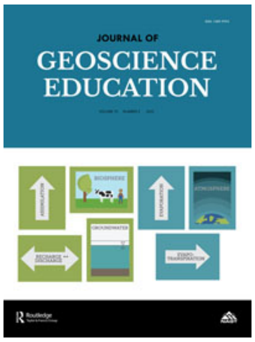

Journal of Geoscience Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ujge20

Learning outcomes of the educational board game "Taphonomy: Dead and Fossilized," evaluated with high school learners in a summertime program

Estefanía Salgado-Jauregui, Rowan C. Martindale, Katherine Ellins, Enrique Reyes & Anna Weiss

JOURNAL OF GEOSCIENCE EDUCATION
2022, VOL. 70, NO. 2, 176–194
https://doi.org/10.1080/10899995.2021.1965828

Learning outcomes of the educational board game "Taphonomy: Dead and Fossilized," evaluated with high school learners in a summertime program

Estefanía Salgado-Jauregui[a], Rowan C. Martindale[a], Katherine Ellins[b], Enrique Reyes[c] and Anna Weiss[d]

[a]Department of Geological Sciences, The University of Texas at Austin, Austin, Texas, USA;[b]Jackson School of Geosciences, The University of Texas at Austin, Austin, Texas; [c]Akins High School, Austin, Texas; [d]The Kimbell School of Geosciences, Midwestern State University, Wichita Falls, Texas, USA

ABSTRACT

Although many have suggested the use of games to motivate active learning, studies that evaluate the learning outcomes of games with high school students are scarce. Here, we present the evaluation of the board game "Taphonomy: Dead and Fossilized" as an active learning tool to teach fossilization and Earth systems thinking with rising 12th grade learners in GeoFORCE Texas, a summertime outreach program of the Jackson School of Geosciences, The University of Texas at Austin. The educational activity was evaluated with two groups ($n1 = 22$, $n2 = 27$). During the activity, an observation protocol was implemented; prompts to evaluate learners' behaviors and instructor behaviors were included in a form that a trained observer filled out while the learners played the game. Learning outcomes were assessed with a 2-page paper survey immediately following gameplay; survey questions are aligned with the Next Generation Science Standards. "Strategizing" was the most common learner behavior observed during the activity and the majority of behaviors can be considered "active learning". The results from the survey show that after playing the game learners were able to apply paleontological knowledge to tasks that involved establishing cause-effect relationships and Earth systems thinking. Our results provide evidence that board games (as educational strategies) are effective active learning tools that foster student development of scientific skills. Cooperative learning was observed, which we suggest is a key benefit for diverse classrooms. Findings were used to guide the refinement of the high school-level version of "Taphonomy: Dead and Fossilized", as well as a scaffolded teaching module with formative and summative questions for use in a classroom setting.

ARTICLE HISTORY
Received 20 January 2021
Revised 4 August 2021
Accepted 5 August 2021

KEYWORDS
Board game; active learning; multicontext theory; paleontology; high school

Purpose and learning goals

Recent studies have emphasized the need to incorporate educational strategies in high school classrooms that go beyond memorization, toward understanding, applying concepts, and developing scientific skills (St. John, 2018). Teaching faculty consistently agree that active learning promotes the development of higher-order thinking in high school learners (Batzri et al., 2015; McConnell et al., 2003; 2005) and recently, serious games have been proposed as active learning strategies to engage students in STEM fields (e.g., Foster, 2008; Martindale & Weiss, 2020; Nadolski et al., 2008). Games provide opportunities for students to participate in a social, educational environment (Hequet, 1995; Ramaley & Zia, 2005) as well as encouraging knowledge retention and creative problem solving (e.g., Bergen, 2009; Kumar & Lightner, 2007; Rieber, 1996; Wilson et al., 2009). That said, there is a paucity of data about the way in which the mechanics of serious games correlate with specific desirable learning outcomes (Wilson et al., 2009). Here, we present the evaluation of the board game "Taphonomy: Dead and Fossilized" with rising 12th grade GeoFORCE learners (i.e. students who have completed the 11th grade and not started 12th grade) as an active learning strategy in a collaborative setting to teach fossilization, the history of life, and Earth system thinking. Evaluation tools were designed and implemented to assess the extent to which the game is an effective active learning strategy for rising 12th grade learners participating in the GeoFORCE program (GeoFORCE Texas). We hypothesize that incorporating games in teaching will engage participants in active learning, facilitating not only an increase in content knowledge, but the development of scientific skills. Our results provide information about the efficacy of using board games with high school learners, in a cooperative setting; these data can be used to support the inclusion of games in instruction in both informal settings and formal Earth Science diverse classrooms.

CONTACT Estefanía Salgado-Jauregui esalgadoj@utexas.edu Department of Geological Sciences, The University of Texas at Austin, Austin, TX 78712, USA.

Geoscience education in Texas, and in the United States in general, is mostly restricted to the college level, with the exception of integration of concepts in other courses (e.g., volcanoes or earthquakes in general science classes). "Taphonomy: Dead and Fossilized" is a board game designed to teach University-aged players about the various factors that influence the process of fossilization and taphonomy (Martindale & Weiss, 2020). In this study, we modified the game for high school students and piloted it with the 2019 12th grade GeoFORCE Texas academy. We chose GeoFORCE Texas as a test case because, as a summertime program, it offers opportunities to implement innovative instructional strategies and new material to extend the learning that occurs in the academic year (National Academies of Sciences, Engineering and Medicine (NASEM), 2019). The game is described below in "Materials and Methods" and Figure 1 presents the materials used (e.g. game board, tokens and cards), as well as, a synopsis of the game's rules. An instructional video is also available at https://www.youtube.com/watch?v=MKcHnbbGtaQ. We posited that the game would help high school learners apply their paleontological knowledge, strengthen scientific skills, and be introduced to the challenges that arise when analyzing an imperfect or incomplete fossil record.

Learning goals

The first version of "Taphonomy: Dead and Fossilized" (hereafter, "the game") was designed as an alternative to a two-hour lab exercise to teach college undergraduate learners about taphonomy (Martindale & Weiss, 2020). Taphonomy, the modification of specimens as they transition from the biosphere into the geosphere (including death, decay, burial, and geological modification) (Efremov, 1940), is complex

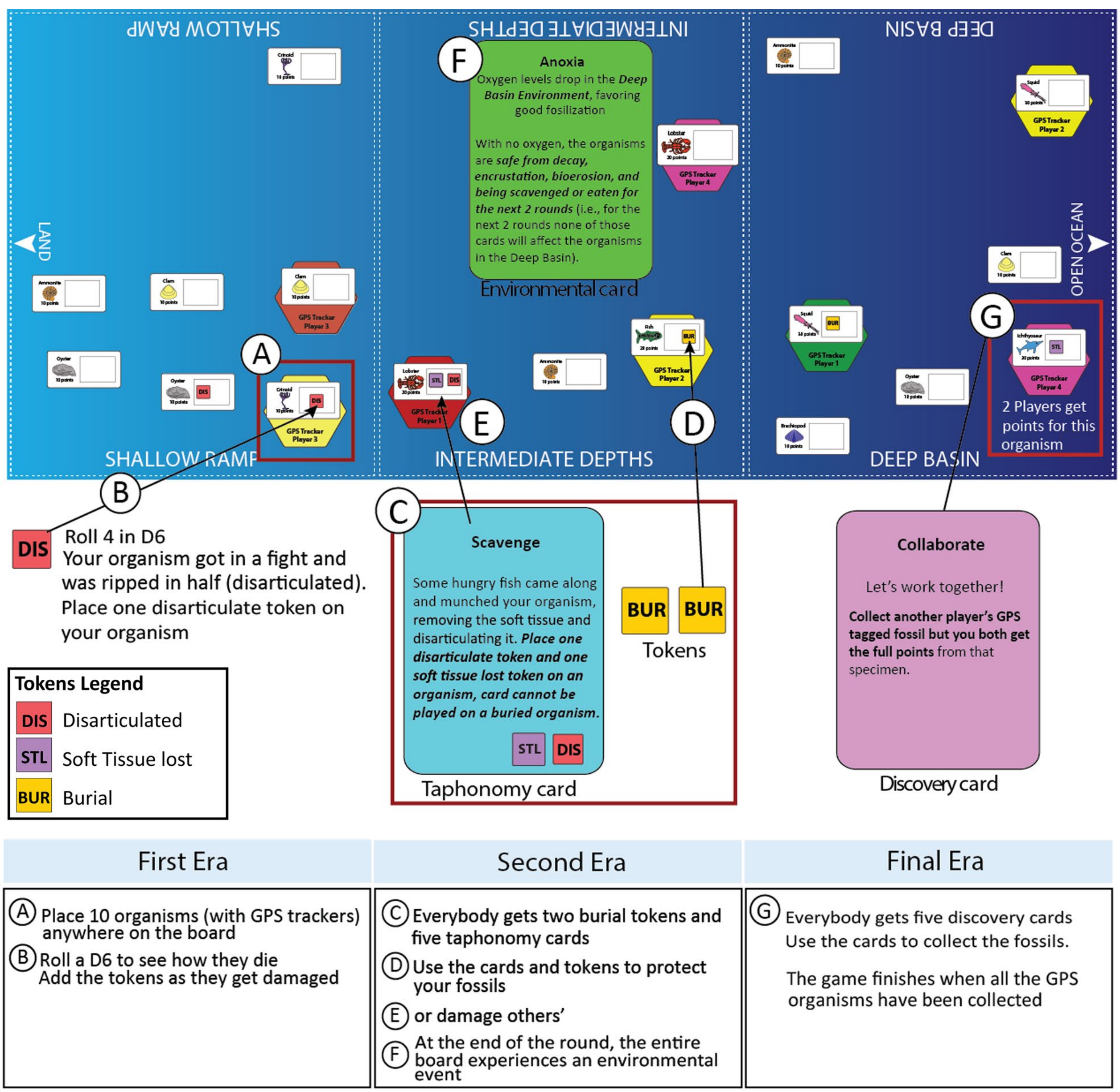

Figure 1. Summary of "Taphonomy: Dead and Fossilized" board game materials and phases of gameplay.

and involves multiple different factors, from an organism's physiological characteristics, to larger-scale environmental conditions (Briggs, 2014; Muscente et al., 2017; Seilacher, 1990), in other words, interconnected Earth systems. Martindale and Weiss (2020) assessed the effectiveness of the game as an educational tool in college-level classes at 20 institutions across the US showing that undergraduate learners had good conceptual knowledge following game play. Student answers on evaluation surveys demonstrated a clear understanding of how multiple taphonomic factors can influence the fossil record and the difficulty of preserving an organism as a fossil (Martindale & Weiss, 2020). In order to adapt the game for high school classes, the game mechanics were simplified. For example, the "Adoption phase" was removed (players start with 10 organisms not five); players also worked in two-person teams.

The game and the associated high school-level activities are aligned with the Next Generation Science Standards (NGSS), a set of research-based, K–12 science standards that capture the expectations for what students should know and be able to do for basic proficiency and continuing study in science (NGSS Lead States, 2013). Table 1 presents the expected learning outcomes of the game (first column); Student Learning Objectives one to four (SLO#1-SLO#4) are taken from Martindale and Weiss (2020), Student Learning Objectives five to nine (SLO#5-SLO#9) were added for this study. This educational board game provides opportunities to apply and develop science practices and crosscutting concepts, such as systems and systems models and differentiating between cause and effect (NGSS Lead States, 2013) (Table 1). Moreover, the learning goals align with the Texas Essential Knowledge and Skills (TEKS) for teaching high school science (Table 1). The Knowledge Skill [8] in section 112.36 Earth and Space Science in Texas education code (2014) states that learners are expected to analyze and evaluate a variety of fossil types and explain how the process of fossilization affects the degree of completeness of the fossil record. After playing the game, participants are expected to be able to identify physiological characteristics that make an organism more or less likely to become a fossil, recognize taphonomic factors that could enhance or diminish the preservation potential, identify environmental events that impact the fossilization process, and understand how chance and sampling biases affect fossil collections (Martindale & Weiss, 2020).

Literature context

Evidence-based research in different disciplines, such as learning sciences, cognitive science, and educational psychology, support active learning and student-centered education over traditional, passive lectures (Michael, 2006). Active learning is often defined as any instructional method where students are involved in more than listening; they are doing meaningful learning activities and thinking about what they are doing (Bonwell & Eison, 1991; Prince, 2004). Emphasis is given in developing students' skills over transmitting information, as well as engaging students in higher order thinking activities (Bonwell & Eison, 1991; Prince, 2004). Games have the potential to engage learners in active learning and all three domains of Bloom's taxonomy: cognitive, affective, and psychomotor (Weigel & Bonica, 2014, and Martindale & Weiss, 2020). For example, a passion for winning (affective) engages learners at the cognitive level, or a collaborative setting engages participants socially as they discuss and strategize moves.

GeoFORCE learners are predominantly Hispanic (~62%) or from underrepresented groups and under-resourced schools in Texas. Multicontext theory seeks to address conflicts that may occur between a student's cultural background and the educational environment in which learning takes place, shifting attention onto a more inclusive academic culture (Ibarra, 1999, 2001). Educational approaches inherited from Germanic and Northern European cultural roots (i.e., low context, *sensu* Weissmann et al., 2019) still prevail in the academic culture of the U.S. (Ibarra, 1999, 2001; Weissmann et al., 2019). Examples of these approaches include activities that value individualism, emphasize predominantly faculty-oriented perspectives, and promote lineal cause-effect thinking. The incompatibility between traditional science cultures and the cultures of underrepresented groups has been largely acknowledged (Aikenhead, 1996, 1997; Riggs, 1998; Riggs & Semken, 2001; Wolfe & Riggs, 2017, as cited in Pfeifer et al., 2021). Multicontext interventions recognize that learning styles can vary depending on students' cultural backgrounds, which are instilled in them during childhood by family and community (Hall, 1977; Pfeifer et al., 2021). Such interventions have shown high levels of student achievement, independent of a person's race, gender, or socio-economic backgrounds (Weissmann et al., 2019). Research suggests that students from high context cultures can be expected to be especially engaged in cooperative learning and systems thinking (Kagan, 1986; Pfeifer et al., 2021; Slavin, 1983; Weissmann et al., 2019). The board game "Taphonomy: Dead and Fossilized" was implemented with GeoFORCE Texas learners as a multicontext educational activity (*sensu* Weissmann et al., 2019), to encourage them to work cooperatively (high context), apply systems thinking (high context), and understand cause-effect relations (low context).

In this study, the board game was played in teams. Two students played collaboratively, competing with other teams around the table to win. Numerous studies dating back to the late 1800s validate cooperative learning (Johnson & Johnson, 2008). Cooperative learning exists when students work together to achieve a shared set of learning goals (Johnson et al., 1992). A meta-analysis examining over 600 research studies by Johnson and Johnson (1986) shows that when students work cooperatively, they enjoy the subject matter more, have higher levels of self-esteem, and are more inclusive and accepting of diversity. "Taphonomy: Dead and Fossilized" provides learners with opportunities to analyze and synthesize ideas cooperatively around the common goal of having the largest and most pristine collection of fossils at the end of the game. Social interdependence theory established that for cooperation to exist, an individual's actions

Table 1. Cross-comparison of students learning outcomes (SLO) with performance expectations for High School Earth and Space Sciences in the NGSS and expectations from the Texas Essential Knowledge and Skills (TEKS).

Learning outcomes		Next Generation Science Standards (NGSS)		Texas Essential Knowledge and Skills (TEKS)
Knowledge				
1	SLO#1: Identify physiological characteristics that make an organism more or less likely to become fossilized	**HS-ESS2-2** Analyze geoscience data to make the claim that one change to Earth's surface can create feedbacks that cause changes to other Earth systems		**(Knowledge skill [8A] in section 112.36 Earth and Space Science)** Analyze and evaluate a variety of fossil types such as transitional fossils, proposed transitional fossils, fossil lineages, and significant fossil deposits with regard to their appearance, completeness, and alignment with scientific explanations in light of this fossil data.
2	SLO#2: Identify the climate, oceanographic, and geological events that occur in different marine environments and describe the effect they have on the preservation potential of fossils in that setting	**HS-ESS2-2** Same as above		**(Knowledge skill [8B] in section 112.36 Earth and Space Science)** explain how sedimentation, fossilization, and speciation affect the degree of completeness of the fossil record.
3	SLO#3: Describe multiple taphonomic factors that would impact an organism as it fossilizes in a marine setting and determine if they would enhance or diminish the preservation potential	**HS-ESS2-2** Same as above		**(Knowledge skill [8B] in section 112.36 Earth and Space Science).**
4	SLO#4: Describe how chance and sampling biases affect fossil collections	**HS-ESS2-2** Same as above		None
Skills				
1	Make basic calculation to determine original and final diversity based on number of taxa.	Scientific and engineering practice	Analyze and interpret geoscience data	None
2	Identify the effects of taphonomic factors and environmental events in fossil preservation.	Crosscutting concepts	Cause and effect: Mechanisms and explanation	None
3	Identify the boundaries, components and interactions in time and space among components of the system presented in the game	Crosscutting concepts	Systems and system models	None
4	**Thinking temporally: retrospection and prediction. Infer past events that can explain current fossil conditions			
5	**Identify short and long-term effects of environmental disturbances on other Earth systems			

** Not incorporated in the implementation/evaluation of the game with GeoFORCE, but included in the new HS version of the game

must be affected by the actions of other individuals in a bidirectional sense (Johnson & Johnson, 2002). Collaboration creates positive interactions (Johnson & Johnson, 2002); in the game, these interactions manifest as the two-person team promoting each other's learning to win the game. Competition creates negative interactions (Johnson & Johnson, 2002); in the game, negative interactions manifest as teams playing against each other since only one team can win the game.

Earth systems thinking is another important high context component of "Taphonomy: Dead and Fossilized" (Weissmann et al., 2019). The game itself represents the system in which fossilization takes place. It has clear boundaries established in time (i.e., First, Second and Final Era) and space (i.e., deep, intermediate, and shallow ocean basin) where components interact (see Figure 1). Systems thinking has been recognized as a desirable scientific skill for high school learners (NGSS Lead States, 2013) and a high context cultural value (*sensu* Pfeifer et al., 2021) with great potential to engage learners from underrepresented groups in sciences (Pfeifer et al., 2021; Weissmann et al., 2019). The game provides learners with opportunities for high context-orientated students to study dynamic relations between components, a very important and usually overlooked characteristic of systems (Assaraf & Orion, 2005). To advance our understanding on how to incorporate and evaluate Earth Systems thinking in geosciences education, we use sketches as an evaluation instrument. Evidence suggests that drawings provide valuable information on the level of the student's understanding of natural phenomena (Dove, Everett, & Preece, 1999). Assaraf and Orion (2005) provide an example of how sketches are useful to evaluate Earth systems thinking, (1) identifying the system's components and processes; (2) the dynamic relationships within

Table 2. Roles and academic background of the educational and evaluation team for group 1 and group 2.

Educational and evaluation team	Role	Group 1	Group 2
Instructor	Lead the activity	Psychologist with extensive experience in elementary education	Geosciences graduate student at the University of Texas at Austin
Educational coaches	Support instructor and students in learning process	Individual recently graduated from college, or currently enrolled at a four-year college or university.	
Observers	Report behaviors of students and instructor in observation protocol	Graduate student in geosciences	A. Earth sciences high school teacher B. Person with a Ph.D. in Geosciences

the system; and (3) the existence of the human aspect. A discussion on how the sketches were used in this study is provided on the "Evaluation" section of the paper.

Finally, beyond student evaluation surveys, we incorporated an evaluation instrument (i.e., the observation protocol based on Smith et al. (2013)) to measure learning effectiveness as recommended by the Association of American Universities (2011). In this study we focused the observation on identifying students' behaviors that reflect active learning. We selected a structured observation protocol, because it provides standardized results that can be compared among classrooms and observers (e.g., Sawada et al., 2002; Smith et al., 2013; Weiss et al., 2003).

Study population and setting

We implemented the board game and associated materials with rising 12th grade GeoFORCE Texas learners. GeoFORCE Texas (https://www.jsg.utexas.edu/geoforce/) is a summertime science outreach program administered by the Jackson School of Geosciences at The University of Texas at Austin; it offers enriched learning experiences to about 320 learners from high schools that serve under-resourced communities from Southwest Texas and Houston. Participants enter the program at the end of 8th grade and are encouraged to remain in GeoFORCE Texas for four years. The students participating are predominantly from underrepresented groups with 62% identifying as Hispanic, 16% White, 13% African American, 7% Asian, and 2% Other (GeoFORCE, 2020). These statistics are a good representation of the 12th grade academy demographics. The 12th grade GeoFORCE academy take place once a year, every summer for nine days. Instruction follows a challenge-based learning model, the STAR Legacy Cycle (Ellins et al., 2018) and supports active learning in outdoor field, indoor laboratory, and classroom settings. Challenges with accompanying goals and specific learning activities that support these goals drive student learning that align with the Next Generation Science Standards (NGSS Lead States, 2013).

Learning in GeoFORCE academies is facilitated by a team of instructors who function as mentors and coaches (Ellins et al., 2018). Table 2 presents the roles and educational background of the team who lead the implementation and evaluation of "Taphonomy: Dead and Fossilized". A session aimed to introduce the game to members of the GeoFORCE Texas educational team took place three months before the activity was implemented with learners. During this session, the instructors, the coordinator of group 1, and two of the observers played the game and familiarized themselves with game mechanics and materials. Observer A from group 2 played the game in a previous opportunity. The protocol for the lecture with a detailed description of the activities to follow, the observation protocol, as well as the lab procedure summarizing the most important rules of the game were shared with instructors and observers one week prior to the 12th grade experience.

The board game was included as one of the in-classroom educational activities of the GeoFORCE Texas 12th grade academy. Forty-nine learners participated in the activity; they were divided in two groups ($n1 = 22$, $n2 = 27$), and each group played the game in a different session. The session started with the educational team setting up the board games at tables, including booklets with instructions, cheat sheets, tokens, cards, dices, and score sheets. GeoFORCE Texas learners played in self-selected teams of two with four teams per table. Group 1 was organized in three tables (two "four-team" tables, and one "three-team" table); group 2 was organized in four "four-teams" tables. Three teams in group 2 had three participants. The instructor guided the participants through each round, following the lab procedure provided; the observers registered the learners' and instructors' behaviors following the observation protocol discussed below. In group 1, the activity lasted one hour and forty minutes; in group 2, one hour and five minutes. As soon as the learners finished playing the game, they answered the evaluation survey provided.

Materials and methods

General game overview (see Martindale & Weiss, 2020, for further details)

This study was conducted under IRB approval from the Office of Research Support & Compliance at The University of Texas at Austin. All the materials used to evaluate the modified game with GeoFORCE Texas learners can be found at DOI: https://doi.org/10.18738/T8/R2KSCY. The board represents a marine environment (ramp) where different organisms lived and died during the Jurassic Period. The game is played through geological time and the participants are time travelers. They start the game with 10 different organisms (represented by tokens), which they track with GPS tokens (Figure 1). Their mission is to protect their organisms and advance through the game with the best collection possible. The modified version of the game played with GeoFORCE Texas rising 12th grade learners consist of three Eras. In the First Era, the organisms die; there are different ways and places to die, and this has a direct impact on the preservation of the organisms in the fossil record. The players roll a dice to determine how each of the organisms die, and then they place them in the board (A and B in Figure 1). In the Second Era, participants have a hand of five randomly drawn taphonomy cards and two burial tokens (C in Figure 1). Players take turns to use the taphonomy cards and burial tokens to protect their specimens or attack and damage the specimens of their opponents (D and E in Figure 1). After one round of taphonomy events, one environmental card is drawn from the deck. This card impacts all the organisms in the board (F in Figure 1). In the Final Era the participants discover the fossils (again, with a randomly drawn hand of cards), learning about how human bias and chance impact fossil collection (G in Figure 1). The winner is the participant that ends with the largest, most diverse, and pristine fossil collection. There are bonuses for different collection achievements, for example, having five or more specimens of one fossil taxa. The materials associated with the game included a 10-minute instructional video to guide the participants through gameplay (available at https://www.youtube.com/watch?v=MKcHnbbGtaQ), the game booklet with all the information needed to play, a cheat sheet with the point value assigned to each fossil and the different ways in which a participant can win or lose points, and a score card.

Why certain modifications were selected for GeoFORCE Texas

Martindale and Weiss (2020) report that the game typically takes one to two hours in a college-level setting but present several options for a shorter game. Given that only two hours were available to implement the game and answer the educational survey in GeoFORCE Texas 12th grade academy, a shorter version of the game was played. We decided to follow the third alternative to shorten the game: omitting the organism's adoption phase (Martindale & Weiss, 2020). In the original version of the game, participants have the opportunity to adopt untagged organisms from the board; this phase of the game allows players to learn from past mistakes or successes and use their new knowledge to adopt the fossils that will make their collection the best. The Fourth Era is essentially the same as the Second Era (participants protect both their original and adopted fossils). In the modified version of the game, played with GeoFORCE Texas, the learners started the game with 10 organisms, and they did not adopt additional organisms during the gameplay. The rest of the Eras were not modified (Eras 2/4 were played as one Era with four rounds total), so that in the Final Era the players could still discover fossils. Although this modification does not allow the players to revise their strategy by adopting new fossils, it does include all the learning outcomes (Martindale & Weiss, 2020). Furthermore, there are still opportunities for learners to apply their knowledge; for example, attacking the most valuable organisms of their opponents or collaborating with other participants to excavate the best-preserved specimens.

We developed a one-page lab procedure synthesizing the goals, materials, setting up, and gameplay. We also provided the GeoFORCE Texas educational team with a protocol describing what 12th grade learners, instructor and observers were expected to be doing before, during, and after playing the game. These, and other materials developed based on the lessons learned from the implementation of the activity with GeoFORCE, are now included in the high school version of the game, which can be found at DOI: https://doi.org/10.18738/T8/USRIGL

Evaluation

The learning outcomes of the game were evaluated following an observation protocol modified from Smith et al. (2013) and an evaluation survey completed by the GeoFORCE Texas learners after playing the game. The observation protocol aims to document learner and instructor engagement during the activity, while the survey includes questions to evaluate the ability of learners to apply paleontological knowledge to recognizing cause-effect relationships and Earth systems thinking.

The original observation instrument (Smith et al., 2013) was selected from the instruments and surveys compiled in the Geoscience Education Researcher Toolbox and adapted to the specific needs of this research. The observation protocol implemented includes 10 prompts describing learners' behaviors and seven prompts describing instructor behaviors (Figure 2). The prompts "asking questions about the rules" and "asking questions about the content" were not originally included in the protocol presented in Smith et al. (2013). They are included here as two separate prompts for three reasons: (1) to evaluate when learners had difficulty understanding the rules, (2) to determine student engagement in their own learning process, and (3) to assess when learners required help understanding the geoscientific concepts that underpin the game (i.e., in which Era of the game). We also added the prompt "learners are distracted by non-related activities", to have an option in the protocol to track learners who were not engaged in their learning process; the only prompt that achieves this goal in Smith et al. (2013) is "learners waiting", which is also included in our observation protocol. Several prompts from Smith et al. (2013) were not included because those behaviors are not expected in the implementation of this activity (e.g., "thinking or discussing clicker questions"). The modified observation protocol, presented here, focuses on the students' experience, including 10 prompts to track students' behaviors, and only seven prompts for instructors' behaviors.

An observer registered the behaviors of the learners and instructor in the protocol provided (Figure 2) while the game was played by GeoFORCE Texas learners. At five-minute intervals throughout the activity, the observer selected the students' predominant behaviors, as well as the instructor's behavior in the form provided (Figure 2).

	Students doing										Instructor doing							Round	Comments
min	L	AnQ	QR	QC	WC	Prd	TQ	W	D	O	L	SM	PQ	AQ	1o1	Adm	O		
0-5																			
5-10																			
10-15																			
15-30																			

12th grade rising learners are doing	
L	Listening to instructor
AnQ	Answering a question posed by the instructor with rest of class listening
QR	Asking questions about the rules of the game
QC	Asking questions about the content of the game
WC	Engaged in whole class discussion by offering explanations, opinion, judgment.
Prd	Making a prediction about the outcome of the game or planning strategies.
TQ	Test or quiz
W	Waiting (instructor late, working on fixing AV problems, instructor otherwise occupied, etc.)
D	Distracted in non-related activities (Looking at the phone, talking about not related topics)
O	Other (explained in comments)
Instructor is Doing	
L	Lecturing (e.g., explaining the rules of the game)
SM	Using supporting material (videos, board, projector)
PQ	Posing a guiding question to learners
AQ	Answering student questions with other learners listening
1o1	One–on–one extended discussion with one or two individuals, not paying attention to the rest of the class
Adm	Administration (e.g., assigning a test)
O	Other (explain in comments)

Figure 2. Sample of the observation protocol used to evaluate "Taphonomy: Dead and Fossilized" and definition of the codes to describe learner and instructor behaviors.

Additionally, the observer was instructed to use the comments section if they noted some relevant behavior that was not included in the form. This form provided valuable quantitative information of what was happening during gameplay.

The evaluation tools used to assess the learning outcomes of the game also included a survey that was completed by the learners immediately after playing the game. The questions included in this survey align with the scientific practices that are expected from high school learners (NGSS Lead States, 2013). We included questions to evaluate the ability of learners to (1) make cause-effect predictions and (2) engage in Earth systems thinking, making explicit a model of a system, specifying its boundaries, components and interactions among these components (NGSS Lead States, 2013).

Martindale and Weiss (2020) implemented an online survey to assess undergraduate students and instructors self-reported opinions about the game. The authors included 15 questions to assess content knowledge. These questions inform the development of the learners' evaluation survey used on this study (Table 3). We adapted the questionnaire to assess learner's scientific skills, along with content knowledge.

Question #1 evaluates students' ability to analyze simulated geological data. Students are asked to determine if the final diversity of their fossils collection was representative of the original diversity, comparing the number of taxa (i.e., types of organism tokens) at the beginning of the game and the number of taxa in their collection at the end of the game. We did not include this question in our results or analyses because it was clear from the learners' responses that they misinterpreted the question. For example, some learners compared their final diversity to the original diversity of their 10 organisms, not the entire board, and others counted the total number of organisms that they had in their collection at the end of the game, as opposed to the number of different types of organisms. Therefore, this question was not used to evaluate learning.

The Geosciences Concept Inventory (GCI) presents 69 multiple-choice validated questions for use in college entry-level geoscience courses (Libarkin & Anderson, 2006). Each GCI question has gone through rigorous reliability and validation studies, including scale development theory, grounded theory, and item response theory (Libarkin & Anderson, 2006). We included question [38] from the GCI in our evaluation survey (question #2 in Table 3) to assess learners' understanding about the incompleteness of the fossil record. In question #2 learners are

Table 3. Questions from the survey implemented by Martindale and Weiss (2020) used as reference for the questions of the survey used in this study.

Martindale and Weiss (2020)	This study
[C12] The diversity of the fossil record reflects the original diversity of the community	1. Is the diversity of your collection representative of the original diversity? A. Original diversity-OD (Number of taxa at the beginning of the game): 9 B. Collection diversity -CD (Number of taxa that you collected) _____ Percentage (%) (CD*100/OD) _____
[C11] The fossil record is complete (i.e., there is no loss of information during fossilization) [C9] It is easy for an organism to become a fossil	2. A scientist collects all of the fossils ever discovered into one room. This room now contains A. Fossils of a few plants and animals that ever lived B. Fossils of most of the plants and animals that ever lived C. Fossils of most of the types of plants and animals that ever lived D. Fossils of all of the plants and animals that ever lived E. Fossils of all the types of plants and animals that ever lived
[C7] Fossils that are protected from scavengers are more likely to fossilize [C10] When an organism is fossilized it is often perfectly preserved [C4] Random events influence whether an organism is fossilized	3. Match the environmental event with the effects that it produces on the system (Note that each event produces several effects. Select all that apply) A. Tsunami ___ Specimens protected from erosion B. Fast sedimentation ___ Specimens moved to a deeper environment C. Decay ___ Specimens disarticulated D. Scavenge ___ Lost of soft tissue
[C2] Burial influences whether an organism is fossilized [C6] Fossils that are buried faster are more likely to fossilize [C5] How an organism dies influences whether it is fossilized [C8] Organism soft parts have a high chance of being fossilized	4. Select all the statements that are true A. How an organism dies does not influence whether is fossilized B. Fossils that are protected from scavengers are more likely to fossilize C. Organisms soft parts have a high chance of being fossilized D. Fossils that are buried faster are more likely to fossilize.

expected to infer that if a scientist had all the fossils ever discovered in one room, this collection will only contain (A) fossils of few of the plants and animals that ever lived. In the game players learn how difficult it is for an organism to be preserved as a fossil by protecting their fossils from different taphonomic and environmental factors and human biases.

Question #3 was posed to evaluate the ability of learners to establish cause-effect relationships from the game; learners were asked to match the impact that different taphonomic factors and environmental events may have on the specimens through the fossilization process. This question is based on three taphonomic cards: fast sedimentation, decay, and scavenge, as well as one environmental event card: tsunami. Each card contains a detailed description of the impacts it will have on the organisms on the board. For example, the tsunami card states: "The Earth quakes and there is a massive tsunami! All organisms in the Shallow Ramp and Intermediate Depth are moved one environment deeper and disarticulated…If an organism is not buried and already disarticulated, they are lost to the fossil record". In question #3, the learners are expected to match the event tsunami with the impact that it causes on the organisms, the specimens are moved to a deeper environment and disarticulated.

In question #4, the learners are expected to infer cause-effect relations based on their experience during gameplay. They are presented with four statements written in a cause-effect format and asked to identify the true statements (question #4 in Table 3). The demonstrated knowledge includes: (1) different ways to die impact the fossilization process, (2) burial protect organisms and (3) hard tissues (skeleton or shell) are more likely to be preserved than soft parts (skin or muscle).

While "Taphonomy: Dead and Fossilized" teaches learners about cause-effect relationships in the fossilization process, the game goes beyond these lineal relations to be a model of a system where complex processes occur in time and space. Question #5 assesses learners' Earth systems thinking, asking them to draw a diagram of the system presented in the game, including the parts of the system, connections among these parts, and labels or annotations to explain how different parts of the game's scenario work together. The rubric designed to grade this question (Table 4) and sample drawings were discussed, evaluated, and agreed upon by three of the authors. The rubric incorporates systems thinking characteristics [1], [2], and [8] (*sensu* Assaraf & Orion, 2005, p. 523). We evaluated learners' abilities to (1) identify the main components of the system (e.g., boundaries and organisms); (2) identify dynamic relations among components (e.g., modification of organisms and environmental events), and (3) think temporally (e.g., basic understanding of change over time and geological time) (Table 4).

Results

Observation protocol of learner and instructor behaviors

Figure 3 displays learner behavior as a percentage of 'total time', calculated with the Equation [Total time = 5 * (number of 5 min slots checked by the observer in the protocol)]. The percentages in the charts (Figures 3, 4, and 5) represents how frequently each observer marked the behaviors in their form. On average the observers reported three students' behaviors and two instructor's behaviors in a 5 minutes slot. Behaviors that reflect active learning (i.e., making strategy, asking questions about the rules, asking questions about the content, engaged in whole class discussion and answering questions) were predominant in both groups (colored red) when compared with behaviors that reflect passive learning, such as listening to instructions, distracted and waiting (grey colors in Figure

Table 4. Rubric proposed to evaluate learners Earth systems thinking in question #5.

Identify the components of the system	Categories: 1. Boundaries (ocean basin, deep, intermediate, shallow) 2. Organisms (labeled?) 3. Other physical components (sediments, water)	Basic Good Excellent	Diagram includes components from one of the categories Diagram includes components from two of the categories Diagram includes components from three categories
Ability to identify dynamic Relationships among Components	Categories: 1. Different ways to die 2. Modification of organism (e.g., movement, burial, dissolution) 3. Environmental events (e.g., Tsunami, Anoxia)	Basic Good Very Good Excellent	Diagram includes dynamic relationships from one of the categories Diagram includes dynamic relationships from two of the categories Diagram includes dynamic relationships from three or more categories or learners identify one or two dynamic relations and they locate them in the space Diagram includes dynamic relationships from three categories and locate them in space (e.g., Tsunami moving organisms from shallow to deeper water)
Thinking temporally (retrospection and prediction)		Basic Good	The diagram includes processes evidencing temporal thinking The diagram includes processes evidencing temporal thinking and a basic understanding of geological time

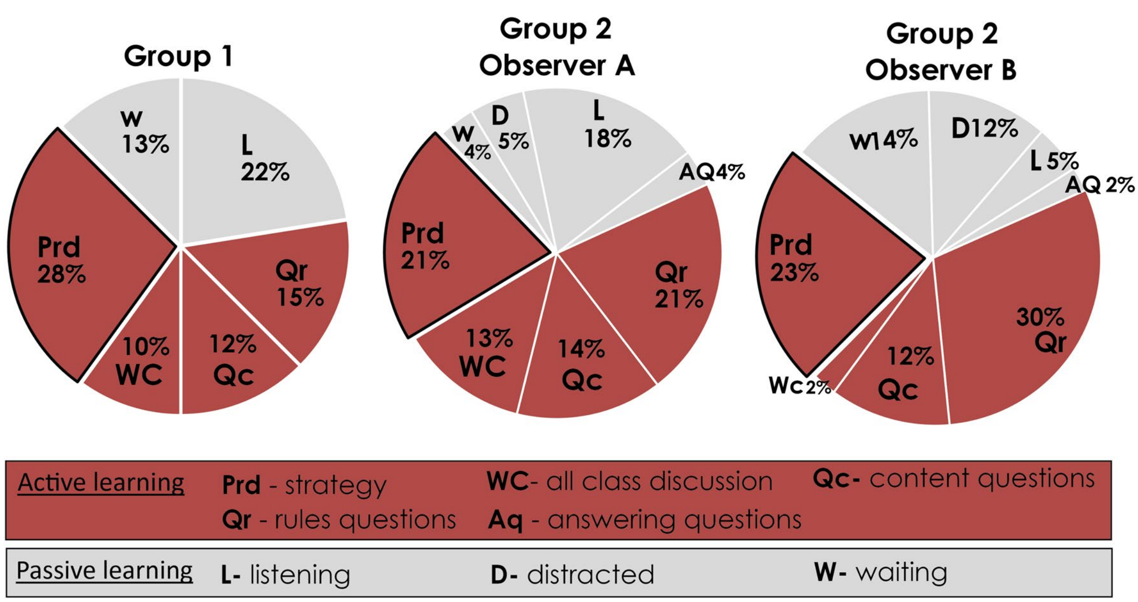

Figure 3. Learners' behaviors while playing the board game observed in two groups of GeoFORCE Texas participants (group 1 n = 22, group 2 n = 27). The figures displayed the results provided **by one** observer in group 1 and two observers in group 2. One hundred percent is the 'total time' learners were engaged in the behaviors reported by the observers in the form (as opposed to the total time that the activity lasted); when the observer reported more than one behavior in a five-minute slot, these were counted independently.

3). In group 1, learners were engaged in behaviors reflecting active learning 65% of the "total time", whereas in group 2, Observer A reported that learners were engaged in active learning behaviors 73% of the "total time", while Observer B reported these behaviors 69% of the "total time". In group 1 the most reported behavior was making predictions about the outcome of the game, or strategizing (Figure 3). Observer A in group 2 also reported strategizing as the most observed behavior, but Observer B reported it as the second most common behavior, after learners asking questions about the rules of the game.

In both groups observers reported that instructors spent less than half of the "total time" lecturing (L in Figure 5). In group 1 the instructor lectured for 43% of the "total time", in group 2, Observers A and B reported that instructors were lecturing for 28% and 33% of the "total time", respectively (Figure 5). In group 1 the observer marked the behavior "lecturing" with the same frequency as the activities that reflect active learning, such as interacting with learners in one-on-one discussions or answering and posing questions. In group 2, both observers reported predominance of behaviors reflecting active learning over lecturing. Differences were observed in the instructor behaviors reported by Observers A and B in group 2. With the exception of lecturing and one-on-one discussion, Observers A and B did not agree on the percentage of total time that the instructor spent in each of the other behaviors. Some of these observations represent substantial differences.

To provide a quantitative estimate on the similarity of observations made by Observers A and B, we calculated Jaccard similarity scores for each code, using Equation [2]. The results are presented in Table 5. Jaccard score values close to one indicate the greatest similarity between observers, while low scores indicate discrepancy among observers (Smith et al., 2013). Our observation protocol closely

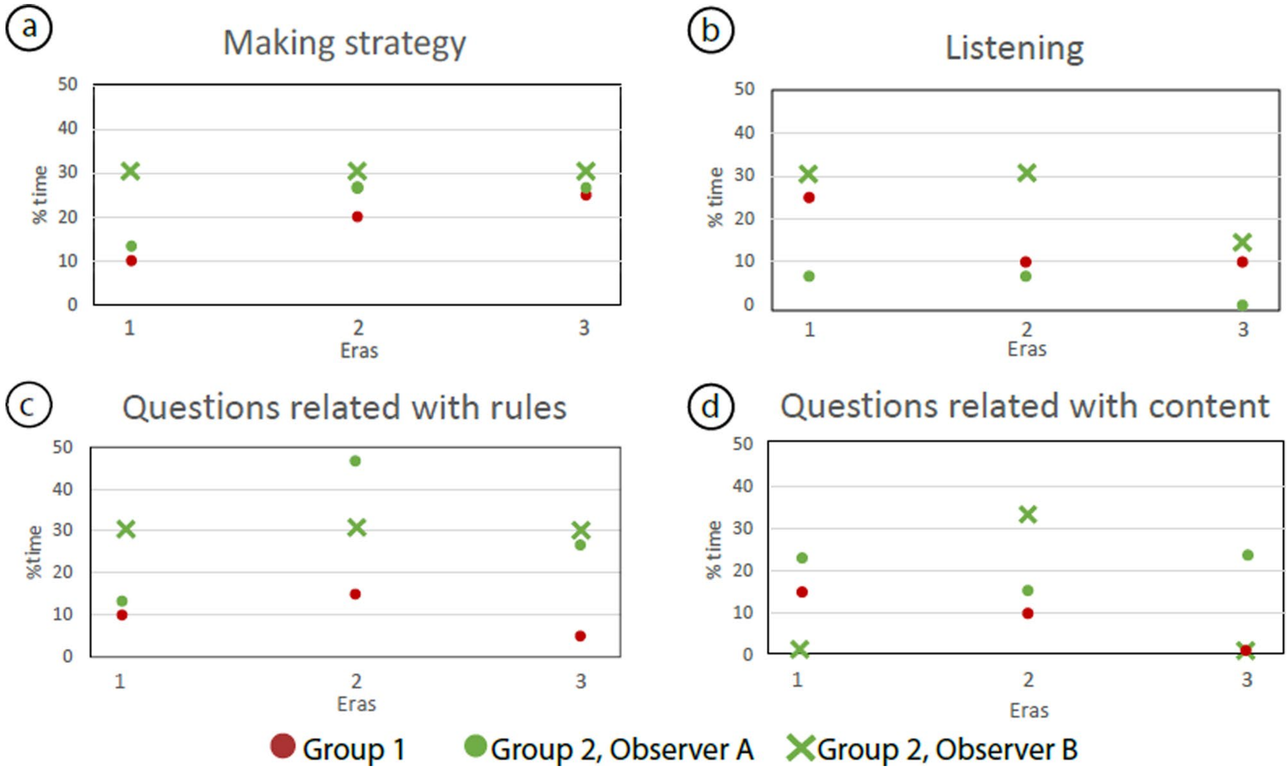

Figure 4. Behaviors observed in each of the three Eras of the game, in two groups of GeoFORCE learners (group 1 n = 22, group 2 n = 27). The figures show the percentage of time that each behavior was observed with respect to the total duration of the game in the First, Second, and Third or Final Era. A) Learners were listening to instructor; B) learners were making a prediction about the outcome of the game or planning strategies; C) learners were asking questions about the rules of the game; D) learners were asking questions about the content of the game.

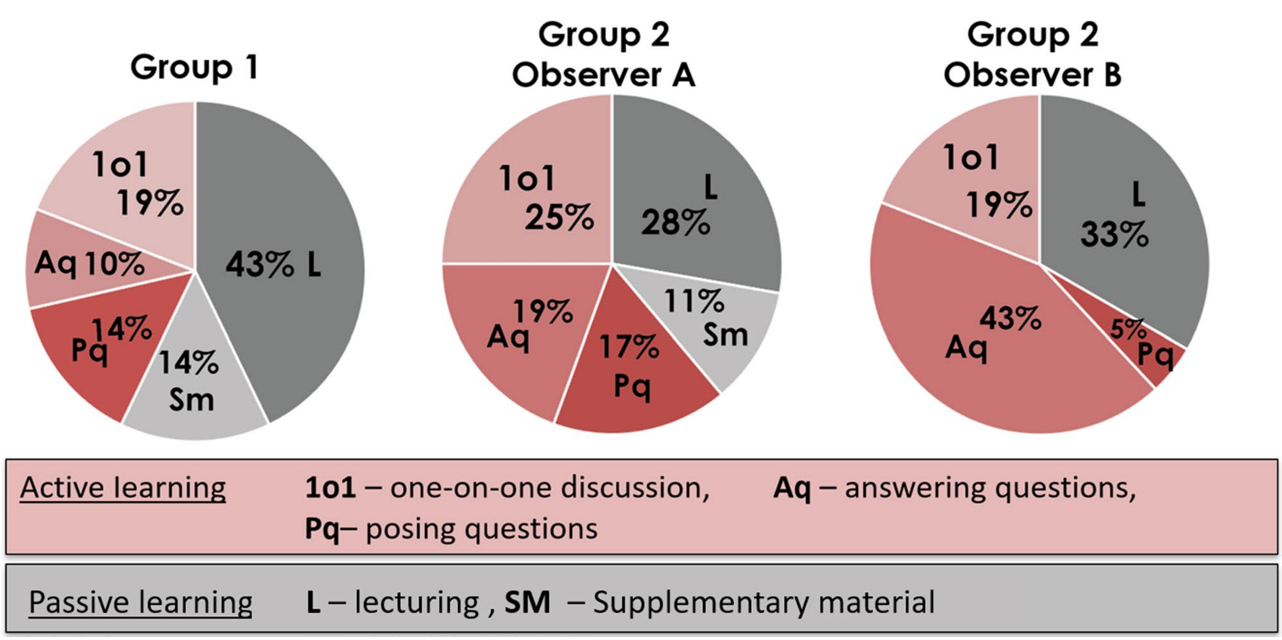

Figure 5. Instructors' behaviors observed in two groups of GeoFORCE Texas learners while playing the board game. The figures displayed the results provided by one observer in group 1 and two observers in group 2. One hundred percent is the total time learners were engaged in the behaviors reported by the observers in the form (as opposed to the total time that the activity lasted); when the observer reported more than one behavior in a five-minute slot, these were counted independently.

Table 5. Jaccard similarity scores (T) for each code of the observation protocol across the two observers reporting for Group 2.

Students behaviors		T
L	Listening to instructor	0.5
AnQ	Answering a question posed by the instructor with rest of class listening	0.8
QR	Asking questions about the rules of the game	0.7
QC	Asking questions about the content of the game	0.4
WC	Engaged in whole class discussion by offering explanations, opinion, judgment.	0.5
Prd	Making a prediction about the outcome of the game or planning strategies.	0.6
TQ	Test or quiz	1
W	Waiting (instructor late, working on fixing AV problems, instructor otherwise occupied, etc.)	0.7
D	Distracted in non-related activities (Looking at the phone, talking about not related topics)	0.6
Instructor is Doing		**T**
L	Lecturing (e.g., explaining the rules of the game)	0.4
SM	Using supporting material (videos, board, projector)	0.7
PQ	Posing a guiding question to learners	0.7
AQ	Answering student questions with other learners listening	0.5
1o1	One-on-one extended discussion with one or two individuals, not paying attention to the rest of the class	0.3
Adm	Administration (e.g., assigning a test)	1

resembles the Classroom Observation Protocol for Undergraduate STEM (COPUS), presented in Smith et al. (2013). The COPUS was designed with the intention of not requiring multiday training periods for observers to achieve valid results (Smith et al., 2013). Our results show an important level of agreement (i.e. average Jaccard score of 0.6 for the students' behaviors and instructor's behaviors), even though the observers did not participate in a training period. Moreover, for increasing the level of agreement we recommend a training period for observers.

$$T = \frac{n_c}{n_a + n_b + n_c} \tag{2}$$

Where T = Jaccard score, nc = number of 5 min slot boxes marked the same by observer A and B, na = nc + number of 5 min slot boxes marked by observer A, not marked by observer B, nb = nc + number of 5 min slot boxes marked by observer B, not marked by observer A.

Survey answered by GeoFORCE learners

Figure 6 illustrates the rising 12th grade learners' answers to question #2: "*A scientist collects all of the fossils ever discovered into one room. This room now contains:*" (from the GCI; Libarkin & Anderson, 2006, 2008). Most of the learners in group 1 selected the correct answer, A: Fossils of a few of the plants and animals that ever lived. Most of the learners in group 2 selected answer C (Fossils of most of the types of plants and animals that ever lived), with A being the second most selected answer. These results show that learners have misconceptions regarding the total number of organisms that ever lived and the percentage of these organisms that fossilized.

Questions #3 and #4 assess the ability of the GeoFORCE Texas learners to apply paleontological knowledge to establish cause-effect relationships. In both groups, learners showed good performance on these questions (Figure 7) with more than 50% answering the four components of question #3 correctly in both groups (more than 59% in group 2). Establishing that a tsunami or the action of scavengers were causes of having a fossil disarticulated was the most challenging concept for learners in both groups (Figure 7). In question #4 the majority of the learners (95% in group 1, 85% in group 2) were correct when selecting the statements that were true: A. Fossils that are protected from scavengers are more likely to fossilize and fossils that are buried faster are more likely to fossilize.

To evaluate question #5, we used the rubric presented in Table 4. The learners were better at identifying the components of the system than establishing relationships among the components or showing evidence of temporal thinking

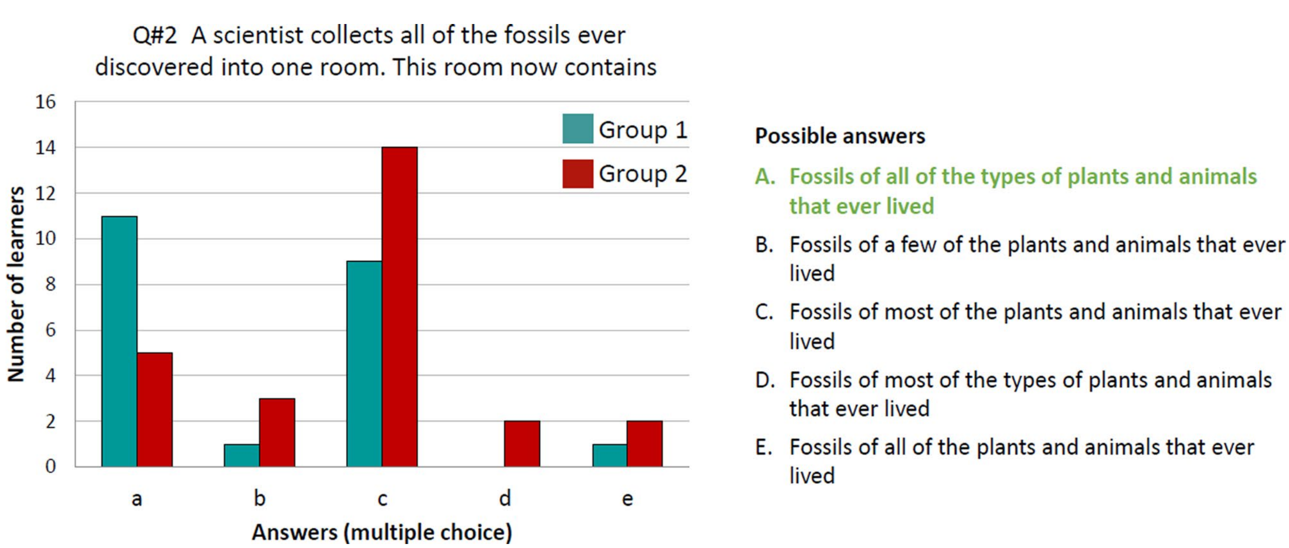

Figure 6. Summary of answers to survey question #2 by GeoFORCE learners. (group 1 n = 22, group 2 n = 27).

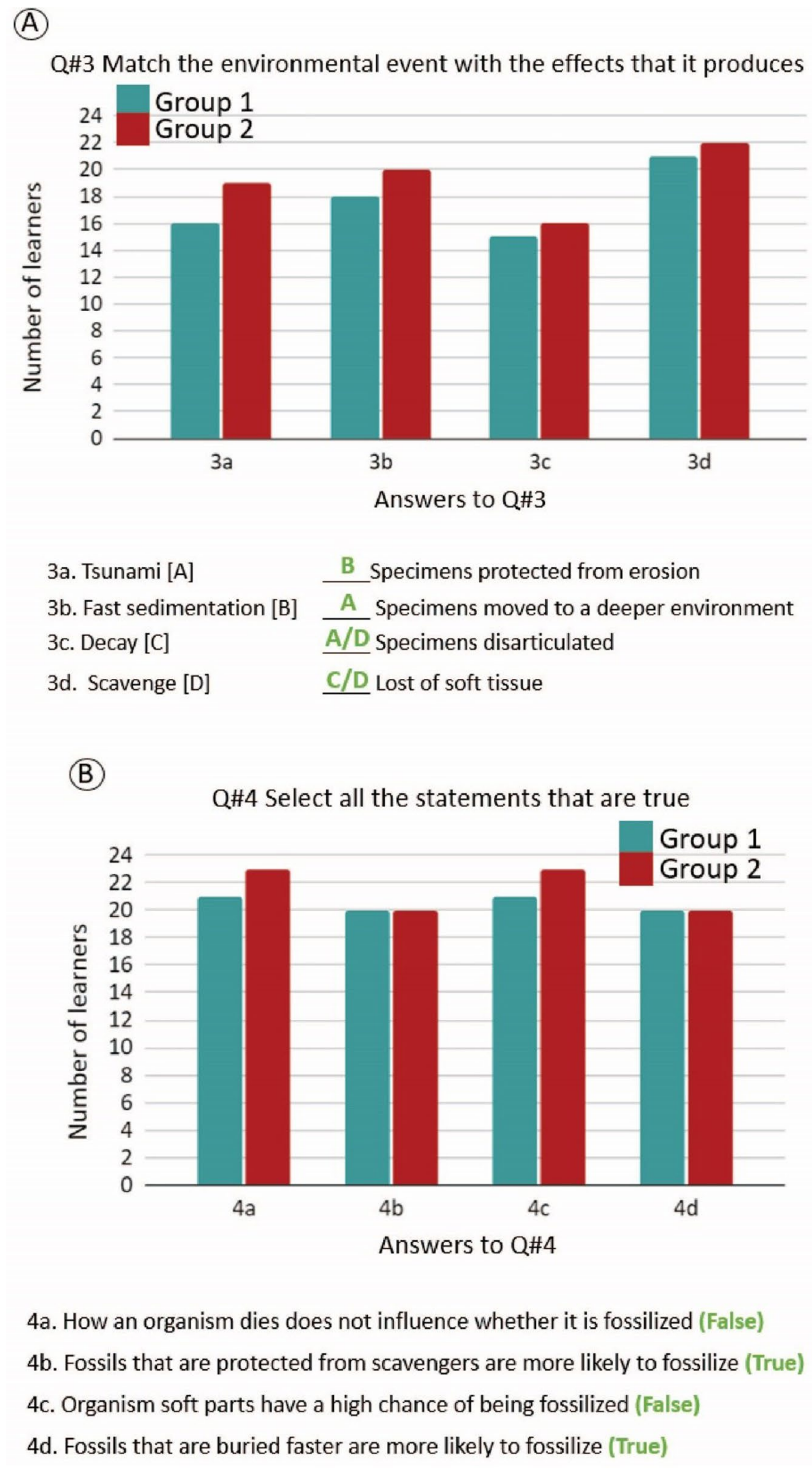

Figure 7. Number of learners who correctly answered A) survey question #3, and B) survey question #4. (group 1 n = 22, group 2 n = 27).

on their diagrams. Figure 8 presents two examples of sketches from the GeoFORCE learners. The sketch from Group 2 (Figure 8A) does a good job identifying the components and borders of the system (e.g., organisms and sediments), and identifying dynamic relationships between them (e.g., organisms dying, modification of organisms-microbes attack, environmental events-tsunami). That said, this sketch is basic in displaying an understanding of time. In contrast, the sketch from Group 1 (Figure 8B) incorporates elements that imply deep time thinking (i.e., sequence of events that take place during fossilization), but the student does not identify the boundaries of the system.

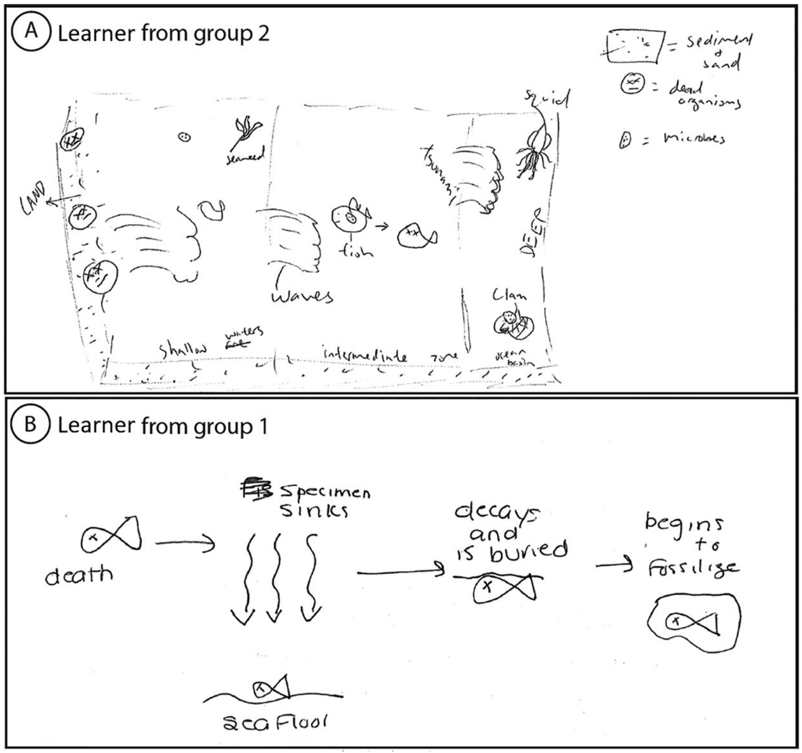

Figure 8. Example of the sketches drawn by GeoFORCE learners to illustrate the system presented in the game. A. A learner from group 2 who identifies the components of the system and dynamic relationships but who also has a basic understanding of time. B. A learner from group 2 who identifies some components of the system but does not establish boundaries; they identify dynamic relationships and incorporate elements that indicate deep time thinking.

Therefore, we do not know where the processes are taking place in their sketch.

In group 1, 95% of the learners identified the components of the system, but most only identified one (Figure 9A). In group 2, 88% of the learners identified the components and most of them identified two or more components (Figure 9B). Fewer learners, only 45% in group 1 and 51% in group 2, identified dynamic relationships among components (Figure 9), but the results from group 2 were more heterogeneous. Most of the learners identified only one dynamic relationship; however, there were some who identified two or three dynamic relationships, even locating them in space (Figure 9B). Finally, only 31% and 26% of learners in groups 1 and 2, respectively, demonstrated temporal thinking on their diagrams, including processes that occur on geological timescales in some cases (Figure 9).

Interpretations and discussion

Student and instructor behaviors

Our results from the observation protocol data show that "Taphonomy: Dead and Fossilized" is effective as an educational strategy that engages players in active learning; active learning behaviors were predominant in the two groups observed (Figure 3). Our results highlight that during the game, GeoFORCE Texas learners were engaged in making strategy, asking questions, and answering questions; in fact, most of the time during the gameplay they were engaged in making strategy. This is an important outcome since it reflects the ability of learners to use the knowledge gained and go beyond memorizing concepts to generate their own conjectures based on learned concepts. The observers reported that learners spent more time making strategy in the Second and Third Eras (Taphonomy and Discovery Eras), when there are numerous possibilities for strategizing. In the Second Era, for example, one possible strategy is protecting the most valuable organisms with burial tokens, while disarticulating an opponent's organisms. It is important to note that the GeoFORCE Texas learners were playing in teams of two (unlike the learners assessed in Martindale & Weiss, 2020), which allowed them to discuss strategy with their partner and thus engage in cooperative learning. Additionally, learners were playing against other teams to win the game, which still allowed them to engage in competitive learning.

While competition increases a student's engagement in the activity (Burguillo, 2010; Martindale & Weiss, 2020), the need for collaborative group work has been emphasized in much of the higher education literature (e.g., Gokhale, 1995). In the game, students spent an important period of time strategizing with their partner. Strategizing included positive interactions expected in cooperative learning, such as group members exchanging resources and information, teaching one's knowledge to others, finding help from a peer, challenging each other's reasoning, and encouraging others to achieve (Johnson & Johnson, 2002). Based on the implementation of the activity with GeoFORCE, we highly recommend having students play in teams to enhance

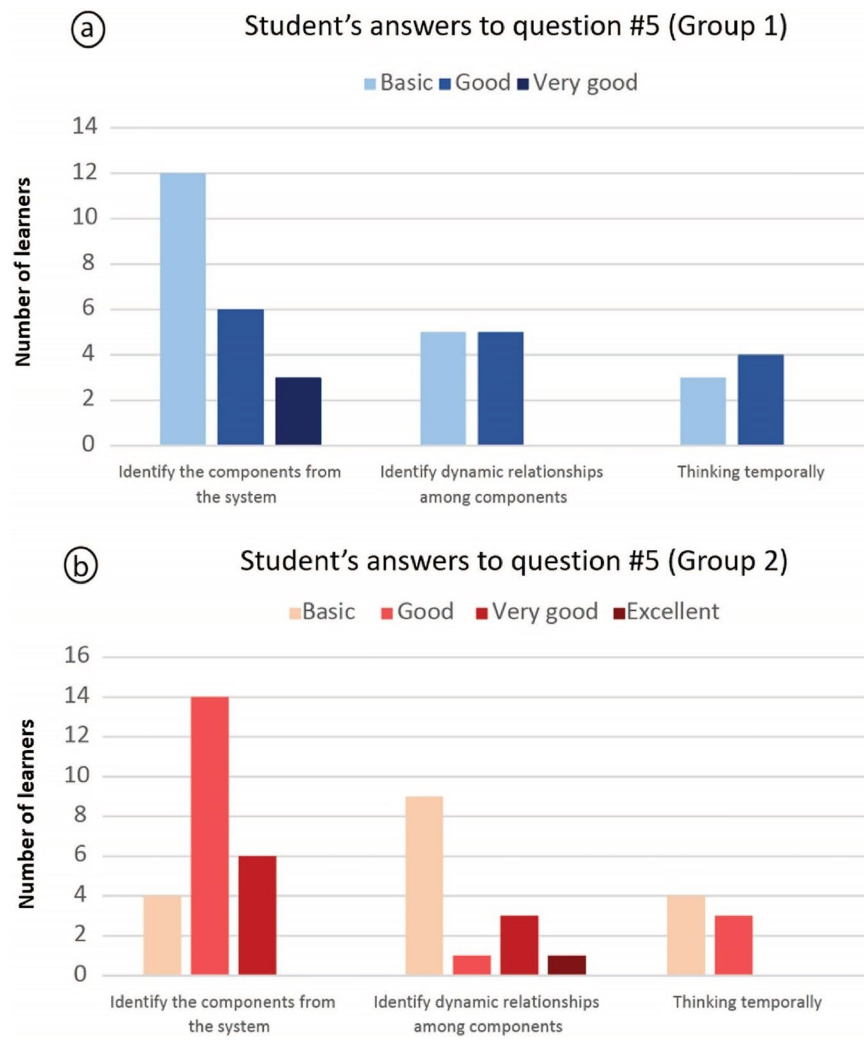

Figure 9. Summary of answers to survey question #5 by GeoFORCE learners; diagrams were graded with rubric presented in Table 4. Note, two students did not complete this question.

cooperative learning. Instructors could enhance cooperation by planning team structure so that groups are heterogenous (bring different perspectives and skills) and/or assigning specific tasks to each group member (e.g., each member needs to play one action per round) so that students cannot accomplish the goal of having the most pristine collection by themselves. Another option would be to have the entire table play as a team and then compete with other tables. This would keep the competition aspect of the game as a whole (although competition would be downplayed in individual interactions or Eras), while focusing on collaboration for the majority of interactions during the game.

Do learners understand that the fossil record is an imperfect representation of biodiversity?

Geosciences Concept Inventory Question [38] (Libarkin & Anderson, 2006, 2008) was included in our evaluation survey as question #2. This question was very useful in revealing the potential for students to develop the misconception that fossil diversity is an accurate representation of original diversity. This finding led us to suggest that instructors who implement the activity emphasize the concept that only few fossils of the plants and animals that ever live are preserved in the fossil record (answer A to the question [38] of the concept inventory; question #1 in Table 3) and to clarify misconceptions about fossil diversity that could arise during gameplay. Participants' responses to this question show that they are able to make inferences about biases in the fossil record; however, there were variable responses, which could be the result of GeoFORCE learners focusing on their outcomes of the game, without discussing the different outcomes of other tables. The players are expected to understand the difference among the initial and final diversity when playing the game (scaffolded by question #1); however, if the student finished with a collection that closely resembled the original diversity at the start of the game, they may have instead thought that answer C was correct: "if they had all of the fossils ever discovered into one room, this collection would be representative of most of the types of plants and animals that ever lived". Martindale and Weiss (2020) recommend having a following up discussion or activity to share the final outcomes of the game among multiple boards, where the learners could see that final diversity will vary from game to game and an activity where participants take pictures of the board once all the pieces have been placed and before they begin the Discovery Era, and that they compare the composition of live versus fossil assemblages. We did not incorporate this activity with GeoFORCE Texas learners due to time limitations. However, this kind of supplementary activity would be ideal for taking full advantage of the game as a learning tool and emphasizing the (possible) incompleteness of the fossil record in a high school classroom setting where instructors could dedicate two or three class sessions to teaching about fossilization, the history of life, and Earth system thinking.

While the Geosciences Concept Inventory is probably the most complete source of validated questions to evaluate geoscience understanding, we only found two questions related to the fossilization process. Three more questions relate to fossils; however, they evaluate the student's understanding of dating techniques used in geosciences, not the process of fossilization. We believe that an update of the inventory would benefit by incorporating questions related to taphonomy, a complex process that, as suggested by our results and discussed below, provides opportunities for learners to develop important scientific skills aligned with the Next Generation Science Standards such as Earth systems thinking.

Earth systems thinking

A key component in our evaluation of the game was assessing how effectively it promoted Earth systems thinking in high school learners. The sketches provided important information about how students identify relationships among the components of the system in time and space, important evidence of systems thinking (Assaraf & Orion, 2005). On the other hand, drawing is integral to scientific thinking; it is an important learning strategy that helps students organize their knowledge and communicate understanding (Ainsworth et al., 2011). The board game is played through the geological time and includes processes and events that impact fossilization. Student sketches varied from literal representations of the board game (i.e., drawing the tokens and cards) to actual representations of an ocean basin (e.g., cross-sectional view), with organisms being impacted by natural events such as tsunamis. As recognized by Assaraf and Orion (2005), one of the limitations of using sketches as evaluation tools is that the drawing aptitude of the students can interfere with their ability to represent their understanding of the natural phenomena. Asking the students to label and annotate their drawings is critical to address this limitation. GeoFORCE Texas students had 20 minutes to complete the five questions survey. Although some of the drawings were very detailed, many did not have annotations and two students did not complete this question; drawing ability could explain these differences. For researchers interested in using the evaluation survey presented in this study, we recommend giving students at least 10 minutes to complete this question as well as having a follow up interview or open question in the survey where the students have the opportunity to explain their diagrams.

The impact of multicontext learning environments within different student demographics

In high context culture, learning is characterized as community focused, subscribed to a holistic worldview, and thinking in terms of systems and connections is encouraged (Weissmann et al., 2019). GeoFORCE participants were mostly Hispanic students, who tend to find high context approaches to learning more salient to their academic experience (*sensu* Weissmann et al., 2019). In contrast, low context cultures (predominant in the USA) are characterized by valuing individual success, categorizing tasks and concepts, and applying a linear logical thought process (Weissmann et al., 2019). "Taphonomy: Dead and Fossilized" provides a learning setting that promotes student interaction, which is favorable to high context students. In the game, isolating the components to understand the system (low context) is as important as understanding the complex interactions among the components (high context). Similarly, understanding linear cause-effect relationships (low context) is as important as Earth systems thinking (high context). The game has the potential to serve as a multicontext educational strategy to engage students from different cultural backgrounds in high school classrooms.

Martindale and Weiss (2020) reported than Latinx students who participated in the evaluation of the game were less likely than white or non-Hispanic students to agree that the game was fun. The authors concluded that there may be a correlation among these results and the socioeconomic status of the students that would impact whether students were familiar with board games. Students who were familiar with, or had grown up playing board games, were more likely to enjoy the game than those who had not (Martindale & Weiss, 2020). This explanation aligns with anecdotical experience shared with us by a high school teacher (co-author ER) who implemented the game in his high school classroom. He observed that students from high socioeconomic status, including several Hispanic students, enjoyed the game more than those from low income backgrounds, regardless of ethnicity. Finally, Martindale and Weiss (2020) reported that Hispanic students and nonwhite students were less likely than white students to agree that the game was fun. In contrast, GeoFORCE Texas students were highly engaged and enjoyed the game. The data collected does not allow us to determine if the Hispanic students were more or less engaged than white students, but everyone was actively participating in the activity. One possible explanation for this divergence is that in GeoFORCE Texas, Hispanic students are not in the minority but make up the majority of student participants. Most of the students are Hispanic and the activity provides an opportunity for them to learn in a high context setting. A setting where Hispanic students are minorities could jeopardize the predominance of high context behaviors (e.g., more collaborative learning style, holistic view of the system) or favor low context behaviors (e.g., more individual, less interactive learning style, more linear thinking). The incorporation of "Taphonomy: Dead and Fossilized" in high school classrooms should consider the demographics of their student population to take full advantage of the activity as a multicontext educational tool.

In the student populations studied both in this paper and in Martindale and Weiss (2020), female students outnumbered male students; GeoFORCE students were 57-60% female between 2016–2018 (GeoFORCE Texas Annual Report, 2016, 2017, 2018), and 53% of students surveyed by Martindale and Weiss (2020) were female compared to 44% male. These numbers are not representative of the number of women in geosciences in USA college institutions (approximately 40% women enrolled in undergraduate degrees in 2015) (American Geosciences Institute (AGI), 2018). Since female students may see greater gains than male students from cooperative learning (when compared to individual or competitive learning tools) (Petersen et al.,1991), this could be a possible explanation of why the use of teams led to students overall feeling more engaged than they were when they played individually in Martindale and Weiss (2020). While male students may be motivated by competition (Dalton, 1990), both male and female students reported being more motivated to learn science when learning in teams or cooperative environments (Dalton, 1990; Lord, 1997; Mergendoller & Packer, 1989; Petersen et al., 1991). Female students in cooperative learning settings also report higher self-esteem following collaborative learning exercises (Johnson, Johnson, & Smith, 1998). Including cooperative and competitive aspects to the game (through "team play") may benefit all learners (Ke, 2020), but it is thought to benefit Hispanic and female students in particular (Petersen et al., 1991).

Limitations

The main limitation of the game evaluation was the time constraints associated with the GeoFORCE Academy, which impeded our ability to incorporate more evaluation protocols. For example, we were unable to conduct a pretest of the learners' prior knowledge; therefore, it is difficult to assess the amount of knowledge gained through playing the game. Our results show that GeoFORCE Texas learners achieved the expected learning outcomes of the game (Table 1); following the activity, participants successfully applied paleontological knowledge about organism physiology, environmental setting, physical and chemical changes during exposure, burial, and decomposition when asked questions that required establishing cause-effect relationships and Earth systems thinking. Although, we do not have a pre-game assessment, paleontology is not emphasized in the student's schools, or the previous GeoFORCE academies, and therefore, learning outcomes are more likely related to gameplay. Nevertheless, without a pretest we cannot differentiate between students who acquire this knowledge by playing the game versus those who had some prior knowledge. Moreover, there was no control group with a different intervention around the same content. Although our results provide evidence of the effectiveness of this activity as an active learning strategy, we were not able to compare the learning outcomes of this activity with those of a traditional lecture. In general, there is evidence showing that active learning works better than passive approaches for allowing students to assess their own degree of understanding of concepts and retain knowledge through participation (Michael, 2006). We expect the game to be more effective than teaching by lecturing. Future studies comparing the learning outcomes of games, such as "Taphonomy: Dead and Fossilized", to those of a traditional lecture are recommended to further validate the effectiveness of game-based education.

One of the goals of evaluating "Taphonomy: Dead and Fossilized" with rising 12th grade learners was to provide evidence to support the incorporation of games in high school classrooms. Although our results show that the game is effective in engaging high school participants in active learning about taphonomy, the game was evaluated in a summertime program (i.e., GeoFORCE Texas), which is different from a high school classroom setting. Transferability studies should accumulate empirical evidence about contextual similarity (Lincoln & Guba, 1985), paying special attention to the differences among the sending and receiving contexts. To make transferability possible, we provided a detailed description of the setting in which the activity was evaluated in this study (i.e., sending context). The evaluation of the educational intervention in new settings will require validation of contextual similarities (Lincoln & Guba, 1985). Some considerations when evaluating the game in a high school classroom are: (1) In GeoFORCE Texas, the activity was led by an instructor with the support of an educational team (4 educational coaches). In a high school classroom, the activity is more likely to be led by a single instructor. In the modified high school version of the game, we have incorporated activities for students to familiarize themselves with the materials of the game prior to playing it (i.e., formative activities and assessments), which will facilitate gameplay in a scenario with only one instructor. (2) Most of the GeoFORCE participants are from underrepresented groups in the geosciences (e.g., Hispanic) with high performance in science and a demonstrated interest in geosciences. By the time they played this game as rising 12th graders, many of them have had exposure to geology in past summer academies. This can explain why their learning outcomes are more aligned with the overall positive opinions of STEM- and Geoscience-majors in Martindale and Weiss (2020). A traditional high school classroom will be more heterogeneous. Instructors could use the differences in skills among students to enhance cooperative learning, planning the groups beforehand (e.g assigning specific tasks to group members, encouraging cooperation). (3) GeoFORCE Texas instructors only interact with the students during the summer academies, whereas high school teachers are engaged in teaching science to their students throughout the academic year, which offers opportunities for follow up on concepts, correct misconceptions and extend learning. (4)

Efforts started to assess the game and evaluation instruments in high school classrooms were interrupted by COVID 2019 after only one trial. Our preliminary results with 9th grade students show that even though they were familiar with paleontology (i.e., the class where the activity was implemented has emphasis in paleontology), students found the game challenging, especially the First Era. We recommend that instructors interested in incorporating the activity in high school classrooms dedicate at least two sessions. We have developed supporting material to help students familiarize with the mechanics of the game, and instructors to support the learning process and prepare students to get the best learning gains from playing the game.

The observation protocol is an important evaluation tool to analyze students and instructors' behaviors during gameplay. Nevertheless, one issue with this study was that two observers reported different results while observing the same group (group 2). Observers were not focused on specific tables, they were tracking all the board games in the classroom. It is likely that in some cases observers were coding behaviors differently; for example, one observer may have selected "engage in one-to-one discussion", while the other selected "answering a question" for the same observed behavior. Similarly, if a team of learners was strategizing and then asked for clarification about a game rule, depending on what part of the conversation an observer witnessed, the activity could be coded differently. Considering that students are expected to engage actively in this activity, having more than one evaluator to report what is happening in the classroom is desirable. In addition, both a training period and post-implementation debrief is recommended to help ensure that the observations are closely aligned. For example, evaluators could watch a video of an activity, fill in the proposed form, code observed behaviors every five minutes and discuss rankings to align observations among the group (Smith et al., 2013).

Implications

"Taphonomy: Dead and Fossilized" is an effective educational strategy to engage rising 12th grade students in active learning. GeoFORCE Texas participants were able to apply their paleontological knowledge to questions that required them to establish cause-effect relationships and engage in Earth systems thinking. Having learners play in teams (two people) and collaborate to achieve the common goal of having the most pristine collection (thus winning the game), engaged students in discussion, sharing knowledge gained while playing the game, and applying this knowledge to design strategies to win.

The results of this study informed the modified version of the game "Taphonomy: Dead and Fossilized" for high school learners. The new version of the game includes discovery options with which the students can better relate. For example, "A high school group wants to learn about fossils, so you take them fossil hunting. Great news, they are really good at finding fossils! Collect two untagged fossils". New environmental events and taphonomic factors are also included to reinforce some geological concepts, such as remineralization.

To support the incorporation of the game in high school settings, we designed a two to three-day lesson plan (depending on class duration). The lesson plan follows the 5E model (Engage, Explore, Explain, Elaborate, and Evaluate; Bybee & Landes, 1990), including learning goals that are aligned with the Next Generation Science Standards (NGSS) and Texas Essential Knowledge and Skills (TEKS), as well as documents that outline student and instructor expectations for each of the activities proposed. We recommend that the students play the game in groups of two to enhance cooperative learning, and teachers are advised to plan the play groups beforehand to take full advantage of the game as a multicontext educational tool. Additional materials, not used with GeoFORCE, but available in the modified high school version of the game (DOI: https://doi.org/10.18738/T8/USRIGL), include a student handout (essentially a game booklet that students are allowed to keep so they can continue learning outside the classroom setting) and teacher notes to guide the instructor through the activity. The latter materials highlight learning goals and suggest different strategies to increase learning gains in each part of the game. We designed a worksheet with activities and formative assessments to help students familiarize themselves with the materials and rules of the game (before and during gameplay), which will scaffold their learning of the scientific concepts and game rules. For example, activity #2 asks students to explain the three taphonomy cards they could use to protect their collection and explain their strategy. In day one, students create a strategy for the First and Second Eras by exploring organism physiology, environments, and possible taphonomic or environmental problems they may encounter. In day two the game allows students to engage, explore, explain, and elaborate on their knowledge. A summative assessment was also created to help students reflect on and galvanize their newly acquired knowledge following gameplay. The results of the GeoFORCE Texas study show that asking the students to draw a diagram is an effective strategy to evaluate whether they are understanding the connections and structure of the Earth system. Therefore, we provided the option for students to sketch their answers for some of the questions in the summative assessment. We also included a number of other questions that challenge the students to synthesize multiple geological concepts and make predictions or model the system.

Our results show that GeoFORCE participants who played the game engaged in active and cooperative learning but also demonstrated an ability to apply their newly gained geological knowledge to tasks that involved synthesis of scientific topics (i.e., establishing cause-effect relations and Earth systems thinking). The GeoFORCE results informed a modified high school version of the game and instructional material, a multicontext activity especially recommended for diverse classrooms, that teacher can download and integrate into a curriculum. All materials, including the game itself, can be found at: https://doi.org/10.18738/T8/USRIGL.

Acknowledgements

We sincerely thank GeoFORCE Texas for taking the time to incorporate this game in the 12th grade summer academy in 2019. Especially, thanks to the instructors, educational coaches, and observers who supported the implementation of the activity. We also thank the GeoFORCE learners who test played the game so enthusiastically. ESJ would like to acknowledge the contributions of the class "Curiosity to Question" (offered in the Jackson School of Geosciences), where the earlier versions of the paper were discussed, and feedback was provided.

Funding

This research was supported in many stages by the National Science Foundation Division of Earth Sciences (NSF EAR) under Grant #1660005.

ORCID

Estefanía Salgado-Jauregui http://orcid.org/0000-0001-7064-7852
Rowan C. Martindale http://orcid.org/0000-0003-2681-083X
Katherine Ellins http://orcid.org/0000-0003-4632-4003
Anna Weiss http://orcid.org/0000-0003-0835-4906

References

Ainsworth, S., Prain, V., & Tytler, R. (2011). Science education. Drawing to learn in science. *Science (New York, N.Y.)*, *333*(6046), 1096–1097. https://doi.org/10.1126/science.1204153

American Geosciences Institute (AGI). (2018). U.S. Female Geoscience Enrollments and Degrees Remain Level in 2015. Retrieved from https://www.americangeosciences.org/

Assaraf, O. B. Z., & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching*, *42*(5), 518–560. https://doi.org/10.1002/tea.20061

Association of American Universities. (2011). *Five-Year Initiative for Improving Undergraduate STEM Education*. AAU. Retrieved from www.aau.edu/WorkArea/DownloadAsset.aspx?id=14357

Batzri, O., Ben-Zvi Assaraf, O., Cohen, C., & Orion, N. (2015). Understanding the earth systems: Expressions of dynamic and cyclic thinking among university students. *Journal of Science Education and Technology*, *24*(6), 761–775. https://doi.org/10.1007/s10956-015-9562-8

Bergen, D. (2009). Play as the learning medium for future scientists, mathematicians, and engineers. *American Journal of Play*, *1*(4), 413–428.

Bonwell, C. C., & Eison, J. A. (1991). ASHE-ERIC Higher Education Reports. ERIC Clearinghouse on Higher Education, The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183.

Briggs, D. E. G. (2014). Konservat-Lagerstätten 40 years on: The exceptional becomes mainstream. *The Paleontological Society Papers*, *20*, 1–14. https://doi.org/10.1017/S1089332600002771

Burguillo, J. C. (2010). Using game theory and competition-based learning to stimulate student motivation and performance. *Computers & Education*, *55*(2), 566–575. https://doi.org/10.1016/j.compedu.2010.02.018

Bybee, R., & Landes, N. M. (1990). Science for life and living: An elementary school science program from Biological Sciences Improvement Study (BSCS). *The American Biology Teacher*, *52*(2), 92–98. https://doi.org/10.2307/4449042

Dalton, D. W. (1990). The effects of cooperative learning strategies on achievement and attitudes during interactive video. *Journal of Computer-Based Instruction*, *17*(1), 8–16.

Dove, J. E., Everett, L. A., & Preece, P. F. W. (1999). Exploring a hydrological concept through children's drawings. *International Journal of Science Education*, *21*(5), 485–497.

Efremov, I. A. (1940). Taphonomy: A new branch of paleontology. *Pan American Geologist*, *74*, 81–93.

Ellins, K. K., Thomas, D. L., Campos, D., George, S. W. M., Goldfarb, E., Kotowski, A., McCall, L., Soltis, N., Stocks, E., & Wright, V. (2018). Using the STAR Legacy Cycle to Promote Student-Centered Field Learning in GeoFORCE and STEMFORCE 12th Grade Summer Academies, Paper No. 252-2, Geological Society of America Abstracts with Programs. Vol. 50, No. 6. https://doi.org/10.1130/abs/2018AM-323816

Foster, A. (2008). Games and motivation to learn science: Personal identity, applicability, relevance and meaningfulness. *Journal of Interactive Learning Research*, *19*(4), 597–614.

GeoFORCE. (2016). 2016 Annual Report. Retrieved from: https://issuu.com/geoforcenextgeneration/docs/2016_geoforce_ar_vnov112016_highres

GeoFORCE. (2017). 2017 Annual Report. Retrieved from: https://issuu.com/geoforcenextgeneration/docs/2017_geoforce_texas_annual_report

GeoFORCE. (2018). 2018 Annual Report. Retrieved from: https://issuu.com/geoforcenextgeneration/docs/digital_geoforce_annualreport_2018_

GeoFORCE. (2020). 2020 Annual Report. Retrieved from: https://www.jsg.utexas.edu/geoforce/files/2020GeoforceAnnualReport.pdf

Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education*, *7*(1), 22–30. https://doi.org/10.21061/jte.v7i1.a.2

Hall, E. T. (1977). *Beyond culture*. Anchor Books/Doubleday.

Hequet, M. (1995). Games that teach. *Training*, *32*(7), 53–58.

Ibarra, R. A. (1999). Multicontextuality: A new perspective on minority underrepresentation in SEM academic fields. *Making Strides*, *1*(3), 1–9.

Ibarra, R. A. (2001). *Beyond affirmative action: Reframing the context of higher education*. University of Wisconsin Press.

Johnson, D. W., & Johnson, R. T. (2002). Cooperative learning and social interdependence theory. In R. Scott Tindale, L. Heath, J. Edwards, E. J. Posavac, F. B. Bryant (Eds.), *Theory and research on small groups* (pp. 9–35). Springer.

Johnson, R. T., & Johnson, D. W. (1986). Cooperative learning in the science classroom. *Science and Children*, *24*(2), 31–32.

Johnson, R. T., & Johnson, D. W. (2008). Active learning: Cooperation in the classroom. *The Annual Report of Educational Psychology in Japan*, *47*, 29–30. https://doi.org/10.5926/arepj1962.47.0_29

Johnson, D. W., Johnson, R., & Holubec, E. (1992). *Advanced cooperative learning*. Interaction Book Company.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works? *Change: The Magazine of Higher Learning*, *30*(4), 26–35.

Kagan, S. (1986). Cooperative learning and sociocultural factors in schooling. Beyond Language: Social and Cultural Factors in Schooling Language Minority Students, 231, 298.

Ke, F. (2020). Collaboration and competition in game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 329–347). MIT Press.

Kumar, R., & Lightner, R. (2007). Games as an interactive classroom technique: Perceptions of corporate trainers, college instructors and students. *International Journal of Teaching and Learning in Higher Education*, *19*(1), 53–63.

Libarkin, J. C., & Anderson, S. W. (2006). The geoscience concept inventory: Application of rasch analysis to concept inventory development in higher education. In X. Liu and W. Boone (Eds.), *Applications of Rasch measurement in science education* (pp. 45–73). JAM Publishers.

Libarkin, J. C., & Anderson, S. W. (2008). Development of the Geoscience Concept Inventory. In D. Deeds & B. Callen (Eds.), *Proceedings of the National STEM Assessment Conference (Assessment of Student Achievement)*, October 19-21, 2006, Washington, DC (pp. 148–158). Drury University.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

Lord, T. (1997). Comparing traditional and constructivist teaching in college biology. *Innovative Higher Education*, *21*(3), 197–217. https://doi.org/10.1007/BF01243716

Martindale, R. C., & Weiss, A. M. (2020). Taphonomy: Dead and Fossilized": A new board game designed to teach college undergraduate students about the process of fossilization. *Journal of Geoscience Education*, *68*(3), 265–285. https://doi.org/10.1080/10899995.2019.1693217

McConnell, D., Steer, D., & Owens, K. (2003). Assessment and active learning strategies for introductory geology courses. *Journal of Geoscience Education*, *51*(2), 205–216. https://doi.org/10.5408/1089-9995-51.2.205

McConnell, D., Steer, D., Owens, K., & Knight, C. (2005). How students think: Implications for learning in introductory geoscience courses. *Journal of Geoscience Education*, *53*(4), 462–470. https://doi.org/10.5408/McConnell_v53p462

Mergendoller, J., & Packer, M. J. (1989). *Cooperative learning in the classroom: A knowledge brief on effective teaching*. Far West Laboratory for Educational Research and Development.

Michael. (2006). How we learn where's the evidence that active learning works?*Advances in Physiology Education*, *30*, 159–167. https://doi.org/10.1152/advan.00053

Muscente, A. D., Schiffbauer, J. D., Broce, J., Laflamme, M., O'Donnell, K., Boag, T. H., Meyer, M., Hawkins, A. D., Huntley, J. W., McNamara, M., MacKenzie, L. A., Stanley, G. D., Hinman, N. W., Hofmann, M. H., & Xiao, S. (2017). Exceptionally preserved fossil assemblages through geologic time and space. *Gondwana Research*, *48*, 164–188. https://doi.org/10.1016/j.gr.2017.04.020

Nadolski, R. J., Hummel, H. G. K., van den Brink, H. J., Hoefakker, R. E., Slootmaker, A., Kurvers, H. J., & Storm, J. (2008). EMERGO: A methodology and toolkit for developing serious games in higher education. *Simulation & Gaming*, *39*(3), 338–352. https://doi.org/10.1177/1046878108319278

National Academies of Sciences, Engineering and Medicine (NASEM). (2019). *Shaping summertime experiences: Opportunities to promote healthy development and well-being for children and youth. Shaping summertime experiences (the nation)*. National Academies Press.

NGSS Lead States. (2013). *Next generation science standards: For states by states*. The National Academies Press.

Petersen, R. P., Johnson, D. W., & Johnson, R. T. (1991). Effects of cooperative learning on perceived status of male and female pupils. *The Journal of Social Psychology*, *131*(5), 717–735. https://doi.org/10.1080/00224545.1991.9924655

Pfeifer, L. S., Soreghan, M. J., Feille, K. K., Soreghan, G. S., Weissmann, G. S., Ibarra, R. A., & Stroud, W. A. (2021). Activation of the Multicontext model in a field-based program for traditionally underserved students. *Journal of Geoscience Education*, *69*(1), 85–95.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Ramaley, J. A., & Zia, L. (2005). *The real versus the possible: Closing the gaps in engagement and learning. EDUCAUSE.*

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development*, *44*(2), 43–58. https://doi.org/10.1007/BF02300540

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics*, *102*(6), 245–253. https://doi.org/10.1111/j.1949-8594.2002.tb17883.x

Seilacher, A. (1990). Taphonomy of fossil-Lagerstätten. In D. E. G. Briggs & P. R. Crowther (Eds.), *Palaeobiology: A synthesis* (pp. 266–270). Blackwell Science.

Slavin, R. E. (1983). *Cooperative learning. Research on teaching monograph series*. Longman Inc.

Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate stem (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, *12*(4), 618–627. https://doi.org/10.1187/cbe.13-08-0154

St. John, K. (2018). *A community framework for geoscience education research*. National Association of Geoscience Teachers. https://doi.org/10.25885/ger_framework/2

Weigel, F., & Bonica, M. (2014). An active learning approach to Bloom's Taxonomy. *U.S. Army Medical Department Journal*, 21–29. Retrieved from: https://ufdc.ufl.edu/AA00062689/00033

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *A study of K-12 mathematics and science education in the United States. Horizon Research.*

Weissmann, G. S., Ibarra, R. A., Howland-Davis, M., & Lammey, M. V. (2019). The multicontext path to redefining how we access and think about diversity, equity, and inclusion in STEM. *Journal of Geoscience Education*, *67*(4), 320–329. https://doi.org/10.1080/10899995.2019.1620527

Wilson, K. A., Bedwell, W. L., Lazzara, E., Salas, E., Burke, S. C., Estock, J. L., & Conkey, C. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation and Gaming*, *40*(2), 217–266. https://doi.org/10.1177/1046878108321866