Published in final edited form as:

J Sci Educ Technol. 2018 June ; 27(3): 236–247. doi:10.1007/s10956-017-9720-2.

Implementation of a Curriculum-Integrated Computer Game for Introducing Scientific Argumentation

Robert C. Wallon, Chandana Jasti, Hillary Z. G. Lauren, and Barbara Hug

University of Illinois at Urbana-Champaign, College of Education, 1310 South 6th St, Champaign, IL 61820

Abstract

Argumentation has been emphasized in recent U.S. science education reform efforts (NGSS Lead States 2013; NRC 2012), and while existing studies have investigated approaches to introducing and supporting argumentation (e.g., McNeill & Krajcik 2008; Kang, Thompson & Windschitl 2014), few studies have investigated how game-based approaches may be used to introduce argumentation to students. In this paper, we report findings from a design-based study of a teacher's use of a computer game intended to introduce the Claim, Evidence, Reasoning (CER) framework (McNeill & Krajcik 2012) for scientific argumentation. We studied the implementation of the game over two iterations of development in a high school biology teacher's classes. The results of this study include aspects of enactment of the activities and student argument scores. We found the teacher used the game in aspects of explicit instruction of argumentation during both iterations, although the ways in which the game was used differed. Also, students' scores in the second iteration were significantly higher than the first iteration. These findings support the notion that students can learn argumentation through a game, especially when used in conjunction with explicit instruction and support in student materials. These findings also highlight the importance of analyzing classroom implementation in studies of game-based learning.

Keywords

game-based learning; scientific argumentation; curriculum

Introduction

Current reform efforts in U.S. science education promote learning that involves three dimensions: disciplinary core ideas, crosscutting concepts, and science and engineering practices. The *Next Generation Science Standards* (NGSS; NGSS Lead States 2013) and *A Framework for K–12 Science Education* (hereafter *Framework*; National Research Council [NRC] 2012), on which the NGSS are based, highlight specific practices such as engaging in scientific argumentation. This emphasis on argumentation has been a growing trend in science education. The *National Science Education Standards* called for a shift in emphasis from teaching “science as exploration and experiment” to teaching “science as argument and explanation” (NRC 1996, p. 113). Furthermore, members of the science education research

community have advocated teaching argumentation in science classes for well over a decade before the NGSS (e.g., Kuhn 1993; Newton, Driver, & Osborne 1999). Thus, longstanding calls to teach argumentation make research of this practice broadly relevant to science educators.

Concurrently with this emphasis on argumentation, there has been growing interest in exploring possibilities for using computer games to achieve a variety of science learning goals (NRC 2011). However, most of the existing research on game-based science learning has focused on goals related to learning scientific content (Li & Tsai 2013), rather than goals related to learning scientific practices. This paper addresses calls to teach science as argumentation while further exploring possibilities for using game-based approaches to teaching science.

Review of Related Research

Our study is informed by previous research on supporting scientific argumentation and game-based science learning. In this section, we briefly describe key literature in each of these areas before presenting our study that is at the intersection of both areas.

First, it is important to clarify our use of the term “scientific argumentation” and its distinction from “scientific explanation.” Both “engaging in argument from evidence” and “constructing explanations” are practices described in the *Framework* and incorporated into the NGSS. While these practices complement one another, some scholars have disagreed about the degree to which the practices of explanation and argumentation should be separated in classrooms (Berland & McNeill 2012; Osborne & Patterson 2011; Osborne & Patterson 2012). For clarity, we use the term scientific argumentation to describe activities in which students do three things: (a) state a claim that answers a question, (b) provide evidence to support their claim, and (c) give their reasoning by using scientific principles to show how their evidence supports their claim. This operationalization of argumentation is based on the claim, evidence, reasoning (CER) framework (McNeill & Krajcik 2012), which was adapted from Toulmin’s (1958) model of argumentation in order to be more accessible to teachers. Because of the variety of terms used in previous research and the overlap between these two practices, our review of the literature includes research on both scientific explanation and argumentation, although we have decided to use the term argumentation within our study. To further clarify our use of the term argumentation, it is important to note that while some previous studies have used the term to refer to discursive engagement in the practice (e.g., Kuhn 2010), we focus on written arguments as products of this argumentation.

Supporting Scientific Argumentation—The practice of scientific argumentation has become widely advocated as a central goal of science education (Driver, Newton, & Osborne 2000; Kuhn 2010; NRC 2012). To address this goal, educational researchers have investigated how to support students in developing argumentation skills through effective instruction and the design of learning environments that support student engagement in the practice (e.g., Berland & Reiser 2009; McNeill & Krajcik 2008; Osborne, Erduran, & Simon 2004; Sandoval & Reiser 2004).

One aspect of research on supporting argumentation focuses on the effects of instruction on student learning of the practice. The quality of student arguments has been shown to improve with explicit instruction in scientific argumentation (Cetin 2014; Khishfe 2014; Nussbaum, Sinatra, & Poliquin 2008; Zohar & Nemet 2002). Studies of specific instructional practices have indicated that students' argumentation skills improve when teachers define the components of an argument, model the practice, and provide a rationale for the practice (Berland & Reiser 2009; Lizotte, McNeill, & Krajcik 2004; McNeill & Krajcik 2008).

Researchers have also studied features of curriculum materials and learning environments that support the development of argumentation skills. For example, Berland and McNeill (2010) recommended beginning with simpler instructional contexts, such as initially providing students with limited and relevant data. With regards to scaffolding argumentation, Kang, Thompson, and Windschitl (2014) reviewed five types of scaffolds embedded within assessment prompts, which included (a) using contextualized phenomena, (b) checklists, (c) rubrics, (d) sentence frames, and (e) prompting students to draw explanatory models along with written explanations. They concluded that, while all five scaffold types were positively linked with the quality of students' arguments, the most impactful factor was the use of contextualized phenomena (p. 674).

In addition to these general features that support argumentation, computer-based epistemic tools have been shown to support students in constructing and evaluating scientific arguments and explanations. For example, *SenseMaker*, a digital tool for graphically representing arguments, helped students construct and evaluate arguments by modeling through expert examples, visualizing claims and groups of evidence, and fostering communication with other students (Bell & Linn 2000). *ExplanationConstructor*, a tool that resembled an electronic journal, provided conceptual scaffolds to guide students' inquiry processes as they organized investigations and constructed explanations (Sandoval & Reiser 2004). Digital learning environments such as *SenseMaker* and *ExplanationConstructor* have shown value in making processes of argumentation explicit to students. In the next section, we discuss computer games as digital learning environments that have also gained the interest of science education researchers because of their ability to provide contextualized and immersive learning.

Digital Game-Based Science Learning—A recent synthesis of research on digital gaming in science education indicated that computer games “have potential to advance multiple science learning goals” (NRC 2011, p. 15). Despite this potential, there has been relatively limited research on game-based approaches in science education, and consequently there is inconclusive evidence for extent to which game-based approaches can influence science learning (NRC 2011). Another limitation of what is known about game-based approaches to science learning is that the majority of studies currently making up the research base have focused on student learning of science concepts and content knowledge, and they have often overlooked research on other science learning goals (Li & Tsai 2013).

From the limited group of gaming studies that have researched learning other than content knowledge, there are several notable examples. One example is *River City* (Ketelhut &

Nelson 2010), which focused on inquiry skills and experimentation. Additional examples include *Quest Atlantis* (Barab, Sadler, Heiselt, Hickey, & Zuiker 2007), *Mad City Mystery* (Squire & Jan 2007), and *Citizen Science* (Gaydos & Squire 2012), all of which included aspects of argumentation. Ketelhut and Nelson (2010) found that students who learned experimentation in the multi-user virtual environment *River City* performed as well as students who learned experimentation in a physical environment on a test that measured understanding of scientific inquiry. Barab and colleagues (2007) reported that students who used *Quest Atlantis* were engaged in argumentation, but the study did not include a systematic analysis of written arguments from individual students. Squire and Jan (2007) used a cross-case comparison to show how students in various grades engaged in argumentation while playing the game *Mad City Mystery*. Their study showed that gaming environments could provide a motivating context for argumentation, and they suggested that future research should investigate how such learning environments could be used to explicitly teach argumentation. Subsequently, Gaydos and Squire (2012) described the explicit incorporation of argumentation in *Citizen Science* with an in-game argument constructor. Although Gaydos and Squire (2012) did not evaluate gains in student learning or quality of arguments, they suggested that games may be useful for introducing foundational concepts within formal learning environments, and the most effective learning may occur in conjunction with curricular activities that reinforce in-game tasks. The present study builds on findings from research on game-based learning and instruction of scientific argumentation by investigating the scaffolds that are useful for introducing the scientific practice of argumentation to students in a game based learning environment.

Research Questions

The present study involved two iterations of enactment of curriculum materials featuring a computer game by a teacher as she introduced scientific argumentation to her students. This study addresses two research questions: (a) How does a high school science teacher use a computer game as part of an introduction to scientific argumentation?, and (b) How do the two iterations compare in terms of the quality of student arguments?

The Present Study

This study took place as part of a larger design-based research (DBR) project focused on developing a computer game integrated with a curriculum unit about the neuroscience of traumatic brain injury (TBI). The project used a design-based research approach (Design-Based Research Collective 2003; Edelson 2002; Penuel, Fishman, Cheng, & Sabelli 2011), which included key features of a DBR approach, including implementation of materials in authentic classroom settings and iterative refinement of materials. Collins, Joseph, and Bielaczyc (2004) have recommended that reports of DBR include five categories: (a) Goals and elements of the design, (b) Settings where implemented, (c) Description of each phase of the study, (d) Outcomes found, and (e) Lessons learned. These categories are used to structure the remaining sections of this paper.

Goals and Elements of the Design

The curriculum was designed to provide a context for science teachers to introduce students to the structure of arguments and to provide students with a context for developing a high quality argument. These goals were supported with lesson plans and scaffolds, both internal and external to the computer game. The design is elaborated on in the rest of this section, which discusses the game, the lesson plan, and the student materials.

Design Elements of The Golden Hour Game—The focus of this study is *The Golden Hour*, a computer game integrated into Lessons 1 and 3 of the high school curriculum unit *Why Dread a Bump on the Head?* The unit and game were developed through an iterative design process (Krajcik, McNeill, & Reiser 2008) that included feedback from scientists, medical professionals, teachers, students, and observations from enactments in authentic classroom settings. The curriculum materials and game are published for educators online at the Project NEURON web site at <http://neuron.illinois.edu>.

Why Dread a Bump on the Head? is a unit with seven lessons through which students learn about the neuroscience of TBI. In Lesson 1, students read case studies to learn about different causes and symptoms of TBIs, followed by conducting sheep brain dissections to explore brain anatomy and physiology in Lesson 2. Next, students analyze computed tomography (CT) scans to identify different types of brain injuries in Lesson 3. Lessons 4 and 5 focus on the cellular level as students learn about neurons and investigate apoptosis and necrosis of neurons after a TBI. In Lessons 6 and 7, students research TBI using publicly available data and communicate their findings to the general public. For each lesson, teachers have access to student materials and a detailed lesson plan to guide instruction.

The Golden Hour is an example of what we are calling a *curriculum-integrated computer game*. We use this term to refer to a computer game that is integral to achieving learning outcomes of a curriculum unit (rather than tangential use of a game to increase motivation, review concepts, etc.). In the context of *The Golden Hour*, a key component of teaching scientific argumentation relies on students playing the game. As described below, features in the game were designed to support students in creating scientific arguments by (a) creating an instructional context through a real-world scenario (Kang, Thompson, & Windschitl 2014; Rivet & Krajcik 2008), (b) providing a rationale (McNeill & Krajcik 2008), and (c) modeling argumentation (Lizotte, McNeill, & Krajcik 2004), which have been shown to be effective instructional strategies.

Contextualization of student learning occurs through the game's settings, characters, tools, and tasks performed by students, which are realistic and plausible within the medical context of TBI. Students play the role of an advanced medical student to help diagnose and treat a patient who has sustained a TBI. The game is comprised of three scenes: (Scene 1) assessing the patient's initial condition on an Emergency Medical Services helicopter, (Scene 2) using a CT scan to examine and diagnose the injury, and (Scene 3) performing neurosurgery to treat the patient. Throughout each scene, students gather data and learn neuroscience-related concepts in order to perform necessary tasks such as treating and assessing the patient with in-game medical tools and devices.

Furthermore, the game provides an implicit rationale for argumentation in two ways. First, it embodies a situation in which the student's argument for treatment could affect the patient's quality of life after TBI. Second, in-game characters stress to the player that he or she is being evaluated for possible acceptance into the medical team based on performance and reasoning skills. Thus, the game provides students with a purpose for constructing a quality argument.

Finally, the game implicitly models scientific argumentation at the end of every scene when students report and interpret their findings by engaging with the lead physician in a dialogue based on the CER framework (Figure 1). Beginning with the question "What should be done next for the patient?" the lead physician prompts students to first make a recommendation (i.e., a claim), then support it with evidence and reasoning. Students answer the physician's step-by-step questions by selecting from multiple choice responses. If a student selects an incorrect response, the physician provides immediate feedback by stating why the answer was incorrect and then asks the student to make another choice. For reference, we provide a flowchart of the dialogue between the physician and student in Scene 1, including the dialogue prompts, possible answers, and feedback (see Online Resource 1). After completing the interactive dialogue with the lead physician, students write a scientific argument in the form of a medical recommendation for the patient's treatment.

Design Elements of the Lesson Plan—These design elements within the game are further supported by the accompanying lesson plan for teachers. Although the lead physician character in the game uses the words "claim," "evidence," and "reasoning" within the dialogue to guide students in developing their medical recommendation or argument, the game does not explicitly define CER components, use the term "scientific argument," nor provide explicit instructions for students on how to construct an argument. The decision to integrate scientific argumentation implicitly within the game was part of a conscious effort to balance the "instructional" and "engaging" aspects of an educational game. Therefore, in order to support teachers in their instruction, the accompanying lesson plan highlights the scientific practice of engaging in argument from evidence, briefly describes the CER framework, and references McNeill and Krajcik's (2012) book as a resource for instructing students on the CER framework.

Design Elements of the Student Materials—Student materials external to *The Golden Hour* computer game also play a role in supporting argumentation. In the first iteration of enactment, students were asked to write their medical recommendation on a sheet of paper following a prompt from the in-game physician of "Can you add a summary of our recommendation, with evidence and reasoning, to the report in the tablet?" In the second iteration, students received a sheet with scaffolds to provide additional support. This sheet included (a) a prompt that instructed students to answer a specific question ("What should be done next for the patient?") using a claim, evidence, and reasoning, and (b) a table to organize the process of constructing an argument using the CER framework. The table contained three columns, one for each of the CER components, and an additional space for a "Medical Recommendation" at the bottom to combine the three components. The sheet instructed students to use the three columns to organize their information and the "Medical

Recommendation” space to write their full argument using complete sentences. Figure 2 shows an example of a student argument from each iteration of enactment.

Setting

We selected the teacher for this case study because she (a) attended professional development workshops for the curriculum unit, (b) was the first teacher to use the computer game with the unit, and (c) used the curriculum materials during two consecutive years with two iterations of the materials. The teacher was an experienced educator, having taught for more than ten years.

The study was conducted at a public high school located in a small urban community in the Midwest. The race/ethnicity composition of students at the school was approximately 50% white, 30% black, 10% Asian, 5% Hispanic, and less than 5% in any other category. Approximately 50% of the school’s students were identified as low-income.

The teacher enacted the first three lessons of the unit and the game in an Anatomy and Physiology elective course that enrolled mostly grade 11 and grade 12 students. Due to teaching assignments set by the school, the teacher had 115 students total (46 males and 69 females) from five class sections during the first iteration, and she had 84 students total (27 males and 57 females) from four class sections during the second iteration. In both iterations, all of the students enrolled in the teacher’s Anatomy and Physiology classes were invited to participate in the research. However, only students who assented and had parental consent were included in the study.

Description of Study Phases

During Iteration One and Iteration Two of the study, data were collected from three sources: (a) observations and audio recordings of classes, (b) documents with written arguments from individual students, and (c) post-test documents from individual students. One researcher observed the classes and took field notes, and audio was recorded with a digital recorder that the teacher wore around her neck. Scene 1 of the game was implemented over two class sessions for all class sections in Iterations One and Two. Initial observations and audio recordings indicated that the teacher was consistent in her instruction for all class sections throughout the day. Therefore, on subsequent days, observations and recordings were only collected and transcribed for a subset of class sections, with at least one observation and recording collected each day. Transcriptions were performed using Express Scribe software. The exact numbers of data in these subsets are detailed in Table 1.

In this section we address the first research question, “How does a high school science teacher use a computer game as part of an introduction to scientific argumentation?” We first describe key stages of instruction during each iteration, and then we compare differences between iterations. These stages were identified based on effective instruction of argumentation, summarized in the review of research earlier in this paper.

Iteration One—An overview of the first iteration is shown in Figure 3. The instruction during this iteration included five key stages:(1) Playing Scene 1 of the game, (2) Defining

CER, (3) Providing a rationale for an argument, (4) Modeling an argument, and (5) Writing arguments.

(1) Playing Scene 1 of the game: The teacher started the instructional sequence by having students play Scene 1 of *The Golden Hour*. When students played the game, they worked in pairs, and each pair played the game on a small laptop computer with a mouse; meanwhile, the teacher walked around the classroom and had discussions with the pairs of students. During this iteration, playing the game was the initial anchor for instruction.

(2) Defining CER: The teacher explained the CER framework after students played Scene 1 of the game and before they wrote their medical recommendations. She provided verbal explanations while projecting supporting written items (e.g., claim, evidence, reasoning and their definitions) onto a screen via a PowerPoint presentation. When defining the C, E, R components in Iteration One, the teacher initially did not explicitly say “claim” and began by telling students that they needed to write their recommendation to answer the question, “What is the best treatment for the patient?” When presenting an example CER argument a few minutes later in instruction, she further defined a claim as “a statement or a conclusion that answers the original question we have.”

The teacher defined evidence as “the scientific data that supports your recommendation” and told students that “the evidence needs to be appropriate for your claim.” The sufficiency aspect of evidence was not as explicitly addressed as the appropriate aspect, but the teacher did mention to “include enough data so you have enough information to really convince someone.” The teacher further explained evidence within the context of the game by saying, “the evidence would be the data that you collected about the patient’s condition, his vital signs and so on.”

Reasoning was explained to students in Iteration One as follows: “How do you connect your claim with the data you’ve collected? In your reasoning statement, you’re going to tell why your data count as evidence to support that claim. So, it’s a process of applying the scientific knowledge to solve that problem of what treatment is needed for [the patient].”

These descriptions of classroom instruction show that the teacher used the context of the game to exemplify CER components.

(3) Providing a rationale for an argument: In Iteration One, the teacher provided a rationale for writing scientific arguments by telling students that their task in writing the medical recommendation was to “convince someone that [the student’s] suggested treatment would be the best treatment for the individual.” During this stage, the computer game was also specifically referenced for context.

(4) Modeling an argument: In Iteration One, the teacher projected an example argument on a screen while providing a verbal explanation to students. The argument was about how polar bears can live in the arctic because of specific adaptations that allow them to survive in the harsh environment. This specific example can be found in McNeill and Krajcik’s (2012) book, which was provided as a reference to participants of the professional development

workshop that the teacher had attended prior to enacting the curriculum. This stage did not make connections back to the game.

(5) Writing arguments: The last stage of Iteration One involved students individually writing their arguments to address the question, “What should be done next for the patient?” Students were able to refer back to notes they took while playing Scene 1 of the game, including any data about the patient that they recorded. During this stage the previous game play provided context for students’ arguments.

Upon noticing that many of her students were struggling to write an argument, the teacher suggested to the researchers that an organizer should be provided to students to remind them of the three parts of an argument that were discussed: a claim, evidence, and reasoning. This was one of the main changes implemented in the next cycle of implementation.

Iteration Two—An overview of the second iteration is shown in Figure 4. The instruction during this iteration included five key stages: (1) Defining CER, (2) Providing a rubric, (3) Modeling an argument, (4) Playing Scene 1 of the game, and (5) Writing arguments.

(1) Defining CER: In Iteration Two, the teacher began by stating that students will need to answer the question “What should be done next for the patient?” using a claim, evidence, and reasoning. Verbally, she described a claim as “a concise statement that relates to the questions.” She included a handout, which defined a claim as “a conclusion about a problem.”

For the evidence component, the teacher asked students to reflect on and restate the data that they would collect from the patient. The handout defined evidence as “scientific data that is appropriate and sufficient to support the claim.”

Finally, the teacher explained reasoning by saying, “You’re going to show how that evidence answers that claim.” The handout defined reasoning as “a justification that shows why the data count as evidence to support the claim and includes appropriate scientific principles.”

During Iteration Two some of the CER components were discussed in terms of the game, but these connections were broad because students had not played the game yet at this stage.

(2) Providing a rubric: During the second stage of Iteration Two, the teacher provided students with a rubric, which further clarified that a claim is a “concise statement” that “relates directly to the question and hypothesis” and “focuses on only the most important features of the experiment.” The rubric stated the evidence should include “several data sources used to explain the claim, including observations and accurate measurements” and “clear connections to the question and hypothesis.”

The rubric did not refer to the game and instead referred to an experimental context, stating that good reasoning “illustrates understanding of how the experiment fits into the ‘big picture,’” and that it “incorporates background knowledge, and makes connections to science concepts studied in class, to draw conclusions about experiment.” The teacher encouraged students to review the handout and rubric for more information on how to write

their arguments. It is noteworthy that the game takes place in a medical context rather than an experimental context, and therefore some of the guidelines on the rubric did not directly align with the game.

(3) Modeling an argument: In Iteration Two, the teacher gave students a written example of an argument that discussed an experiment in which the results indicated a positive relationship between pumpkin plant growth rate and temperature. This example was not related to the game.

(4) Playing Scene 1 of the game: The teacher continued the instructional sequence by having students play Scene 1 of The Golden Hour. As was the case during Iteration One, students worked in pairs, and each pair played the game on a small laptop computer with a mouse. The teacher circulated around the classroom and had discussions with the pairs of students while they worked on playing the game.

(5) Writing arguments: During the last stage of Iteration Two, students individually wrote their arguments as they had in Iteration One to address the question, “What should be done next for the patient?” However, students were provided with an organizer to help them structure the claim, evidence, and reasoning components of their arguments. See Figure 2 for a comparison of a student response with this organizer from Iteration Two to a student response without the organizer from Iteration One. Once again, the game play provided a context for student arguments.

Comparison of Iteration One and Iteration Two—Both iterations involved five key stages. However, not all five stages were the same, and the order in which they were implemented differed. Both iterations included the stages of (a) Defining CER, (b) Modeling an argument, (c) Playing Scene 1 of the game, and (d) Writing arguments. Iteration One included the stage of explicitly providing a rationale for an argument, which was not included in Iteration Two. Iteration Two included the stage of providing a rubric, which was not included in Iteration One. One main difference in the order of stages in the iterations was the timing of game play. In Iteration One, students played the game *before* receiving explicit instruction about argumentation. During Iteration Two, however, students played the game *after* explicitly discussing argumentation.

Although both iterations included the stages of defining CER and modeling an argument, these stages were enacted differently. One main difference was that during Iteration Two, many more paper supports were provided to students. Definitions of CER components and an example argument were provided to students during Iteration Two, while these stages were discussed verbally during Iteration One. More written support was also provided during Iteration Two in the form of the paper rubric.

An additional main difference between iterations was the organizer that was provided to support students with writing their arguments during Iteration Two. As shown in Figure 2, students were prompted to consider individual CER components before synthesizing them into an argument for what should be done next for the patient.

In summary, the main differences in implementation during two iterations were timing of game play and the amount of supports for students outside of the game. The game was played at the beginning of the instructional sequence during Iteration One, and it was played near the end of the instructional sequence during Iteration Two. The difference in timing of game play seemed to affect the extent to which the game could be used to contextualize other stages such as defining CER. During the second iteration more supports were provided to students, including supports presented in revisions to the curriculum such as the organizer and also including supports that were supplemented by the teacher such as the written definitions of CER components, the written example argument, and the rubric.

Outcomes: In this section, we address research question 2: How do the two iterations compare in terms of the quality of student arguments? Written arguments were collected from individual students who had parental consent and who assented to participate in the research. This resulted in 49 arguments that were collected from 5 class sections in the first iteration and 39 arguments that were collected from 4 class sections in the second iteration. Student argument data consisted of students' medical recommendations from Scene 1 of *The Golden Hour* where students measured the patient's vital signs and consciousness and used these data as evidence to determine the next step in medical treatment of the patient.

Students' arguments were scored using a rubric (see Online Resource 2) that was adapted for this specific argumentation task based on a rubric from McNeill and Krajcik (2012). The rubric was used to generate a score for each of the individual Claim, Evidence, and Reasoning components of the argument. The scores on each component were summed to produce a total score, with a maximum of ten points. A claim was evaluated on accuracy and completeness on a scale of zero to two points. Evidence was scored for appropriateness (how well it supported the claim) and sufficiency (detail and amount of evidence) on a scale of zero to four points. Reasoning was scored for appropriateness (detail and accuracy of justification) and sufficiency (how well it connected all pieces of evidence) on a scale of zero to four points. Scores in a given category should be used to compare within the category rather than across categories (e.g., comparing one claim to another rather than comparing a claim to evidence).

All student arguments were scored independently by two researchers, and inter-rater agreement was calculated using Cohen's kappa, which differs from directly reporting percentage agreement because kappa factors out agreement due to chance (Cohen 1960). A kappa value of 0 indicates agreement no better than chance, and a kappa value of 1 indicates perfect agreement (Cohen 1960). While evaluating kappa values is somewhat arbitrary, Fleiss, Levin, and Paik (2003) suggest values higher than 0.75 indicate excellent agreement, values between 0.40 and 0.75 indicate moderate agreement, and values less than 0.40 indicate poor agreement (p. 604). Using these guidelines, the calculated kappa values of 0.90 for claim, 0.74 for evidence, and 0.74 for reasoning indicate substantial inter-rater agreement above chance. Additionally, all discrepancies were resolved by discussion.

Means for the student arguments (individual Claim, Evidence, and Reasoning scores and total CER scores) and post-test scores were each analyzed by iteration group with independent samples *t*-tests. Caution should be taken to avoid generalizing the results

beyond these specific cases because students were nested within class periods. All other assumptions were checked, and in cases when Levene's test for equality of variances was violated, Welch's *t*-test statistic was used instead. Effect size was calculated with Cohen's *d*; a value of 0.3 or less was considered a "small" effect, around 0.5 was considered a "moderate" effect, and 0.8 or greater was considered a "large" effect (Cohen 1988). All analyses were performed in SPSS 22 (IBM, Inc.) except Cohen's *d*, which was calculated using an online tool (Becker 1998).

The average total CER score on students' arguments between iterations was significantly different with a large effect size ($t(86) = 4.23, p < 0.001; d = 0.92$). Out of a maximum of ten points, students from Iteration Two ($M = 7.15, SD = 2.60$) scored higher than those in Iteration One ($M = 4.39, SD = 3.36$). The distribution of scores in both iterations was spread across the whole range, from zero to ten, but the distribution in Iteration Two was mainly concentrated in the higher scores (Figure 5).

Students in Iteration Two also scored significantly higher for each of the CER components than students in Iteration One. With a maximum of two points for the Claim component, the average score for students in Iteration Two was significantly higher with a moderate effect size ($t(86) = 3.13, p = 0.002; d = 0.65$) than the average scores in Iteration One. With a maximum of four points for the Evidence component, the average score for students in Iteration Two was also significantly higher with a moderate effect size ($t(86) = 3.42, p = 0.001; d = 0.72$) than those for students in Iteration One. Out of a maximum of four points for the Reasoning component, the average score for students in Iteration Two was also significantly higher with a large effect size ($t(86) = 4.20, p < 0.001; d = 0.90$) than those for students in Iteration One. Descriptive statistics are reported in Table 2.

We became curious whether students who participated in Iteration Two systematically outperformed students who participated in Iteration One, and therefore we compared scores on an additional data source collected from individual students at the end of the curriculum unit, a post-test. Note that the unit post-test focused on content knowledge and did not involve an argument task. The post-test was developed as a unit assessment included with the curriculum materials, and it was iteratively refined with feedback from scientists, teachers, and science education experts to ensure validity. Because post-test data were not available for all students who completed a written scientific argument due to absences, we only considered post-test scores from the subset of students who completed both the scientific argument task and the post-test, which included 32 students in the first year and 31 students in the second year. The researcher-designed post-test (see Online Resource 3) measured science content knowledge covered by the unit and consisted of 5 questions with a maximum score of 15 points. The test was scored by researchers using an answer key, and scores were scaled as percentages.

The average post-test score for students in Iteration One ($M = 52.50\%, SD = 11.55\%$) was not significantly different ($t(61) = 0.74, p = 0.46$) than the average score for students in Iteration Two ($M = 50.22\%, SD = 12.91\%$).

Lessons Learned

The purpose of this study was to explore how a high school teacher used a curriculum-integrated computer game to introduce scientific argumentation to students. Current research has demonstrated the possibility for game-based approaches to foster student participation in broader science learning goals such as argumentation (e.g., Squire & Jan 2007; Gaydos & Squire 2012). However, studies have not specifically looked at how teachers implement a curriculum-integrated game to introduce argumentation, nor students' ability to construct arguments from using such an approach. Our findings highlight implications for curriculum and game design and inform the need to support teachers and students with argumentation in science classrooms.

Our initial design conjecture for *The Golden Hour* learning environment was that contextualization provided by the game would help students write high quality arguments. However, the outcome of student argument scores showed that the contextualization of the game was not sufficient to support students with writing high quality arguments during Iteration One. Students who participated in Iteration Two were able to write higher quality arguments, and they were supported with multiple scaffolds. Our interpretation is that the scaffolds supported students with writing better arguments because students who participated in Iteration Two did not simply systematically outperform students in Iteration One, as evidenced by no difference in the comparison of post-test scores of students in Iteration One and Iteration Two. This supports the idea that a game-based approach can help introduce students to argumentation when it is used in conjunction with supports external to the game. Curriculum developers and game designers should plan for these external supports that make game play more instructionally valuable for teachers and students.

Existing literature emphasizes the vital role of teachers in implementing educational games and the importance of considering teacher instruction and authentic classroom settings as components of educational game design (Sadler, Romine, Menon, Ferdig, & Annetta 2015; Watson, Mong, & Harris 2011; Webb, Bunch, & Wallace 2015). The teacher involved in this study used the game during an introduction to argumentation in ways that extended beyond having students play the game and using the game as a context for writing arguments. The game was also used to illustrate aspects of explicit instruction about argumentation such as defining CER components. This highlights the need to further study how teachers can use games to support instruction, other than obvious occasions such as when students play the game.

Iteration One involved a stage of discussion a rationale for argumentation, while Iteration Two did not include this stage. One possibility is that a game-based approach that is highly contextualized provides a tacit rationale for argumentation, which makes it less important for this stage to be explicitly addressed by the teacher.

Another lesson learned from this study is that there are various ways in which a computer game may be used to introduce students to argumentation. Two possibilities are that a game may be used to start an instructional sequence, as was the case during Iteration One, or a game may be used to conclude an instructional sequence, as was the case during Iteration

Two. This highlights the need for further research into the efficacy of game-based approaches that investigate differences in implementation rather than simply comparing a game-based approach to a non-game-based approach.

Limitations of the study

Caution should be taken to avoid interpreting causal relationships between factors explored in the study. Although the addition of the written scaffolds in Iteration Two was expected to support and improve student arguments, other factors may have influenced student performance. For instance, the rubric provided by the teacher in Iteration Two may have helped students create stronger arguments. Other differences in implementation included the timing of the CER instruction (after playing Scene 1 versus before playing the game), the content of the CER examples (polar bear adaptations versus plant experiment results), and how the CER examples were provided to students (projected on a screen versus on a physical handout). Any of these differences in implementation should not be viewed as directly causing changes in student performance.

Other limitations of this study include the measures of learning. First, this intervention was intended to introduce argumentation to students, and we assumed that all students had no previous experience constructing scientific arguments based on information provided by the teacher. Having a baseline of students' initial abilities to make scientific arguments would be a good way to check this assumption. However, measures of learning scientific argumentation are problematic in that performing a scientific practice cannot be devoid of content; the extent of students' content knowledge on the topic of a particular argument may confound an instrument's precision. Therefore, it may be better to analyze relative performance on argumentation tasks unrelated to the content of the intervention in future studies.

Future Research

Research suggests that high-quality curriculum (Eastwood & Sadler 2013), activities external to the game (Gaydos & Squire 2012), and highly prepared teachers (Sadler et al. 2015) are factors that may positively influence game-based learning in authentic classroom environments. In our own study, we observed the inter-relatedness of factors that could affect learning, including the teacher, curricula, game, and students. Future research is needed to determine the relative impact of these components on learning and how and when game-based learning is an effective practice to choose among other instructional practices.

Conclusions

This study highlights integral connections between games, curriculum development, teacher instruction, and student learning that are important to consider when designing game-based learning environments. In addition, this study supports existing literature that emphasizes the importance of supporting teachers in their instruction and integration of games in curriculum. Furthermore, with the continued emphasis on practices in science education, cultivating teacher instruction that effectively supports scientific practices will likely take repeated and long-term efforts through various professional development opportunities.

Future research should continue to investigate the potential of curriculum-integrated computer games for supporting students in argumentation and should extend to investigating game-based approaches that integrate additional scientific practices.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research and curriculum materials described in this publication were supported by Office of the Director, National Institutes of Health (NIH) under Award Number R25OD011144 and a supplement to the parent grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH). We thank everyone—Project NEURON members and others—who developed, tested, and offered feedback on the *Why Dread a Bump on the Head?* curriculum unit and *The Golden Hour* game. We realize this study would not be possible without teachers who are willing to test our materials, and we are extremely grateful to the teacher in this study for welcoming us into her classroom. We are also grateful to Claire Scavuzzo and Emily Serblin, for assistance with data transcription and analysis.

References

- Barab SA, Sadler TD, Heiselt C, Hickey D, Zuiker S. Relating narrative, inquiry, and inscriptions: Supporting consequential play. *Journal of Science Education and Technology*. 2007; 16(1):59–82.
- Becker, LA. Effect size calculators. University of Colorado Colorado Springs. 1998. Retrieved from <http://www.uccs.edu/~lbecker/>
- Bell P, Linn MC. Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education*. 2000; 22(8):797–817.
- Berland LK, McNeill KL. A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*. 2010; 94(5):765–793.
- Berland LK, McNeill KL. For whom is argument and explanation a necessary distinction? A response to Osborne and Patterson. *Science Education*. 2012; 96(5):808–813.
- Berland LK, Reiser BJ. Making sense of argumentation and explanation. *Science Education*. 2009; 93(1):26–55.
- Cetin PS. Explicit argumentation instruction to facilitate conceptual understanding and argumentation skills. *Research in Science & Technological Education*. 2014; 32(1):1–20.
- Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960; 20:37–46.
- Cohen, J. Statistical power analysis for the behavioral sciences. 2nd. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
- Collins A, Joseph D, Bielaczyc K. Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*. 2004; 13(1):15–42.
- Design-Based Research Collective. Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*. 2003; 32(1):5–8.
- Driver R, Newton P, Osborne J. Establishing the norms of scientific argumentation in classrooms. *Science Education*. 2000; 84(3):287–312.
- Eastwood JL, Sadler TD. Teachers' implementation of a game-based biotechnology curriculum. *Computers and Education*. 2013; 66:11–24.
- Edelson DC. Design research: What we learn when we engage in design. *Journal of the Learning Sciences*. 2002; 11(1):105–121.
- Fleiss, JL., Levin, B., Paik, MC. Statistical methods for rates and proportions. 3rd. New York: John Wiley; 2003.
- Gaydos MJ, Squire KD. Role playing games for scientific citizenship. *Cultural Studies of Science Education*. 2012; 7(4):821–844.

- Kang H, Thompson J, Windschitl M. Creating opportunities for students to show what they know: The role of scaffolding in assessment tasks. *Science Education*. 2014; 98(4):674–704.
- Ketelhut DJ, Nelson BC. Designing for real-world scientific inquiry in virtual environments. *Educational Research*. 2010; 52(2):151–167.
- Khishfe R. Explicit nature of science and argumentation instruction in the context of socioscientific issues: An effect on student learning and transfer. *International Journal of Science Education*. 2014; 36(6):974–1016.
- Krajcik J, McNeill KL, Reiser BJ. Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Science Education*. 2008; 92(1):1–32.
- Kuhn D. Science as argument: Implications for teaching and learning scientific thinking. *Science Education*. 1993; 77(3):319–337.
- Kuhn D. Teaching and learning science as argument. *Science Education*. 2010; 94(5):810–824.
- Li MC, Tsai CC. Game-based learning in science education: A review of relevant research. *Journal of Science Education and Technology*. 2013; 22(6):877–898.
- Lizotte, DJ., McNeill, KL., Krajcik, J. Teacher practices that support students' construction of scientific explanations in middle school classrooms. In: Kafai, Y.Sandoval, W.Enyedy, N.Nixon, A., Herrera, F., editors. *Proceedings of the Sixth International Conference of the Learning Sciences*. Mahwah, NJ: Lawrence Erlbaum; 2004. p. 310-317.
- McNeill KL, Knight AM. Teachers' pedagogical content knowledge of scientific argumentation: The impact of professional development on K–12 teachers. *Science Education*. 2013; 97(6):936–972.
- McNeill KL, Krajcik J. Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching*. 2008; 45(1): 53–78.
- McNeill, KL., Krajcik, J. Supporting grade 5–8 students in constructing explanations in science. New York, NY: Pearson Allyn & Bacon; 2012.
- McNeill KL, Lizotte DJ, Krajcik J, Marx RW. Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*. 2006; 15(2):153–191.
- National Research Council. National Committee on Science Education Standards and Assessment Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press; 1996. *National Science Education Standards*.
- National Research Council. Learning science through computer games and simulations. Committee on Science Learning: Computer Games, Simulations, and Education. In: Honey, MA., Hilton, ML., editors. *Board on Science Education, Division of Behavioral and Social Sciences and Education*. Washington, DC: The National Academies Press; 2011.
- National Research Council. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press; 2012. *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas Committee on a Conceptual Framework for New K–12 Science Education Standards*.
- Newton P, Driver R, Osborne J. The place of argumentation in the pedagogy of school science. *International Journal of Science Education*. 1999; 21(5):553–576.
- NGSS Lead States. *Next Generation Science Standards: For states, by states*. Washington, DC: The National Academies Press; 2013.
- Nussbaum EM, Sinatra GM, Poliquin A. Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education*. 2008; 30(15):1977–1999.
- Osborne J, Erduran S, Simon S. Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*. 2004; 41(10):994–1020.
- Osborne JF, Patterson A. Scientific argument and explanation: A necessary distinction? *Science Education*. 2011; 95(4):627–638.
- Osborne JF, Patterson A. Authors' response to "For whom is argument and explanation a necessary distinction? A response to Osborne and Patterson" by Berland and McNeill. *Science Education*. 2012; 96(5):814–817.

- Penuel WR, Fishman BJ, Cheng B, Sabelli N. Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*. 2011; 40(7):331–337.
- Rivet A, Krajcik J. Contextualizing instruction: Leveraging students' prior knowledge and experiences to foster understanding of middle school science. *Journal of Research in Science Teaching*. 2008; 45(1):79–100.
- Sadler TD, Romine WL, Menon D, Ferdig RE, Annetta L. Learning biology through innovative curricula: A comparison of game- and nongame-based approaches. *Science Education*. 2015; 99(4):696–720.
- Sandoval WA, Reiser BJ. Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*. 2004; 88(3):345–372.
- Squire KD, Jan M. Mad City Mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology*. 2007; 16(1):5–29.
- Toulmin, S. The uses of argument. Cambridge, UK: Cambridge University Press; 1958.
- Watson WR, Mong CJ, Harris CA. A case study of the in-class use of a video game for teaching high school history. *Computers and Education*. 2011; 56(2):466–474.
- Webb AW, Bunch JC, Wallace MFG. Agriscience teachers' implementation of digital game-based learning in an introductory animal science course. *Journal of Science Education and Technology*. 2015; doi: 10.1007/s10956-015-9571-7
- Zohar A, Nemet F. Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*. 2002; 39(1):35–62.

Fig. 1. Screenshot from *The Golden Hour* game depicting the lead physician, Dr. Picotte, who engages the students in dialogue based on the CER framework. Dialogue from the physician appears at the upper-left corner, and the player can choose a response from the options at the bottom of the screen.

I recommend the patient have a CT scan done. Evidence that supports this recommendation is he had a GCS score of 10, which isn't especially low. All his vitals were normal. He did have a slightly low blood pressure, so I would recommend you monitor that. His pupillary reflexes are fine, so we know there wasn't an severe brain trauma ~~caused~~ caused by the accident.

Complete this page AFTER you have submitted the report above and spoken with Dr. Picotte in The Golden Hour game.
To complete your report, you must write a scientific explanation to answer "What should be done next for the patient?" Use the table below to organize your information. Under "Medical Recommendation", use this information to write your full scientific explanation using complete sentences. Be sure to: make a CLAIM, provide EVIDENCE and explain your REASONING.

CLAIM	EVIDENCE	REASONING
I think that the patient needs a CT scan of the brain.	My evidence is the GCS score. The GCS score was a total of 10. That means that it's moderate.	My Reasoning is that we need more conclusive information about the head trauma. It's stable enough for the patient to survive the CT scan since his GCS score is moderate.
MEDICAL RECOMMENDATION I recommend that this patient gets a brain scan. I recommend that since the patient has a head injury, we need to know more about the injury so we can know how to treat the patient and if he needs surgery or not.		

Fig. 2.

Examples of student arguments that scored close to the mean in each group, also exemplifying the (a) open-ended format in Iteration One and (b) scaffolded format in Iteration Two

Fig. 3.
Sequence of stages of enactment in Iteration One.

Fig. 4.
Sequence of stages of enactment in Iteration Two.

Fig. 5.
Distribution of student argument scores in Iteration One and Iteration Two.

Table 1

Summary of Data Collected

Data Type		Iteration One		Iteration Two	
		Day 1	Day 2	Day 1	Day 2
Classroom	Observations	2 periods	3 periods	2 periods	1 period
	Audio recordings	5 periods collected; 3 transcribed	3 periods; 2 transcribed	2 periods collected and transcribed	1 period collected and transcribed
Student	Arguments (CER scores)	49 artifacts		39 artifacts	
	Post-tests	32 artifacts		31 artifacts	

Table 2

Mean (M) and standard deviation (SD) statistics for each Claim, Evidence, and Reasoning component and total scores for students in Iteration One and Iteration Two

	<i>n</i>	Claim (0–2 possible)		Evidence (0–4 possible)		Reasoning (0–4 possible)		Total (0–10 possible)	
		M	SD	M	SD	M	SD	M	SD
Iteration 1	49	1.45	0.87	1.63	1.52	1.31	1.45	4.39	3.36
Iteration 2	39	1.90	0.45	2.64	1.25	2.62	1.46	7.15	2.60