



Pedagogy, Partnership, and Collaboration: A Longitudinal, Empirical Study of Serious Educational Gameplay in Secondary Biology Classrooms

Georgia W. Hodges¹ · J. Steve Oliver¹ · Yoonsun Jang² · Allan Cohen¹ · David Ducrest³ · Tom Robertson³

Accepted: 18 September 2020 / Published online: 4 January 2021
© Springer Nature B.V. 2021

Abstract

The use of serious educational games has the potential to increase student learning outcomes in science education by providing students with opportunities to explore phenomena in ways that vary from traditional instruction; yet, empirical research to support this assertion is limited. This study aimed to explore deeply what learning gains were associated with the use of three serious educational games (SEGs) created for use in secondary biology classrooms that partner teachers implemented during a 2-week curriculum unit. This longitudinal, mixed method study includes a control year, in which we examined how six highly qualified teachers taught students ($n = 407$) a 2-week curriculum unit addressing cellular biology *without* the SEGs, followed by 2 years in which the teachers integrated the SEGs into the curriculum unit with students ($n = 871$). Data were collected from multiple sources, including a validated content pre- and post-test measure, embedded gameplay data, participant observation, teacher interviews, and focus groups. Quantitative findings showed significant learning gains associated with students who experienced the game condition during year 2, when compared with the control condition. During the replication year (year 3), learning gains increased again, compared with year two. Although the SEGs did not change between years 2 and 3, teachers were provided real-time access to students' performance during gameplay. Thematic analysis of observation notes, teacher interviews, and student performance in-game identified four affordances teachers identified related to the use of serious educational games in their classrooms and the extended partnership model employed. Implications for researchers and game designers are discussed.

Keywords Embedded gameplay assessment · Longitudinal serious game studies · Teacher role in SEG environment · Serious educational game design

Neither an excessive focus on rote memorization in the science classroom, nor the hope that all students engage

in inquiry-based learning experience in the classroom is new (Dewey, 1910). Throughout the decades, printed curriculum materials, laboratory exercises, and teacher trainings have focused on developing learning environments in which students are tasked with thinking like a scientist (DeBoer, 1991) as they actively explore phenomena, using

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10956-020-09868-y>) contains supplementary material, which is available to authorized users.

✉ Georgia W. Hodges
georgiahodges@uga.edu

J. Steve Oliver
soliver@uga.edu

Yoonsun Jang
todekfr@gmail.com

Allan Cohen
acohen@uga.edu

David Ducrest
David.ducrest@explorellearning.com

Tom Robertson
Tom.robertson@explorellearning.com

¹ University of Georgia, Athens, USA

² Associate Research Fellow/Statistics & Panel Research Office, National Youth Policy Institute, Seoul, South Korea

³ Explore Learning, Charlottesville, USA

the scientific method to construct explanations and engage in argumentation. Recent reform documents (e.g., Australian Curriculum Assessment and Reporting Authority [ACARA], 2013; National Research Council, 2012; United Kingdom Department for Education, 2014) continue to push for students to experience active science learning environments. Yet, the National Research Council (Honey & Hilton, 2010, p. 22) states:

A growing body of research indicates that engaging students in science processes (inquiry) can motivate and support science learning. However, because inquiry approaches can be difficult for students, teachers, and schools, they are rarely implemented.

Since then, serious educational games (SEGs) have emerged as a promising tool that may equip secondary science teachers to implement active learning environments in which players engage with real-world science phenomena, using scientific practices, such as collecting and analyzing data, simulating the work that scientists do (Ching & Hagood, 2019). There is a growing body of scholarship that supports the use of SEGs as instructional tools in science (Riopel et al., 2019; Vitale, McBride, & Linn, 2016) as researchers have created and tested these contextually rich environments where students conduct experiments and practice the skills that characterize science. Yet, many questions remain unanswered regarding *how*, *why*, and with whom serious games improve science learning outcomes and engagement. To deeply study whether or not students are learning from gameplay, scholars have advocated for more longitudinal, mixed method studies of SEG gameplay (Girard, Ecalle, & Magnan, 2013; Kapp 2012) that take place within authentic classroom settings.

In addition to studying learning gains associated with gameplay, longitudinal, mixed method studies in authentic environments offer researchers an opportunity to study the strategies utilized by teachers to implement the gaming environments. While many scholars (Remillard, 2005) have identified the important role that teachers play in successfully using novel curricular materials in classrooms, there is need for more work in this area regarding the role of the teacher in the use of serious games (Molin, 2017; Shah & Foster, 2015). History provides ample examples of using technology to *teacher proof* curriculum, as seen post World War II as well-meaning advocates of education advocated for replacing instruction by trained teachers with video lectures from scientists (Rudolph, 2002) to improve science education. The view that transmitting scientific knowledge to students will lead directly to significant learning outcomes oversimplifies the complexities of using any pedagogical tool within a given context and is salient to studying the use of immersive learning technologies, such as the SEGs developed and tested in this study. This historical example of *teacher proofing*

curriculum informed the philosophical underpinnings of each aspect of this research project, from the design of the SEG to the research study presented here, as our team of researchers and developers agree that the teacher is the most important variable found in a classroom. As such, the ways in which a teacher interacts with a given pedagogical tool, such as the SEGs in the study, may influence the subsequent learning gains associated with the intervention and should, therefore, be studied as well. In-depth evaluation of the use of SEGs in school settings is understudied, due to the complexity of working in schools and the newness of this area of scholarship; yet, the findings generated from this research could inform a variety of stakeholders, ranging from game developers to teachers and researchers (Molin, 2017). As such, we designed this study to examine two research questions in an effort to enrich the knowledge and understanding that we have of learning gains associated with SEG gameplay. Our research team collaborated with a biology department in a large public high school in the Southern United States to explore how science teachers taught with and without the SEGs developed by our collaborative team across a 3-year timespan. During year 1, partner teachers taught a 2-week curriculum unit that addressed cellular biology concepts *without* the use of the SEGs, followed by 2 years in which the teachers taught the same curriculum unit *with* the SEGs. Two research questions guided this study:

RQ1: What learning gains are associated with the use of three SEGs in secondary biology classrooms?

RQ2: What affordances do qualified science teachers identify related to SEG integration in classrooms?

Background and Conceptual Framework

To orient readers to our study, we first introduce the biological conceptual knowledge addressed in our SEGs, and we make the case that there is a need for new learning tools to support instruction of the concepts. Next, we define our use of the term serious educational game, provide a relevant summary of secondary biology games, and examine the manner in which learning gains associated with gameplay have been measured by other research teams. We conclude with an analysis of what our field currently knows about the ways in which teachers use serious games for learning in classrooms.

Choosing Science Content for Game Design

The *Next Generation Science Standards* (NGSS) identify and characterize what and how students in the USA should learn science (NGSS, 2013). In kindergarten, students are

introduced to four Disciplinary Core Ideas (DCIs) that span science: physical, life, earth and space, and engineering, technology, and applications of science. Throughout the K-12 learning experience, teaching and learning is associated with the four DCIs, as, over time, students construct a deep understanding of the content as well as the scientific practices and crosscutting concepts articulated in the *Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (NRC, 2012). Within the life sciences, four core ideas were articulated to unify the vastness of this content domain. The first of the core ideas is identified as LS1: From Molecules to Organisms: Structures and Process, and requires a deep understanding of the cell. At the high school level, performance expectations should help students to answer the question, “how do the structures of organisms enable life’s functions?” (NGSS, 2013, p.261) emphasizing cellular activities such as nutrient uptake and water movement. Osmosis is the net movement of free water through a selectively permeable membrane from a region of lower solute concentration to a region of higher solute concentration. Odum’s (1995) scholarship identified that high school and college students lack an understanding of osmosis. Fisher, Williams, and Lineback (2011) assert that “part of the challenge may be due to the fact that these processes result from the constant, random motion of invisible particles, and a significant fraction of students struggle to comprehend such abstract ideas” (p.426). We assert that SEGs can provide visualizations that support students’ conceptual development of phenomena by zooming in to the microscopic, *invisible* nature of molecular movement, then zooming back out to the macroscopic, which is more familiar to the students’ lived experience.

SEGs and Science Learning

Lamb, Annetta, Firestone, and Etopio (2018) define serious educational games as “a specific form of video game played within a virtual immersive three-dimensional environment used for educational purposes that includes a directed and a priori pedagogical approach” (p. 159). Within these environments, players “engage with an artificial conflict, defined by rules, that results in a quantifiable outcome” (Salen & Zimmerman, 2004, p.80). Often, these learning environments require players to use specific content knowledge to move forward in the game, a characteristic that demarcates SEGs from simulations and other computer learning experiences (Lamb et al., 2018). Serious educational games differ from games for fun due to the use of learning theories and learning objectives that guide game design and the subsequent use of embedded assessment items to measure learning (Loh, Sheng, & Ifenthaler, 2015) during gameplay. To prove the effectiveness of a SEG is to demonstrate that

it enhances the learning of the players (Girard, Ecalle, & Magnan, 2013) through embedded assessment points (Loh, Sheng, & Ifenthaler, 2015), while also engaging the learner with the game (Marsh, 2011).

Multiple reviews (Boyle Hainey et al., 2016) and meta-analysis (Clark et al., 2016) have identified the affordances of using SEG for instructional purposes across a variety of content domains. In science specifically, Riopel and colleagues (2019) recently conducted a meta-analysis to examine the impact of serious games on learning in the natural sciences. The authors analyzed 15 moderator variables that focused on 3 main aspects: the context (subject area, grade level, intervention duration, comparison group activities), game qualities (ludic content, level of realism, level of player control), and methodology employed (experimental design, randomization, publication status, and year). Five moderator variables were associated with significant learning effects: grade level, intervention duration, user control, publication year, and publishing status. Science learning gains were significantly higher for students in high school in which the intervention lasted less than 1 week and the player felt they had control over gameplay. The ludic, or entertainment value, of the game was not associated with increased learning, nor was the level of realism associated with the game (Riopel et al., 2019). Overall, Riopel and colleagues found that serious games were more beneficial to students in the natural sciences, when compared with traditional instruction for measures related to declarative knowledge gain, knowledge retention, and procedural knowledge gain. Collectively, this research base documents well that well-designed SEGs support learning science.

In the biological sciences, researchers have developed multiple role-playing games and examined learning gains associated with the gameplay. For example, Rosenbaum et al. (2007) created an augmented reality environment named *Outbreak @ The Institute* where players take on the role of medical professionals trying to stop a viral outbreak. Data analysis from pre- and post-experience surveys suggested that students’ application scientific content knowledge improved and that students perceived the learning experience as an authentic. Similarly, in *Quest Atlantis*, students play the role of a scientist, where human impact is studied in a variety of settings. Analysis of pre- and post-test scores confirmed that the gaming experience supported student learning of science concepts (Barab, Sadler, Heiselt, Hickey, & Zuiker, 2007; Hickey et al., 2009). During *River City* gameplay, middle school students explore disease transmission while practicing the science skills of hypothesis formation and experimental design (Ketelhut, Dede, Clarke, Nelson, & Bowman, 2007). Researchers (Ketelhut, Nelson, Clarke, & Dede, 2010) found that *River City* supported students’ development of inquiry practices as reflected in students’ lab reports constructed upon completion of the

experience. Collectively, these science SEGs have shown that students learn from gameplay, as evidenced by significant learning gains based on pre- and post-test measures, but these studies did not compare the learning outcomes to a comparison condition.

Sadler and colleagues (2015) conducted a rigorous study that utilized a quality comparison condition to examine the efficacy of *Mission Biotech*, a role-playing game addressing the cause of an outbreak. Their team created a control curriculum to ensure that all students were exposed to the same learning objectives and they trained teachers on the use of both interventions (Sadler, Romine, Menon, Ferdig, & Annetta, 2015). Students in both groups experienced significant learning gains, and there was no significant difference in performance between the two groups. Sadler et al. (2015) did find that students who had lower interest in science experienced slightly higher learning gains than their more engaged peers and hypothesized that the game provided a more motivating environment for learning for those students in particular than the typical classroom interventions. Sadler and colleagues conclude their paper by advocating for more longitudinal study of serious gameplay in classrooms, and they discuss the difficulties that characterize this type of research, including teacher comfort with the technology, struggles implementing interventions in schools, and the high cost of developing and testing these environments.

Teacher Use of SEGs in Classrooms

Silseth (2012) and Ulicsak and Williamson (2010) have described struggles that teachers face in using SEGs related to access and resources. While we do not intend to minimize these issues, we believe that it is well documented that lack of computer access, limited bandwidth, and minimal administrative support will bottleneck novel technology integration. Any team conducting research in schools know of these limitations and understand that availability and access vary by geographic region and each specific school. This study was conducted in a school that addressed the aforementioned barriers, allowing our team to focus on the use of SEGs by teachers who have adequate resources to implement the novel technology.

To explore gaming environments, we must first acknowledge the inextricable linkage of teachers and the learning environments created in a given classroom. Educational research has identified and confirmed this finding (Lave & Wenger, 1991; Tsai & Chai, 2012) in multiple classroom settings, as seminal research has identified the teacher as a key variable in the learning environment, accounting for 30% of the variance in student learning, second only to individual student factors (Hattie, 2003). More recently, a review of educational effectiveness by Reynolds et al. (2014)

identified the need for longitudinal, context-specific study of teachers to investigate the ways in which teachers interact in classrooms that lead to student growth. Evidence exists of the importance of the teacher in learning environments; yet, there is a lack of research exploring teachers' interactions with students in a technology-centered learning experience, such as a SEG (Jong, Dong, & Luk, 2017).

Various roles have been identified that teachers may play in SEG learning environments. Hanghøj and Brund (2011) describe four distinct roles teachers may play during SEG implementation in classrooms: instructor, playmaker, guide, and evaluator. Shah and Foster (2015) describe three roles: (1) the expert who connects learning goals to the experience; (2) the facilitator who integrates a variety of pedagogical strategies such as discussion and observation to encourage reflection; and (3) the connector who ensures students understand the importance of the concepts *beyond* the classroom experience. More recently, Kangas, Koskinen, and Krokfors et al. (2017) conducted a literature review of educational games in classrooms to explore the roles that teachers play in a gaming environment. They analyzed 15 years of research and identified five key roles for teachers: planning, playing, orienting, assessing, and reflecting. While these roles have been identified, there is a lack of scholarship on the interplay between the teacher, the gaming environment, and student learning outcomes.

The Interventions

During this study, our research team used three SEGs that were designed as a stand-alone, 45-min learning experiences, during which students roleplay a specific scientist, who has been tasked with solving a problem. Each decision that the player makes is captured by the gaming system and has been designed to assess the player's use and understanding of specific disciplinary science content and science practices that are outlined as propositional statements (Appendix 1). Each SEG includes approximately 35 assessment items during gameplay. The three SEGs tested in this study addressed the fundamental biological process of diffusion, osmosis, and filtration. Due to page limitations, a detailed description of one of the three SEGs, *Clark the Calf: Osmosis*, is provided as well as screen shots of the immersive environment (Fig. 1) and examples of the formative assessment items.

During *Clark the Calf: Osmosis*, students play the role of a veterinarian and are presented with a patient, *Clark the Calf*, who is having a seizure (a). To prepare for the arrival of the calf at their clinic, the player is tasked with completing an interactive guide in which they learn the key concepts underlying the system being studied (b) and complete multiple simulations to test a player's understanding of the

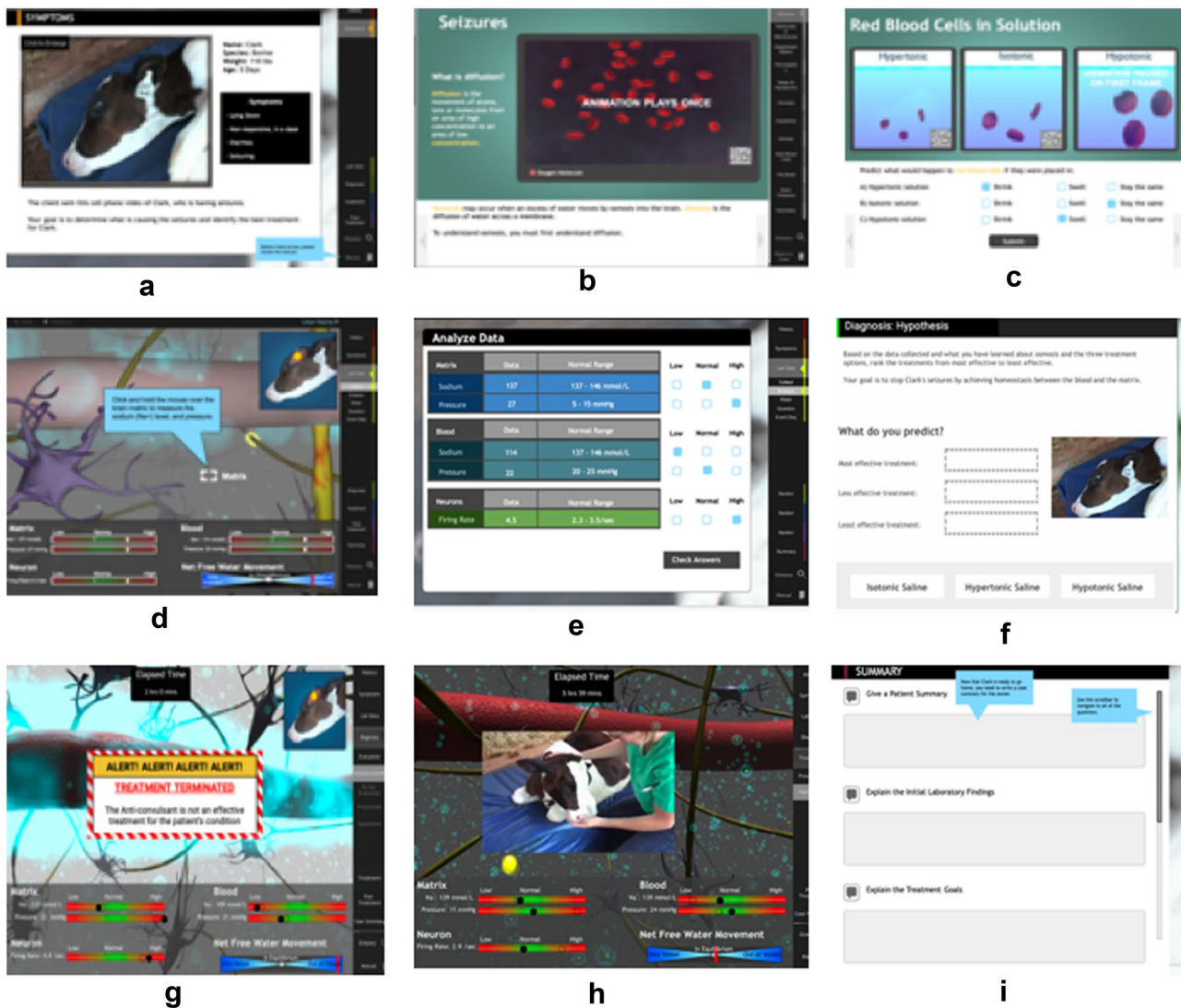


Fig. 1 Clark the Calf

concepts (c). After the patient arrives, the player “flies” into the brain of the calf, where they collect pertinent data from the cells and fluids in the brain (d). The player then analyzes the data (e) and forms a hypothesis that could explain why the calf is having a seizure (f). Next, the player must predict what treatment would best help the calf recover. The player is then returned to the brain and asked to apply their treatment of choice. The data change, as they would do in the body, based on the treatment applied. If the player’s hypothesis and treatment choices are incorrect, the treatment is stopped, and the student is asked to reflect on their choices and revise their hypothesis (g). If the player’s hypothesis is correct, then the data return to normal values, and a video appears showing that they have saved the calf’s life (h). The player then communicates their findings by writing a case report (i). Finally, the player is shown a

“behind the scenes” video that shows how the calf’s seizures were faked by gently shaking his hindquarters while filming his head. A brief video of the immersive gameplay is provided (Appendix 2).

The Teacher Dashboard

Between years 2 and 3 of data collection, the research and development team collaborated with partner teachers to develop a teacher dashboard, based on interviews with teachers and observation of the actual learning environments during year 2. Based on iterative feedback throughout these years, a dashboard was created that equipped teachers to access student responses to embedded assessment items in real time (Fig. 2). Student names are listed in one column, with each adjacent column representing student responses to embedded case study questions.

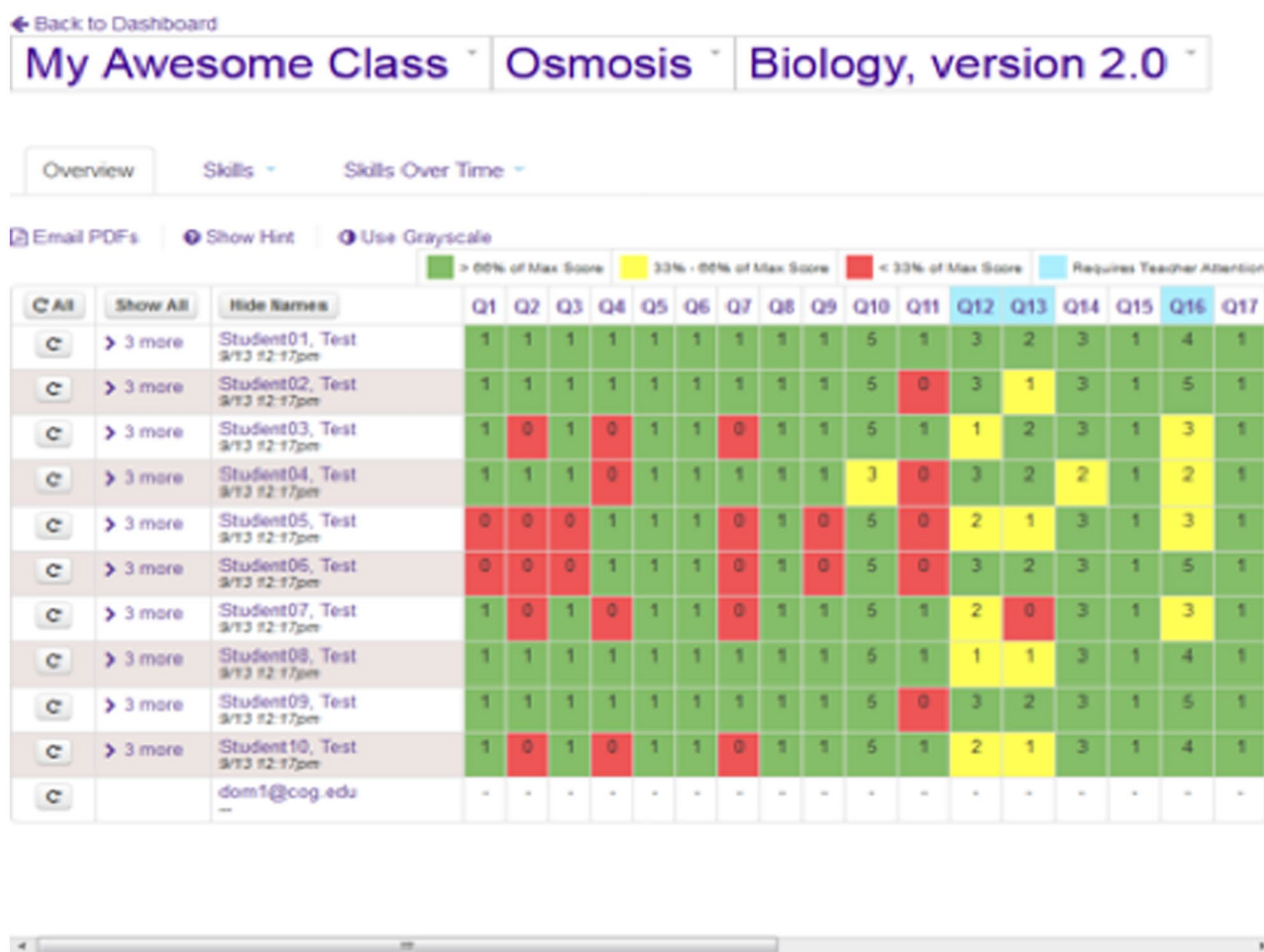


Fig. 2 The teacher dashboard: SABLE (skills and assessment-based learning)

Responses to forced choice questions (e.g., analyzing data) are auto-graded by the system, and constructed responses items are left for teachers to evaluate. The system then analyzes the data and produces a “heat map” of performance using the colors red, yellow, and green. Color-coding of student response patterns was used to assist teachers in identifying student response patterns immediately. Individual student data are accessed by clicking on student names, and the dashboard includes a screen shot of each question from the SEG, the specific science skill being practiced, a suggested rubric, and an exemplar response generated by the collaborative team. More detail regarding the design and development process is discussed elsewhere (Authors, 2017; Authors, 2018).

Methodology

Study Design and Data Collection

Two foundational beliefs regarding educational research influenced the design of the study: the importance of

testing in a school context and the primacy of detailed observation and analysis of individual teacher action in classroom. Thus, this study was designed to take place in a public school, in the context in which the intervention was designed for use, with *all* students who attended the school.

Participants and Context

This study was conducted in a large, suburban school in southeastern USA that serves approximately 3000 students, with a demographic composition characteristic of the nation, with 62% White, 13% African-American, 7% Asian, 13% Latino, and 5% did not identify; 22% of the student population received free or reduced lunch. At this school, introductory biology was taught at four different levels: gifted (identified by results on an aptitude test), honors (identified by course grades or recommendation), college preparatory (CP-general biology), and CP-collaborative (students meeting special education

guidelines). CP-collaborative classrooms are co-taught by a science teacher and a special education teacher. Teachers were recruited by the researchers and written consent was obtained for each teacher and student in the study. All students completed an assent form, agreeing to participate in the study, and a guardian for each student also provided written consent.

Year 1 Six biology teachers agreed to participate in the study during year 1. The teachers attended a 5-day workshop hosted by the research and development team aimed at enriching their content knowledge of cellular biology. During this time, the teachers co-planned a 2-week curriculum unit (Table 1) to address cellular biology concepts that characterize general introductory biology courses. To minimize diffusion (Cook & Campbell, 1979), the teachers were not exposed to the SEGs during the 2-week co-planning time. Teachers were informed of the learning objectives (Appendix 1) addressed by the SEGs as well as the amount of time required for gameplay so that they could plan instruction for years 1 and 2. During year 1, teachers built three additional learning experiences to address the learning objectives addressed by the gameplay so that during year 2, they would simply remove the three lessons and implement the SEGs. Data collected from the 407 students in the study during year 1 included a pre- and post-test that measured cellular biology content

knowledge. This was administered by the teachers in the study before and after the unit of instruction. Researchers observed teachers as they implemented the instruction throughout the curriculum unit.

Year 2 The same six biology teachers attended a 5-day workshop during which university researchers led discussions with the teachers that addressed the content in each of the SEGs. Teachers then played each of the three games and the team discussed questions. Next, teachers implemented the same curriculum unit, replacing 3 days of traditional instruction with the SEGs. Teachers were interviewed before, during, and after the curriculum unit. Next, 393 completed the same pre- and post-test as administered in year 1.

Year 3 Five of the same biology teachers from the years 1 and 2 and one new teacher implemented the cellular biology unit and the three SEGs during year 3. The teachers were provided access to the newly created teacher dashboard that provided real-time feedback on student performance during gameplay. The year 3 sample consisted of 478 students who completed the same pre- and post-test used in years 1 and 2. Teachers were interviewed before, during, and after the implementation of the curriculum unit, and researchers observed teachers' classrooms during the curriculum unit.

Table 1 An overview of learning experiences utilized

Day	Control year 1	Treatment year 2	Treatment year 3
0	Pre-test	Pre-test	Pre-test
1	<ul style="list-style-type: none"> •Student lab set up 	<ul style="list-style-type: none"> •Reading guide: cell history, structure and function •Reading: how to use microscope 	<ul style="list-style-type: none"> •No change
2	<ul style="list-style-type: none"> •Collect data for lab •Reading guide: cell history, structure and function 	<ul style="list-style-type: none"> •Microscope Lab and quiz •Lecture on cell theory, history and types •Reading: Life Cell Lab 	<ul style="list-style-type: none"> •No change
3	<ul style="list-style-type: none"> •Collect data for lab •Lecture on cell theory, history and types •Prokaryote vs. eukaryote diagram and question sheet 	<ul style="list-style-type: none"> •Begin cell ID lab •Prokaryote vs. eukaryote diagram and question sheet •Assign cell project 	<ul style="list-style-type: none"> •No change
4	<ul style="list-style-type: none"> •Collect data for lab •Complete microscope lab 	<ul style="list-style-type: none"> •Complete cell ID lab •Cell membrane notes 	<ul style="list-style-type: none"> •Complete cell ID lab •Cell types lab
5	<ul style="list-style-type: none"> •Collect data for lab •Assign cell project 	<ul style="list-style-type: none"> •Cell membrane and cell transport notes 	<ul style="list-style-type: none"> •Diffusion SEG •Diffusion SEG debrief
6	<ul style="list-style-type: none"> •Collect data for lab •Cell ID Lab •Cell membrane lecture and book assignment 	<ul style="list-style-type: none"> •Diffusion SEG 	<ul style="list-style-type: none"> •Osmosis SEG •Osmosis SEG analysis
7	<ul style="list-style-type: none"> •Collect data for lab •Cell types lab 	<ul style="list-style-type: none"> •Osmosis SEG 	<ul style="list-style-type: none"> •Filtration SEG •Diffusion, osmosis, filtration jigsaw
8	<ul style="list-style-type: none"> •Cell transport lecture and worksheet 	<ul style="list-style-type: none"> •Filtration SEG 	<ul style="list-style-type: none"> •Gummy bear osmosis lab
9	Test review and project presentations		
10	Post-test	Post-test	Post-test

Data Sources

Data sources presented in this research study include teacher interviews and focus groups from years 2 and 3, pre- and post-test results from years 1 through three, and embedded gameplay assessments from years 2 and 3.

1. *Pre-test and post-test.* In advance of data collection, a team of science content experts and science educators created a set of items to assess the primary learning objectives associated with the content addressed in SEGs. High school science teachers edited the items, and we conducted cognitive interviews with students to validate the items. Items were then validated by examination of student responses from over 400 students not included in this study. We created two versions of the test using comparable multiple-choice items to minimize memory effects that occur when assessments use the same questions (Wooldridge et al., 2014). Each form of the test included a set of identical items that were designed so that we could use differential item functioning analysis (Pine, 1977) to create a common scale so that scores between the pre-test and post-test and also between the two forms
2. *Embedded gameplay assessments:* Strategic embedded assessment occurs throughout the SEGs tested. Each embedded assessment aligns with a specified propositional statements (Appendix 1). Students were formatively assessed throughout the gameplay 34 times. Of the 34 items, 6 items were designed as constructed responses, which require a teacher or researcher to assess, while the remainder of the questions is graded by the program and provides the student with real-time feedback. Players are required to correctly answer embedded gameplay items that are scored by the computer prior to moving forward in the gameplay.
3. *Interviews and classroom observations.* Our team conducted multiple semi-structured interviews (Seidman, 2013) with each teacher during every year of the study. Teachers were interviewed before and after teaching during the

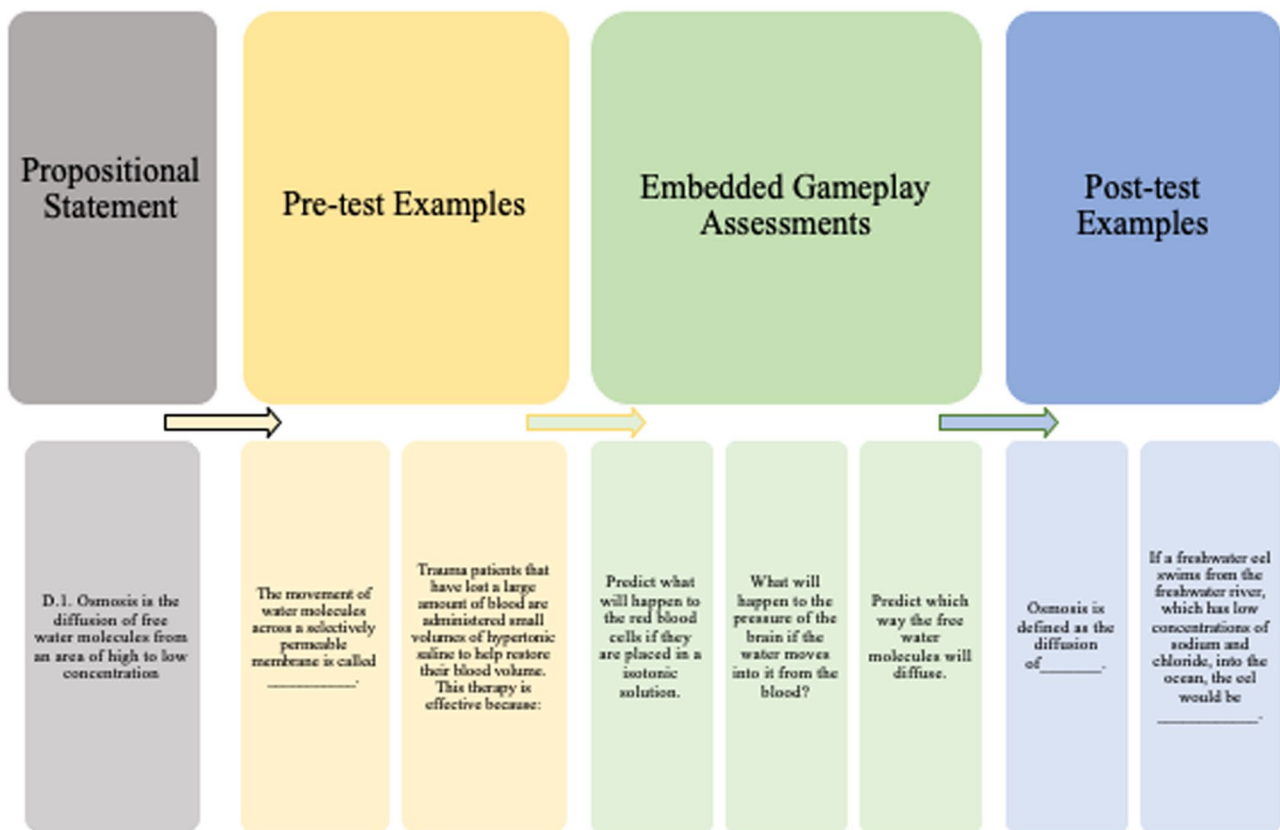


Fig. 3 Connecting the propositional statements, assessment items, and embedded gameplay items

Table 2 Embedded gameplay assessment item examples, alignment and description

Embedded gameplay assessments and scoring ranges	Propositional statement alignment	Bloom's taxonomy item level	Forced choice or constructed response
Predict what will happen to the red blood cells if they are placed in an isotonic solution. (0–2)	E3	3	FC
What will happen to the pressure of the brain if the water moves into it from the blood? (0–2)	D3	4	FC
Using the sodium data you collected and what you learned from the Seizure Manual, [answer the following question] why are the free water molecules diffusing out of the vessel?(0–3)	D2	5	CR
Predict the effect of your hypertonic saline on: blood sodium concentration–decrease–no change–increase.(0–2)	D3	4	FC
Predict the effect of your hypotonic saline on: brain matrix pressure–decrease–no change–increase. (0–2)	D3	3	FC
Based on what you have learned, summarize the relationship between solute concentrations on opposite side of a semi-permeable membrane and the direction of movement of free water molecule. (0–5)	D1	4	CR
Describe the initial lab findings. (0–3)	D1 and D3	3	CR
Describe how osmosis was involved in causing Clark's seizures. (0–8)	D1 and D3	4	CR

intervention. In addition, all classes were recorded. These interviews and classroom observations were recorded, then transcribed, and analyzed by coding passages inductively.

Data Analysis

1. *Pre- and post-test.* A multivariate analysis of variance (MANOVA) was conducted to examine simple and main effects as well as the interactions using student responses to the pre- and post-test as the dependent variable.
2. *Embedded gameplay assessment.* Our early work (Authors, 2017; Authors, 2018), coupled with continual examination of the literature, informed the design of our data capture and analysis framework from the onset. We knew that the amount of data generated through gameplay would quickly overwhelm our team if we did not strategically choose which data sources to transform into analysis items. Although our data capture system tracked all student movement within the game environment and generated log files of the gameplay, we created a deductive framework that prioritized student responses to specific content and skill questions, while we excluded extraneous variables such as length of time on screen, movement within game, and the number of times a student repeated a simulation. While this framework limits the inferences we can make regarding time on task, our primary focus was evidence of student learning. Next, we created rubrics to analyze student responses for each question. While many items were automatically scored by the software, any question that generated an open ended (i.e., constructed) response was analyzed utilizing inductive content analysis (Elo and Kyngäs 2008) and deductive content analysis methods (Polit & Beck, 2012). All student responses were scored by two raters and discussed until there was 100% agreement (Table 2).
3. *Interviews.* Thematic analysis was applied to all interviews and focus groups conducted with teachers (Ezzy, 2002). We inductively analyzed each line of the transcripts of teacher discourse, then used axial coding (Ezzy, 2002) to identify themes, processes

Table 3 Summary of MANOVA results

Predictors	Wilks' Lambda	F value	Hypothesis df	Error df	p value
Year	.956	14.31	4	2,514	< .0001***
Type of class	.867	97.76	2	1,257	< .0001***
Teacher	.908	10.33	12	2,514	< .0001***
Year × type of class	.981	6.01	4	2,514	< .0001***
Year × teacher	.937	5.99	14	2,514	< .0001***
Type of class × teacher	.992	1.67	6	2,514	.125

*** $p < .001$

Table 4 Simple effects of teacher for each year

Level of year	Pre-test		Post-test	
	<i>F</i> value	<i>p</i> value	<i>F</i> value	<i>p</i> value
Year 1	39.39	< .0001***	44.28	< .0001***
Year 2	17.07	< .0001***	27.50	< .0001***
Year 3	21.35	< .0001***	60.97	< .0001***

****p* < .001

and relationships among the codes to address the research questions.

Findings

Two research questions guided the study presented here:

RQ1: What learning gains are associated with the use of three SEGs in secondary biology classrooms?

RQ2: What affordances do qualified science teachers identify related to SEG integration in classrooms?

Findings 1 through 3 address the specific learning gains associated with the SEG intervention, using year 1 as a comparison year. Finding 4 presents an overall thematic analysis of the affordances that partner teachers identified as salient to the success of the SEG intervention presented in this study.

Finding 1: Significant Learning Gains Associated with SEGs

As expected, learning gains were associated with each year of the intervention. A MANOVA was conducted to examine the change in students' performance from the pre- to post-test across three treatments years (Table 3). Data from the 1278 students who participated in years

1, 2, and 3 were included in this analysis. The three predictor variables in this analysis included year, type of class, and teacher. The year indicates the specific treatment conditions: one comparison group (year 1) and two treatment groups (year 2 with SEGs and year 3 SEGs + teacher dashboard). The type of class indicates two groups: the combined sample of students in the college preparatory (CP) and CP-collaborative (Collab) classes and the combined sample of students in the gifted and honors classes. The third predictor is teacher: in year 1, there were five teachers (T1, T2, T3, T4, and T5); in year 2, one teacher (T6) was added resulting in a total of six teachers in this study. Finally, in year 3, four teachers (T3, T4, T5, and T6) one new teacher (T7) joined in the project.

Since the variables of year, type of class, and teacher were significant at alpha level = 0.05, Wilks' lambda, a measure of the proportion of unexplained variance in the dependent variables by the predictor variables, was also performed. In this case, type of class explained the largest proportion of the variance (13.3%), followed by the teacher (8.2%), and intervention year (4.4%). Since the interaction between year and teacher was also significant, the simple effects of teacher for each level of year were tested (Table 4).

All simple effects for "teacher" were significant across all years of the intervention so pairwise post hoc tests were conducted using a Bonferroni correction, along with the post hoc pairwise test for the type of class. The mean ability for all 3 years increased from the pre-test to the post-test, with the largest growth in year 3 (Table 5).

Learning outcomes at the class level (CP/Collab or gifted/honors) increased between years 2 and three, while the instructional sequence remained comparable.

Table 5 Descriptive statistics for ability by year

Year	Type of class		Mean	SD	Minimum	Maximum
Year 1	CP and Collab (<i>n</i> = 178)	Pre	-.701	.610	-2.58	.81
		Post	-.526	.644	-2.19	1.08
	Gifted and honors (<i>n</i> = 229)	Pre	-.005	.592	-1.61	1.29
		Post	.383	.603	-1.87	1.62
Year 2	CP and Collab (<i>n</i> = 141)	Pre	-.534	.563	-2.22	1.00
		Post	-.083	.564	-1.65	1.47
	Gifted and honors (<i>n</i> = 252)	Pre	.123	.530	-1.68	1.62
		Post	.761	.578	-0.68	2.41
Year 3	CP and Collab (<i>n</i> = 134)	Pre	-.507	.503	-1.65	1.04
		Post	.030	.654	-1.66	1.72
	Gifted and honors (<i>n</i> = 344)	Pre	.055	.599	-1.92	1.58
		Post	1.175	.805	-1.06	3.33

Finding 2: Teacher Interaction During Gameplay Matters

To explore the increased learning gains between years 2 and 3, we further examined the learning outcomes associated with individual teachers in more detail. As reported in Table 5, student growth, as measured by the pre- and post-test, increased each year of the study. When we disaggregated the data by individual teacher (Fig. 4), we found that students whose teachers had participated in the study for 3 years (i.e., teachers T3, T4, and T5) experienced more growth than students whose teachers were newer to the project. This suggests that the way in which teachers interact with and utilize a gaming environment may influence student learning gains.

Participant observation and teacher interviews revealed a pattern of offloading instruction onto the SEG by teachers (T6 and T7), while other teachers (T3, T4, T5) increased their interaction with students during gameplay. Specifically, T7 explained, “when I know that a learning experience is good, like these games, I choose to use the time to plan, grade, and address other teaching tasks. They (the students) don’t need my help during the game.” Participant observation confirmed this pattern of offloading, as evidenced by the number of interactions between teachers and students identified in different classrooms. Teachers who offloaded their instruction ensured that students were logged in the gaming environment, then left their students to complete the activity. These teachers (T6 and T7) intervened when a student approached them with a question or when there was a classroom management issue that needed attention. Conversely,

Table 6 Embedded gameplay percent correct by year

Class type	Year	SEG 1 diffusion	SEG 2 osmosis	SEG 3 filtration
CP	Year 2	50%	45%	60%
	Year 3	54%	50%	70%
Honors	Year 2	54%	54%	71%
	Year 3	64%	63%	82%
Gifted	Year 2	61%	62%	87%
	Year 3	79%	72%	93%

n = 100 for each group

teachers (T4,T5,T6) initiated interaction with their students throughout the learning experience. For example, T5 introduced the concepts addressed in each SEG before students played, then intervened with specific students to address performance, provided whole class instruction during gameplay, and then summarized the storyline of the gameplay afterward. T3 chose a different approach as she displayed the heat map (while hiding student names) on the smartboard during the class so that all students could see their progress, and compare their progress to the class as a whole. Finally, T4 watched student responses populate on a tablet, and she graded the first constructed response for each student. Next, she provided each student feedback individually, walking over to the student and discussing their response. This feedback ranged from simple acknowledgement of a quality response to requiring a student to restart the gaming experience and increase their effort. These data suggest that teachers who provide elaborated feedback to students during gameplay add value to the students’ learning experience.

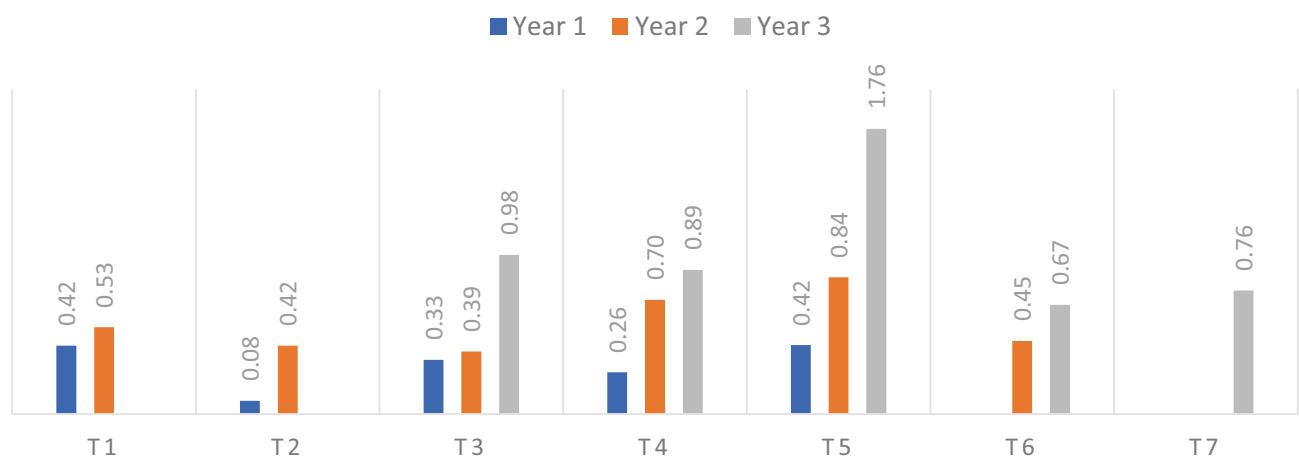


Fig. 4 Student growth by year and teacher

Table 7 Comparing embedded gameplay data scores from osmosis: Clark the Calf

Embedded gameplay assessments and scoring ranges	Propositional statement alignment	Bloom's taxonomy item level	Forced choice or constructed response	Percent correct	
				Year 2	Year 3
Predict what will happen to the red blood cells if they are placed in an isotonic solution.(0–2)	E3	3	FC	75%	78%
What will happen to the pressure of the brain if the water moves into it from the blood? (0–2)	D3	4	FC	85%	87%
Using the sodium data you collected and what you learned from the Seizure Manual, [answer the following question] why are the free water molecules diffusing out of the vessel?(0–3)	D2	5	CR	47%	54%
Predict the effect of your hypertonic saline on: blood sodium concentration–decrease–no change–increase.(0–2)	D3	4	FC	66%	70%
Predict the effect of your hypotonic saline on: brain matrix pressure–decrease–no change–increase. (0–2)	D3	3	FC	70%	70%
Based on what you have learned, summarize the relationship between solute concentrations on opposite side of a semi-permeable membrane and the direction of movement of free water molecule. (0–5)	D1	4	CR	25%	33%
Describe the initial lab findings. (0–3)	D1 and D3	3	CR	28%	35%
Describe how osmosis was involved in causing Clark's seizures. (0–8)	D1 and D3	4	CR	46%	58%

Finding 3: Students' In-game Performance Improves During Year 3

In order to more deeply explore the increased learning gains between years 2 and 3, we compared the embedded gameplay data from a sample of 100 students from three ability bands to determine *if* student responses improved during year 3. Students who completed the three SEGs were randomly selected from each ability bands. Aggregate student performance during year 3 surpassed aggregate student performance during year 2 (Table 6) based on the overall percent correct during each of the three SEGs.

When we analyzed specific embedded gameplay items, we found more difference in student performance on constructed response items than on forced choice items (Table 7), regardless of the item level difficulty that was determined. The feedback provided to students during years 2 and 3 did not change on the forced choice items, as the system provided immediate feedback to these items. However, during year 3, teachers were able to see student responses to the constructed response items. This equipped teachers to provide students feedback on these responses either during gameplay or afterward. Thus, students were provided more feedback on their performance during year three.

Finding 4: Teacher-Identified Affordances of the Gaming Environment

Theme 1: Prioritizing Science Phenomena. Teachers indicated that the real-world contextualization of the problem, the quality of the visualization, and the integration of macro- and micro-views of the problem enhanced students'

learning opportunities. Teacher 3 explained that, "In class, students have worked on tonicity problems, where they have drawn arrows to show the direction of water flow. When they played the *Clark the Calf*, they actually watched ions move. Instead of working out problems and drawing arrows, students actually engaged with the phenomena." Teacher 5 added, "the SEGs seamlessly showed students the macro- and the micro-view of osmosis taking place. I *cannot* simulate this in my classroom; this is something only technology can do." Teacher 4 explained, "when students complete a wet lab, they only get one chance. Here, they get to try again."

Theme 2: Empowering Teachers Through Real-Time Data. After year 2, teachers expressed their disappointment with one aspect of gameplay: student interaction. This led to the research team developing the teacher dashboard, providing teachers with real-time access to student performance within the gaming environment. Teacher 3 explained, "with the dashboard, the students fully recognize that we are in this together. I am *in* this experience with them, so they give it their best." Teachers explained that the ability to intervene in real-time with students equipped them to support high achievers and struggling students more efficiently. Participant observation clearly documented T4 and T5 examining student response patterns and providing students specific feedback. Teachers also highlighted the value of the embedded assessments as they explained that this provided students feedback in the moment instead of days or weeks later, after grading.

Theme 3: Identification of Lack of Student Effort and Student Struggle. As teacher 6 watched the heat map populate for the first time, she was shocked to see the lack of effort put forth by many of her students. "It's disappointing

to see that many of the students are not trying to explain what is happening. They are writing a few words and moving on.” Teacher 6 identified individual students who were not responding with complete explanations and walked over to the students and showed them their work. Students continually responded to teachers with an apology for their lack of effort and surprise that the teacher was monitoring student progress. Student 1’s response exemplified this student response pattern: “I’m sorry, Teacher 6. Usually teachers just put us on the computer and never do anything with the work we do. Had I known you really cared, that this wasn’t just busy work, I would have tried harder.”

Theme 4: Teacher Ownership. the Affordances of Partnership. When teachers implemented the SEGs during year 2, they relied on researchers to support the students, even though the teachers attended professional learning with the SEGs and were knowledgeable of how to use the SEGs. Specifically, during year 2, teacher 5 introduced the SEGs as university-developed games designed for high school use that address science concepts in innovative ways. Afterward, teacher 5 did not interact with students while they played the game. During year 3, teacher 5’s subsequent ownership of the learning experience was evident from the way in which she introduced the game stating, “the lab you are going to complete today is on a computer. You are playing the role of a veterinarian, and you have approximately 45 min to save the life of a calf. You will see that osmosis plays the starring role in this story, and this is real-life. I cannot stress this enough that what you are learning in science really matters, and you get to see this today. I will watch your responses as you go, so give this your best. You will see these concepts again on your test on Friday.” Teacher 5 explained that she now felt equipped to use the games and that she no longer needed any support. She could monitor student responses and provide instructional feedback, just as she would during a traditional learning experience.

Discussion and Implications

Prior reviews of the role the teacher plays in technology environments (Shah & Foster, 2015) identified the following as roles teachers may play: (1) connecting the learning goals of the class to the technology enhanced learning experience, (2) facilitating strategies to encourage reflection on the experience, and (3) connecting the experience to the lives of students *beyond* the classroom. We agree that these roles are valuable in a classroom, and as a result, these attributes were embedded in the SEG design. In this regard, the primary role our teachers played differed from those identified in the literature. Teachers using SABLE SEG’s primary role was *differentiator*, which was made possible by *teacher presence* in the gaming experience. Depending on the need of a given

student, a partner teacher provided individual feedback to the student, due to the use of the teacher dashboard. This feedback ranged from identifying students who were struggling with the experience to students who were not taking the learning experience seriously. By designing the game to align with science standards and using a real-world problem to introduce the concept to students, teachers were free to focus on individual student needs, thus leveraging technology support teacher in differentiating the support and feedback students were provided.

When we compared student growth in year 2 to year 3, we found that students across all levels learned more during the third year of the study. Results also suggested that the teachers who were using the SEGs for the second time had higher student growth than the teachers new to the SEGs. Deeper analysis of teacher interaction during gameplay highlights the importance of what teachers are doing during gameplay. Teachers new to the use of the SEGs interacted less with students during gameplay, *offloading* instruction onto the SEG. As teachers who offloaded their instruction onto the SEG explained, they perceived that the game was sufficient support and that students did not need any direction during the gameplay, as the SEG encapsulated instruction, engagement, and assessment. This notion of encapsulation resonates with researchers in the past who asserted that technology alone could change the US educational landscape. In the 1950s, educators struggled with how to most appropriately utilize televisions in classrooms, as they were a novel form of technology. Well-meaning researchers asserted that the *best* scientists should be filmed teaching concepts; then, this teaching could be shared through the television, thus providing more students with access to high-quality science education experiences. While speaking at a National Manpower Council Meeting in 1954, Henry Chauncey focused on using scientific and professional people power most appropriately to improve teaching and learning as he asserted that “instructional films can do as good a job in the respect—if not better—than the average classroom teacher” (as cited in Rudolph, 2002). As stakeholders in education move forward in exploring how to integrate novel technologies into instruction, it is important that we learn from the past, as novel technologies, such as SEGs become available to more students and teachers. Thus, we assert that, while valuable, we must explore *how* teachers actually use tools such as these in classroom settings with students to determine if the technologies enrich learning experiences.

In a commencement address to Stanford University graduates, Steve Jobs (2005) offered the following remark based on his lived experience: “you can’t connect the dots looking forward; you can only connect them looking backwards.” This remark resonates with this research teams as we collaborated over a 7-year timeframe, conducting a variety of research studies throughout the years, as we sought to

understand how to create, test, and then scale the use of immersive technologies in high school biology classrooms. Based on this experience, we have identified four specific suggestions for other researchers and game designers to consider:

- 1). SEGs should offer novel learning opportunities that prioritize relevant science problems for students to explore that teachers may not readily provide otherwise;
- 2). SEGs should equip teachers to interact and intervene in gaming environments so that they play the role of the teacher during gameplay;
- 3). SEG design teams must understand that students think of computer games “busywork” unless they are proved wrong. Interactivity must drive gameplay; and
- 4). SEGs must have teacher input in the design and subsequent research.

In 2003, Maddux reviewed 20 years of research in education technology and concluded that “the value of integrating technology lies in *how*, not whether, it is used” (p. 45). Our research supports this assertion as we found that the way in which teachers interacted with students during gameplay appears to have led to increased learning outcomes for students. Perhaps as important is the ownership embodied by partner teachers with the use of the SEGs. None of the partner teachers self-identified as “gamers” or as technology experts, but all of the partner teachers who collaborated with the team for 3 years continue to use the SABLE SEGs, 5 years later. While we strive to provide quality learning experiences to students across the world using digital technologies, due to the COVID-19 pandemic, the importance of ensuring these learning environments are connected to measurable learning has never been as important to the teaching world as we consider the likelihood that more virtual teaching will characterize our future. Not only must researchers create, refine, and iterate these learning environments, we must ensure that the role of the teacher is understood deeply so that teachers may be trained to leverage these immersive environments for the most meaningful learning experience possible.

Funding This research has been partially funded by the National Institutes of Health grant no. R25RR025061 but should not be construed to represent the opinions or positions of the National Institutes of Health (NIH) nor the Science Education Partnership Awards (SEPA) program. Responsibility for the content of this document rests entirely with the authors. National Science Foundation grant no.1021RR246076)

Compliance with Ethical Standard

Conflict of Interest There are no conflicts of interest among the researchers, and we have identified the funding agencies that supported the work.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (University of Georgia +MOD00001890) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was gathered from each teacher as well as each student whose data are analyzed in the study. A guardian signed a letter of consent for each student who participated in the research study. In addition, each student signed a letter of assent.

Appendix 1

Detailed propositional statements outlining the scientific processes addressed in the three SEGs addressed in this research study.

Category A: general diffusion principles

1. Diffusion is the net movement of particles from an area of high to low concentration
2. Particles move continuously, even at equilibrium
3. Diffusion does not require energy
4. Diffusion occurs across semi-permeable membranes

Category B: Factors that affect the rate of diffusion

1. An increase in the concentration difference leads to an increased rate of diffusion
2. An increase in the diffusion distance decreases the rate of diffusion
3. A decrease in the concentration gradient leads to a decrease in the rate of diffusion
4. A decrease in surface area decreases the rate of diffusion

Category C: Anatomical understanding

1. Lungs function to exchange gas
2. Gas exchange occurs at the intersection of alveoli and capillaries
3. Red blood cells contain hemoglobin, which changes color from maroon to bright red when oxygenated
4. Kidneys remove waste from the blood

Category D: general osmosis principles

1. Osmosis is the diffusion of free water molecules from an area of high to low concentration
2. When the concentration of particles (such as salt) is increased, the amount of free water decreases
3. Free water will diffuse from an area of higher concentration to an area of lower concentration until a dynamic equilibrium is reached

- Pressure will increase in a bounded system if amount of water increases

Category E: the effect of tonicity on red blood cells

- Red blood cells placed in a hypertonic solution shrink
- Red blood cells placed in a hypotonic solution swell
- Red blood cells placed in an isotonic solution stay the same

Category F: Filtration

- Filtration is utilized to regulate the concentration of a variety of solutes in the body
- Filtration is utilized to keep the body in homeostasis by changing the concentration of albumin, sodium, and potassium as well as other solutes
- Change in the pore size in a filter affects the solutes that diffuse
- Filtration in parallel flow
- Counter current exchange in filtration creates a gradient throughout the process, increasing the rate of diffusion
- Filtration of blood occurs in the kidney

Category G: Systems

- Homeostasis is the result of a balanced internal environment
- Structure and function relationship
- Scientists use specific language to discuss relative sizes and amounts of materials
- Scientists must interpret data to solve problems
- Scientists must analyze data to solve problems
- Use of the CER framework

Appendix 2

Clark the Calf Video.

See ESM video.

References

- Australian Curriculum Assessment and Reporting Authority [ACARA]. (2013). *Australian curriculum: Science*. Retrieved from <https://www.australiancurriculum.edu.au/> on June 14, 2017.
- Authors (2017). Supporting High School Student Accomplishment of Biology Content Using Interactive Computer-Based Curricular Case Studies, *Research in Science Education*, p. 1–26. <https://doi.org/10.1007/s11165-017-9675-6>.
- Authors (2018). An exploratory study of blending the virtual world and the laboratory experience in secondary chemistry classrooms.
- Barab, S. A., Sadler, T., Heiselt, C., Hickey, D., & Zuiker, S. (2007). Relating narrative, inquiry, and inscriptions: Supporting consequential play. *Journal of Science Education and Technology*, 16(1), 59–82.
- Boyle, E., Hainey, T., Connolly, T., Gray, G., Earp, J., Ott, M., et al. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, 94, 178–192. <https://doi.org/10.1016/j.compedu.2015.11.003>
- Cheng, C., & Hagoood, D. (2019). Activity monitoring gaming and the next generation science standards: Students engaging with data, measurement limitations, and personal relevance. *Journal of Science Education and Technology*, 28(8), 589–601.
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: Systematic review and meta analysis. *Review of Educational Research*, 86(1), 79–122.
- Cook, T. D., & Campbell, D. T. (1979). The design and conduct of true experiments and quasi-experiments in field settings. In R. T. Mowday & R. M. Steers (Eds.), *Reproduced in part in Research in Organizations: Issues and Controversies Santa Monica*. CA: Goodyear Publishing Company.
- DeBoer, G. E. (1991). *A history of ideas in science education*. New York: Teachers College Press.
- Dewey, J. (1910). *The influence of Darwin on Philosophy*. Bloomington: Indiana University Press.
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of advanced nursing*, 62(1), 107–115.
- Ezzy, D. (2002). *Qualitative analysis: Practice and innovation*. St Leonards, New South Wales: Allen & Unwin.
- Fisher, K. M., Williams, K. S., & Lineback, J. E. (2011). Osmosis and diffusion conceptual assessment. *CBE life sciences education*, 10(4), 418–429. <https://doi.org/10.1187/cbe.11-04-0038>
- Girard, C., Ecalte, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, 29(3), 207–219.
- Hanhøj, T., & Brund, C. (2011). Teachers and serious games: teacher roles and positionings in relation to educational games. In B. Meyer, Bh.H. Sorensen, & S. Nielsen-Egenfeldt(Eds.). *Serious games in education*, p. 125-136, Denmark: Aarhus Univrrsity Press
- Hattie, J.A.C. (2003, October). Teachers make a difference: What is the research evidence? Paper presented at the Building Teacher Quality: What does the research tell us ACER Research Conference, Melbourne, Australia. Retrieved from https://research.acer.edu.au/research_conference_2003/4/
- Honey, M. A., & Hilton, M. (Eds.). (2010). *Learning science through computer games and simulations*. Washington, DC: National Academies Press.
- Hickey, D. T., Ingram-Goble, A. A., & Jameson, E. M. (2009). Journal of Science. *Education, and Technology*, 18, 187–208. <https://doi.org/10.1007/s10956-008-9143-1>
- Jobs, S. (2005). Commencement address delivered by Steve Jobs, CEO of Apple Computer and of Pixar Animation Studios on June 12, 2005. Retrieved from <https://news.stanford.edu/2005/06/14/jobs-061505/> on January 5, 2019.
- Jong, J., Dong, A., & Luk, E. (2017). Design-based research on teacher facilitation practices for serious gaming in formal schooling. *Research and Practice in Technology Enhanced Learning*, 12(19). <https://doi.org/10.1186/s41039-017-0056-6>
- Kangas, M., Koskinen, A., & Krokfors, L. (2017). A qualitative literature review of educational games in the classroom: the teacher's pedagogical activities. *Teachers and Teaching*, 23(4), 451–470. <https://doi.org/10.1080/13540602.2016.1206523>
- Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. San Francisco, CA: Pfeiffer.
- Ketelhut, D. J., Dede, C., Clarke, J., Nelson, B., & Bowman, C. (2007). Studying situated learning in a multi-user virtual environment. In

- E. Baker, J. Dickieson, W. Wulfbeck, & H. O'Neil (Eds.), *Assessment of problem solving using simulations*. NJ: Lawrence Erlbaum Associates, Mahwah.
- Ketelhut, D., Nelson, B., Clarke, J., & Dede, C. (2010). A multi-user virtual environment for building and assessing higher order inquiry skills in science. *British Journal of Educational Technology*, 41(1), 56–68. <https://doi.org/10.1111/j.1467-8535.2009.01036.x>
- Lamb, Annetta, Firestone, & Etopio (2018). A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations *Computers in Human Behavior* 80. <https://doi.org/10.1016/j.chb.2017.10.040>
- Lave, J., & Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious games analytics: Theoretical framework. In Loh C., Sheng Y., Ifenthaler D. (eds) *Serious Games Analytics. Advances in Game-Based Learning*. Springer, Cham. https://doi.org/10.1007/978-3-319-05834-4_1
- Marsh, T. (2011). Serious games continuum: Between games for purpose and experiential environments for purpose. *Entertainment Computing*, 2(2), 61–68.
- Molin, G. (2017) The Role of the Teacher in Game-Based Learning: A Review and Outlook. In: Ma M., Oikonomou A. (eds) *Serious Games and Edutainment Applications*. Springer, Cham https://doi.org/10.1007/978-3-319-51645-5_28
- National Research Council. (2012). *A framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New L-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- National Research Council. (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/18290>.
- Odom, A. (1995). Secondary & college biology students' misconceptions about diffusion & osmosis. *The American Biology Teacher*, 57(7), 409–415. <https://doi.org/10.2307/4450030>
- Pine, S.M. (1977). Applications of item characteristic survey theory to the problems of test bias. In D.J. Weiss (Ed.), *Applications of computerized adaptive testing: Proceedings of a symposium presented at the 18th annual convention of the Military Testing Association* (Research Rep. No. 77-1 (pp. 37-43). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.
- Polit, D.F., & Beck, C.T. (2012). *Nursing research: Generating and assessing evidence for nursing practice*. Philadelphia, PA: Lippincott Williams and Wilkins.
- Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, 75(2), 211–246. <https://doi.org/10.3102/00346543075002211>
- Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, & Stringfield. (2014). Educational effectiveness research (EER): a state-of-the-art review. *School Effectiveness and School Improvement*, 25(2), 197–230. <https://doi.org/10.1080/09243453.2014.885450>
- Riopel, M., Nenciovici, L., Potvin, P., Chastenay, P., Charland, P., Sarrasin, J., & Masson, S. (2019). Impact of serious games on science learning achievement compared with more conventional instruction: an overview and a meta-analysis. *Studies in Science Education*, 55(2), 169–214. <https://doi.org/10.1080/03057267.2019.1722420>
- Rosenbaum, E., Klopfer, E., & Perry. (2007). On location learning: Authentic applied science with networked augmented realities. *Journal of Science Education and Technology*, 16(31), 31–45. <https://doi.org/10.1007/s10956-006-9036-0>
- Rudolph, J. L. (2002). *Scientists in the Classroom: The Cold War Reconstruction of American Science Education*. New York: Palgrave Macmillan, Rudolph.
- Sadler, T. D., Romine, W. L., Menon, D., Ferdig, R. E., & Annetta, L. (2015). Learning biology through innovative curricula: a comparison of game- and nongame-based approaches. *Science Education*, 99(4), 696–720. <https://doi.org/10.1002/sce.21171>
- Salen, K., & Zimmerman, E. (2004). *Rules of play: game design fundamentals*. Cambridge, MA: MIT Press.
- Seidman, E. (2013). *Interviewing as qualitative research* (4th ed.). Teachers College Press.
- Shah, M. & Foster, A. (2015). Developing and assessing teachers' knowledge of game-based learning. *Journal of Technology and Teacher Education*, 23(2), 241–267. Waynesville, NC USA: Society for Information Technology & Teacher Education. Retrieved August 16, 2019 from <https://www.learntechlib.org/primary/p/147391/>.
- Silseth, K. (2012). The multivoicedness of gameplay: Exploring the unfolding of a student's learning trajectory in a gaming context at school. *International Journal of Computer-Supported Collaborative Learning*, 6(1), 63–84.
- Tsai, C., & Chai, C. (2012). The “third”-order barrier for technology-integration instruction: Implications for teacher education. *Australasian Journal of Educational Technology*, 28(6), 1057–1060.
- Ulicsak, M., & Williamson, B. (2010). *Computer games and learning*. Bristol: Futurelab. Retrieved from <https://www.nfer.ac.uk/publications/FUTL01>
- Vitale, M. R., McBride, E., & Linn, M. (2016). Distinguishing complex ideas about climate change: knowledge integration vs. specific guidance. *International Journal of Science Education*, 38(9), 1548–1569. <https://doi.org/10.1080/09500693.2016.1198969>
- Wooldrige, C. L., Bugg, J. M., McDaniel, M. A., & Liu, Y. (2014). The testing effect with authentic educational materials: A cautionary note. *Journal of Applied Research in Memory & Cognition*, 3, 214–221. <http://dx.doi.org/10.1016/j.jarmac.2014.07.001>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.