Gea2: A Serious Game for Technology-Enhanced Learning in STEM

Lauren S. Ferro, Francesco Sapio, Massimo Mecella, Marco Temperini, and Annalisa Terracina

***Abstract*—Education and training are among the fields taking advantage of serious games (SGs). In this paper we present *Gea 2: A New Earth*, a digital SG developed as an immersive three-dimensional virtual learning environment, integrating several educational resources, and including multimedia learning material, communication tools, and intelligent tutoring support. The game aims to complement traditional classroom activities in science, technology, engineering, and mathematics (STEM), for high school students and teachers. It incorporates an intelligent pedagogical agent that can converse in natural language with the student and provides unsolicited hints during gameplay. The paper presents the game and its evaluation based on experiments involving about 100 participants. We think that the results presented here add to the research on game-based learning for STEM, by proposing a complex game system, where artificial intelligence techniques are integrated to support students' learning, and by confirming that the game experience can be attractive for the learners, also in very constrained classroom environments such as those we operated in.**

***Index Terms*—Serious games, game-based learning, technology-enhanced learning, STEM education.**

I. Introduction

In the last two decades the world has become significantly more digital. information and communication technology (ICT) has shown to have actual, and potential, application in several fields of human activity. Research in technology-enhanced learning (TEL) has fostered the design and implementation of computer-based/network-based instructional tools, supporting a higher degree of individualization and custom-tailored content delivery, hardly affordable in traditional classrooms. Intelligent educational software [1], such as intelligent tutoring systems (ITSs), focuses on modeling and applying, through a digital system, the processes used in the educational interactions happening in a real world class [2]. A virtual learning environment (VLE) further extends ITSs capabilities by providing the learner with additional educational resources, such as multimedia material, communication tools, recommendation systems, and more. The possibility of interaction with other learners paves the way to new scenarios of pedagogy on the Internet [3].

The attractive bouquet of learning material and tools offers support to educational methods on the network, while presenting teachers with new challenges and opportunities. Among such challenges/opportunities is the use of gamification, game-based learning (GBL) [4], [5], and digital educational serious games. They all surged ahead in the last decades as means to face the evolution of learners, and to foster motivation and engagement in them.

Serious games (SGs) are in general games used to train skills connected to real life tasks, with a negligible interest in entertainment. Their digital implementations are known to have wider educational aims (such as applying a given pedagogic approach), and providing an opportunity of studying instructional content, beside the general aim of training more practical skills [6]. They also confront educators with significant challenges, spanning from teacher's training, to the need of varied professional expertise, and several organizational issues [7], [8], so they are not so frequently used in schools.

In this paper, we present the game *Gea 2: A New Earth*, *Gea2* henceforth, which we designed and implemented with the intent to cope with the technological and pedagogical challenges we briefly mentioned earlier. The game aims to support education on topics in science, technology, engineering, and mathematics (STEM), for students of the initial year in high school.

In particular, we designed the game to be a complement and enhancement of the usual and more traditional frontal lecture, rather than supplanting, and substituting it. In *Gea2*, students move within a three-dimensional (3-D) environment, ultimately reaching three planets, by space travel, to explore them while solving quizzes and looking for objects to unlock interactive experiments. The interactive experiments are conducted in simulation rooms where evaluation activities are practiced on each planet, with the aim to select the best one for human life. To positively affect the learning experience, points and badges can be obtained through player actions. Individual and team tasks are solved by the players, who are also assisted by an interactive virtual tutor capable of interacting in natural language. Underneath the game, an intelligent pedagogical agent (IPA) monitors the progress of the player, and is capable of autonomously deciding whether and when to provide unsolicited help. Players can also share knowledge in their team's "Knowledge Cloud," and consult insights on a didactic cloud platform and chat with other teammates. Lastly, the teacher can follow the game's progress, and coordinate the session from a multimedia whiteboard.

We intended to experiment with SG technology, in a high school setting; we also aimed to add to the research work on artificial intelligence (AI) in education and GBL, in particular with the use of the IPA and the system to deliver unsolicited hints operating in *Gea2*. In addition, from a game-production-

Manuscript received June 10, 2020; revised April 30, 2021 and August 17, 2021; accepted November 14, 2021. Date of publication TBD; date of current version November 14, 2021. This work was partly supported by the European Space Education Resource Office, Italy (ESERO Italia) through project SPATIO. *(Corresponding author: Francesco Sapio.)*

The authors are with the Department of Computer, Control, and Management Engineering, Sapienza University of Rome, 00185 Rome, Italy (e-mail: lsferro, sapio, mecella, marte, terracina@diag.uniroma1.it).

process viewpoint, we wanted to apply a methodology of progressive development of *Gea2*, based also on the collaboration with teachers involved in the experimentation. This last aspect of "co-design" is out of the scope of this paper, basically for reasons of space, and yet was one of the initial drives for our work.

Beside presenting *Gea2* and its experimentation, in this paper we aim to analyze the experimental data, to verify how the *Gea2* learning experience was 1) appreciated by the students, from the point of view of perceived usefulness and engagement; and 2) effective as an instructional means.

We will see that the learners have been very appreciative and engaged, while the summative assessment of their proficiency was less than satisfactory. With regard to the just mentioned negative result we will propose interpretations, relating it to some hard constraints, partially unexpected, we had to confront during the experimentation (in particular hard time slots, and the use of the game detached by actual lectures).

Summarizing the whole contribution of the paper, we think that our experience added to the current research area on GBL for STEM, by 1) proposing a complex game system, where AI techniques are integrated to support students' learning (with the use of the IPA and of the unsolicited hints subsystem; 2) experimenting it in a high school setting; 3) obtaining a confirmation that the game experience can be attractive and engaging for the learners, also in very constrained classroom environments such as those we operated in; and 4) unveiling some threats to the use of this kind of systems, represented, in our case, by unforeseen difficulties offered by the above mentioned constraints: in this respect, we think that our experience can be useful, to us and others, in the organization of other experimental activities.

After introducing current research in GBL (Section II), we give an account of the design (Section III), development (Section IV), and evaluation (Sections V, VI, and VII) of *Gea2*. Then, Section VIII proposes a discussion about the findings, and a description of the limits of this study. Conclusions, and plans for future research are drawn in Section IX.

II. Related Work and Motivations

Digital GBL draws from the constructivist educational theory, and from the fields of computer games, visualization, and human–computer interaction, to administer learning experiences and content, through videogames [3]. It aims to provide and motivate the learners by means of engaging interactive instructional activities, ranging from simple tasks to accomplish, to the development of skills and solutions for complex tasks. This educational approach is expected to have success with current generations of learners, and guide them towards the challenges of the technological society in the 21st-century. Games are structured contexts, with clearly defined rules, where players need to overcome challenges and face opponents to achieve victory. Games can offer incredibly immersive and engaging environments where users can learn by doing and improve skills and competences related to decision making, strategy, teamwork, as well as social skills, leadership, and collaboration. SGs focus on the design and development of games with not only entertainment purposes.

Research in GBL, and SG, provides evidences of success [9], however, there are also studies recommending prudence about the actual effectiveness [10], [11].

In [12] a study is presented showing the current status of empirical research on mobile GBL in STEM education. The conclusions of this survey are that, on the one hand a great majority of the experiments provided positive results from using GBL, and, on the other hand, "the fundamental question of when mobile GBL is an appropriate approach for learning in STEM education and when it is not remains to be answered," suggesting the need for further empirical studies "to identify how and why certain designs work in particular circumstances." Another example, conducive to similar conclusions, is in [13], which describes a use of *game based science learning*, in middle school. The study focused on both learning outcomes (using pre–post tests on the digital game and non-digital groups) and students' self-efficacy. Results were mixed, whereas self-efficacy was reported as better in the digital group than in the non-digital, while no significant differences were detected in relation to learning outcomes.

There are concrete obstacles to a wider use of games: for instance, adopting them could be challenging due to existing stigmas around videogames, that can still influence the educational environment; moreover, integrating the game medium within the existing syllabus may be difficult. Against the adoption of a SG, there may also be issues of cost, as the development of a SG may involve the collaboration of different professional roles and stakeholders. In addition, other challenges can come from the difficulty to find usable off-the-shelf games, to suit the learning needs of the class. Also other factors can propose troubles, such as the quality of the adopted games, the need for specific teacher training to use them, the problem of fitting activities (possibly long ones) into the time span of the lecture, and eventually the difficulties with equipment (or lack thereof) [7], [8].

In fact there is still a limited use of SG [7], [8], and a need for further study [6]. Also the use of intelligent pedagogical agents (IPAs) in immersive virtual learning environments is a promising research topic [14], but there are not very many SGs implementing such combination. Of the several prior studies in this field, many focus on motivation and engagement, and a need is felt in recent years to study how games are eventually able to increase the cognitive or achievement outcomes in students [6], especially with respect to how such outcomes are in agreement with the curricular content the school expects to be acquired by the student. An additional aspect of interest, about SG in education, is in the target of the experimental activity: Although research and application of educational games in the high school context are present [15], this level of the educational system seems to be less frequently met, and in need of further efforts [16]–[18]. Furthermore, an aspect of interest in the use of GBL is its connection with the other activities planned in the course: The use of a game remains often as a single albeit removed experience; what happens in the game and what flows in the classroom often does not connect/crossover. In some instances though, as in Ludwig [19], classroom materials are also provided so that the learning component of the game can extend beyond the interactive

entertainment.

The summary proposed above allows to list some topics, motivating the usefulness of the study presented in this paper. Firstly, further investigation seems needed on the impact of GBL on students, in terms of how the learning outcomes are able to cover the curricular aims of the courses in which they are applied. In particular, such curricular aims are usually quite strictly and formally specified by the general national educational system, and can provide quite narrow constraints. Secondly, while previous studies focused mainly on children and higher education, the impact of GBL during the critical time of adolescence has been less explored. Hence, additional research may be useful at high school level, where keeping students motivated to sharpen their competences can be harder a task than elsewhere [20]. Finally, we observed that, in most studies, games are meant to supplant, rather than complement, the usual classroom lecture, delivered by the teacher: The study of the uses of GBL to flank traditional lectures can then be an interesting topic for additional research [21].

In *Gea2* we tried to meet the above mentioned lines of work, aiming at developing an SG that would 1) complement the classroom activity of learners and teacher, 2) have learning outcomes strictly connected to the ones planned in the first year of a high school, and 3) allow to deepen topics met in the other classroom activities, such as implementing collaborative learning among the students and with the teacher.

III. Designing *Gea2*

In this section we firstly present a description of the design process of the game, discussing also technical and aesthetic aspects. Then we focus on two important pedagogical aspects related to the design of *Gea2*, namely the narrative setting and the nature of *Gea2* as a role-playing game (*RPG*). These aspects are meant to provide the player with an attractive immersive experience, and foster a cooperative approach to her/his learning. Then we present shape and significance of the game components and elements.

Creating *Gea2* required a series of steps that incorporated different considerations from both a game and educational perspective. *Gea2* consisted of three main parts 1) Serious Game, Interactive Board, and Professor App.

In order to ensure that during gameplay the players can achieve the planned learning objectives, we adopted the following criteria during the design phase of *Gea2*:

- *Gea2 has to be a supporting educational tool integrated within a classroom environment*. The idea is that it can be used by teachers and students in class, as a completing/complementing part of the lecture.
- *It is an RPG* that addresses 21st-century skills, such as meta-competences, soft skills, communication, and collaboration skills.
- *It supports students' learning of how to inhabit the headspace of someone other than their primary ego identity*, thus affording the student an opportunity to develop a stronger sense of empathy.

As a result, *Gea2* encourages the development of a sense of community, by training individuals to function as a group (like in [22]) by assimilating with an unfamiliar role (i.e., one never assumed in reality) in the team.

The design of *Gea2* game environment utilized an iterative process. We first began with a survey of existing SGs that were situated within the STEM context. After narrowing the scope, we focused specifically on those that teach concepts of physics. We also considered the accessibility of the tool via mobile devices. This was an important consideration as we wanted that the final game could function on tablet devices.

Next, we created a mood board to refine the aesthetic, with the aim for it to run efficiently on a mobile device. In other words, we wanted to select an aesthetic style that would not be graphically demanding in terms of computational resources (e.g., high poly/tris count). As a result, we settled on a low poly aesthetic that was simplistic enough to require low computational resources, while still allowing to create interesting, large, and explorable environments for the player. In addition, we provided each one of the planets featured in the game, and the spaceship, with an individual color scheme and context. So, planet Maya features trees with a somewhat "Earth"-style aesthetic, planet Violet has a more isolated crystal aesthetic, with purple predominant, and planet Melissa shows quite an imaginary martian environment. Lastly, the spaceship presents the player with a futuristic aesthetic, with illuminated corridors and minimalistic décor, agreeing with the general low-poly aesthetic. The design of the user interface (UI) followed the same lines of simplification, with the additional requirement of adapting the interface to the size of the tablet screen, and managing the visual hierarchy of information, so to make important information readily available for the players, and to ease their navigation of the environment.

Another important aspect that we have taken into account during the design of *Gea2*, is the experience for both the students and teachers. Unlike most SGs, we believe that teachers should not be reduced to an assisting role, or excluded from the gameplay process. Instead, teachers should play an active role, to help and guide students within the SG, much the same way they would within a classroom environment. Therefore, we also made the same considerations for what the teacher required on their end in terms of UI and information. We described in full detail the game design in [23].

A. Narrative

A decision taken in early stages of the project was to set the plot around the hottest topic in astronomy: *Exoplanets and the search for extraterrestrial life*. Astronomy and astrobiology perfectly support interdisciplinary and cross topics discussion, so they are a good match for our interest in STEM education. Exoplanets are planets similar to Earth, orbiting a star other then the Sun. Both adults and young persons are since long passionately interested in this topic, with some further interest arisen in the very last years, after the February of 2017, when the NASA's Spitzer Space Telescope unveiled the first-known star system with Earth-sized planets, some placed in the habitable zone of the system, where a rocky planet is most likely to have *liquid water* (the key to life as we know it) [24]. *Gea2*'s plot is given in the game as follows:

A meteorite is approaching the Earth and very soon there will be an impact provoking earthquakes and tsunamis that will devastate continents. Students are part of a space mission and each one of them has a specific role with assigned tasks. The team's goal is to identify, among three, an exoplanet that is suitable for human life. The single objectives are related to the specific role. For example, physicists have to calculate the gravity on the three planets and compare it to the Earth's gravity. Chemists have to derive the planet's atmosphere composition and so on. Once the student/player has accomplished his/her tasks, the team should compare the obtained individual results and discuss together to infer which is the best planet among the three available ones.

B. Role-Playing Game

Another important decision that we have made, from a pedagogical perspective, is to implement *Gea2* as an RPG.

*RPG*s exist in many forms, from virtual role-playing, to tabletop, to live action. While each type of role-playing offers a unique experience, these games provide a compelling escape from the mundane reality, attracting millions of players worldwide. Unlike the passive experience of watching a film or reading a book, these games encourage players to actively take part in the adventure, sometimes even developing their own stories and characters. *RPG*s also offer a safe, relatively consequence-free space where players can develop certain aspects of themselves. Through role-playing, players learn how to inhabit the headspace of someone other than their primary ego identity, offering them the chance to develop a stronger sense of empathy. The shared, formative experience of *RPG*s provides a ritual atmosphere for players to enact compelling stories or perform unusual, extraordinary deeds. In this way, *RPG*s help encourage a sense of community, by teaching individuals to function as a group. Experiences transpiring in *RPG*s allow players to develop a deeper understanding about themselves and one another during the adventure [22].

Another important reason for implementing *Gea2* as an *RPG* is the collaborative/cooperative aspect of learning, that is, one of the 21st-century skills. Borich [25] asked, "What good are critical-thinking, reasoning, and problem-solving skills if your learners cannot apply them in interaction with others?". Cooperative learning activities train learners to reason and perform in an adult world. Our attitudes and values are among the most important outcomes of schooling, and are formed through social interaction. They provide the framework for thinking independently inside and outside of the classroom, and for guiding our actions outside the classroom. We believe that *RPG* can strongly support collaborative learning, exactly for their intrinsic nature to develop a deeper understanding of other players.

A crucial pedagogical aspect, that found easy application in *RPG*, is bound to interdisciplinary issues. In the last few decades, many reform initiatives have shaped teaching and learning in science. These reform efforts include a shift from teaching students to remember and execute isolated facts/skills, to having students experience learning as scientists, engineers, and mathematicians do. Pedagogues argue that students should engage in learning that allows them to explore, inquire, solve problems, and think critically. To this end, reform efforts within each of the STEM disciplines have focused on such strategies as inquiry learning [26], project-based learning, constructivist learning [27], problem-based learning [28], and the integration of technology across all STEM disciplines. Although these efforts have fostered improved learning outcomes within each of the STEM disciplines [29], many researchers argue that in order for students to be fully prepared for careers in the new millennium, they must be capable of thinking across disciplinary boundaries [30]. Frykholm and Glasson suggest that schools must begin to veer away from treating each STEM discipline as a silo and embrace an approach that blurs the boundaries of these disciplines. They argue that students who engage in rich cross-disciplinary experiences will have a deeper conceptual understanding of science and mathematics content [31], which will improve their achievement in each of the disciplines. Further, interdisciplinary learning can foster an understanding of STEM concepts in their application to real world problems, which are interdisciplinary by nature. In traditional school settings, the compartmentalization of scientific knowledge creates boundaries so rigid that they often serve as barriers to any efforts to develop integrative science and mathematics programs [32].

Gea2 embraces interdisciplinary and integrated science concepts. Table I reports the description and example tasks for each of the four roles appearing in the game: *physicist*, *geologist*, *chemist*, and *astrobiologist*. Each role has three tasks (one on each planet) to accomplish.

C. Game Components and Elements

Here we describe the components that players can use during gameplay.

a) Registration: At game start the learner is greeted by an introductory video, then replaced by the login/registration screen. The registration operation collects data about the learner (name, email, chosen password), and about the school and class (which are selected by a drop down menu, as only the schools where the game was deployed appear). Each learner will be operating in one of the teams defined by the teacher. From here, the data is sent to a dedicated web service, which sends back an authorization token for the player to use to gain access to the game.

b) Team creation: To achieve their own and teams' objectives within *Gea2*, players have to explore the 3-D environment. The environment consists of a spaceship and three simulation rooms—one for each planet. By exploring these environments, players can:

- collect specific (task-related) objects around the ship.
- gain access to the planets' simulation rooms, by correctly answering questions.
- solve their assigned task using interactive panels *(within the planets' simulation rooms)*.
- seek for help in any moment, namely asking questions to the IPA, or consulting some ad-hoc materials external to the game (e.g., video, articles, insight notes, simulations, etc.).

TABLE I
PLAYERS' ROLES IN *Gea2*

Role	Description	Example Tasks
Physicist	The *Physicist* studies a wide range of phenomena, spanning from subatomic particles, which all ordinary matter is made of, to molecular length scales of chemical and biological interest, to cosmological length scales encompassing the universe as a whole.	Calculate the value of gravity on the three exoplanets and make a comparison with the terrestrial gravity. Discuss with other team members, evaluate if the discovered values are suitable for human life.
Geologist	The *Geologist* works to understand the history and future of our planet. The better they understand Earth's history the better they can foresee how events and processes of the past might influence the future.	Calculate the axial tilt and magnetic field and understand the characteristics of the water cycle on the different planets. Discuss with other team members whether the discovered values are compatible with human life.
Chemist	The *Chemist* masters composition of matter and its properties, dealing with molecules and atoms. They carefully measure substance proportions, reaction rates, and other chemical properties.	Calculate at which temperature water boils, the atmospheric composition and the volume of the three planets. Discuss with team members whether humans can breathe in these different atmospheres.
Astrobiologist	The *Astrobiologist* is interested in the study of the origin, evolution, distribution, and future of life in the universe.	Calculate the atmospheric temperature and pressure on the three planets. Discuss human survival conditions with team members.

Therefore, the players are encouraged to explore and interact with the whole virtual environment (classroom, team, teacher, and game), which mimics the ways they would use to obtain information in the real world, in order to solve problems and complete their tasks. Once the players start solving their own objectives, the team can share information to form a complete vision of the overall team task for selecting the best planet suitable for life. As a result, *Gea2* supports cooperation and collaboration (requirement of a successful RPG) via chat systems and in-class discussions.

c) Inventory system: The inventory is the repository where the items collected by the player are stored, to be removed when they are used during an experiment in the game (see letter i below). An item is collected by collision with it by the player's avatar. The inventory items are displayed in the top left hand corner of the user interface. Items are specified in the system through fields for the item id, and for a sound effect. The component also implements a 3-D animation function that makes the 3-D model float in the air to make it more visible to the player.

d) Localization system: The localization system of the game tracks the players' (avatars) movement and identify their location within the game's space. It is implemented through a script, replicated in all the game locations, that intercepts the presence of a player and interact with a web service listening on the server. The web service feeds the Professor Virtual Board (see later), so to make an updated live map continuously available.

e) Note system: Sharing notes is an important part of the students gameplay. This subsystem gives access to a note composing and consulting panel, where it is possible to to 1) write a comment, selecting the subject from a drop-down panel; 2) publish the composed note; 3) list the notes in the team's Knowledge Cloud; and 4) see feedback received about the note. The note is sent to the server, which makes it available to users, through the general Knowledge Cloud of the game. It also applies validation checks, and administers related score and badges.

f) Unsolicited hints system: Based on learner's in-game behavior, the IPA provides a pedagogical intervention through unsolicited hints (see Section IV-E). To compute such hints the IPA monitors the learner's advancement within the game, and her/his emotions:

- Player's advancement: To analyze the progress in gaming sessions, we decided to check two other important values. The first value is the overall score obtained by the player, representative of the actions undertaken during gameplay. The other important value is how many questions the player has already asked to the IPA.
- Player's emotions: We try to infer player's feelings based on the conversation with the IPA, and the participation in chats with teammates. In chat posts we monitor emoticons usage. Emoticons are visual representations of emotions such as a smiley face, which are commonly found on social networking and messaging platforms. To facilitate emoticons usage, we decided to use a limited numbers of them: eight, representative of the basic emotions.

The system can send the monitored data to the web service delegated to analyze it and provide the unsolicited hint if needed. This function is called periodically, with an interval we have empirically set to three minutes. The unsolicited hints system hosts a panel for displaying the hint to the player. To assess the validity of the provided unsolicited hints, that is, the correctness of our algorithm, we added the possibility for the user to provide feedback on the hint they receive, marking it as helpful or not.

g) Chat system: To encourage communication between teammates, we gave players a way to exchange text messages and emoticons in a dedicated chat group for each team. The UI features an interactive panel that hosts an input field, a text container for displaying messages, a smaller panel allowing the selection of emoticons, and a send button. The chat is implemented through the use of the Photon Chat engine [33], which provides several functions used as callback. Photon Chat provides "channels" to make separate chats; we used

this feature to allow for intra-team communication (with one channel for each team).

h) Quiz system: Whenever the player accesses a quiz panel, a script is called that loads the appropriate questions, and answer options, dynamically from a text file: The teacher manages questions and answers by editing this text file. If the player answers the questions correctly, the script plays a success sound and shows a panel with the objects that (s)he will need in the simulation room of the related planet. If the answer is wrong, a "wrong answer panel" is shown. The score of an answer depends also on the number of attempts the player made before providing the correct answer.

i) Animated experiment panels: The 3-D game is multi-component and offers many player–game and player–player interaction possibilities, however a core aspect is the possibility to perform interactive experiments. When the player first interacts with an experiment monitor, if their inventory contains all the items needed to unlock that specific experiment, a panel is displayed, showing the inventory. The panel allows the player to select the objects (s)he think are needed for the experiment. If the selection is correct, the experiment is unlocked: the selected objects are removed from the inventory, and the interactive experiment panel is opened, with a "success" sound. Otherwise, an on-screen notification, with sound is issued.

IV. Developing *Gea2*

For this project we developed a three-tier architecture, with a presentation, a logic, and a data tier.

A. Presentation Tier (Teacher and Student Workspaces)

The teacher workspace consists of three different applications, named *Professor App*, *Professor Virtual Board*, and Opedia®.

Professor App (PA): *a web application developed to manage gaming sessions and classrooms, and to view the collected data about the students' in-game activities, in both past and ongoing sessions*

In the PA, a teacher can: register/login, select her/his school, add/Select a class, and, more importantly, start/stop a gaming session for the selected class and monitor students' results in ongoing or previous sessions.

Professor Virtual Board (PVB): *the teacher's tool used in the classroom to manage the gaming session, to interact with the students, and to control their progresses*

A gaming session can last up to four hours. When a teacher starts a gaming session, the PA generates a unique code (PIN) for the session, That PIN can be used to start the PVB application and link it to that gaming session in the classroom (see below). The PVB is designed to run on an interactive multimedia whiteboard. As a result, teachers can:

- Launch the PVB and link it to the current gaming session using the PIN already generated using the PA
- Add new teams

Information about teams is permanently stored, and retrievable during the successive gaming sessions. A student can join a team during registration. After teams have been created teachers can: access information about the players (such as role, assigned task, score), check the position of the players on the map, and validate the notes of the players in the Knowledge Cloud.

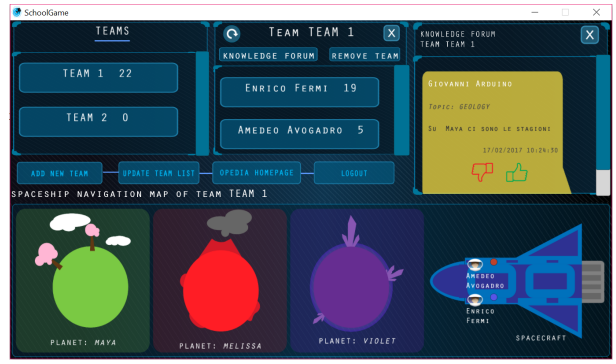

Fig. 1. PVB visualization.

To enable a better cooperation among students, we have added the possibility for them to share ideas, thoughts, and in-game discoveries they make, with the team they belong to. Published notes are at first visible only by the teacher, who can validate them (accept/reject). In the PVB, the live 2-D map of the game is visible. The position of the players in the current gaming session is continuously updated and displayed by icons with each player's name and role (the latter represented by a colour). The PVB also allows to scroll through the list of teams for the current class. For each team the teacher can consult the list of its students/players, and is allowed to focus on each student's information. In particular the teacher can check the student's detailed score, or read her/his assigned task. The teacher can also access, through the PVB, the team's Knowledge Cloud and moderate the notes published by each player of the selected team.

Opedia®: *a cloud platform allowing the sharing of multimedia didactic material between teachers and students*

Opedia® was integrated into our VLE thanks to an agreement we made with its owner. Its use can be also supported by the teacher, who could help understand the retrieved material, or suggest further search queries. We could not involve the teachers in the construction of a whole scaffolding system in *Gea2*, so we used Opedia® as an external means to offer some of the functionalities of a dedicated scaffolding system.

The *Student Workspace* consists of two applications: a 3-D virtual game (3DVG), and Opedia®.

3DVG: *the core of the student workspace. Students become players when they first register to the game, join an existing team for their class, and choose a role*

Once in an active game session, students can :

- Login to the 3DVG and explore the 3-D Environment
- Recover 3-D objects they need to solve the assigned task
- Gain access to the simulation rooms answering quizzes
- Enter the simulation rooms and find the interactive panels
- Solve assigned tasks using interactive panels
- Ask help in natural language to the IPA
- Access insights on Opedia®
- Add notes and chat with teammates

All the actions listed above can be performed by players in any sequence they like. In other words, they can create and follow their own strategies subject to the game's mechanics. To further push the students' problem solving skills, we developed a quiz mechanics. Throughout the game, students are asked to correctly answer various task-related questions, in order to progress in the game and gain access to the simulation rooms. Each simulation room can be accessed by a student when he or she has identified and collected those objects that are defined as necessary, according to her/his task.

Most of the actions in the game are driven via a point-based system, which operates in the background. Students are not aware of this point-based system, and it is an important part of the 3DVG because it contributes to the students evaluation. However, students can see their score by asking the teacher to view it in the PVB. Beside points, players can obtain badges by completing certain tasks. These badges can be accessed in the spaceships' trophy room.

The IPA is part of the 3DVG and guides the player trough the game. The IPA that we have designed has two main functionalities. The first is to reply to questions asked in natural language: it can be thus classified as a Dialogue Management System (DMS). The second is to evaluate game progress and emotions expressed by the players during the game, in order to infer if the player needs help and, if this is the case, provide unsolicited hints. The latter ability provides pedagogical intelligence to our IPA.

B. Logic Tier

The logic tier offers the web services managing players and teachers, teams, gaming sessions (save/resume), and all the functionalities described in Section III-C. It communicates, via network, with the presentation and the data tiers. In-game progress is automatically saved, by uploading to the server and storing in the database (e.g., at each player's achievement, such as unlocking a badge, collecting an object, or scoring points). Progresses are reloaded in the game at log in, so the game can be played seamlessly across different gaming sessions in class.

To implement the web services we took advantage of Jersey, an open source Java framework created by Oracle for developing RESTful web services according to Java™API for RESTful Web Services standard. The *authentication* web services use a security token mechanism. The *class* services are related to all the operations that treat the class as a whole, and allow the creation and deletion of classes, starting or stopping a gaming session for a class, and listing the teams in a class. The *team* services allow the creation and deletion of teams, listing the notes in that team's Knowledge Cloud, and listing the students that have joined that team. The *professor* services are used to manage teachers and the classes they teach to. They allow the creation, deletion, update or retrieval of teachers, and the addition or deletion of a class from their personal list of classes. The *note* services are related to the management of notes in a team's Knowledge Cloud, providing their deletion, retrieval, and validation by the teacher, who can accept or reject them as previously mentioned. The *school* services are used to recover information about available schools and classes for each school. In order to provide an identification number for schools we have used the unique key generated by the government to identify Italian schools. Finally, the *student* services deal with recovering and updating information about players during gaming sessions. In particular they allow the creation, deletion, and retrieval of a student, the creation and update of the game progress for each student, the deletion and addition of a team for each student, the creation and retrieval of notes, the creation and update of a player's position, the addition of the question asked by the player to the IPA, the creation and update of the score for each player, and the provision of unsolicited hints.

C. Data Tier

The data tier consists of a complex database where all data related to the different actors of the system are stored. In particular, the database also stores the knowledge of the IPA. We refer to the knowledge about a specific subject as being held by a non-player character (NPC). Specifically, we have four NPCs, one for each role that a player can have. In particular, the NPC knowledge templates consists of a simple sequence of questions, answers and possible suggestions.

The data tier is implemented as a relational database containing information about all the entities discussed in the previous sections. Logic and data tier are strictly connected since many of the web services in the logic tier can access the database both for reading and writing data.

D. Playing the Game

Once the student is logged in, they begin the game from the spaceship (see Fig. 2). The *heads-up* display contains the

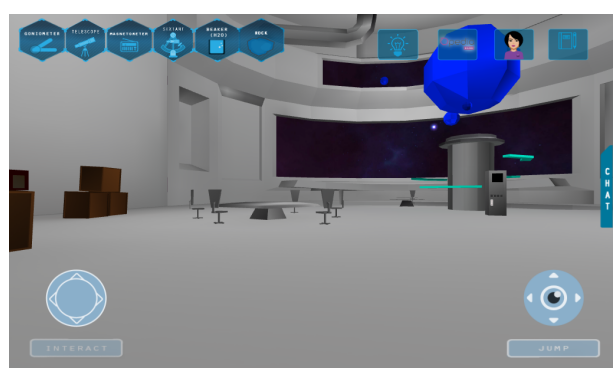

Fig. 2. View of the command centre in the spaceship.

controllers that the player can use to play the game. At the bottom there are the two joypads, one to navigate the other one

to look at the virtual environment. Next to those, we find the *jump* and *interact* buttons. On the right side of the screen the player can access the chat. On the top left corner of the screen is housed the inventory, which shows all the objects the player currently carries. On the top right corner of the screen there are four buttons, to access 1) a panel for inserting answers to the player's tasks (assigned based on the role), 2) Opedia®, 3) the interactive virtual tutor interface, and 4) the note system panel.

The player can explore the 3-D environment, solving quizzes and looking for any information to help her/him solve the assigned task. To ensure a positive reinforcement of learning, we provided the possibility to unlock badges as a result of some actions. When the player approaches a badge, an on-screen message describes the actions to be completed in order to unlock that badge. When a badge is unlocked, the player is notified and the badge will be displayed in the trophy room. To progress in the game, the player has to enter the simulation rooms. The player has first to find the interactive monitors and answer some quizzes, then the entrance to the related simulation room is unlocked. Once a simulation room is accessed, the player finds another interactive monitor, so to perform an experiment: Experiments are the key to solve the players' tasks.

E. Intelligent Pedagogical Agent

In *Gea2*, the communication is implemented via a natural language processing (NLP) algorithm, based on an ad-hoc text retrieval problem solver and on a Naïve Bayes text classifier equipped with an inner product-based threshold criterion (see later) [34]. The problem of selecting the right answer from a knowledge database can be expressed similarly to an information retrieval problem [35]. Following Conati in [36], there are two main challenges:

1) *Assessing students' knowledge and learning from the interaction with the game* (whereas the connection between learner's game action and understanding of the underlying domain might be not clear).
2) *Providing individualized interventions without interfering with the experience* (as, often, to get success, the system ought to provide help in ways that will not resemble, or remind the learner of, traditional educational activities).

In [34], we describe how did we try to tackle these challenges: The IPA was developed, and integrated in *Gea2*, to provide the learner with unsolicited hints based on the learner's state of interaction in the game. The IPA is basically characterized by two abilities:

- It is an *intelligent virtual tutor* (IVT), that is, a virtual agent capable of a conversation in natural language (NL). The IPA can reply in NL to questions written in NL, and is implemented as a DMS [34].
- It provides *unsolicited hints* to the learner, by evaluating the game progression and the player's emotions (through sentimental analysis and emoticons usage) [34]. This makes of it an IPA, that is, an IVT capable of actuating interventions in the game, to support specific user needs.

The core of our system is an implementation of Naïve Bayes text classification [37], which is a probabilistic classification method based on language modelling under the hypothesis of words' conditional independence. This algorithm is an application of the Bayes theorem, which allows the estimation of the probability that a collection's document is relevant for a query, given the sequence of words that make up the query itself. We then validate the answer chosen by the classifier by applying a threshold criterion.

The implementation of the IPA in *Gea2* is done through a script providing the user with a panel allowing to submit a question and visualize the IPA's textual answer. The IPA builds the answer by processing the question and querying the NPC database for the most pertinent answer.

V. Evaluating *Gea2*: Methodological Framework

In this section, and in the following two, we describe the evaluation of *Gea2*. , which occurred during the two phases of development and experimentation of the game. The evaluation regarded two main aspects: On the one hand, we wanted to measure how the students liked the experience, and perceived the game as useful to support learning. On the other hand we considered the effectiveness of the game, namely how it allowed to learn STEM topics in a playful way, and to acquire 21st-century skills.

Assessment describes the process of using data to demonstrate that stated learning goals and objectives are actually being met. Michael and Chen [38] state that "Serious games, like every other tool of education, must be able to show that the necessary learning has occurred." Thus, SGs, to be considered a viable educational tool, must provide some means of testing and progress tracking; testing must be recognizable within the context of the education or training they are attempting to impart. However, learning is a complex construct, difficult to measure, and determining whether a simulation or a SG is effective at achieving the intended learning goals is a complex, time-consuming, expensive, and difficult process [39], [40].

Generally speaking, an assessment can be described as either (i) *summative* whereby it is conducted at the end of a learning process and tests the overall achievements, or (ii) *formative* whereby it is implemented and presented throughout the entire learning process and continuously monitors progress and failures [41]).

Considering the specific SG domain, Michael and Chen describe three primary types of assessment: (i) completion assessment, (ii) in-process assessment, and (iii) teacher evaluation [38]. Basically the first type corresponds to an automated summative assessment, concerned with whether the player is able to successfully complete the game. In-process assessment is related to observing the player behavior (progressively stored in logs) by the teacher or automatically. If an intervention is managed, based on such data, this method could be functioning like a continuous formative assessment. The teacher evaluation is based on the her/his direct observations and judgements of the students, while they are playing; it can function as a combination of the other two methods, but typically aims at evaluating those factors that the functionalities/logic of the game are not able to capture.

The most common method of post assessment currently consists in testing players' knowledge about what they learned by way of a test or teacher evaluation. This method is frequently employed because it is the simplest to implement.

Coming to the experimentation, the tests were administered in two different schools. Each test involved playing the game along four lectures, in class, with no extra time allowed beyond the normal time span of lectures (hence with the limits imposed by a real school environment). Teachers were present, monitoring the activity and active whenever help was needed. Summative evaluations were conducted after the tests. Moreover, we evaluated the *likeability*, *learnability*, *usability*, and *preferences*, as stated/perceived by the learners, through a questionnaire.

The tests are described in the following two sections.

VI. Evaluating *Gea2*: Test No. 1

We had the collaboration of an Italian High School, that provided two classes and a room hosting our tests, thus the testing happened in a true context of education. The school offers different five-years formation tracks, oriented towards scientific or technical/industrial subjects. The two classes had in fact different "specialization topic". The test had 42 participants, as each class was with 21 members (details are in Table II. The lectures schedule associated to the test is shown in Table III.

TABLE II
Test No. 1—Details on Classes

School name	Classes	Students
ITIS Pascal	1F (*Scientific* institute)	21 (9 ♀, 12 ♂)
ITIS Pascal	1N (*Industrial/Technical* institute)	21 (6 ♀, 15 ♂)

TABLE III
Test No. 1—Lectures Schedule

Date	Classes	Method
Play Meeting 1	1F and 1N	Game introduction, pretest
Play Meeting 2	1F and 1N	Play game
Play Meeting 3	1F and 1N	Play game
Play Meeting 4	1F and 1N	Game conclusions, posttest

The test spanned over four lectures, described in the following. During the first lecture the teacher introduced the game to the class and briefly explained roles, tasks and objectives. Furthermore, the pretest was administered. Students formed the teams and chose their roles. We valued the importance of letting students self-organize and, most importantly, decide their team name, in order to ensure a stronger sense of cooperation and individuality in the learning process.

During the second lecture the students had credentials to access the game via premade player accounts. They grouped by team and played the game for about 40 minutes, while the teacher followed the game progress using the Interactive Board.

During the third lesson we divided students per role, in order to encourage collaboration and knowledge sharing among teams. During this meeting, the students started using some functionalities for the first time, such as notes publishing, or asking questions to the IPA, or getting useful insights from Opedia®. This is probably due to the fact that they were then already familiar with the basic game mechanics, so they felt confident to explore further game features.

In the last lecture (held four weeks later because of Christmas holiday) the students had to finalize their answers to the tasks they had been assigned, and each team had to produce a "final decision" about what was the most suitable planet for human life. Then the players' scores were compared on the PA. Unfortunately, due to the unavoidable time constraints of the school, students were not able to discuss their findings during the game, so they could not finalize a team answer to the overall game task.

In the next two subsections we describe the questionnaires we used for the evaluation.

A. Students' Perception of the Experience

We used a Likert scale, in which the format of the five-level items ranged from *Yes a lot* to *Not at all*. The internal consistency of the questionnaire has been evaluated by Split-Half Reliability [42]. We acknowledge the low number of questions, however, considering the difficulties in administering different questionnaires to a whole class, we had to limit the number of questions to 18, in order to make the process manageable. For each student, we partitioned the questions in *odd* and *even*, and summed the score for each half. Then, we computed the correlation between the two halves. Finally, in order to get a better estimate of the reliability of the full test, we applied the Spearman–Brown correction (see Table V), thus leading to the final value of the *Cronbach's Alpha* at .92, which is an excellent result.

The first three questions, showed in Table IV (Part 1) aimed to establish if students liked the game and would recommend it to other students (likeability). 64% of students enjoyed playing the game while 15% disliked it. 81% of the students think that *Gea2* is more engaging then a traditional lecture, but jut 54% would recommend the game to other students while 27% are indifferent.

The second set of questions in Table IV (Part 2) aimed at verifying students perception of learning outcomes (learnability). First, we asked the students if they learned by playing the game: just 36% replied in a positive way. We then asked if *Gea2* is better than a face-to-face lecture, and if they think they learned in a playful way. 67% of the students prefer this type of learning, but just 33% think to have learned in a pleasant way.

We believe that the above result, about learnability, can be interpreted based on the two following observations. First, we had many positive feedback from students "about the game", but they were very upset by the language used in the game. The game was fully in English, and the students revealed great difficulties in understanding the wording of quizzes and tasks. The language factor was also explained by the students as a reason for a reduced interaction with the IPA. These feedback let us decide to rewrite a good deal of the game (quizzes and

TABLE IV
TEST NO. 1—QUESTIONNAIRE ABOUT STUDENTS' PERCEPTION OF THE EXPERIENCE

Part 1: Likeability

Qn	Questions	Yes, a lot	Yes, enough	Indifferent	Not much	Not at all
Q1	*Did you enjoy playing the game?*	2 (6%)	19 (58%)	7 (21%)	2 (6%)	3 (9%)
Q2	*Do you think that playing the game is more engaging than a traditional lesson?*	11 (33%)	16 (48%)	2 (6%)	2 (6%)	2 (6%)
Q3	*Would you recommend to students from other classes to try the game?*	7 (21%)	11 (33%)	9 (27%)	2 (6%)	4 (12%)

Part 2: Learnability

Qn	Questions	Yes, a lot	Yes, enough	Indifferent	Not much	Not at all
Q4	*Do you think you have learned by playing the game?*	1 (3%)	11 (33%)	7 (21%)	6 (18%)	8 (24%)
Q5	*Do you think that playing the game has allowed you to learn in a more interesting way?*	1 (3%)	10 (30%)	12 (36%)	3 (9%)	7 (21%)
Q6	*Do you think that playing the game has allowed you to learn faster?*	1 (3%)	17 (55%)	2 (6%)	7 (23%)	4 (13%)

Part 3: Usability

Qn	Questions	Yes, a lot	Yes, enough	Indifferent	Not much	Not at all
Q7	*Were the user interfaces clear enough?*	1 (3%)	16 (53%)	2 (7%)	7 (23%)	4 (13%)
Q8	*Do you think the interactive board is useful?*	7 (23%)	8 (27%)	8 (27%)	5 (17%)	2 (7%)
Q9	*Do you like seeing team positions on the interactive board?*	10 (31%)	6 (19%)	11 (34%)	3 (9%)	2 (6%)
Q10	*Do you like seeing scores on the interactive board?*	11 (34%)	12 (38%)	7 (22%)	0 (0%)	2 (6%)
Q11	*Did you use Opedia?*	1 (3%)	14 (44%)	1 (3%)	6 (19%)	10 (31%)
Q12	*Were the insights of Opedia useful?*	4 (13%)	8 (25%)	7 (22%)	3 (9%)	10 (31%)
Q13	*Was the Intelligent Pedagogical Agent useful?*	0 (0%)	10 (31%)	9 (28%)	6 (19%)	7 (22%)
Q14	*Was creating and sharing personal notes with team members useful?*	7 (19%)	8 (22%)	19 (51%)	2 (5%)	1 (3%)
Q15	*Do you like that the teacher is responsible for note validation?*	2 (6%)	4 (13%)	17 (53%)	4 (13%)	5 (16%)
Q16	*Would you like to have a personal customizable avatar?*	14 (50%)	8 (29%)	5 (18%)	1 (4%)	0 (0%)

Part 4: Preferences

Qn	Questions	Yes, a lot	Yes, enough	Indifferent	Not much	Not at all
Q17	*Do you like videogames?*	23 (74%)	0 (0%)	0 (0%)	6 (19%)	2 (6%)
Q18	*Do you think Gea 2: A New Earth can be considered a videogame?*	13 (41%)	0 (0%)	14 (44%)	5 (16%)	0 (0%)

TABLE V
TEST NO. 1—QUESTIONNAIRE RELIABILITY

Correlation Coefficient	.859435333
Spearman–Brown correction	.9244046488

tasks) in Italian, for the next test. (We, though, had to leave the IPA in English, as we had no availability for a reliable version of the NLP software to use for Italian language). Secondly, another important remark coming from students was that there were many bugs in the game and sometimes they wasted time while trapped in a collision box, or jumping in places with no exit, or just having fun. In the final version of the game we spent a lot of time fixing malfunctions and bugs.

Finally, we asked other questions aimed at evaluating single components of the game (usability), as reported in Table IV (Part 3). We used these feedback from Table IV to iterate the game interface, and Opedia® contents, in view of test no. 2.

The last part of the questionnaire regarded the attitude of the participants toward gaming, however we did not unveil any correlation of these two answers with the likeability or learnability parts of the questionnaire.

B. Summative Assessment: Pre–Post Tests

The summative evaluation was performed by pre–post competence questionnaires: 22 questions tried to measure STEM knowledge. They were scored 1 point for each correct answer, 0 otherwise. We grouped the questions in six categories, five of them representing specific subject matters, and one collecting general knowledge (see Table VI). Some questions were suitable to be part of more than one category. The questions and their partition in categories are reported in the supplementary material for this article.

The results of the summative assessment suggest that students feebly learned. However, as explained by the teachers, both classes were showing already a low level of proficiency. In particular, about the topics of interest for the test, only 30% of the students was reaching sufficiency at the time of the test, so that another test, given after only face-to-face lectures, would have been likely to show very poor results. Another important aspect, for the test evaluation, is in that students

TABLE VI
TEST NO. 1—AVERAGE GROWTH PER CATEGORY

Category	All	Males	Females
General	14,42%	14,28%	14,06%
Biology	17,53%	15,65%	17,26%
Environment	16,27%	23,81%	-2,08%
Astronomy	13,23%	14,88%	8,33%
Chemistry	23,48%	20,24%	19,45%
Physics	5,25%	0,24%	13,33%
All Categories	13,91%	12,97%	13,38%

had also a very poor preparation in general, as again put by the teachers. The game's contents have been prepared having in mind an audience more familiar with the prescribed syllabus of the target school grade, and this played against better results in the majority of the game's tasks.

VII. EVALUATING *Gea2*: TEST NO. 2

The second test followed the same organization of the previous one, and was performed with two classes in another school. The classes had the same disciplinary specializations as the two classes in Test 1. We had an overall of 53 participants (the classes were of 24 and 29 students) (Table VII shows the details). Table VIII shows the lecture schedule and methods of work.

TABLE VII
TEST NO. 2—DETAILS ON CLASSES

School Name	Classes	Students
ITIS Pascal	1H (*Scientific* institute)	29 (9 ♀, 20 ♂)
ITIS Pascal	1D (*Industrial/Technical* institute)	24 (12 ♀, 12 ♂)

TABLE VIII
TEST NO. 2—LECTURES SCHEDULE

Date	Classes	Methods
Play Meeting 1	1H and 1D	Game introduction, pretest
Play Meeting 2	1H and 1D	Play game
Play Meeting 3	1H and 1D	Play game
Play Meeting 4.1	1H	Play game, game conclusions
Play Meeting 4.2	1D	Play game, game conclusions
Conclusion.1	1H and 1D	Posttest, questionnaire
Conclusion.2	1H	Posttest, questionnaire

In test no. 2, we collaborated with two researchers from the pedagogy department of our university: One of them followed all the lectures, taking notes about the learning aspects of the game and about social dynamics among students, as related to the game. With their help we prepared a questionnaire, deemed to support analysis about the acquisition of side skills, such as the ability to apply problem decomposition, to perform progressive solution of a problem, and to put in practice various aspects of a collaborative approach (see Table XII).

According to test no. 1's feedback we got into a second development phase, where several improvements were introduced. We produced several minor usability enhancements, but the main changes were as follows.

- Quizzes and tasks were proposed in Italian;
- The amount of objects to collect was reduced to 6 per player;
- The use of IPA, notes and chat let now gain less points than in test no. 1;
- The IPA was now providing unsolicited hints.
- Added a further step necessary to unlock the experiment panel in the simulation rooms.

In test no. 1, in order to access the experiment panel it was enough to have the correct objects in the inventory, which are automatically selected by the game. Now, the player had to select the object(s) explicitly, to show more comprehension of the process. As mentioned above, the unsolicited hints system was a very important improvement coming for test no. 2. With this upgrade we enriched the game with the IPA, able to decide on its own whether to help the player or not [34].

In the following subsections we analyze the results of test no. 2, following the same lines adopted for the previous test. On the other hand, during this test, we added an analysis related to 1) the perception of the students about the soft skills related to the pedagogical questionnaire in Table XII, and 2) the students' reactions to the unsolicited hints. Two additional subsections consider these last aspects.

A. Students' Perception of the Experience

We used the same questionnaire used in test no. 1. Table IX reports the overall results. The internal consistency for this questionnaire was evaluated by Split-Half reliability [42], as seen for test no. 1. Table X) shows the results. We had a final *Cronbach's Alpha* at .91, which is still an excellent result.

We found that the percentage of likeability was more or less the same as for the first run. However, the learnability perception had improved a lot (from 37% to 64%). That is probably due to the use of Italian language, and to the improved quality of the game and contents. Unfortunately the student's perceptions had not a confirmation from the summative assessment, as we will see later. Students found the interface more clear, and that is due to the effort we have made in correcting errors in the game. We believe that during test no. 2 students used the interactive board more successfully, participating and exchanging with the teacher to a greater extent. We also saw that Opedia® was not used very much: even if its content is targeted to the game, students preferred going on the Internet. The IPA was under-used, as expected, considering that it remained in English: many students felt, again, not confident enough to ask questions in English. This could of course hinder also the appreciation for the unsolicited hints: we see later that, actually, the majority of the hints were positively appreciated, although the language problem remains. Finally, we saw that a strong majority wished to have a personal avatar to customise in the game. Also in test no. 2, the final two questions did not appear to be particularly revealing, or connected to the other results.

B. Summative assessment: Pre–post test

The summative evaluation in test no. 2 proceeded by the same protocol of test no. 1. We administered pre–post tests: Table XI shows the average knowledge growth per category.

TABLE IX
TEST NO. 2—QUESTIONNAIRE ABOUT STUDENTS' PERCEPTION OF THE EXPERIENCE

Part 1: Likeability

Qn	Questions	Yes, a lot	Yes, enough	Indifferent	Not much	Not at all
Q1	*Did you enjoy playing the game?*	5 (24%)	8 (38%)	5 (24%)	1 (5%)	2 (10%)
Q2	*Do you think that playing the game is more engaging than a traditional lesson?*	8 (38%)	7 (33%)	3 (14%)	2 (10%)	1 (5%)
Q3	*Would you recommend to students from other classes to try the game?*	3 (14%)	8 (38%)	4 (19%)	3 (14%)	3 (14%)

Part 2: Learnability

Qn	Questions	Yes, a lot	Yes, enough	Indifferent	Not much	Not at all
Q4	*Do you think you have learned by playing the game?*	4 (9%)	26 (55%)	0 (0%)	8 (17%)	9 (19%)
Q5	*Do you think that playing the game has allowed you to learn in a more interesting way?*	4 (9%)	26 (55%)	0 (0%)	8 (17%)	9 (19%)
Q6	*Do you think that playing the game has allowed you to learn faster?*	13 (28%)	19 (40%)	2 (4%)	5 (11%)	8 (17%)

Part 3: Usability

Qn	Questions	Yes, a lot	Yes, enough	Indifferent	Not much	Not at all
Q7	*Were the user interfaces clear enough?*	6 (13%)	23 (49%)	6 (13%)	6 (13%)	6 (13%)
Q8	*Do you think the interactive board is useful?*	12 (25%)	18 (38%)	6 (13%)	6 (13%)	6 (13%)
Q9	*Do you like seeing team positions on the interactive board?*	12 (26%)	16 (34%)	12 (26%)	3 (6%)	4 (9%)
Q10	*Do you like seeing scores on the interactive board?*	12 (26%)	16 (34%)	12 (26%)	4 (9%)	3 (6%)
Q11	*Did you use Opedia?*	6 (12%)	14 (28%)	8 (16%)	16 (32%)	6 (12%)
Q12	*Were the insights of Opedia useful?*	4 (9%)	11 (23%)	16 (34%)	8 (17%)	8 (17%)
Q13	*Was the Intelligent Pedagogical Agent useful?*	4 (9%)	13 (28%)	6 (13%)	12 (26%)	12 (26%)
Q14	*Was creating and sharing personal notes with team members useful?*	10 (21%)	15 (32%)	12 (26%)	5 (11%)	5 (11%)
Q15	*Do you like that the teacher is responsible for note validation?*	8 (17%)	13 (28%)	10 (21%)	6 (13%)	10 (21%)
Q16	*Would you like to have a personal customizable avatar?*	26 (55%)	12 (26%)	5 (11%)	2 (4%)	2 (4%)

Part 4: Preferences

Qn	Questions	Yes, a lot	Yes, enough	Indifferent	Not much	Not at all
Q17	*Do you like videogames?*	33 (70%)	0 (0%)	0 (0%)	10 (21%)	4 (9%)
Q18	*Do you think that Gea 2: A New Earth can be considered a videogame?*	20 (43%)	0 (0%)	20 (43%)	7 (15%)	0 (0%)

TABLE X
TEST NO. 2—QUESTIONNAIRE RELIABILITY

Correlation Coefficient	.8361431978
Spearman–Brown correction	.9107603359

TABLE XI
TEST NO. 2—AVERAGE GROWTH PER CATEGORY

Category	All	Males	Females
General	25,02%	26,39%	22,17%
Biology	32,57%	33,61%	28,57%
Environment	35,14%	37,86%	27,08%
Astronomy	16,55%	22,48%	9,03%
Chemistry	38,23%	37,77%	36,11%
Physics	21,36%	26,86%	16,67%
All Categories	25,55%	28,34%	20,96%

The general result is not so good, from the point of view of effectiveness of the game as a learning tool. In fact, the main problem was in the conditions in which the game was used. We expected the game would improve the students understanding of the topics that had been already explained in class, while the game ended up being used as a replacement of face-to-face lectures on the related topics. On the contrary, according to teachers' information, the students were not at all prepared in advance on the topics of the game, and they were showing a low proficiency in general (more than half of the students, 16 in 1H and 14 in 1D, were with insufficient marks in STEM subject matters).

We can, however, report a positive result here, with the substantial improvement of the knowledge growth with respect to test no. 1 (average growth is 25.4% in test no. 2 and was 13.9% in test no. 1).

C. Pedagogical Questionnaire

As mentioned above, we worked with two pedagogues and, beside the annotations during game sessions, they helped preparing a questionnaire (using a Likert scale) aimed at verifying how certain side skills were possessed, such as approaches to problem solving (problem decomposition and progressive solution), and orientation toward a collaborative

approach. Table XII shows the results (all statements started with "*I learned to:*").

Plain positive responses ("Yes ...") cover an average of 32.7% of the answers (SD 7.58). Non negative answers ("Yes ..." and "Indif.") have a mean of 65.2% (SD 7.65). Two interesting aspects in the questionnaire are about the perception of the importance of *Collaboration* and *Interdisciplinary*. We think that questions Q2, Q5, Q6, Q7, and Q8 can better reveal the perception about *Collaboration*: in this case, we have for non negative answers, a mean of 67.93 (SD 4.82). Similarly, we think that questions Q4 and Q10 can be representative of the *Interdisciplinary* perception: in this case, we have a mean of 66.67. We conclude that these aspects seem to be felt by the students as the more positive outcome of the game.

D. Assessment of Unsolicited Hints

The hints mechanism has been just tested jointly with the game itself, so we could not part its specific achievements from those of the game. From a formal point of view we can only see the perception of the students, as provided by the answers to the questionnaires (Tables IV and IX) and from the feedback during the experimentation, reported in the next paragraph. From these data we inferred that this would be a core feature for the success of the game.

In order to evaluate it, instead of asking players if they found the hints useful, we gave them the opportunity to *vote* thumb up or down.

In one game session, for all users (14 because they played in pairs when possible), we provided 72 hints, 47 were judged positive (65%) and 25 negative (35%). Interestingly, we saw that when players' emotion was negative (so, some frustration was felt, and help was much needed) they would always reply the hint with thumb up, whereas when emotion was positive the reply would be always thumb down.

To have a better understanding of unsolicited hints provisioning we have to make the following annotations:

- All the provided hints were in English and students complained about that
- Emotion detection was limited because of the limited usage of IPA and chat.

With all the above limits, we have an indication that emotions' detection can be a valuable mechanism to provide help during game sessions.

VIII. Discussion of Results and Limits of the Research

Beside presenting the design and implementation of *Gea2*, the two main motivations of this paper were in 1) evaluation of how high school students would appreciate a game activity integrated in their normal study program, and 2) analysis of the effectiveness of such an activity on learning outcomes. We had the possibility to perform a two-phases development and experimentation, also enjoying the collaboration of teachers. On the other hand, we had also to deal with rigidity and constraints during the experimentation, derived by the administrative organization of the schools.

We think that, after the experience, we can present some positive results, some controversial findings, and a description of the difficulties to overcome in our educational system, and maybe others', while undertaking experiments with GBL technology.

A. Discussion

Sections V, VI, and VII have shown positive results with respect to the students' perception of *Gea2* usefulness.

A large majority of participants liked using the game, with an average of positive or very positive likeability staying above 60% in both the test phases. There is, in fact, a decrease from test no. 1 to test no. 2, mainly related to answers to Q2, which we cannot explain, unless referring to the the fact that we had to work with different classes, in test no. 2, than in test no. 1 (see a discussion of limits of this research in VIII-B).

Regarding learnability, we have witnessed a growth in the perception of usefulness shown by the students, and, in the second phase, almost two out of three participants found *Gea2* an effective support to learning: This perception grew from 42% of test no.1 till 65% of test no.2.

We think that the analysis of usability did not provide great insight: in general the average positive or very positive feedback rise from 55% to 63% along the test phases, but we saw issues with some functionalities and especially with the system's language, on which we worked in between phases, and that needs additional work before before being ready for further experiments. Notably there are three questions where a decrease of feedback is witnessed: Q11 and Q12 are related to the use of Opedia®, which prompts us to make changes in future; on the other hand, Q10 (concerning the reaction at seeing the team score made public) is conceivably related to different availability to competition of the involved classes, but we don't think we can offer a final explanation.

Regarding the effectiveness of *Gea2* as a learning enhancer, we proposed pre–post questionnaires. As seen in Sections VI and VII we partitioned the questions in categories, and in each test we computed the average growth of students' proficiency on the categories. The result is not rich, but not negative at all in absolute terms: the average growth over all the categories was 14% in test no. 1, and raised to 25% in test no. 2.

One aspect of *Gea2* is in the implicit interdisciplinary learning experience it fosters. In the game, the players face the need to use knowledge coming from different fields, such as chemistry, physics, astrobiology, or geology, not to mention the concepts connected to life sustainability. We think that *Gea2* was able to foster such an interdisciplinary learning, considering the different roles that the students in the same team impersonates, and also in view of the results described in the previous paragraph, coming from questionnaires where questions related to the various fields of interest were comprised.

The use of English language had been a continuing problem in our tests. This note is supported by the growth of learnability (as perceived in the tests) and by the fact that the interactive board was used more and more successfully, during the second test, while the part of the game that remained in English was less used.

TABLE XII
TEST NO. 2—PEDAGOGICAL QUESTIONNAIRE (ANSWERS TO "I LEARNED TO ...")

Qn	Questions	Yes, a lot	Yes, enough	Indifferent	Not much	Not at all
Q1	*Break-down tasks into small problems*	3 (6%)	9 (19%)	16 (33%)	13 (27%)	7 (15%)
Q2	*Appreciate collaborative work*	6 (13%)	15 (31%)	10 (21%)	12 (25%)	5 (10%)
Q3	*Appreciate technology potential*	2 (4%)	10 (21%)	13 (27%)	14 (29%)	9 (19%)
Q4	*View topics from a different perspective*	1 (2%)	12 (25%)	14 (29%)	14 (29%)	7 (15%)
Q5	*Achieve complex tasks in a collaborative way*	5 (10%)	13 (27%)	15 (31%)	8 (17%)	7 (15%)
Q6	*Plan collaborative work*	5 (10%)	10 (21%)	21 (44%)	9 (19%)	3 (6%)
Q7	*Understand requirements coming from different people*	7 (16%)	8 (18%)	14 (31%)	12 (27%)	4 (9%)
Q8	*Develop an idea in a collaborative way*	4 (8%)	13 (27%)	15 (31%)	9 (19%)	7 (15%)
Q9	*Understand the importance of work in progress*	1 (2%)	10 (21%)	22 (46%)	11 (23%)	4 (8%)
Q10	*Understand the importance of interdisciplinarity*	6 (13%)	16 (33%)	15 (31%)	7 (15%)	4 (8%)

This subsection presented the quantifiable results of the proposed research. They can be summarized as being 1) related to a part of the students population which occurs less frequently in studies of GBL, 2) confirmatory of other results in literature about the attractiveness and appreciation of students for SGs integrated in their study activity, and 3) not firmly concluding on the effectiveness of the game as a learning tool.

In the next subsection we explore the limits of this research (that produced, for instance, the above point 3). Among such limits, we analyze also some factors of disturbance, acting on the experiment, trying to give a description of experience that could help (hopefully not only us) in future experimental activities.

B. Limits

The first aspect worth noticing in this subsection, is in that *Gea2* effectiveness, as a learning tool, was measured only by a non in-depth analysis of pre–post tests on an overall limited number of students. This of course means that we cannot claim a whole result on this side, neither positive nor negative. As an explanation we can submit that, during the experimentation, we have been given limited allowance by the teachers, who were in turn subject to administrative and formal constraints. For instance, we could experiment only on a few classes and only by exposing the whole of each class to *Gea2*, and we could not use data from other classes, or from previous years. Other constraints were in the timing of the activities, that could be performed only during official lecture hours allotted for the teacher's subject matter, according to an inflexible time plan. And extra school activities were excluded by the nature of the experiment, that was supposed to be conducted in classroom.

This also brought the consequence that the time the teacher was giving us (i.e., to the game playing) was subtracted by the normal lecture time, which implied that the game ended up being a substitute of the lecture on the topics of interest, and not, as planned in our design, as a supporting resource for the lecture.

Regarding the effectiveness issue, we collaborated with the teachers to examine the results and concluded that the unsatisfactory results mostly depended on three factors:

1) The objectively low level of students' previous proficiency in STEM subject matters, which might have made the learners, on average, ill equipped to deal with *Gea2*'s topics;
2) The use of the game as a substitute, rather than a complementary tool, for the lectures hosting the game sessions. The game could not have been an additional activity, as a consequence of the tight schedule a class have during the school year;
3) The presence of some implementation issues in the current software version of the game, and in particular the extensive use of English language in the user interfaces: on average, the students had greater difficulties with this language issue than expected.

The first two issues as listed above were also true in the second test, so they represent a continuing limitation of this research. In regards to the third issue above, we could learn from the feedback and let the game undergo an implementation upgrade. We solved several software and interface issues, and localized the interface language, except than for the IPA since we had no reliable NLP software to be used in place of the current one.

Two additional aspects could be pointed out as limiting the overall effectiveness of *Gea2* in the experiments. One aspect is the persistent need for interaction in English with the IPA. The other aspect is related to the lack of a complete scaffolding system in *Gea2*: we couldn't involve the teachers in the implementation of such a sub-system (especially with regard to the development of adequate contents), so we opted for the inclusion in *Gea2* of Opedia®. Eventually Opedia® was not used (nor appreciated—see answers to Q11) as much as we hoped, so future versions of *Gea2* will have to deal with a proper solution. Although still suffering from the issues mentioned above in point 1) and 2), the second test produced better results than the first, as seen in Section VIII-A. So we hope that, despite the limitation described above, this paper can provide the reader with useful results, and with useful information about threats to avoid during an experimentation with SG, and GBL in general.

IX. CONCLUSION

Digital games are going to be used more and more extensively for educational purposes. Research in AI is giving useful insights for the development of VLEs, allowing for intelligent, adaptive, and effective behavior of tutoring systems. Following

these premises, we have developed *Gea2*, an SG to be used in class for teaching STEM in high school, as a supporting and completing part of the lecture. The pedagogical aims of *Gea2* are to encourage the acquisition of meta competences and 21st-century skills, beside learning STEM in a playful way. Students can explore a 3-D environment that includes a spaceship and three simulation rooms, one for each planet that they have to evaluate, looking for objects that unlock interactive experiments in the simulation rooms, and solving quizzes.

Overall, we can conclude that the game proved to be enjoyable and students would recommend it to others. As of the AI algorithm performance underneath the IPA, students assessed it to be effective. The students in the sample found the game more engaging than traditional lectures, and felt that they were learning while playing, although not necessarily in a faster way. In addition, the game interfaces received positive feedback, as students found them easy to understand and useful.

An educational game such as *Gea2* has the development of problem solving and critical-thinking skills among its natural aims; we did not measure the effectiveness of *Gea2*, out of the experiments presented in this paper, and we regard further investigation in this direction as a topic for future work. On the other hand, the first interventions on the game will be those directly based on the students' feedback we had after the tests. The most requested feature was to make an avatar available in the game. It would represent the player in the virtual environment, and should be customizable by them, to reflect their personal traits. We believe that this could strengthen the gaming experience and students' engagement.

Lastly, another future development that we see useful for the game will be a text-to-speech engine, which would improve the modularity and the chance to extend the use of the IPA.

Acknowledgment

The authors thank the instructors and students of the schools *ITI Pascal* and *ITI Bernini*, for their support in testing and validating *Gea2*.

References

[1] N. Pinkwart, "Another 25 years of AIED? Challenges and opportunities for intelligent educational technologies of the future," *Int. J. Artif. Intell. Educ.*, vol. 26, pp. 771–783, 2016, doi: 10.1007/s40593-016-0099-7.

[2] E. Wenger, *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge*. Morgan Kaufmann, 2014.

[3] C. Aldrich, *Learning Online with Games, Simulations, and Virtual Worlds: Strategies for Online Instruction*. Hoboken, NJ, USA: Wiley, 2009, vol. 23.

[4] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining 'gamification'," in *Proc. 15th Int. Academic MindTrek Conf. (MindTrek'11)*. ACM, Tampere, Finland, Sept. 28–30 2011, pp. 9–15, doi: 10.1145/2181037.2181040.

[5] K. M. Kapp, *The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education*. Hoboken, NJ, USA: Wiley, 2012.

[6] R. L. Lamb, L. Annetta, J. Firestone, and E. Etopio, "A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations," *Comput. in Human Behav.*, vol. 80, pp. 158–167, 2018, doi: 10.1016/j.chb.2017.10.040.

[7] V. M. Perez-Colado, I. J. Perez-Colado, M. Freire-Moran, I. Martinez-Ortiz, and B. Fernandez-Manjon, "Simplifying the creation of adventure serious games with educational oriented features," *Educ. Technol. & Soc.*, vol. 22, no. 3, pp. 32–46, 2019.

[8] D. Tuparova, V. Veleva, and G. Tuparov, "About some barriers in usage of educational computer games by teachers in STEM," in *Proc. 42nd Int. Conv. Information and Communication Technology, Electronics and Microelectronics (MIPRO'19)*. IEEE, Opatija, Croatia, May 20–24 2019, pp. 727–730, doi: 10.23919/MIPRO.2019.8756999.

[9] S. de Freitas, "Are games effective learning tools? A review of educational games," *Educ. Technol. & Soc.*, vol. 21, no. 2, pp. 74–84, 2018.

[10] J. Kirriemuir and A. McFarlane. "Literature review in games and learning,". NESTA Futurelab, Bristol, UK, Rep. No. 8, 2004. [Online]. Available: https://www.researchgate.net/publication/32231341_Literature_Review_in_Games_and_Learning

[11] R. van Roy and B. Zaman, "Need-supporting gamification in education: An assessment of motivational effects over time," *Comput. & Educ.*, vol. 127, pp. 283–297, 2018, doi: 10.1016/j.compedu.2018.08.018.

[12] F. Gao, L. Li, and Y. Sun, "A systematic review of mobile game-based learning in stem education," *Educ. Technol. Res. Develop.*, vol. 68, pp. 1791–1827, 2020, doi: 10.1007/s11423-020-09787-0.

[13] M. Wang and X. Zheng, "Using game-based learning to support learning science: A study with middle school students," *The Asia-Pacific Educ. Researcher*, vol. 30, pp. 167–176, 2021, doi: 10.1007/s40299-020-00523-z.

[14] M. Soliman and C. Guetl, "Intelligent pedagogical agents in immersive virtual learning environments: A review," in *Proc. 33rd Int. Conv. Information and Communication Technology, Electronics and Microelectronics (MIPRO'10)*. IEEE, Opatija, Croatia, May 24–28 2010, pp. 827–832.

[15] A. Serrano-Laguna, B. Manero, M. Freire, and B. Fernández-Manjón, "A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes," *Multimedia Tools and Appl.*, vol. 77, pp. 2849–2871, 2018, doi: 10.1007/s11042-017-4467-6.

[16] P. Backlund and M. Hendrix, "Educational games — Are they worth the effort? A literature survey of the effectiveness of serious games," in *Proc. 5th Int. Conf. Games and Virtual Worlds for Serious Applications (VS-GAMES'13)*. IEEE, Poole, UK, Sept. 11–13 2013, pp. 1–8, doi: 10.1109/VS-GAMES.2013.6624226.

[17] M. Liu, J. Kang, S. Liu, W. Zou, and J. Hodson, "Learning analytics as an assessment tool in serious games: A review of literature," in *Serious Games and Edutainment Applications*, M. Ma and A. Oikonomou, Eds. London, UK: Springer-Verlag, 2017, vol. II, pp. 537–563, doi: 10.1007/978-1-4471-2161-9.

[18] M. A. Miljanovic and J. S. Bradbury, "A review of serious games for programming," in *Proc. 4th Joint Conf. Serious Games (JCSG'18)*, ser. LNCS, vol. 11243. Springer, Cham, Darmstadt, Germany, Nov. 7–8 2018, pp. 204–216, doi: 10.1007/978-3-030-02762-9_21.

[19] Ludwig. (2013) (Last accessed on 2021-08-15). [Online]. Available: http://www.playludwig.com/en/info/info-for-teachers/

[20] W. Hagborg, "Grades and motivational orientation among high-school students," *J. Psychoeducational Assessment*, vol. 10, no. 4, pp. 355–361, 1992, doi: 10.1177/073428299201000405.

[21] M. Virvou, G. Katsionis, and K. Manos, "Combining software games with education: Evaluation of its educational effectiveness," *Educ. Technol. & Soc.*, vol. 8, no. 2, pp. 54–65, 2005.

[22] S. L. Bowman, *The Functions of Role-Playing Games: How Participants Create Community, Solve Problems and Explore Identity*. Jefferson, NC, USA: McFarland & Co., 2010.

[23] A. Terracina *et al.*, "Conquering an exo-planet through the use of a virtual role playing game assisted by an emotionally intelligent pedagogical agent," in *Proc. 10th Eur. Conf. Games Based Learning (ECGBL'16)*, Paisley, Scotland, UK, Oct. 6–7 2016, pp. 666–675.

[24] NASA. (2017) Last accessed on 2021-08-15. [Online]. Available: https://www.nasa.gov/press-release/nasa-telescope-reveals-largest-batch-of-earth-size-habitable-zone-planets-around

[25] G. D. Borich, *Effective Teaching Methods: Research-Based Practice*, 7th ed. Pearson Education India, 2011.

[26] R. Bybee, "Teaching science as inquiry," in *Inquiring into Inquiry Learning and Teaching in Science*, J. Minstrell and E. H. van Zee, Eds. Washington, DC, USA: American Association for the Advancement of Science, 2000, pp. 20–46.

[27] R. E. Mayer, "Should there be a three-strikes rule against pure discovery learning?" *Amer. Psychologist*, vol. 59, no. 1, pp. 14–19, 2004, doi: 10.1037/0003-066X.59.1.14.

[28] K. Goodnough and M. Cashion, "Exploring problem-based learning in the context of high school science: Design and implementation

issues," *School Sci. & Math.*, vol. 106, no. 7, pp. 280–295, 2006, doi: 10.1111/j.1949-8594.2006.tb17919.x.

[29] D. Cichon and J. Ellis, "The effects of MATH Connections on student achievement, confidence, and perception," in *Standards-based school mathematics curricula*. Evanston, IL, USA: Routledge, 2003, pp. 345–374, doi: 10.4324/9781003064275.

[30] R. Q. Berry III, P. A. Reed, J. M. Ritz, C. Y. Lin, S. Hsiung, and W. Frazier, "STEM initiatives: Stimulating students to improve science and mathematics achievement," *The Technology Teacher*, vol. 64, no. 4, pp. 23–29, 2004.

[31] J. Frykholm and G. Glasson, "Connecting science and mathematics instruction: Pedagogical context knowledge for teachers," *School Sci. & Math.*, vol. 105, no. 3, pp. 127–141, 2005, doi: 10.1111/j.1949-8594.2005.tb18047.x.

[32] S. Nikitina and V. B. Mansilla. "Three strategies for interdisciplinary math and science teaching: A case of the Illinois Mathematics and Science Academy,". Harvard Graduate School of Education, Cambridge, MA, USA, Goodwork Project Rep. Ser. no. 21, Mar. 2003. [Online]. Available: https://www.researchgate.net/publication/228379792_Three_Strategies_for_Interdisciplinary_Math_and_Science_Teaching_A_Case_of_the_Illinois_Mathematics_and_Science_Academy

[33] Photon. (2017) Photon chat engine. Last accessed on 2021-08-15. [Online]. Available: https://www.photonengine.com/en/chat

[34] L. Ferro, F. Sapio, M. Mecella, M. Temperini, and A. Terracina, "Intelligent pedagogic agents (IPAs) in GEA2, an educational game to teach STEM topics," in *Proc. 10th Int. Conf. Methodologies and Intelligent Systems for Technology Enhanced Learning (MIS4TEL'20)*, ser. AISC, vol. 1236. Springer-Verlag, Online, Oct. 6–8 2020, pp. 226–236, doi: 10.1007/978-3-030-52287-2_23.

[35] M. Lease, "Natural language processing for information retrieval: The time is ripe (again)," in *Proc. Conf. Information and Knowledge Management (CIKM'07), 1st Ph.D. Workshop in CIKM (PIKM'07)*. ACM, Lisbon, Portugal, Nov. 6–10 2007, pp. 1–8, 10.1145/1316874.1316876.

[36] C. Conati, "Probabilistic Assessment of User's Emotions in Educational Games," *J. Appl. Artif. Intell.*, pp. 555–575, 2002, doi: 10.1080/08839510290030390.

[37] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[38] D. R. Michael and S. L. Chen, *Serious Games: Games that Educate, Train, and Inform*. Boston, MA, USA: Thomson Course Technology, 2005.

[39] R. T. Hays. "The effectiveness of instructional games: A literature review and discussion,". Naval Air Warfare Center Training Systems Division, Orlando, FL, USA, Rep. no. 2005-004, Nov. 2005. [Online]. Available: https://apps.dtic.mil/sti/pdfs/ADA441935.pdf

[40] J. Enfield, R. D. Myers, M. Lara, and T. W. Frick, "Innovation diffusion: Assessment of strategies within the diffusion simulation game," *Simul. & Gaming*, vol. 43, no. 2, pp. 188–214, 2012, doi: 10.1177/1046878111408024.

[41] C. Boston, "The concept of formative assessment," *Practical Assessment, Res., & Eval.*, vol. 8, Article 9, 2002, doi: 10.7275/kmcq-dj31.

[42] N. J. Salkind, *Encyclopedia of research design*. Newbury Park, CA, USA: SAGE, 2010, vol. 1, doi: 10.4135/9781412961288.

Lauren S. Ferro obtained the Ph.D. degree in Media and Communication, specializing in player and user modeling and profiling, from the Royal Melbourne Institute of Technology, Melbourne, Australia, in 2017. She is currently pursuing a second Ph.D. from the Department of Computer, Control, and Management Engineering, Sapienza University, Rome, Italy, focusing on human factors in cybersecurity. Her research activity centers primarily around understanding human and player behavior in various interactive experiences in the areas of gamification, game-based learning, cultural heritage, cybersecurity, and psychology.

Massimo Mecella obtained the Ph.D. degree in engineering in computer science from Sapienza University of Rome, Rome, Italy, in 2002. He is currently a full professor with the Department of Computer, Control, and Management Engineering, Sapienza University of Rome, Rome, Italy. His research focuses on business process management, cyber-physical systems and the Internet of Things, advanced human–computer interfaces, software architectures, and service-oriented computing, with applications in multiple fields including digital government, smart spaces, Industry 4.0, healthcare, disaster/crisis response and management, cybersecurity, and the digital humanities. He has published over 180 research papers and chaired various conferences in the above-mentioned areas.

Francesco Sapio obtained the PhD degree in Engineering in Computer Science from Sapienza University of Rome, Rome, Italy, in 2021. He is currently a researcher with the Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy. His research focuses on Serious Games, Cultural Heritage, Social and Collaborative Learning, Game Based Learning, Technology Enhanced Learning, AI applied to videogames, Gaming Software Architectures, Human-Computer Interaction, Industry 4.0, with applications in 3-D Visualization, Architecture, Gaming, Healthcare and Entertainment.

Marco Temperini obtained the Ph.D. degree in computer science from Sapienza University of Rome, Rome, Italy, in 1992. He is currently an associate professor with the Department of Computer, Control, and Management Engineering at the same university. His research activity in the education arena is on adaptive e-learning, social and collaborative learning, game-based learning, automated assessment, and learning analytics. He has organized several international workshops and chaired a number of international conferences in the field of technology-enhanced learning. He has been a work-package leader and research unit coordinator for several European Union research projects.

Annalisa Terracina obtained the Ph.D. degree in computer science from Sapienza University of Rome, Rome, Italy, in 2017. She has been working in the field of Earth observation and space missions for many years, previously serving as a consultant for the European Space Agency (ESA) and the Italian Space Agency (ASI) as well as a technical manager for several European Union projects. Since 2018, she has been a project manager for ESERO Italia, a collaboration between ESA and ASI focusing on space-related STEM education.

SUPPLEMENTARY MATERIAL

The list of questions used for the Summative assessments (translated from local language) were:

- Q1: What is an ESO-Planet.
- Q2: What does integrated science in school means?
- Q3: What is it a NEO?
- Q4: What does ASI stand for?
- Q5: With which experiment is it possible to calculate the gravity acceleration?
- Q6: The density of an object is defined as ...
- Q7: What are the 3 main gases that make up the earth's atmosphere?
- Q8: What instrument can I use to study the composition of a gas?
- Q9: The inclination of the Earth's axis is important for which reason?
- Q10: How long does it take for humans to die of hypothermia if exposed to a temperature of -20° ?
- Q11: What is the atmospheric pressure of the earth's atmosphere at sea level?
- Q12: Why is the Earth's magnetic field important?
- Q13: Have we discovered liquid water on other planets in the solar system?
- Q14: Does the inclination of the Earth's axis affect the length of the day?
- Q15: What do all the solar system planets have in common?
- Q16: What is an astro-biologist?
- Q17: Are volcanoes important for the origin of life?
- Q18: Who should be part of a team of scientists studying an exoplanet?
- Q19: What is the acceleration of gravity on Earth worth?
- Q20: What is a conceptual map useful for?
- Q21: What the water cycle defines?
- Q22: What is the difference between a virus and a bacterium?

Table XIII shows the partition of the questions in categories, used for the analysis of effectiveness of *Gea2*. The names of the categories are self-explanatory. Notice that each question can pertain to, and so being included in, different categories.

TABLE XIII
QUESTIONS IN EACH CATEGORY

Category	Questions
General	Q1, Q2, Q3, Q4, Q13, Q16, Q17, Q18, Q20
Biology	Q21, Q7, Q8, Q10, Q22, Q16, Q17
Environment	Q21, Q7
Astronomy	Q3, Q9, Q13, Q12, Q15, Q16
Chemistry	Q6, Q7, Q8, Q22
Physics	Q5, Q6 Q11, Q12, Q14, Q15, Q19