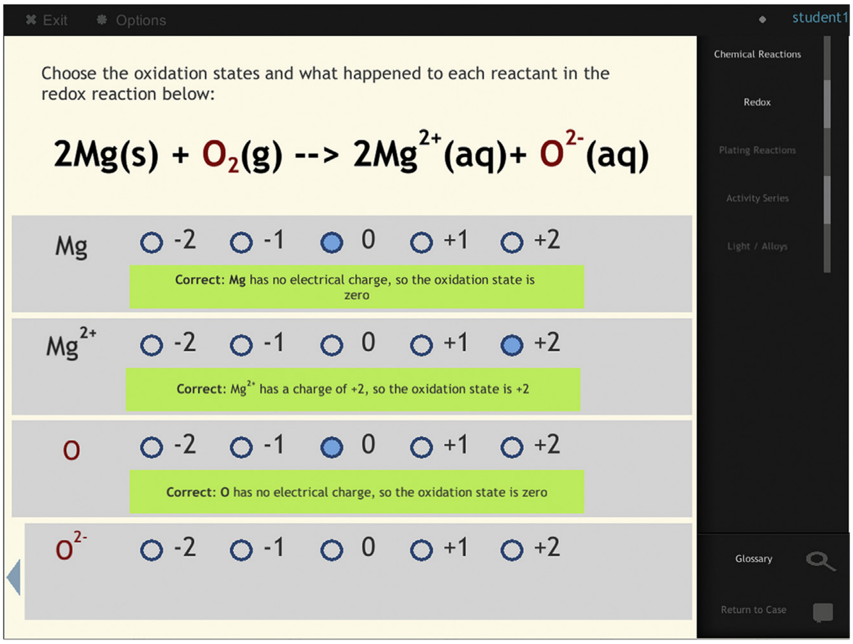
The final stage of the experiment involves another chemical reaction that changes the silver looking penny to appear gold in color. Students return to the laboratory manual where they interact with an atomic simulator that allows the student to observe the effects




**Picture 1.** *Engaging* students with a coherent storyline in the BRE.



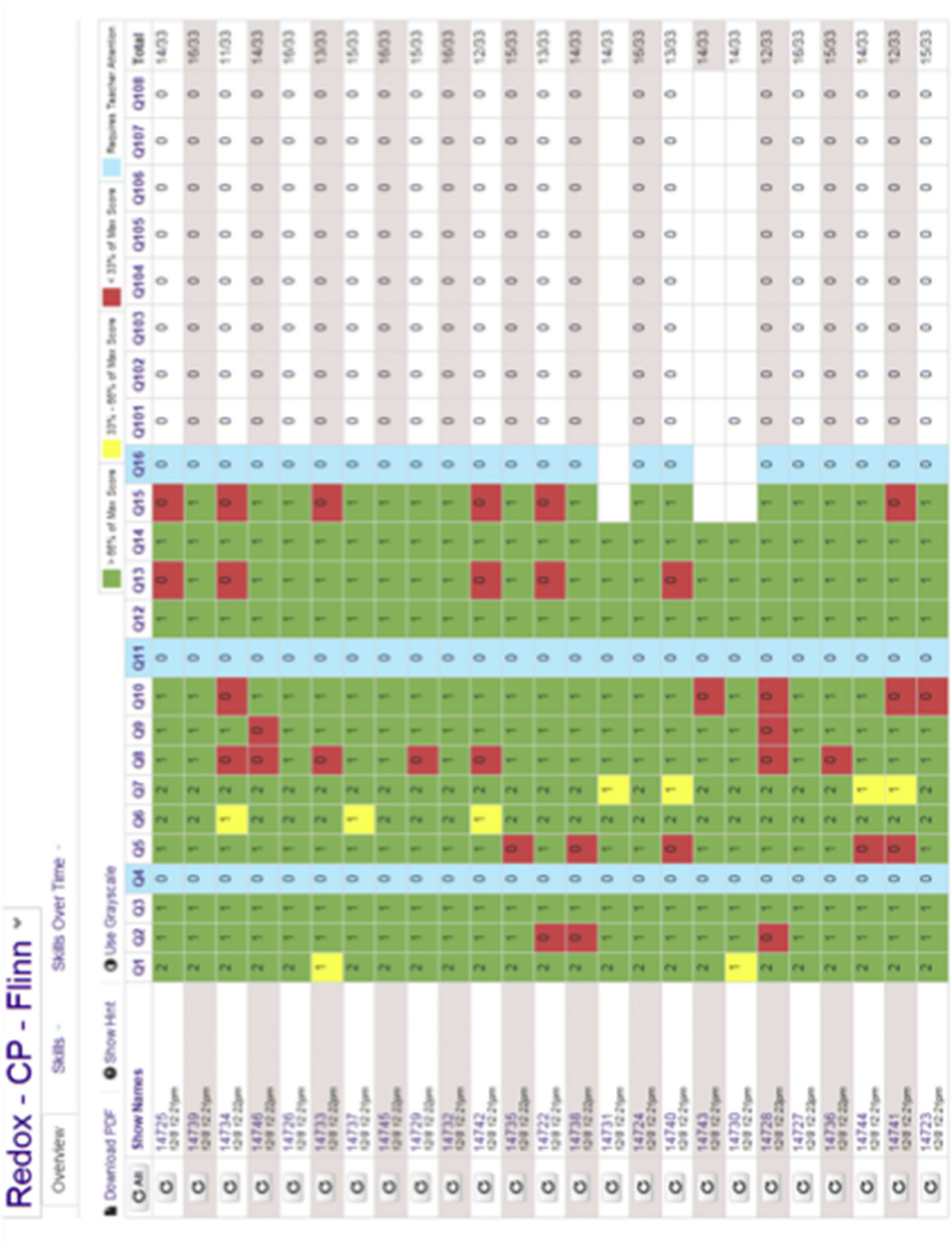

**Picture 2.** Embeddingassessment in the BRE.



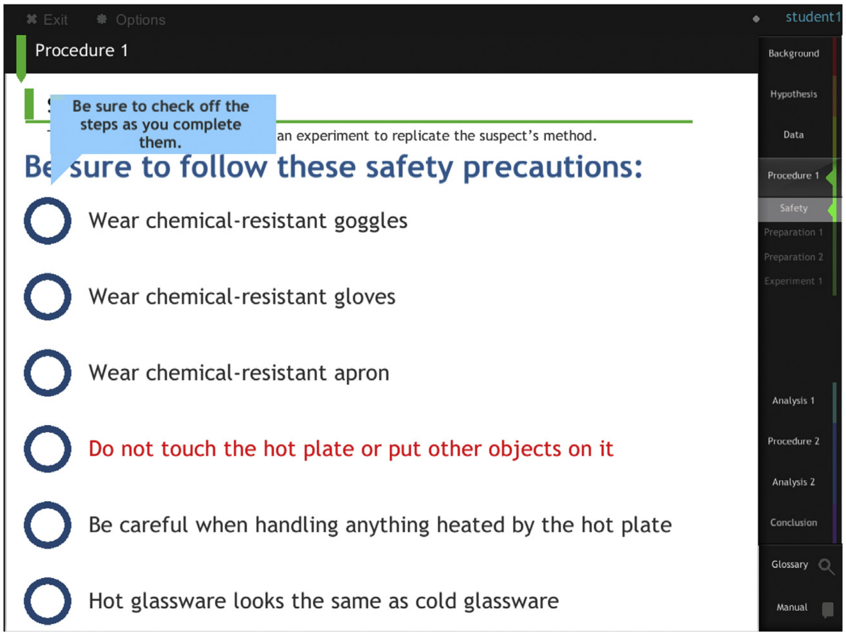

**Picture 3.** Providing elaboratedfeedback in the BRE.

of applying heat to a plated metal and the subsequent formation of an alloy. After reviewing alloy formation, students work through a safety and procedural checklist as they place the zinc-coated penny onto a hotplate and observe the color change. Again, a video of the reaction is provided should their experiment not work.

Upon completion of the experiment, students are tasked with writing a case summary of their investigation that requires them to construct a meaningful explanation of the phenomena studied. To assist students in writing an appropriately structured report, students are asked to complete specific sections and are provided with suggestions for key terms and concepts to include in these sections to scaffold the process. After submitting the report, students are taken to a debrief page that shows their report submitted to the Secret Service where they receive text messages that congratulates the student on their investigation and informs them that the evidence obtained was used to successfully prosecute the suspects.




**Picture 4.** Providingreal time data to teachers on student performance in BRE.




**Picture 5.** Reminding students ofimportant safety concerns before beginning the wet lab.

### 1.5.2. *The telemetry assessment*

Concurrent with this newfound acceptance of SEGs as powerful pedagogical tools is the hope that fine-grained data from gameplay may be captured and transformed into a measure of student learning (e.g. Loh et al., 2015; Shute & Ke, 2012; U.S. DOE, 2010). Our team created an analysis system (see Hodges et al., 2015) that monitors student progress and performance on the embedded assessment items. Data from this system are sent in real time to a Teacher Dashboard accessible via the IS3D websitewww.cogenteducation.com. Data from forced choice items are automatically transformed into any easy-to-read heat map with green, amber and red cells in the table correlating to good, intermediate and poor performance, respectively. This enables teachers to know exactly where each student is within the case and to immediately identify students that may be struggling and need their help. Text responses appear as blue cells, and teachers can see this text by clicking (on a computer) or tapping (on a tablet) on a cell – a pop up displays the question being asked, the student's text, a suggested scoring rubric, and an exemplar answer. Teachers can assign scores for each text response and leave comments for their students. Each assessment item is "tagged" with a specific scientific practice (aligned to standards), and upon scoring their students' text responses, teachers can click a button that sorts all of the assessments into steps of the scientific method of problem-solving, namely: 1) the core science concepts, 2) data analysis, 3) data interpretation, 4) making a hypothesis and providing a reasoned justification, 5) predicting the effect of their intervention upon the data under consideration, and 6) communicating their findings in a summary report at the end of the game. Upon completion of a case, students can view and download their game report, including teacher comments from their IS3D account. As students complete more cases, teachers are able to formatively assess the development of *skills* over time. Short overview videos of the real time data system can be viewed at: http://tinyurl.com/UGASable.

In summary, our BRE seamlessly "blends" hands-on laboratory with a SEG, that: 1) relates science to the real world; 2) engages students in learning science, and 3) immerses students in a virtual world. This learning environment enables students to: 1) acquire the requisite knowledge; 2) collect data; 3) analyze and interpret data; 4) form and test a hypothesis; 5) refine the hypothesis according to the data; and 6) communicate their findings. The portion of the study addressed in this article addresses the following research questions:

1. How does the Blended Reality Environment (BRE) improve students understanding of redox reaction?
2. How does the Blended Reality Environment (BRE) support student learning of molecular concepts in chemistry?

## 2. Method

This study is a mixed-methods research. We utilized qualitative and quantitative methods to examine students learning gains and the efficacy toward use of the BRE.

### 2.1. *Research design and participants*

We conducted the research at a large public high school (4000 students) in the southeast with demographics that align with those of the US collectively. We met with the chemistry department at the school and all of the teachers (n = 6) agreed to participate in the study. The teachers use a collaborative approach in planning and implement the same activities across all classrooms. Due to limits in the number of computers available, two lab spaces were supplied with the technology required to run the BRE. As such, the two teachers who taught six sections of chemistry served as our treatment teachers, while the rest of the teachers agreed to serve as the control sample for the study. All students were 10th grade students in introductory chemistry classes that completed the same pre- and post-test. There were 578 students that participated the study, 351 students finished both pre- and post-tests. We only included students in the analysis who completed both the pre- and post-tests as well as the intervention or the business as usual (BAU) lab experience.

Students in both groups were asked to complete the same online pre-test created that assessed their knowledge of the redox reaction. The next day, all of the students completed the *gold rush* lab. Thus, the only difference involved the use of the BRE by the treatment students. Upon completion of the laboratory experience, all the students finished the same post-test within three days.

### 2.2. Assessments

#### 2.2.1. Pre- and post- tests

Students' understanding of redox reaction was measured through pre- and post-tests developed by a team of scientists, teachers, and educational researchers to address specific learning objectives associated with the redox reaction (Appendix 1).

#### 2.2.2. Telemetry data assessment

To analyze student learning within the BRE learning environment, we applied logic derived from the knowledge integration theoretical perspective to create analysis frames for each question. The knowledge integration framework (Linn, Lee, Tinker, Husic, & Chiu, 2006) provides a theoretical perspective for evaluating the processes by which students connect ideas and evidence in order to develop more coherent and comprehensive understandings of science and science principles. Knowledge integration is a “dynamic process where students connect their conceptual ideas, link ideas to explain phenomena, add more experiences from the world to their mix of ideas and, restructure ideas with a more coherent view” (Bell & Linn, 2000, p. 797). Most of the work done with the Knowledge Integration Framework has used digital tools such as *SenseMaker* (Bell & Linn, 2000) or the Web-based Inquiry Science Environment (WISE) (Linn et al., 2006). Bell and Linn (2000) have examined the development of middle school students' science arguments using these supports. We applied this construct to create our own rubrics for the research team to apply to the data set. Use of these rubrics led to an interrater reliability of 85%.

#### 2.2.3. Interviews

The two teachers who taught the treatment group students were interviewed before, during, and after the intervention. We also conducted semi-structured interviews with students while they completed the BRE experience, then we conducted whole class focus groups upon completion of the experience.

We used qualitative methods including thematic analysis to examine students' work and talk (Ezzy, 2013). We used line-by-line analysis of transcripts from students' discourse during lab group work and focus groups (Ezzy, 2013). The coding scheme was deductive since it began with rubrics previously described in the literature (e.g. Clark & Sampson, 2008; Zembal-Saul, McNeill, & Hershberger, 2013; Linn, and Elyon, 2011), but it included inductive attributes as well. For example, we modified rubrics as additional themes emerged from the data. We then used axial coding (Ezzy, 2013) to identify themes, processes and relationships among codes to address the research questions.

## 3. Results and discussion

### 3.1. The pre- and post-tests data analysis

We administered the pre-test and post-test to measure students understanding of redox reaction (Table 1). A total of 351 students completed both the pre- and post-tests. Students in the treatment group (BRE) and control group (BAU) started with similar levels of prior knowledge as the mean of total scores for pre-tests were 7.55 and 7.07 respectively. Since all teachers co-planned and utilized the same lesson plans with all the students prior to the intervention, it is not surprising that students' knowledge of redox was comparable across classes. After the treatment, students in the BRE group outperformed the BAU group. The mean of total scores for the BRE group increased to 8.67 while the mean of total scores for the BAU group decreased silently to 6.50. This pattern is plotted in

**Table 1**
Descriptive statistics of total scores.

| Group | Test | N | Mean | SD |
|---|---|---|---|---|
| BRE | Pre-test | 184 | 7.55 | 2.79 |
| | Post-test | | 8.67 | 3.18 |
| BAU | Pre-test | 167 | 7.07 | 2.45 |
| | Post-test | | 6.50 | 2.18 |

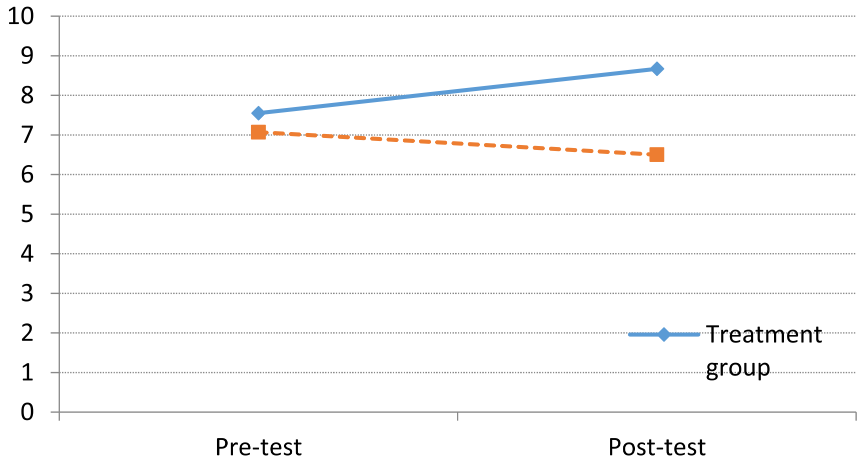



**Fig. 2.** Means of the total scores for the pre- and post-test by group.

Fig. 2.

Since students of BRE and BAU groups scored similarly on the pre-test, we can compare the learning gains across the groups. We calculated gain scores by subtracting the pre-test score from post-test score (i.e., gain scores = post-test total score – pre-test total score). We use gain scores to explore the significance of difference of treatment effects. We also examined the effect of gender on students' performances on pre- and post- tests. A two-way Analysis of Variance Analysis (ANOVA) was employed and the descriptive statistics of gain scores and the results of the two-way ANOVA are summarized in Table 2.

The means of gain scores for the BRE group students were positive (mean = 1.12), while the means of gain scores for the BAU group were negative (mean = −0.56) as shown the Table 1 and Fig. 1. Moreover, the effect of the treatment on the gain scores was statistically significant at the significant level 0.05 (F-value = 32.44), that is, the gain scores of BRE group students were significantly higher than the gain scores of BAU group students. The effect size of treatment was about 9% ($\eta^2$ = .09). On the other hand, the means of gain scores for the male students and female students were similar, and the effect of gender on gain scores was not significant, which means that gender did contribute to students' differences in performances. Research has shown that in traditional science and chemistry instruction, more female students tended to drop out and female students expressed less interests in chemistry than their peers (Levine, Serio, Radaram, Chaudhuri, & Talbert, 2015; Simpkins, Price, & Garcia, 2015). But our data showed no gender differences, which suggests that our BRE instruction engaged both male and female students equally. Additionally, the interaction between the treatment and gender was not significant.

After identifying the significant ($p < .05$) learning gains associated with use of the BRE redox case, we conducted a deeper analysis to identify where the learning gains occurred. We did this by comparing the learning gains of each question (Fig. 3). This analysis identified six questions where students who utilized BRE experienced learning gains that varied significantly from the control group. The largest differences between the two groups occurred on questions 5, 10, and 13. These three questions are associated with three major concepts of redox reaction: defining reduction and oxidation; identifying the relative strength of reducing and oxidizing agents; and writing complex redox equations. Students are required to understand redox concepts at the molecular level, specifically the mechanism of redox reaction. The BRE students' performances on these questions showed that this learning environment supported students learning of molecular level concepts. This is a clear advantage of the BRE intervention as it makes visible the processes and structures which cannot be seen through naked eyes. This supports the use of visualizations to teach molecular concepts that are more difficult to teach using traditional methods.

### *3.2. Analysis of telemetry data*

To assess how the BRE support students learning of redox reaction, we examined the telemetry data, which recorded students' responses across the items that are embedded within the gameplay. Within the BRE, students are assessed at 15 points throughout the

**Table 2**
Descriptive statistics of gain scores by gender and group.

| Independent variable | | N | Mean | SD | *F* | Effect size ($\eta^2$) |
|---|---|---|---|---|---|---|
| Treatment | BRE | 179 | 1.12 | 2.85 | 32.44* | .09 |
| | BAU | 164 | −0.56 | 2.44 | | |
| Gender | Male | 156 | 0.32 | 2.84 | 0.01 | .00 |
| | Female | 177 | 0.34 | 2.75 | | |

Note. *$p < .05$.

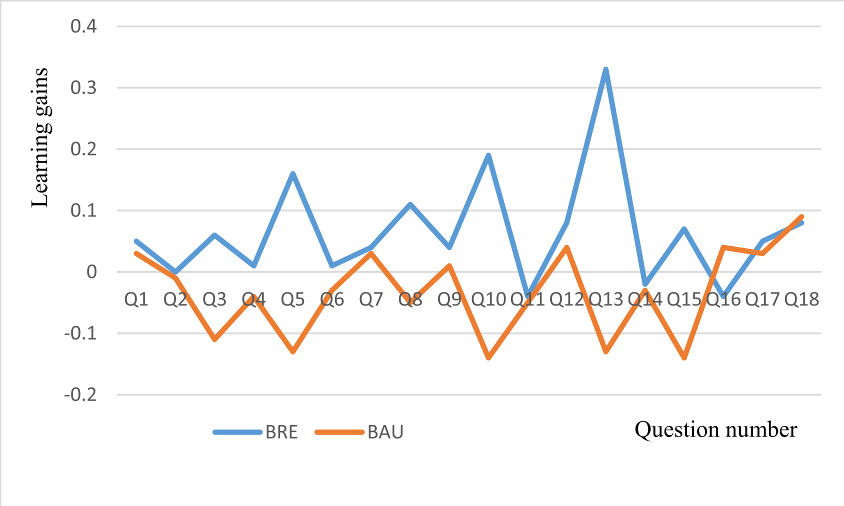



**Fig. 3.** Learning gains of BRE and BAU groups of individual questions.

experience. This data set consists of two types of questions, multiple choice questions and constructed response questions, which are designed to monitor student progress within the gameplay.

### 3.2.1. *The multiple-choice questions*

The multiple-choice questions were constructed to provide immediate feedback to students and teachers regarding location within gameplay or to provide conceptual feedback. For example, two multiple choice items require students to show evidence that they understand the safety procedures required to conduct the lab experiment. Multiple choice items also identify when students reach milestones within the game, such as when chemicals are being used, thus enabling teachers to monitor students more closely when needed. Other multiple-choice items formatively assess students' content knowledge as they progress through the game. For example, the post-test data identified significant learning gains associated with the concept of redox. Telemetry data collected during gameplay explains *how* students constructed this new knowledge. Here, we only presented the questions that are related directly to core concepts of redox reaction (Table 3). The high correct percentage of students' answers shows that the students learned the concept of redox reaction and could subsequently use their knowledge to solve practical problems. For example, question 3 tests students' understanding of electron transfer by asking the number of electrons transferred in a calcium and oxygen reaction. We asked the same concepts with question 10 in our pre- and post- tests. The students' mean score for this question in the pre-test were 0.31 (treatment) and 0.28 (control). This comparison shows that the BRE support students understanding the molecular concepts of redox reaction. Students' different performances on post-tests further confirms the efficacy of BRE (with mean scores of 0.51 and 0.14 respectively).

For question 15, since it is the last question in the SEG, the lower correct percentage is due to the low percentage of students who attempted to answer the question.

### 3.2.2. *The constructed response questions*

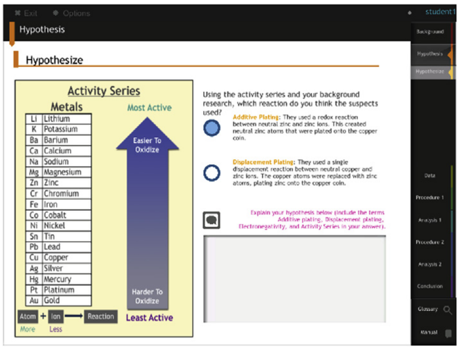
Constructed response items require students to provide a written response to a question within the gameplay. Two constructed response questions focused on the four major concepts related to understanding the redox reaction. In this section, we analyze one question (Question E11) to highlight the efficacy of BRE. Question E11 takes place before students have conducted the lab, when they are asked to make a hypothesis related to the redox reaction phenomenon. After students develop their hypothesis, the SEG provided additional data that could not be obtained through the *Gold Rush* lab. Then students were offered an opportunity to revise their hypothesis and conduct the lab to test if their hypotheses were correct. From the telemetry data, we can see students' original hypotheses, their use of evidence to support their hypotheses, and *revisions to* students' hypotheses after experimenting.

To assess E11, we developed a rubric based on Linn's research (Linn, M.C., 2000; Linn et al., 2006). The rubric and sample student responses are presented in Table 4. Specifically, question E11 asked students to make a hypothesis of the reaction that happened and then use two pieces of evidence to support their claim. One piece of evidence involves using an activity series to predict if a redox reaction could happen. The other piece of data requires students to utilize the mass of the penny. We examined student performance on all four components of this item.

**Table 3**
Students performances on example multiple choice questions.

| Questions | E3 | E5 | E15 |
|---|---|---|---|
| Concept assessed | Electrons flow between atoms | Half reaction | Application of redox reaction |
| Mean | 0.83 | 0.58 | 0.20 |

**Table 4**
Rubric for grading question E11.

| | |
|---|---|
| Concept Criteria<br>1) Predict that displacement plating could not have occurred.<br>2) Using activity series to support that it is additive plating.<br>3) Using mass data as the second evidence to support an additive plating hypothesis. |  |
| Level | 1. Additive plating; 2. Activity series; 3. Copper ion cannot react with Zinc; 4. Mass data. |
| No Answer | |
| Irrelevant/Incorrect | “They used Displacement Planting because it will show whether it's really copper or not. if you have a neutral copper with a copper ion then you could tell whether it reacts and if it reacts then you will know that it is fake or real.” |
| Partial connects | “I meant to select additive plating because the criminals used the neutral zinc atoms and zinc ions to create the fake coins. In additive plating the zinc metal and zinc ions react, which is allowed since they are in the same position on the activity series. Then the new neutral zinc atoms bond metallically with the copper atoms. If the criminals used displacement plating there would be no need for zinc ions.”-1, 2. |
| Basic | “Because there are neutral zinc and copper ions, the neutral zinc in more active than copper. Since these metals can react according the Activity Series, the suspects used Additive Plating. It cannot be Displacement Plating because the neutral copper cannot react with the zinc atoms.”-1, 2, 3. |
| Complex | “Additive plating is what they used because if they used displacement plating, then the zinc and copper would not react due to their relative placement on the activity series. Since the mass of the suspect's coin is greater than the copper coin, it supports the idea that they used additive plating. (additive plating increases mass)”-1, 2, 3, 4. |

Students' performances on this question (see Table 5) indicated that 50% of the students made the correct hypotheses. Among the students who responded with an incorrect hypothesis originally, almost half changed their hypotheses after being provided the additional data that could not be obtained through their *Gold Rush* lab experience. There were 29% and 26% students who used the concepts of activity series and place of metals in the activity series to predict if a redox reaction could happen. Following the BRE, we found that half of the students who had incorrect hypothesis revised their hypothesis.

### 3.2.3. *Interview and participant observation triangulation*

Interviews, focus groups, and participant observation of teachers and students involved in the study further confirmed the efficacy of the BRE. From this qualitative research data, we identified three main themes that demarcate the BRE from the traditional lab: (1) a coherent explanation of scientific phenomena, (2) an authentic science inquiry, and (3) a connection between the student and teacher within the learning environment.

*Coherence*. When conducting the experiment, students continually returned to the SEG when they experienced a phenomena change, such as the penny color change. Students explained, “we often get to conduct experiments in chemistry, but we typically talk about what is actually happening before or after we finish. Well, that's not really when I need the information. I need to understand when I'm in the middle of the experiment, not when it's over. And when teachers try to tell me everything before I begin, it doesn't even matter to me, and I can't remember when I'm in the middle of it.” Teacher interviews provided a different perspective on this same concept. Ms. Smith explained, “this experience meets the students where they are, when they arrive, and supports them in real-time.” Unlike a typical lab experience where teachers front load instruction regarding how students should complete a lab, the BRE supports students when they arrive at a new concept or need to follow a set of instructions.

*Authentic inquiry*. In addition to a coherent explanation, students experienced authentic scientific inquiry due to the shift in instructional support. Ms. Jones explained, “Students actually got to experience the lab the way it's designed, where they get to figure something out rather than me telling the students what is going to happen, and what to look for as they progress.” Teacher interviews

**Table 5**
Students performances on question E11.

| Original hypothesis | | Revised hypothesis |
|---|---|---|
| Correct | Correct-correct: 46% | Correct |
| 50% | Incorrect-correct: 30% | 76% |
| Incorrect | Correct-incorrect: 3% | Incorrect |
| 50% | Incorrect-incorrect: 21% | 24% |

after the lab confirmed the interest that the new design added to a lab often conducted by the teachers. “In eighteen years of using this lab, I have never seen my students so engaged.” The BRE supported students so that they solved the problem by explaining the phenomena using their chemistry knowledge. When interviewed prior to using the BRE case study, 100% of the teachers explained that students were conducting a legitimately inquiry-based lab. During a collaborative debrief with the six teachers, their beliefs regarding the traditional lab had shifted. Both of the teachers who used the BRE case study said that they now believed that the lab without the BRE supports was an exercise in following directions rather than discovery. Ms. Smith explained, “I never have had the opportunity to let my students go with a lab like this. I had to tell them step-by-step what they needed to do because I did not have a tool to do otherwise. Now they get to figure out what is going on and I get to coach them as they go.”

*Connecting*. Teachers and students conveyed support for the BRE learning platform. Students explained, “to be honest, I can typically slack off when we use computers because my teacher doesn't know what I'm doing. Ms. Smith came up and asked me why I was spending so much time on part of this lab-I had no idea she knew what I was doing!” Similarly, teachers explained that the real-time feedback functionality of BRE kept them involved in the learning experience. “I knew where my students were, how they were progressing through the lab, and when I needed to go talk to specific lab groups. Usually I am just winging it, trying to spot students who may need my help. BRE enabled me to systematically examine what my students are doing and utilize my time efficiently to support them.”

## 4. Conclusion and future study

This research investigated the efficacy of BRE in teaching the redox reaction to high school students. Data from this mixed method study showed that students in the BRE condition outperformed students in the BAU condition in understanding the key concepts of redox reaction. For the first research question, the BRE is more effective in promoting students understanding of redox than completion of the lab experience only (BAU group). The BRE provides students opportunities to examine redox reactions at microscopic level, explaining the change of electrons between atoms and the mechanism of why and how electrons change, during the lab experience. For the second research question, the BRE supports students learning by providing a context, just as scientists conduct research to answer questions and explain phenomena. This research suggests that the SABLE SEG tested, which blends the virtual and the laboratory world enriched students' understanding of the redox phenomena.

Although we found significant learning gains associated with the novel SEG, limitations exist in our experimental design. First, due to time constraints and limited resources, our team did not conduct interviews with students in the BAU condition. As a result, we cannot compare the knowledge construction process that took place in the BAU condition to the treatment condition. Future research should include interviewing and observing students who participate in the BAU condition to deepen our understanding of gaps in conceptual knowledge. Subsequent research should also compare the scientific reasoning of students across groups to determine if practicing the scientific method and receiving feedback in real-time better equipped BRE students than BAU students who received feedback upon completion of the laboratory experience. We also envision future studies to investigate use of BRE environments across more lab experiences and students from varying backgrounds and age groups.

Finally, as our team continues to develop, refine, and assess engaging, immersive environments designed for learning, we struggle to name and describe the actual experiences. While it is accurate for our team to apply Marsh's 2011 definition of serious education games to the *Blended Reality Environment* examined in this paper, it does not fully capture or conceptualize the experience. Many other scholars (e.g. Connolly et al., 2012; Marsh, 2011; Ratan & Ritterfeld, 2009) have addressed the struggle to adequately articulate, define, and bound the phrase *serious educational game*. Within the BRE, players are *challenged* by the tasks found within gameplay as they must master concepts before moving forward in the gameplay. Students *play* through the deliberate use of the science practices of analyzing and interpreting data, hypothesizing, and predicting what will happen at the atomic level during the experiment, and the *purpose* of the gameplay is evident as it dovetails the real-world lab experience. As such, Salen and Zimmerman's (2003) definition of a serious educational game, “… a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome” (p. 80) and Prensky's (2003) six structural elements of games that aptly describe attributes of the BLE. As such, we acknowledge that Marsh's definition of serious gameplay does not provide a robust, detailed checklist of the attributes of the gaming environment, yet we chose to use Marsh's definition due to the way in which he conceptualized serious gameplay as an inclusive continuum.

In 2011, Marsh conceptualized an inclusive continuum to characterize the varying types of serious games including immersive environments, digital media, and simulations. Marsh's continuum is divided into three ranges, beginning with games for purpose. Games for purpose include traditional gameplay attributes of challenge, play, and fun, with the additional requirement of a specific *purpose* associated with the gameplay (e.g. America's Army and Full Spectrum Warrior). The second range of games defined by Marsh has fewer traditional gaming qualities due to the prioritization of conveying purpose within the gameplay. For example, Human Mixer, an environment that blends the real world and the virtual world to teach music theory or Fatworld, a game that focuses on the politics of nutrition are characterized by Marsh within this range. The narrowest end of this continuum consists of games with little or no traditional gaming attributes. Rather, these games aim to provide players exposure to cultural or emotional events to help enrich understanding of different perspectives in order to make meaning. For example, Scalable City, developed by Sheldon Brown's team engages players with a variety of landscapes, ranging from urban to rural that the player explores and disrupts to explore current cultural trends associated with these environments. Marsh (2011) asserts that “not all game characteristics, such as, challenge, fun and play are appropriate descriptions or labels for all serious games. The consequence of this is a fragmented and unconnected movement wherein, at best, factions or sub-groups begin to be formed and at worst, separate or detach themselves completely from the wider serious games community (p.67).

As we consider the BRE presented in this article, we situate the experience within the second range identified by Marsh due to the educative curricular design that guided the game development. Specifically, educators and scientists collaborated to create a storyline that clearly addresses the concept oxidation and reduction reaction and strategically integrates the science practices into the experience. In addition, players are required to show understanding of specific concepts before moving forward in the experience. We agree with Abt who explained that "We are concerned with *serious games* in the sense that these games have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement." (Abt, 1975, p. 9). When addressing how to define and characterize serious games, Breuer and Bente (2010) suggest conceptualizing gameplay attributes on "sticky notes" instead of "stone tablets" (p.18). This sentiment resonates with our research team as we endeavor to create immersive gameplay environments as pedagogically sound tools that maximize learning of difficult concepts that serve as gatekeepers to the STEM pipeline.

## Funding

This work was supported by the National Science Foundation through (1021RR246081), (1021RR246076), and (1021RR246080).

## Acknowledgements

The research team would like to thank Cogent Education for creating the blended reality tool that was used in the research. Cogent Education is a small business that was created at the University of Georgia to create and implement novel instructional tools in the science classroom.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.compedu.2018.03.003.

## References

Abt, C. C. (1975). *Serious games.* New York: Viking Compass.
Hodges, G. W., Oliver, J. S., Jeong, S., Cohen, A. S., Jang, Y., Robertson, T., et al. (2015). Leveraging technology to measure student learning during inquiry based virtual case studies in introductory biology. *Presented at the national association of research and science teaching, Chicago, Illinois on April 13, 2015*.
Barke, H. D., Hazari, A., & Yitbarek, S. (2009). *Misconceptions in chemistry: Addressing perceptions in chemistry.* Berlin: Springer-Verlag.
Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education, 22*(8), 797–817.
Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning, 19*(1–2), 205–220.
Bousbia, N., & Belamri, I. (2014). Which contribution does EDM provide to computer-based learning environments? In A. Pena-Ayala (Ed.). *Educational data mining: Applications and trends* (pp. 3–28). Switzerland: Springer.
Bowen, C. W. (1998). Item design considerations for computer-based testing of student learning in chemistry. *Journal of Chemical Education, 75*, 1172–1175.
Boyle, E., Connolly, T. M., & Hainey, T. (2011). The role of psychology in understanding the impact of computer games. *Entertainment Computing, 2*(2), 69–74.
Breuer, J., & Bente, G. (2010). *Why so serious?* On the relation of serious games and learning eludamos. *Journal for Computer Game Culture, 4*(1), 7–24.
Chung, G. K. (2015). Guidelines for the design and implementation of game telemetry for serious games analytics. *Serious games analytics* (pp. 59–79). Springer International Publishing.
Clark, D. B., Nelson, B. C., Chang, H. Y., Martinez-Garza, M., Slack, K., & D'Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education, 57*(3), 2178–2195.
Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching, 45*(3), 293–321.
Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*, 661–686.
De Jong, O., Acampo, J., & Verdonk, A. (1995). Problems in teaching the topic of redox Reactions: Actions and conceptions of chemistry teachers. *Journal of Research in Science Teaching, 32*(10), 1097–1110.
Ellington, H., Adinall, E., & Percival, F. (1982). *A handbook of game design.* London: Kogan.
Ezzy, D. (2013). *Qualitative analysis: Practice and innovation.* London: Routledge.
Hawkes, S. J. (2004). Chemistry is not a laboratory science. *Journal of Chemical Education, 81*, 1257.
Honey, M. A., & Hilton, M. H. (2011). *Learning science: Computer games, simulations, and education.* Washington, DC: The National Academies Press.
Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning, 19*(1–2), 221–240.
Jagodziński, P., & Wolski, R. (2015). Assessment of application technology of natural user interfaces in the creation of a virtual chemical laboratory. *Journal of Science Education and Technology, 24*(1), 16–28.
Johnstone, A. H. (1993). The development of chemistry teaching. *Journal of Chemical Education, 70*(9), 701–705.
Leite, L., & Afons, A. S. (2002). Prospective physical sciences teachers' use of laboratory activities: An analysis of its evolution due to a science teaching methods course. *Rev Elect Ensen Ciencias, 1*, 153–179.
Lenhart, A., Kahne, J., Middaugh, E., Macgill, E. R., Evans, C., & Vitak, J. (2008). *Teens, video games, and civics.* Washington, DC: Pew Internet & American Life Project.
Levine, M., Serio, N., Radaram, B., Chaudhuri, S., & Talbert, W. (2015). Addressing the STEM gender gap by designing and implementing an educational outreach chemistry camp for middle school girls. *Journal of Chemical Education, 92*(10), 1639–1644.
Linn, M. C. (2000). Designing the knowledge integration environment. *International Journal of Science Education, 22*(8), 781–796.
Linn, M. C., & Eylon, B.-S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration.* New York, NY: Routledge.
Linn, M. C., Lee, H. S., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science, 313*(5790), 1049–1050.
Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious games analytics: Theoretical framework. *Serious games analytics* (pp. 3–29). Springer International Publishing.
Marsh, T. (2011). Serious games continuum: Between games for purpose and experiential environments for purpose. *Entertainment Computing, 2*(2), 61–68.
McKeachie, W. J. (1999). *Teaching Tips: Strategies, research, and theory for college and university teachers* (10th ed.). Boston: Houghton-Mifflin (Chapter 12).
National Research Council (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Washington, DC: The National Academies

Press.
Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist, 50*(4), 258–283.
Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment, 1*, 21.
Ratan, R., & Ritterfeld, U. (2009). Classifying serious games. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.). *Serious games: Mechanisms and effects*. New York/London: Routledge.
Robertson, T., & Ducrest, D. (2016). Retrieved from http://www.cogenteducation.com/products.
Salen, K., & Zimmerman, E. (2003). *Rules of Play: Game design fundamentals*. Cambridge: The MIT Press.
Simpkins, S. D., Price, C. D., & Garcia, K. (2015). Parental support and high school students' motivation in biology, chemistry, and physics: Understanding differences among Latino and Caucasian boys and girls. *Journal of Research in Science Teaching, 52*(10), 1386–1407.
Shute, V. J., & Ke, F. (2012). Games, Learning, and Assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.). *Assessment in Game-Based Learning*. New York, NY: Springer.
Squire, K. (2011). Video games and learning. *Teaching and participatory culture in the digital age*New York, NY: Teachers College Print Cerca con Google.
Steinkuehler, C., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology, 17*(6), 530–543.
Tezcan, H., & Bilgin, E. (2004). Affects of laboratory method and other factors on the student success in the teaching of the solvation subject at the high schools. *Gazi University Journal of Gazi Educational Faculty, 24*, 175–191.
U.S. Department of Education. (2010, November). Transforming American education: Learning powered by technology. Retrieved from http://www.ed.gov/sites/default/files/netp2010-execsumm.pdf.
Xie, Q., & Pallant, A. (2011). The molecular workbench software: An innovative dynamic modeling tool for nanoscience education. In M. S. Khine, & I. M. Saleh (Eds.). *Models and modeling: Cognitive tools for scientific enquiry* (pp. 121–132). New York: Springer.
Zembal-Saul, C., McNeill, K. L., & Hershberger, K. (2013). *What's your evidence?: Engaging K-5 children in constructing explanations in science*. Pearson Higher Education.