Study of the Learning Performance and User Experience of the Interactive Systems of the Catavento Cultural Museum

Ana Grasielle Dionisio Corrêa
Programa de Pós-Graduação em Distúrbios do Desenvolvimento (PPG-DD)
Universidade Presbiteriana Mackenzie, Brazil

Marcelo Nogueira Rossi
Universidade Federal de São Paulo (UNIFESP)
Brazil

Irene Karaguilla Ficheman
Escola Politécnica da Universidade de São Paulo (EP-USP)
Brazil

Roseli De Deus Lopes
Escola Politécnica da Universidade de São Paulo (EP-USP)
Brazil

Abstract

Educational interactive technologies are placed in museums to enhance the visitor experience. Great time and money and hard work hours are spent to design, develop and deploy these systems. But do these systems actually achieve their pedagogical goals? Do they contribute to greater effectiveness and user experience (UX)? In this article we report the use of the technological attraction "Adventure in the Solar System", from the Catavento Cultural Museum, in São Paulo, Brazil. The attraction simulates an adventure in space through a spaceship composed of collaborative games designed to teach about astronomy, physics and mathematics. 129 students participated in this empirical study and were classified into three groups (ESI - Elemental Scholl I; ESII - Elemental Scholl II; HS - High Scholl), all aged between 7 and 16 years. The main objective of the study was to analyze learning performance and user experience (UX). The results showed that there were no statistically significant results betwen groups, but higher proportions of positive scores were observed in the ESII and SH groups. Regarding the user experience, ESI and ESII presented more expressive results than SH. The main results are discussed and ideas for future research are presented.

Keywords: Museums, interactive systems, educational games, user experience.

1. Introduction

Traditionally, museums were created to store and catalog real objects, works of art and important artifacts to convey information to visitors. Traditionally, in this scenario the information is passively transmitted, that is, the visitor behaves like a mere spectator. This scenario has changed in the last two decades [1, 2]. There is an increase in the interest of museums in the implementation of exhibitions that allow greater involvement of visitors through innovative and interactive technologies [3]. Interactive systems in science museums have basically two main objectives [4]: to attract more visitors and to transmit information in a more participatory, engaging and effective way.

Museums are open to any kind of media that the new communication and information technologies force them to fulfill. The expectations of the millennium generation in particular, following the latest technological innovations, are driving museums to a technological dependency with unpredictable consequences [1]. Using multimedia technologies and supported by social media, visitors move from passive learning consumers to active learning consumers, where they participate as co-authors in a learning process [4, 5, 6].

But do these interactive systems really achieve their goals? Do they provide a good user experience (UX)? We know that developing an interactive museum system requires a lot of time, team effort, and a high cost. So is it worth investing in this kind of technology? In this article we evaluate the learning performance and the user experience (UX) of the attraction "Adventure in the Solar System", installed in the Catavento Cultural Museum in São Paulo, Brazil. The objective of this evaluation was to verify if the attraction is adherent to its purpose of teaching concepts related to astronomy, physics and mathematics through collaborative digital games.

The study was carried out with 129 elementary school students from public schools in the city of São Paulo - Brazil. A research carried out in the literature served as a theoretical reference for the construction of an evaluation model that made it possible to identify: (a) the level of motivation of the students when using the games; (B) whether the games offer a good user experience; And (c) the learning performance. To investigate whether the positive scores differed before and after the cognition questionnaire applications, we run General Linear Models (GLM) taking into account the grade and the gender of the students.

In the next section we present some works related to interactive systems installed in science museums, showing that there was a tendency to implant such systems to improve the experience of their visitors. Next, we present the evaluated system, the attraction "Adventure in the Solar System". The article then presents the research method, results and discussions. Finally, we present the conclusions and perspectives of future work.

2. Related Work

Many modern museums around the world have explored new technologies as a tool to attract and engage their visitors and to pass information more authentically [2, 3]. Technologically advanced systems can engage the user synestesically in various possibilities through audio, video and other directions, to convey information more effectively. Such systems have (or should at least have) a theoretical basis for transmitting scientific knowledge to as many people as possible. In this section we present a brief discussion of museum learning, and then we review some works involving interactive museum systems.

Museums have changed over time and many museums have gone from being a simple arts presentation space to being an extension of the classroom [7, 8, 9]. According to Andre and Volman [10], museums and scientific centers stimulate students' curiosity and offer the opportunity to improve, at least in part, the needs of schools lacking laboratories, audiovisual resources, among other recognized resources to stimulate learning. In science museums, visitors have the opportunity to perform actions and discoveries, that is, they become active subjects, as they have a direct relationship with the devices through the manipulation or observation of how others manipulate them [11, 12].

Learning in museums should extend beyond cognitive gains. There is research emphasizing attitudinal, affective and social outcomes [13, 14, 15]. Thus, a successful outcome of a museum visit should be an enjoyable experience that cultivates positive attitudes toward the museum and its purpose. Upon leaving the museum, students can remove the knowledge learned. However, they must have the will and enthusiasm to continue their investigation outside the walls of the museum, giving the teacher a solid platform to base the post-visit school work.

A review of learning aspects obtained through science museum technology was done by Hawkey [16]. The author discusses learning experiences in museums and asks the question: should museums propose to offer content or engagement? Should reasoning be passive (information transmission only) or active (constructivist)? The author then proposes different learning taxonomies that were considered relevant to his time.

In recent years, emphasis has been placed on the construction of interactive museum systems [3, 4, 5, 6, 17, 18). Many of these systems place visitors in a constructive process where they are faced with several possible paths and learning possibilities [19]. These new developments take into account the learning taxonomies of the digital age as an active process where they can build new knowledge.

Zaharias [4] classifies interactive museum systems into two categories: those accessible remotely through multimedia systems or Virtual Reality and those within the physical spaces of the museum. The first case puts the participant in contact with a virtual museum, with representations of works and exhibitions in 3D, for example [3, 20, 21, 22). In the second case, the interactive systems that are located within the spaces of the museum, aim to attract more visitors and give them a different way of obtaining information pertinent to that museum. Such systems provide innovative ways of learning in order to generate pleasure and satisfaction in your visitors. Several systems allow 3D visualization including applications running on Virtual Reality and Augmented Reality systems [3, 5]. These technologies combine, to a large extent, entertainment, education and learning.

Several studies have investigated issues such as engagement and motivation of learners, increased learning performance and knowledge retention [6, 12, 23, 24]. The evaluation of these types of systems is a relevant question of research and should be valued, since they are used by different profiles of visitors with different age, knowledge and cultures. Despite the growing interest in the evaluation of interactive technologies in science museums, studies on the user experience assessment (UX) issue are still scarce.

Vavoula et al. [11] presented an evaluation of the Myartspace system that uses mobile phones to support learning in museums. This system allowed students to collect data during an out-of-school experience. The data was automatically sent to a site where students could later obtain this information and share it with others. The evaluation of this study was based on questions of usability, learning effectiveness and impact of this technology in the context of museums.

In the MuseumScoutts project [12] a visitor-centered approach to museums was adopted. In this project, students use collaborative authoring tools to collect information during the museum visit and create multimedia presentations. The focus is on the possibility of providing a tool for creating authentic works developed by students.

Reynolds et al. [25] presented a three-track evaluation program at the Victoria and Albert Museum in London. Evaluations were successful, but usability issues with device and network usage were found from these assessments. However, students' feedback enabled the authors to note that there was an improvement in learning and increased their interest in exploring the museum's other spaces.

Zaharias et al. [4] developed an interactive 3D holographic projection system for use in experimental learning based on physiology. Students manipulate three-dimensional learning objects (targets) through the non-tactile somatosensory mode to learn about the characteristics of the physiological structure in the 3D holographic projection environment. The study explored the usability factors of the system to improve human-computer interaction. Four important usability factors of the system were proposed through Principal Component Analysis (PCA): Labeling, Continuity, Backlash, and Ambiences.

The MuseumEye project explored the user experience (UX) of Augmented Reality applications in museums. The study contributes to synthesize a UX design model for AR applications to achieve the optimal levels of user interaction required, which ultimately reflect the entire museum experience [5].

Studies of human-computer interaction in museums can not only show the wisdom of these tools to the experimenter, but also present new methods of design and evaluation for the scientific community and professionals. In addition, these studies should be valued as they allow us to see improvements that can impact the usability of the system and consequently the learning. This work meets this need and proposes an evaluation model to analyze learning performance and user experience (UX) with the interactive system "Adventure in the Solar System".

3. The attraction "Adventure in the Solar System"

The Catavento Cultural Museum, located in the city of São Paulo, emerged through the movement of interiorization of science in Brazil. It was inaugurated in 2009, was designed with the function of integrating science and social problems in an attractive and interactive way. It has about 250 attractions in four sections whose subjects interrelate: Universe, Life, Engineering and Society [26].

The "Adventure in the Solar System" is an installation of the Universe section of the Catavento Museum designed to simulate an interactive space journey through a projection system and interactive activities presented in game format. The activities of the museum are carried out in three moments: setting, simulation and reflection. The scenario occurs at the beginning of the activity and allows familiarization with the interaction spaceship and the purpose of the space mission. The simulation proposes the interaction of the participants with the environment, through collaborative games. The moment of reflection aims at provoking in students questions about the subjects addressed during their visit to space

Through a set that reproduces the interior of a spaceship, using Virtual Reality technologies, narration, soundtrack and sound effects, the installation increases the sensation of immersion in the activities (Figure

1). The trajectories of the spaceship through the universe are visualized through a set of televisions that display images of the spaceship, simulating its windows.

Figure 1. Images of the spaceship "Adventure in the Solar System"

Inside the spaceship, a group of up to 24 participants is distributed on four workstations. Each station is sized to accommodate six members, who interact with the system through touch screens (blue team), joystick (red time) or a set of buttons (yellow time) show in Figure 1.

2.1 Summary of the spaceship's activities

Hélio, the character who represents the spaceship's commander explains the procedures for making the space voyage. Before the spaceship flies into space, it is necessary to charge the spaceship's batteries and power the solar power systems. This procedure is implemented in the game of calibrating the solar panels (Game 1). After the spacecraft takes off, students are assigned the task of helping the crew of the International Space Station who need supplies, so by means of a game they must dock the spaceship with the International Space Station and transfer the supplies (Game 2). Then the students learn that there was a problem with the Hubble telescope. At that time, students learn that there is a telescope that captures images a thousand times more than the human eye and must interact with a game to mend the telescope (Game 3). Students begin to receive information about Mars: its color is red, it has been discovered that there is water in the liquid state and that, therefore, there may be life on Mars. A robot was sent to Mars to check for water, but it gets stuck in the rocks (Game 4). Then students are given a mission to help the robot (Game 5). Students also receive information about Venus who is known as Earth's twin brother. It is reported that there is garbage in space, so students are given the mission to ward off space junk (Game 6). Then the commander informs them that they will return to Earth. The spaceship is landed and the game ends. Depending on the performance of all students, the game may end with a mission accomplished or they should come back next time to be able to win the game.

2.2 Pedagogical aspects of the attraction "Adventure in the Solar System"

The main objective of the attraction is to provide more interactivity and increase the engagement and participation of the visitors. The attraction has been developed to provide different ways to explore the universe and solve problems related to astronomy, physics and mathematics. The experience lasts 25 minutes and was built from a guided pedagogical approach. Astronaut Marcos Pontes interacts with students all the time through videos, guiding them throughout the experiment, always inquiring questions and providing clues to solving problems through games. Learning is based on collaborative games and videos with concepts and information relevant to the content being explored. Educational games are considered important strategies for museums since they entertain at the same time that they allow to simulate situations and transmit and reinforce concepts [27].

3. Methods

3.1 Research design

It is a quantitative and exploratory empirical study. Three groups of students (ESI - Elemental Scholl I; ESII - Elemental Scholl II; HS - High Scholl) were compared that interacted with the attraction "Adventure in the Solar System" in the Catavento Cultural museum. The objectives of the research were defined, aiming to obtain information about the affective (user experience) and cognitive (learning performance) gains of "Adventure in the Solar System". A study carried out in the literature served as a theoretical reference for the construction of an evaluation model that made it possible to identify (Bloom 1956; Keller 2010; Savi et al., 2010): (a) the level of students' motivation to use the games; (B) whether the games offer a good user experience (UX); and (c) the learning performance. The following research questions were investigated: are there differences in learning performance between the EI, EII and ES groups? Is there a difference in user experience between EI, EII and ES?

3.2 Participants

The students were selected randomly by the time of arrival at the museum. Students entered the ship in groups of up to 24 students, usually students of the same class, but this was not a rule. There were rounds in which there were students of different classes and/or series of teaching. 129 children participated in the study, 62 boys and 67 girls, ranging in age from 7 to 16 years (Table 1). All participants are university students from public schools in the city of São Paulo - Brazil. In order to improve the analysis, the participants were graduated in three groups (Table 1): ESI - Elementary School I (2nd, 3rd and 4th grades); ESII - Elementary School II (5th, 6th, 7th and 8th grades) and HS - High School (11th, 12th, 13th grades).

Table 1. Distribution of students by group

School degree	**N**	**% Sample**	**Age group**	**Boys**		**Girls**	
				N	%	N	%
ESI (2ª, 3ª e 4ª)	44	34,1%	7-9	23	52,27	21	47,72

ESII (5ª, 6ª, 7ª, e 8ª)	56	43,4%	10-13	25	44,64	31	55,35
HS (11ª, 12ª, 13ª)	29	22,5%	14-16	14	48,27	15	51,72

3.3 Learning assessment models

The model proposed by Keller [28], defined as ARCS (Attention, Relevance, Confidence and Satisfaction) was used to assess the level of student motivation. According to the author, motivation in the educational context is related to voluntary engagement in continuing to learn about a particular subject and, for this to happen, the student must: maintain a satisfactory level of attention during the experiment (A-Attention); Students must realize that the educational proposal is consistent and helps them connect the content with the professional or academic future (R-Relevance); Students need to succeed in the experiences derived from their skills (C-Confidence); And let them experience positive learning experiences accompanied by recognition and rewards (S-Satisfaction).

To evaluate if the user experience is satisfactory, the models proposed by Savi et al. [29] was used. Experiences resulting from interaction with games, such as changes in people's emotional state. Thus, the evaluation of the experience of use can be made through the evaluation of the elements of interaction, such as: fun, immersion, challenge, control and social interaction. Fun provides feelings of pleasure, relaxation, distraction, and satisfaction. For the authors, a fun experience is usually accompanied by the desire to rejoin and to recommend it to friends. Immersion is related to engagement and deep involvement with the game. Time is a factor that makes it possible to check the immersion in the game, for example, where the player does not notice the time passing and remains hours or weekends playing uninterruptedly. The challenge is related to the levels of difficulty of the game and these should be adequate to the level of skill of the player. New obstacles, situations and variations of activities should be appropriate to minimize player fatigue and provide an experience that maintains their willingness to continue participating. Control is related to the sense of autonomy over the interface that should be easy to learn to use. Social interaction is measured by engaging the participant with others to achieve group success.

The Bloom's Taxonomy [30] was used to assess the degree of retention of information. It is a framework that can be applied to plan, design and evaluate learning effectiveness defined at six levels: Knowledge (remembering information about facts, dates, words, places, procedures, etc.), Understanding (grasping the meaning of Information), Application (applying knowledge in real situations), Analysis (identifying the parts and their interrelationships), Synthesis (combining the parts to form a whole) and Assessment (judging the value of knowledge).

In the evaluation model proposed in this paper, only the first level of the Bloom's Taxonomy was considered to evaluate a learning performance: Knowledge, which involves memory. According to cognitive theories of learning [31, 32], accessing the information that is recorded in memory, brings direct influences on learning, since the structure and material to be learned are largely dependent on the knowledge retained in memory, that is, what the individual already knows and can remember. Current knowledge not only influences the learning of new knowledge and information, but also the way that knowledge and information is organized, so that it can be retrieved in the future.

Data Collection

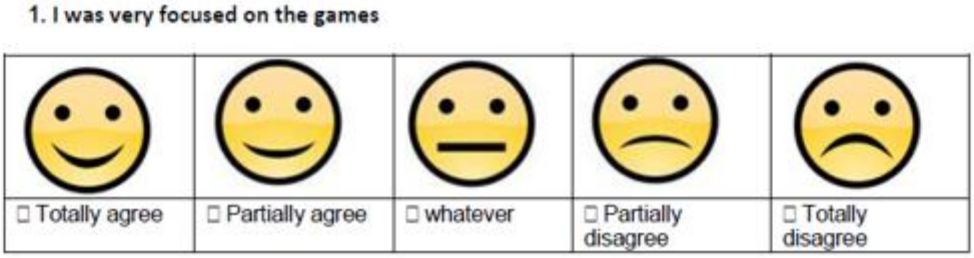
Quantitative data were collected and observations were made by the authors who accompanied the evaluations in the museum. To obtain more expressive statistical data, valid in the context of quantitative research, it was established that the sample size should be greater than 100 students. At the same time, the evaluation model was elaborated, in which Keller's ARCS model [28] was included to evaluate the motivation; The model of Savi et al. [29] to evaluate the experience with games (fun, immersion, challenge, control and social interaction); And Bloom's Taxonomy [30], to assess the learning performance.

Based on the designed model, pre-test and post-test questionnaires were created based on the parameters that make up the evaluation model (Table 2). In order to evaluate the level of knowledge (cognition), questions were asked related to the contents covered in the games with the following scale of answers: “Yes”, “No” and “Do not Know”. These questions were responded individually, before (pre-test) and after (post-test) to each participate in the experiment.

Table 2 - Data Collection Questionnaires

Pre-test and Post-test (cognition)	**Post-test (affective)**
1. Do spacecraft use solar energy? 2. Does the International Space Station receive fuel and supplies carried by a small unmanned spacecraft? 3. Is there a telescope in space that sees a thousand times more than the human eye? 4. Are robots sent to Planet Mars to do research? 5. Is there life on Mars? 6. Is there trash in space?	1. Attention: I was very focused on the games. 2. Relevance: I learned about the solar system. 3. Confidence: I scored in the games. 4. Satisfaction: I feel happy after playing. 5. Fun: I find the games very fun. 6. Immersion: I did not even see the time pass as I played. 7. Challenge: I wanted to keep playing. 8. Control: I find it easy to use game control. 9. Social Interaction (optional question): My colleague and I were able to interact and play together. 10. Ranking of favorite games (optional question). Mark with an X the games you liked best: (a) Calibrate the solar panels; (B) Bring food to the International Space Station; (C) Repair the Hubble telescope; (D) Decrypt distress message; (E) Save the robot on Mars; (F) Take away the space junk.

The post-test questionnaire (affective) was applied shortly after the visitation. Items related to motivation (attention, relevance, confidence and satisfaction) and experience of interaction with games (fun, immersion, challenge, control and social interaction) were added. These items were presented in the form of statements, see Table 2, so that students could indicate their degree of agreement or disagreement based on the 5-point Likert scale: ranging from "5-Totally agree" to "1-Totally disagree". To facilitate the understanding of the answers, considering the elementary school children, we use the smiles images (Figure 2). The last two questions in the questionnaire are optional: the first to identify whether students would like to play on another team and the second to rank preferred games.



1. I was very focused on the games

☐ Totally agree	☐ Partially agree	☐ whatever	☐ Partially disagree	☐ Totally disagree

Figure 2. Post-test questionnaire responses (affective)

3.4 Procedure

The responsible researcher announced the research objectives and procedures for the students and invited them to volunteer. They then completed the questionnaire with demographic information (gender, age and school grade) and the cognition questionnaire (pre-test). Then a group of up to 24 students was invited to board the spaceship. In cases where there were a greater number of students, they were invited to participate in another activity with assembly of laymen, while waiting their turn to participate.

The experience on the spaceship is a simulation of a journey in space. Soon, all the groups went through all the stages that included: a) a greeting of the astronaut Marcos Pontes contextualizing the experience and explaining that the group would have to take off the spaceship for a mission in the space; b) interaction with games to solve problems; c) presentation of the commander Marcos Pontes on the victory or failure of the group in the mission, followed by a greeting.

Students were free to choose their accents that directly impacted on the use of interaction devices. For example, there were students who preferred to use the touchscreen interface (blue accent) rather than the buttons (yellow accent) or joystick (red accent). They were free to negotiate between them their accents (Figure 3a).

Figure 3. a) Children interacting in spaceship; b) children responding to post-test questionnaire

A researcher was always present in a session observing and making notes. This showed that there was great enthusiasm and curiosity of the students. At the end of the experiment, the researcher led the students to express and share their opinions. The students were then given 20 minutes to complete the cognition questionnaire (post-test, Figure 3b).

3.5 Data analysis

To investigate whether the positive scores differed before and after the cognition questionnaire applications (named as "period"), we run General Linear Models (GLM) taking into account the grade (Elemental Scholl I, Elemental Scholl 2 and High Scholl) and the gender (girls and boys) of the students. Since data was analyzed as proportions, the GLMs were structured with binomial errors [33]. We started the modeling process fitting the most complex model, containing the interaction of all explanatory variables. After that, the interaction terms were either removed or kept in the model throughout model simplification. The Akaike's information criterion (AIC) was used to find the minimal adequate model. The AIC penalizes unnecessary parameters in the model (penalized log-likelihood), and the minimal adequate model is credited to the smallest AIC found after all simulation steps[33].

The same analysis (GLM with binomial errors) and modeling simplification procedure described above were done to compare the type of games preferred by students, indicated by the frequencies (proportions) with which the answers were marked in the affective questionnaire. The analyses were conducted taking into account the team, grade, gender and type of game (explanatory variables). Finally, following the likert scale, we compared the frequencies of the type of answers provided by students for each of the nine questions of the affective questionnaire, according to the team, grade, gender and likert (explanatory variables). In this case, however, there were an excess of warning messages during modeling due to null values, and frequencies were compared as a contingency table [34], using count data. Therefore, the GLMs were structured with Poisson errors [33]. In both cases described above, modeling simplification started with the most complex model, containing the interaction among the four explanatory variables.

When the structure of the minimal adequate models contained interactions between the explanatory variables, the results were summarized to observe whether the interactions should be kept in the models. When significant results were not found for the interaction terms, they were removed from the final models. All modeling process was carried out in the R System for Windows, version 3.4.1 (R Development Core Team, 2017), using the "step" function.

4. Results and Discussions

4.1 Pre-test and Post-test (cognition)

When the proportion of positive scores in the questionnaires was compared considering the different periods of evaluation, taking into account the grade and gender, we found that the minimal adequate model contained the different periods, grades and genders within its structure, without the interaction term (Table 3). Therefore, results from GLM showed differences between the period, the grade and gender (Table 4). As expected, the highest proportion of positive scores was observed after students have undergone the experience in the spaceship (Figure 4a). However, it is interesting to note that the greatest proportions of positive scores were observed on grades ESII and HS (Figure 4b), which did not differ from each other (estimate = 0.270; Std. error = 0.166; z-value = 1.623; P = 0.105). In addition, the boys had a slightly higher proportion of positive scores than the girls (Figure 3c).

Although the level of attention has been satisfactory in all groups, there may have been distraction from the ESI at the end of the game, where the subject "Life on Mars" is dealt with: the game "saving the robot on Mars" lasts approximately 13 minutes. In it, players have the expectation that the robot found life on Mars. There are several mini-games and videos around this idea. However, at the end of this game the commander says there was no life on Mars through a short explanatory video of only a few seconds. This may have led the participants to keep the first information (of which there may be life on Mars), but not the second (that there is no life on Mars). Another point to consider is that, when the game reveals that there is no life on Mars, the younger students are exalted, celebrating the victory of the game and the video of the commander goes unnoticed. It is therefore necessary to correct this explanatory video that there is no life on Mars to last longer or to add a new game that reinforces this concept.

Table 3. Results of positive scores in the questionnaires, before and after the students have undergone the experience in the spaceship (Period), according to the grade and gender

AIC for each step	Model complexity (main effects kept in the model)	DF	Deviance	AIC
711.55	Period*Grade*Gender	2	391.91	709.88
709.88	Period* Gender	1	392.45	708.42
	Grade* Gender	2	394.80	708.77
	Period*Grade	2	395.60	709.58

708.42	Grade* Gender	2	395.16	707.14
	Period*Grade	2	396.07	708.05
707.14	Period*Grade	2	398.60	706.58
	Gender	1	400.55	710.52
706.58	Gender	1	403.95	709.93
	Grade	2	441.55	745.52
	Period	1	608.18	914.16

Each model was fitted using a general linear model (GLM) with binomial errors. The Akaike's information criterion (AIC) was used to find the minimal adequate model.

Table 4. Results from the general linear model (GLM) comparing the proportion of positive scores in the questionnaires, before and after the students have undergone the experience in the spaceship

Source of variation*	Estimate	Std. error	z-value	P†
Intercept	−0.354	0.121	−2.933	0.003
After	1.789	0.134	13.348	< 0.001
Fundamental 2	0.669	0.139	4.816	< 0.001
Médio	0.972	0.169	5.745	< 0.001
Boys	0.286	0.124	2.305	0.021

*For contrasts (estimate), the ESI was compared with ESII and HS, the period "before" was compared with "after", and the gender "girls" was compared with "boys". †Significant effects are shows at P < 0.05.

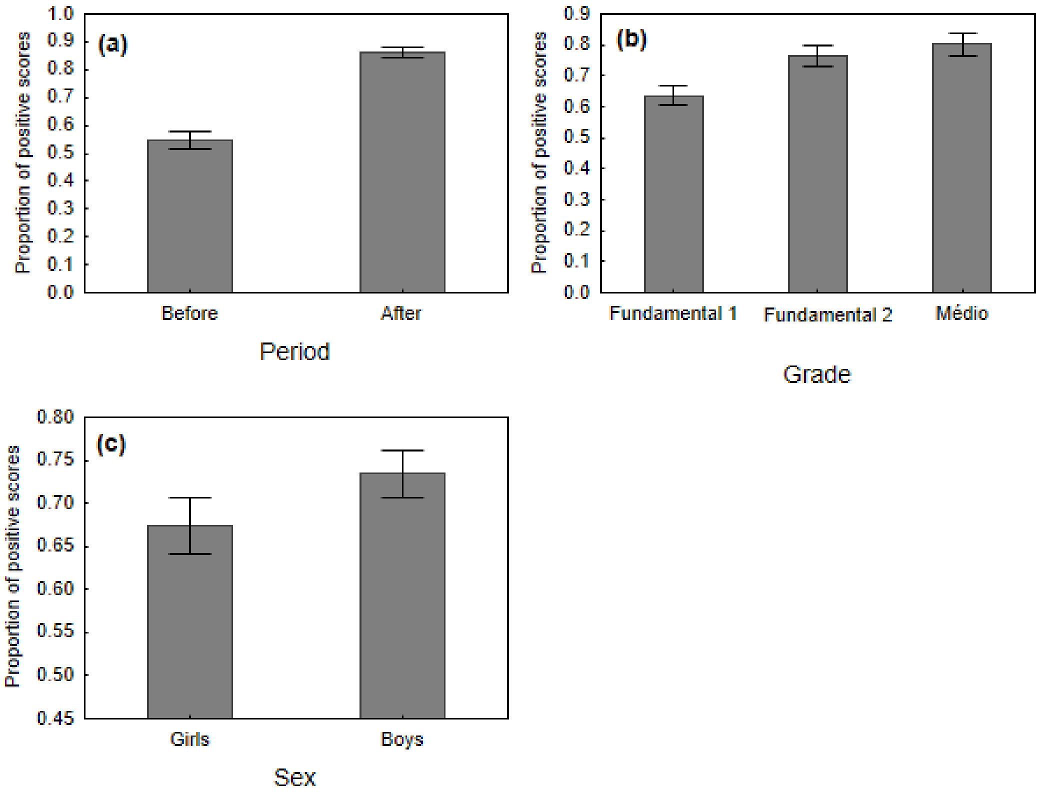


Figure 4. Comparisons of the mean proportion (± 1.0 SE) of positive scores in the questionnaires, before and after the students have undergone the experience in the spaceship (a), among grades (b) and between the genders (c). For grades, ESI differed from ESII and HS, but ESII did not differ from HS.

4.2 Post-test (affective)

With respect to the games preferred by the students, it was observed that minimal adequate model took into account the interaction between grade and the type of game (Tables 1S and 2S; supplementary material in Appendix 1). For the ESI, for example, game 3 (Bring food to the International Space Station) and 4 (Fix the Hubble telescope) presented the highest frequencies marked in the forms, followed by games 2 (Calibrate the solar panels) and 1 (Take away the space junk), with games 5 (Save the robot on Mars) and 6 (Decrypt distress message) showing the lowest proportions. On the other hand, the Game 1 was the most preferred by students on ESII, with the other Game 5 presenting similar preferences (Figure 5); in this case, the lowest preference was observed on Game 5 (Figure 5). The students on HS showed preferences for Game 1, 2 and 3 in similar magnitude, with lower preferences for games 4, 5 and 6 (Figure 5).

Figure 5 sows games preferred by the students according to grades, indicated by the frequency (mean proportions ± 1.0 SE) with which they were marked in the forms. Significant effects were found for the interaction Grade*Type of games, after applying a general linear model (GLM) with binomial errors (AIC = 333.44; df = 10; deviance = 153.23; $P < 0.05$; see Tables 1S and 2S, supplementary material in Appendix 1). The numbers scribed to the games are: 1 = Take away the space junk; 2 = Calibrate the solar panels; 3 = Bring food to the International Space Station; 4 = Fix the Hubble telescope; 5 = Save the robot on Mars; 6 = Decrypt distress message.

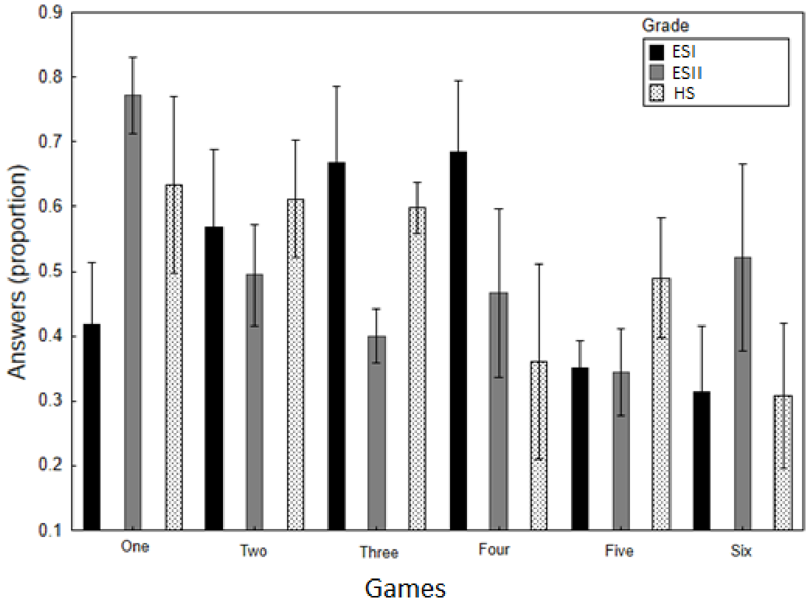


Figure 5. Games preferred by the students according to grades

The minimal adequate model for the answers provided in effectives questionnaire, considering the likert scale, varied according to the question under investigation. For example, for questions one to three, and questions five, seven and eight, the best model contained the different grades and the likert within its structures, without the interaction term (Table 5). However, for questions four, six and nine the interaction term was kept in the model (Table 5); the interaction between grade and likert was kept for questions four and six, and the interaction between gender and likert (plus grade) was maintained in the models for question nine (Table 5). Detailed results obtained during the whole modeling process are shown as supplementary material in Appendix 1 (Table 3S to Table 15S).

Table 5 showing the final minimal adequate models throughout the modeling process using the step function, considering the frequencies about the type of answers provided by students for all questions proposed, according to the team, grade, gender and likert. Each model was fitted using a general linear model (GLM) with Poisson errors. The Akaike's information criterion (AIC) was used to find the minimal adequate model.

Table 5. Effectives questionnaire considering the likert scale

Questions	Final model	Deviance	AIC
One	Grade + Likert	50.360	164.61
Two	Grade + Likert	53.084	175.11
Three	Grade + Likert	45.714	172.59
Four	Grade*Likert	34.252	166.32

Five†	Grade*Likert	50.812	179.94
	FMAM = Grade + Likert		
Six†	Gender *Likert + Grade*Likert	46.869	183.88
	FMAM = Grade*Likert		
Seven†	Grade*Likert	47.065	183.48
	FMAM = Grade + Likert		
Eight	Grade + Likert	51.445	203.67
Nine	Grade + Gender *Likert	56.672	184.94

Those models with “ + ” sign means that both fixed effects must be retained in the model, and those with “ * ” sign means that the interaction between fixed effects are important. †After summarizing the models, it was verified that the interactions Grade*Likert, Gender*Likert and Grade*Likert for questions five (Table 13S), six (Table 14S) and seven (Table 14S), respectively, did not need to be maintained in the minimal adequate models. FMAM = Final Minimal Adequate Model. Details of all the modeling process can be found in Tables 3S-15S (supplementary material in Appendix 1).

Considering all grades, “totally agree” was the most common answer, which confirms the differences among likerts (Figure 6). Differences were also observed for the answers provided among the grades, particularly between ESI (Figure 6a), ESII (Figure 6b) and HS (Figure 65c), confirming that the grade level was an important explanatory variable for all questions (i.e., either considering or not the interaction between variables). Results for the type of answers (mean proportions ± 1.0 SE) provided by students for all questions proposed. For each question, variations for each likert are shows for the ESI (a), ESII (b) and HS (c). Indifferent means neither agree nor disagree.

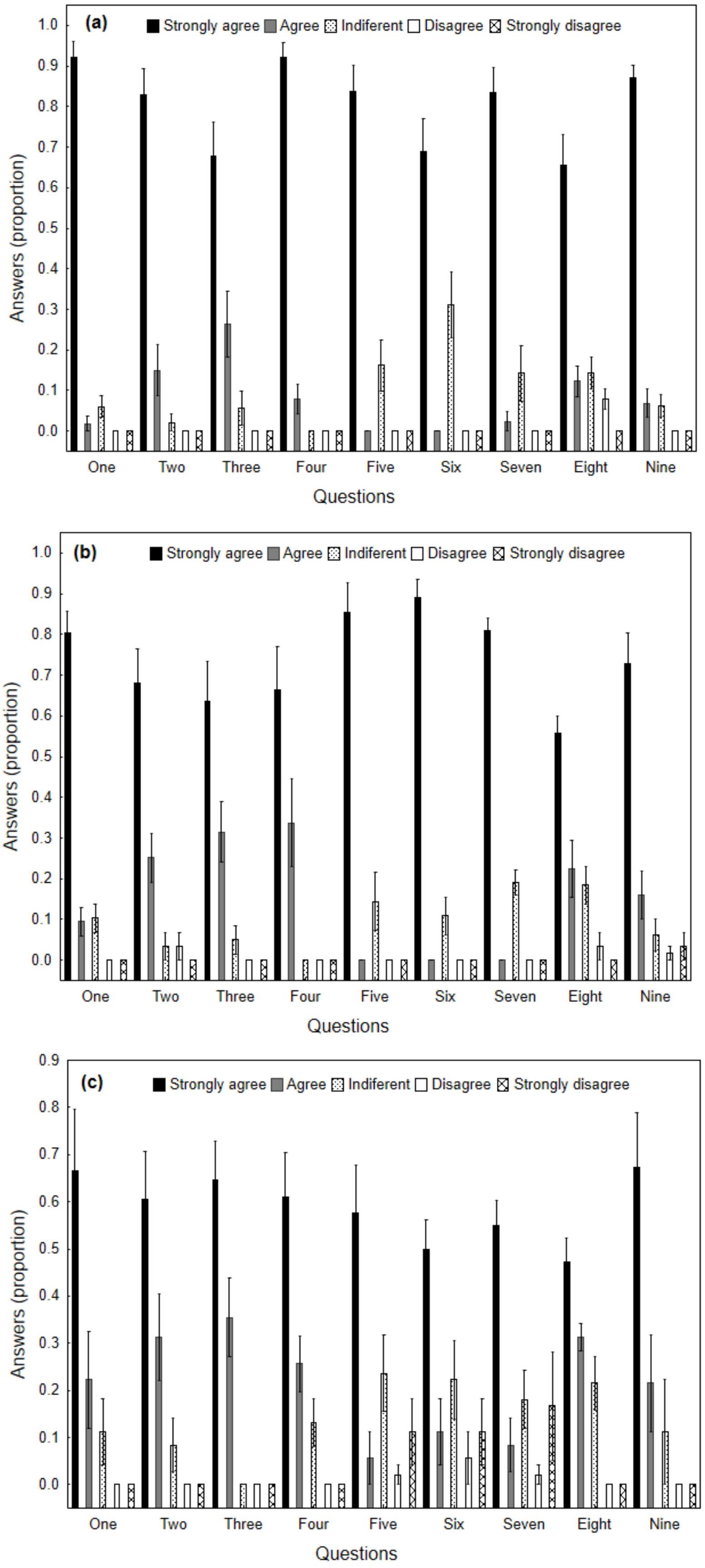


Figure 6. Results for the type of answers (mean proportions ± 1.0 SE) provided by students particularly for question nine, according to sex and likert. SA = strongly agree; A = agree; I = indifferent (neither agree nor disagree); D = disagree; SD = strongly disagree.

It was found that ESI students presented the highest levels of attention, followed by ESII and HS students, and the same pattern was observed with respect to relevance (Figure 6, questions one and two). However, the results showed that high school students showed greater confidence than students in elementary school (Figure 6, question three).

Regarding satisfaction and fun (questions four and five, respectively), clearly the ESI and ESI students presented more expressive results than the students in high school (Figure 6). Interestingly, ESII students were the ones most immersed in games (question six), and ESI and ESII students found games more challenging than HS (Figure 6; question seven).

Regarding the control, the students of HS were the ones that presented greater ease in the handling (Figure 6, question eight). Finally, social interaction was more present in ESI and ESII than in high school (Figure 6, new question). In particular, for question nine, comparing the responses provided by girls and boys on the likert scale, it was found that the "agree" response was more accentuated for the girls, differing from the responses provided in the other likerts (Figure justifying the difference found in the interaction between gender and likert (Table 5).

5. Conclusion and Future Works

This study aimed to analyze learning performance and user experience (UX) with the interactive games of the "Adventure in the Solar System" attraction of the Catavento Cultural Museum, located in the city of São Paulo - Brazil. For this purpose, a sample of 129 students from public education - 44 from Elementary School I; 56 from Elementary School II and 29 High School - were sampled. Through the construction of an evaluation model, it was possible to obtain results on: (a) the level of student motivation in using the games; (b) the experience of using interactive games; and (c) the learning performance. We provide an evaluation model of a system that can be reused and / or intensified in searches by us in search of interactive systems in museums.

This research showed that the attraction "Adventure in the Solar System" reached its purpose of learning that is to transmit and reinforce concepts of astronomy, physics and mathematics within a technological space installed in the museum. This was verified through the cognition questionnaire (containing questions related to the subjects mentioned) before and after the experience in the spaceship. Although learning performance increased statistically significantly in both groups (comparing knowledge before and after the visit), there was no significant difference in learning performance between the three groups, but higher proportions of positive scores were observed in the ESII and SH grades.

As far as user experience issues are concerned, the results of this study show that the ESI has an experience in higher education levels. They were the most experienced with environmental spaces which other students of ESII and HS, being engaged and expressed the largest plan object of general visit. These findings also confirm relevant data from other studies. But the usability study also showed the need for an adjustment at the end of the experiment with the video about "does life exist on Mars?". It is necessary to increase the time of this video with an explanation as to why there is no life on Mars, or propose a new game that reinforces this concept. Such enhancements can lead to a greater contribution to the learning effectiveness and the visitor experience. This could be investigated in a future study.

The study has some specific limitations. The knowledge test we developed (cognition questionnaire) was quite short and focused on factual type of knowledge and short-term retention. In a future study, we intend to carry out another type of study with retention of knowledge in the longer term. To this end, future studies require post-museum visit activities so that students reflect on what was learned during the museum visit and assess whether this knowledge can be maintained and propagated.

To conclude, this study demonstrates that the use of new types of interactive systems can contribute to a better experience of museum visitors, increasing their level of participation and engagement and their intention to repeat visits. As for the learning gains that the visitor is expected to capture with museum visits, it can be argued that such interactive technologies provide learning experiences no less important than conventional display methods. Past the novelty effect these technologies provide, one can provide more authentic learning and entertainment simultaneously

6. References

[1] H. Kraemer. “Media Are, First of All, for Fun: The Future of Media Determines the Future of Museums”, The Future of Museums, Springer, Cham, 2018, pp. 81-100.

[2] H. Setiadi. “Designing a Supplementary Reading Using Cultural Language Learning Approach (CLLA)”, International Journal for Innovation Education and Research, 5(8), 2017, pp. 111-128.

[3] H.C. Huang and M.C. Chuang. “An Aesthetic Practice in Higher Education through Exhibition Transformation”, International Journal for Innovation Education and Research, 3(3), 2015, pp. 117-131.

[4] P. Zaharias, D. Michael and Y. Chrysanthou. “Learning through multi-touch interfaces in museum exhibits: An empirical investigation”, Journal of Educational Technology & Society, 16(3), 2013, pp. 374-384.

[5] R. Hammady, M. Ma and A. Powell, A. “User Experience of Markerless Augmented Reality Applications in Cultural Heritage Museums:‘MuseumEye’as a Case Study”, In: International Conference on Augmented Reality, Virtual Reality and Computer Graphics. Springer, Cham, 2018, pp. 349-369.

[6] H. Huang, C.W. Chen, and Y.W. “Hsieh. Factors Affecting Usability of Interactive 3D Holographic Projection System for Experiential Learning”, In: International Conference on Learning and Collaboration Technologies, Springer, Cham, 2018, pp. 104-116.

[7] C.C. Chen, an T.C. Huang, “Learning in a u-museum: Developing a context-aware ubiquitous learning environment”, Computers & Education, 59(3), 2012, pp. 873–883.

[8] K.E. Chang, C.T. Chang, H.T. Hou, Y.T. Sung, H.L. Chao and C.M. Lee. “Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in

an art museum”, Computers & Education, 71(1), 2014, pp. 185-197.

[9] M.S. Horn, B.C. Phillips, E.M. Evans, F. Block, J. Diamond and C. Shen, C. “Visualizing biological data in museums: Visitor learning with an interactive tree of life exhibit”, Journal of Research in Science Teaching, 53(6), 2016, pp. 895-918.

[10] L. Andre, T. Durksen and M.L. Volman. “Museums as avenues of learning for children: A decade of research”, Learning Environments Research, 20(1), 2017, pp. 47-76.

[11] G. Vavoula, M. Sharples, P. Rudman, J. Meek and P. Lonsdale. “Myartspace: Design and evaluation of support for learning with multimedia phones between classrooms and museums”, Computers & Education, 53(2), 2009, pp. 286-299.

[12] J. Wishart and P. Triggs P. “MuseumScouts: Exploring how schools, museums and interactive technologies can work together to support learning”, Computers & Education, 54(2), 2010, pp. 669-678.

[13] L.J. Rennie and T.P. McClafferty.” Science centres and science learning”, Studies in Science Education, 27(1), 1996, pp. 53–98.

[14] J.H. Falk and M. Storksdieck. “Using the Contextual Model of Learning to Understand Visitor Learning from a Science Center Exhibition”, Science Education, 89(1), 2005, pp. 744–778.

[15] S.D. Coll, R. Coll and D. Treagust. “Making the most of out-of–school visits: How does the teacher prepare? Part II: Implementation & Evaluation of the Learner Integrated Field Trip Inventory (LIFTI)”, In: International Journal of Innovation in Science and Mathematics Education (formerly CAL-laborate International), 26(4), 2018, pp. 1-19.

[16] Hawkey, R. Learning with digital technologies in museums, science centres and galleries [Report 9, Futurelab Series]. London, UK: King’s College, 2004.

[17] K.F. Lee, Y.L. Chen, H.C. Hsieh and K.Y. Chin, K. Y. “Application of intuitive mixed reality interactive system to museum guide activity”, In: 2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), IEEE, 2017, pp. 257-258.

[18] Z. Mu, M. Liu and Z. Tan. “On the Interactive Display of Virtual Aerospace Museum Based on Virtual Reality”, In: 2018 IEEE International Conference on Robots & Intelligent System (ICRIS), IEEE, 2018, pp. 144-146.

[19] A. Caraban, T. Paulino, R. Pereira, R. Spence, M.I.M.I. Madeira-ITI and P. Campos). “The Monarch

Room: An Interactive System for Visualization of Global Migration Data", In: Proceedings of the 32nd International BCS Human Computer Interaction Conference. BCS Learning & Development Ltd., 2018. p. 1-5.

[20] T. Kunkel, M. Averkiou and Y. Chrysanthou, Y. "A web-based virtual museum application", In M. Ioannides, A. Addison, A. Georgopoulos, L. Kalisperis (Eds.), Proceedings of the 14th International Conference on Virtual Systems and Multimedia Budapest, Hungary, 2008, pp. 275-277.

[21] P. Patias, Y. Chrysanthou, S. Sylaiou, C. Georgiades, D. Michael and S. Stylianidis. "The development of an e-museum for contemporary arts", In M. Ioannides, A. Addison, A. Georgopoulos, L. Kalisperis (Eds.), Proceedings of the 14th International Conference on Virtual Systems and Multimedia Budapest, Hungary, 2008, pp. 1-7.

[22] Y. Day. "Visitor centered design in contemporary art museum. Research on the interactive development trend and digital strategy of contemporary art museum", Tesi Specialistiche/Magistrali in Design della comunicazione, Politecnico di Milano, 2018.

[23] A. Origli, E. Leone, A. Sorgente, P. Vanacore, M. Parascandolo, F. Mele and F. Cutugno. "Designing interactive experiences to explore artwork collections: a multimedia dialogue system supporting visits in museum exhibits", In: Proceedings of the 10th AI*IA 2016, 1772, 2016, pp. 26–33.

[24] R. Rosales, M. Castañón-Puga, F. Lara-Rosano, J.M. Flores-Parra, R. Evans, N. Osuna-Millan and C. Gaxiola-Pacheco. "Modelling the Interaction Levels in HCI Using an Intelligent Hybrid System with Interactive Agents: A Case Study of an Interactive Museum Exhibition Module in Mexico", Applied Sciences, 8(3), 2018, pp. 1-21.

[25] R. Reynolds, K. Walker and C. Speight. "Web-based museum trails on PDAs for university-level design students: Design and evaluation", Computers & Education, 55(3), 2010, pp. 994-1003.

[26] M.L. Hallack and M.C. Lavalhegas. "Discurso Científico: Um estudo de caso do Museu Catavento Cultural no Município de São Paulo", Sensos-e, 4(1), 2017, 88-97.

[27] G.G. Bitter, J. Puglisi, A. Gorges and H.P. Uppal. "The effects of an online collaborative elementary math program using team-based games to improve student math achievement, attitude and motivation", International Journal for Innovation Education and Research, 4(7), 2016, pp. 113-138.

[28] Keller, J.M. Motivational design for learning and performance: the ARCS model approach. New York, Springer, 2009.

[29] R. Savi, C.G. von Wangenheim and A.F. Borgatto. "A model for the evaluation of educational games

for teaching software engineering", In: 2011 25th Brazilian Symposium on Software Engineering (SBES), IEEE, 2011, pp. 194-203.

[30] Bloom, T. M. E. Bloom's taxonomy of educational objectives. Longman, 1965.

[31] Moore, B. J. Situated cognition versus traditional cognitive theories of learning. Education, 1998.

[32] Moore, B. J. Cognitive Theories of Learning, 1999.

[33] Crawley, M.J. The R book. John Wiley & Sons, Chichester, UK, 2007.

[34] Zar, J.H. Biostatistical analysis. Prentice Hall, Upper Saddle River, NJ, 1999

Appendix 1

Table 1S. Results from the modeling process using the step function, considering the games preferred by students (proportions), according to the team, grade, sex and type of game. Each model was fitted using a general linear model (glm) with binomial errors. The main effect results are presented with or without interactions throughout the modeling simplification. The Akaike's information criterion (AIC) was used to find the minimal adequate model, which is credited to the smallest AIC observed after all simulation steps.

AIC for each step	Model complexity (main effects kept in the model)	DF	Deviance	AIC
393.75	Team*Grade*Sex*Type of game	20	37.897	391.65
391.65	Team*Grade*Sex	4	38.065	383.82
	Team*Sex*Type of game	10	53.293	387.05
	Grade*Sex*Type of game	10	56.914	390.67
	Team*Grade*Type of game	20	78.856	392.61
383.82	Team*Sex*Type of game	10	53.321	379.08
	Grade*Sex*Type of game	10	56.922	382.68
	Team*Grade*Type of game	20	78.860	384.61
379.08	Grade*Sex*Type of game	10	68.405	374.16
	Team*Grade*Type of game	20	88.861	374.62
	Team*Sex	2	53.322	375.08
374.16	Team*Sex	2	68.410	370.16
	Grade*Sex	2	68.419	370.17
	Sex*Type of game	5	75.334	371.09
	Team*Grade*Type of game	20	105.771	371.53
370.16	Grade*Sex	2	68.423	366.18
	Sex*Type of game	5	75.334	367.09

	Team*Grade*Type of game	20	105.771	367.53
366.18	Sex*Type of game	5	75.334	363.09
	Team*Grade*Type of game	20	105.775	363.53
363.09	Team*Grade*Type of game	20	111.941	359.70
	Sex	1	75.334	361.09
359.70	Team*Type of game	10	119.69	347.44
	Team*Grade	4	111.94	351.70
	Sex	1	111.94	357.70
	Grade*Type of game	10	146.82	374.57
347.44	Team*Grade	4	119.69	339.44
	Sex	1	119.69	345.44
	Grade*Type of game	10	153.23	360.99
339.44	Team	2	119.69	335.44
	Sex	1	119.69	337.44
	Grade*Type of game	10	153.23	352.99
335.44	Sex	1	119.69	333.44
	Grade*Type of game	10	153.23	348.99
333.44	Grade*Type of game	10	153.23	346.99

Table 2S. Results (coefficients) from the general linear model (glm) with binomial errors comparing the games preferred by students, including the interaction between the main effects grade and type of game (minimal adequate model).

Source of variation*	Estimate	Std. error	z-value	$P^{\dagger}$
Intercept	−0.747	0.405	−1.847	0.065
Fundamental 2	0.097	0.539	0.179	0.858
Médio	0.673	0.559	1.205	0.228
Game (2)	1.183	0.560	2.112	*0.035*
Game (4)	1.664	0.582	2.858	*0.004*
Game (6)	−0.169	0.582	−0.291	0.771
Game (3)	1.494	0.572	2.611	*0.009*
Game (1)	0.312	0.560	0.557	0.577
Fundamental 2*Game (2)	−0.589	0.745	−0.791	0.429
Médio*Game (2)	−0.578	0.788	−0.733	0.463
Fundamental 2*Game (4)	−1.301	0.763	−1.705	0.088
Médio*Game (4)	−2.283	0.809	−2.823	*0.005*
Fundamental 2*Game (6)	0.992	0.762	1.301	0.193
Médio*Game (6)	−0.807	0.825	−0.978	0.328
Fundamental 2*Game (3)	−1.132	0.756	−1.497	0.134
Médio*Game (3)	−1.046	0.793	−1.318	0.187

Fundamental 2*Game (1)	1.555	0.776	2.004	*0.045*
Médio*Game (1)	0.627	0.800	0.784	0.433

*For contrasts (estimate), the grade Fundamental 1 was compared with Fundamental 2 and Médio, and the game (5) was compared with games (1), (2), (3) and (6).

†Significant effects at $P < 0.05$.

Table 3S. Results from the modeling process using the step function, considering frequencies of the type of answers provided by students for question one, according to the team, grade, sex and likert. Each model was fitted using a general linear model (glm) with Poisson errors. The main effect results are presented with or without interactions throughout the modeling simplification. The Akaike's information criterion (AIC) was used to find the minimal adequate model, which is credited to the smallest AIC observed after all simulation steps.

AIC for each step	Model complexity (main effects kept in the model)	DF	Deviance	AIC
280.25	Team*Grade*Sex*Likert	16	$3.864\text{x}10^{-10}$	248.25
248.25	Team*Grade*Likert	16	21.275	237.52
	Team*Sex*Likert	8	10.748	243.00
	Grade*Sex*Likert	8	14.650	246.90
	Team*Grade*Sex	4	7.556	247.80
237.52	Team*Sex*Likert	8	24.407	224.66
	Grade*Sex*Likert	8	27.682	227.93
	Team*Grade*Sex	4	24.914	233.16
224.66	Team*Likert	8	26.183	210.43
	Grade*Sex*Likert	8	31.007	215.61
	Team*Grade*Sex	4	27.364	219.61
210.43	Grade*Sex*Likert	8	33.338	201.59
	Team*Grade*Sex	4	29.458	205.71
201.59	Grade*Likert	8	40.904	193.15
	Sex*Likert	4	34.292	194.54
	Team*Grade*Sex	4	36.614	196.86
193.15	Sex*Likert	4	41.723	185.97
	Team*Grade*Sex	4	44.180	188.43
185.97	Team*Grade*Sex	4	45.00	181.25
	Likert	4	317.94	454.19
181.25	Team*Grade	4	46.10	174.34
	Grade*Sex	2	45.62	177.87
	Team*Sex	2	47.32	179.56
	Likert	4	321.22	449.47
174.34	Grade*Sex	2	46.76	171.01
	Team*Sex	2	48.46	172.70

	Likert	4	322.32	442.56
171.01	Team*Sex	2	49.12	169.37
	Grade	2	54.44	174.69
	Likert	4	322.98	439.23
169.37	Team	2	50.29	166.54
	Sex	1	49.19	167.44
	Grade	2	56.80	173.05
	Likert	4	325.34	437.59
166.54	Sex	1	50.36	164.61
	Grade	2	57.97	170.22
	Likert	4	326.51	434.76
164.61	Grade	2	58.04	168.28
	Likert	4	326.58	432.83

Table 4S. Results from the modeling process using the step function, considering frequencies of the type of answers provided by students for question two, according to the team, grade, sex and likert. Each model was fitted using a general linear model (glm) with Poisson errors. The main effect results are presented with or without interactions throughout the modeling simplification. The Akaike's information criterion (AIC) was used to find the minimal adequate model, which is credited to the smallest AIC observed after all simulation steps.

AIC for each step	Model complexity (main effects kept in the model)	DF	Deviance	AIC
288.02	Team*Grade*Sex*Likert	16	9.999	266.02
266.02	Team*Grade*Likert	16	18.774	242.80
	Team*Sex*Likert	8	10.486	250.51
	Grade*Sex*Likert	8	15.863	255.89
	Team*Grade*Sex	4	14.066	262.09
242.80	Team*Sex*Likert	8	22.100	230.12
	Grade*Sex*Likert	8	26.475	234.50
	Team*Grade*Sex	4	22.484	238.51
230.12	Team*Likert	8	26.466	218.49
	Grade*Sex*Likert	8	29.408	221.43
	Team*Grade*Sex	4	25.702	225.73
218.49	Grade*Sex*Likert	8	33.459	209.48
	Team*Grade*Sex	4	29.742	213.77
209.48	Grade*Likert	8	40.195	200.22
	Team*Grade*Sex	4	36.735	204.76
	Sex*Likert	4	38.190	206.21
200.22	Team*Grade*Sex	4	43.471	195.50
	Sex*Likert	4	44.448	196.47
195.50	Team*Grade	4	44.569	188.59

	Sex*Likert	4	47.724	191.75
	Grade*Sex	2	44.097	192.12
	Team*Sex	2	45.789	193.81
188.59	Sex*Likert	4	48.821	184.84
	Grade*Sex	2	45.236	185.26
	Team*Sex	2	46.928	186.95
184.85	Grade*Sex	2	49.489	181.51
	Team*Sex	2	51.181	183.20
	Likert	4	279.526	407.55
181.51	Team*Sex	2	51.849	179.87
	Grade	2	57.167	185.19
	Likert	4	280.194	404.22
179.87	Team	2	53.014	177.04
	Sex	1	51.918	177.94
	Grade	2	59.527	183.55
	Likert	4	282.554	402.58
177.04	Sex	1	53.084	175.11
	Grade	2	60.692	180.72
	Likert	4	283.719	399.74
175.11	Grade	2	60.762	178.79
	Likert	4	283.789	397.81

Table 5S. Results from the modeling process using the step function, considering frequencies of the type of answers provided by students for question three, according to the team, grade, sex and likert. Each model was fitted using a general linear model (glm) with Poisson errors. The main effect results are presented with or without interactions throughout the modeling simplification. The Akaike's information criterion (AIC) was used to find the minimal adequate model, which is credited to the smallest AIC observed after all simulation steps.

AIC for each step	Model complexity (main effects kept in the model)	DF	Deviance	AIC
292.88	Team*Grade*Sex*Likert	16	3.287	264.17
264.16	Team*Grade*Likert	16	23.173	252.05
	Grade*Sex*Likert	8	9.272	254.15
	Team*Sex*Likert	8	12.164	257.04
	Team*Grade*Sex	4	9.047	261.93
252.05	Grade*Sex*Likert	8	24.403	237.28
	Team*Sex*Likert	8	27.232	240.11
	Team*Grade*Sex	4	26.729	247.61
237.28	Grade*Likert	8	27.927	224.81
	Team*Sex*Likert	8	28.574	225.45
	Team*Grade*Sex	4	28.115	232.99

224.81	Team* Sex*Likert	8	31.696	212.57
	Team*Grade*Sex	4	31.203	220.08
212.57	Team*Likert	8	34.831	199.71
	Sex*Likert	4	34.030	206.91
	Team*Grade*Sex	4	34.972	207.85
199.71	Sex*Likert	4	37.079	193.96
	Team*Grade*Sex	4	38.107	194.99
193.96	Team*Grade*Sex	4	40.355	189.23
	Likert	4	257.181	406.06
189.23	Team*Grade	4	41.452	182.33
	Grade*Sex	2	40.981	185.86
	Team*Sex	2	42.673	187.55
	Likert	4	260.457	401.33
182.33	Grade*Sex	2	42.119	179.00
	Team*Sex	2	43.812	180.69
	Likert	4	261.554	394.43
179.00	Team*Sex	2	44.479	177.36
	Grade	2	49.798	182.68
	Likert	4	262.221	391.10
177.36	Team	2	45.645	174.52
	Sex	1	44.549	175.43
	Grade	2	52.158	181.04
	Likert	4	264.747	389.46
174.52	Sex	1	45.714	172.59
	Grade	2	53.323	178.20
	Likert	4	265.747	386.62
172.59	Grade	2	53.393	176.27
	Likert	4	265.817	384.69

Table 6S. Results from the modeling process using the step function, considering frequencies of the type of answers provided by students for question four, according to the team, grade, sex and likert. Each model was fitted using a general linear model (glm) with Poisson errors. The main effect results are presented with or without interactions throughout the modeling simplification. The Akaike's information criterion (AIC) was used to find the minimal adequate model, which is credited to the smallest AIC observed after all simulation steps.

AIC for each step	Model complexity (main effects kept in the model)	DF	Deviance	AIC
282.07	Team*Grade*Sex*Likert	16	6.649	256.72
256.72	Team*Grade*Likert	16	10.812	228.88
	Team*Sex*Likert	8	13.762	247.83
	Grade*Sex*Likert	8	14.993	249.06

	Team*Grade*Sex	4	12.485	254.56
228.88	Team*Sex*Likert	8	15.413	217.48
	Grade*Sex*Likert	8	18.303	220.37
	Team*Grade*Sex	4	14.448	224.52
217.48	Team*Likert	8	16.854	202.93
	Grade*Sex*Likert	8	22.992	209.06
	Team*Grade*Sex	4	19.209	213.28
202.93	Grade*Sex*Likert	8	24.527	194.60
	Team*Grade*Sex	4	20.130	198.20
194.60	Sex*Likert	4	25.616	187.69
	Team*Grade*Sex	4	27.803	189.87
	Grade*Likert	8	49.871	203.94
187.69	Team*Grade*Sex	4	28.892	182.96
	Grade*Likert	8	51.001	197.07
182.96	Team*Grade	4	29.989	176.06
	Grade*Sex	2	29.518	179.59
	Team*Sex	2	31.210	181.28
	Grade*Likert	8	54.277	192.35
176.06	Grade*Sex	2	30.657	172.73
	Team*Sex	2	32.349	174.42
	Grade*Likert	8	55.374	185.44
172.73	Team*Sex	2	33.016	171.09
	Grade*Likert	8	56.042	182.11
171.09	Team	2	34.182	168.25
	Sex	1	33.086	169.16
	Grade*Likert	8	58.402	180.47
168.25	Sex	1	34.252	166.32
	Grade*Likert	8	59.567	177.64
166.32	Grade*Likert	8	59.637	175.71

Table 7S. Results from the modeling process using the step function, considering frequencies of the type of answers provided by students for question five, according to the team, grade, sex and likert. Each model was fitted using a general linear model (glm) with Poisson errors. The main effect results are presented with or without interactions throughout the modeling simplification. The Akaike's information criterion (AIC) was used to find the minimal adequate model, which is credited to the smallest AIC observed after all simulation steps.

AIC for each step	Model complexity (main effects kept in the model)	DF	Deviance	AIC
279.13	Team*Grade*Sex*Likert	16	6.521	253.65
253.65	Team*Grade*Likert	16	21.686	236.82

	Grade*Sex*Likert	8	6.781	237.91
	Team*Sex*Likert	8	10.863	242.00
	Team*Grade*Sex	4	8.860	247.99
236.82	Grade*Sex*Likert	8	21.960	221.09
	Team*Sex*Likert	8	26.587	225.72
	Team*Grade*Sex	4	24.700	231.83
221.09	Team*Sex*Likert	8	26.853	209.99
	Team*Grade*Sex	4	25.058	216.19
	Grade*Likert	8	39.439	222.57
209.99	Team*Likert	8	35.274	202.41
	Team*Grade*Sex	4	30.707	205.84
	Sex*Likert	4	34.583	209.72
	Grade*Likert	8	45.089	212.22
202.41	Team*Grade*Sex	4	38.549	197.68
	Sex*Likert	4	42.177	201.31
	Grade*Likert	8	52.618	203.75
197.68	Team*Grade	4	39.647	190.78
	Grade*Sex	2	39.158	194.29
	Team*Sex	2	40.868	196.00
	Sex*Likert	4	54.452	196.59
	Grade*Likert	8	55.893	199.03
190.78	Grade*Sex	2	40.316	187.45
	Team*Sex	2	42.006	189.14
	Sex*Likert	4	46.550	189.68
	Grade*Likert	8	56.991	192.12
187.45	Team*Sex	2	42.675	185.81
	Sex*Likert	4	47.217	186.35
	Grade*Likert	8	57.658	188.79
185.81	Team	2	43.841	182.97
	Sex*Likert	4	49.577	184.71
	Grade*Likert	8	60.018	187.15
182.97	Sex*Likert	4	50.743	181.88
	Grade*Likert	8	61.183	184.32
181.87	Sex	1	50.812	179.94
	Grade*Likert	8	68.085	183.22
179.94	Grade*Likert	8	68.155	181.29

Table 8S. Results from the modeling process using the step function, considering frequencies of the type of answers provided by students for question six, according to the team, grade, sex and likert. Each model was fitted using a general linear model (glm) with Poisson errors. The main effect results are presented with or without interactions throughout the modeling simplification. The Akaike's information criterion (AIC) was used to find the minimal adequate model, which is credited to the smallest AIC observed after all simulation steps.

AIC for each step	Model complexity (main effects kept in the model)	DF	Deviance	AIC
287.01	Team*Grade*Sex*Likert	16	4.984	259.99
259.99	Team*Grade*Likert	16	10.665	233.67
	Team*Sex*Likert	8	8.478	247.49
	Grade*Sex*Likert	8	9.778	248.79
	Team*Grade*Sex	4	12.164	259.17
233.67	Team*Sex*Likert	8	12.976	219.98
	Grade*Sex*Likert	8	14.525	221.53
	Team*Grade*Sex	4	17.467	232.47
219.98	Grade*Sex*Likert	8	16.062	207.07
	Team*Likert	8	25.262	216.27
	Team*Grade*Sex	4	18.714	217.72
207.07	Team*Likert	8	27.693	202.70
	Team*Grade*Sex	4	21.165	204.17
	Sex*Likert	4	28.419	211.43
	Grade*Likert	8	48.495	223.50
202.70	Team*Grade*Sex	4	30.969	197.98
	Sex*Likert	4	38.234	205.24
	Grade*Likert	8	57.348	216.36
197.98	Team*Grade	4	32.066	191.07
	Grade*Sex	2	32.331	195.34
	Team*Sex	2	33.287	196.29
	Sex*Likert	4	41.509	200.52
	Grade*Likert	8	60.624	211.63
191.07	Grade*Sex	2	33.596	188.60
	Team*Sex	2	34.426	189.43
	Sex*Likert	4	42.607	193.62
	Grade*Likert	8	61.721	204.73
188.60	Team*Sex	2	35.956	186.96
	Sex*Likert	4	43.274	190.28
	Grade*Likert	8	62.389	201.40
186.96	Team	2	37.122	184.13
	Sex*Likert	4	45.634	188.64
	Grade*Likert	8	64.749	199.76
184.13	Sex*Likert	4	46.000	185.81

	Grade*Likert	8	65.914	196.92

Table 9S. Results from the modeling process using the step function, considering frequencies of the type of answers provided by students for question seven, according to the team, grade, sex and likert. Each model was fitted using a general linear model (glm) with Poisson errors. The main effect results are presented with or without interactions throughout the modeling simplification. The Akaike's information criterion (AIC) was used to find the minimal adequate model, which is credited to the smallest AIC observed after all simulation steps.

AIC for each step	Model complexity (main effects kept in the model)	DF	Deviance	AIC
286.41	Team*Grade*Sex*Likert	16	6.710	261.12
261.12	Team*Grade*Likert	16	8.964	231.38
	Grade*Sex*Likert	8	10.286	248.70
	Team*Sex*Likert	8	13.073	251.49
	Team*Grade*Sex	4	10.801	257.21
231.38	Grade*Sex*Likert	8	13.389	219.80
	Team*Sex*Likert	8	18.028	224.44
	Team*Grade*Sex	4	14.503	228.92
219.80	Team*Sex*Likert	8	22.278	212.69
	Team*Grade*Sex	4	19.085	217.50
	Grade*Likert	8	38.035	228.45
212.69	Team*Likert	8	30.236	204.65
	Team*Grade*Sex	4	26.980	209.39
	Sex*Likert	4	32.067	214.48
	Grade*Likert	8	45.955	220.37
204.65	Team*Grade*Sex	4	33.512	199.92
	Sex*Likert	4	38.429	204.84
	Grade*Likert	8	52.241	210.65
199.92	Team*Grade	4	34.609	193.02
	Grade*Sex	2	34.480	196.89
	Team*Sex	2	35.830	198.24
	Sex*Likert	4	41.705	200.12
	Grade*Likert	8	55.516	205.93
193.02	Grade*Sex	2	35.713	190.13
	Team*Sex	2	36.969	191.38
	Sex*Likert	4	42.802	193.22
	Grade*Likert	8	56.614	199.03
190.13	Team*Sex	2	38.072	188.49
	Sex*Likert	4	43.470	189.88
	Grade*Likert	8	57.281	195.69
188.49	Team	2	39.238	185.65

	Sex*Likert	4	45.830	188.24
	Grade*Likert	8	59.641	194.05
185.65	Sex*Likert	4	46.995	185.41
	Grade*Likert	8	60.806	191.22
185.41	Sex	1	47.065	183.48
	Grade*Likert	8	68.564	190.98
183.48	Grade*Likert	8	68.633	189.05

Table 10S. Results from the modeling process using the step function, considering frequencies of the type of answers provided by students for question eight, according to the team, grade, sex and likert. Each model was fitted using a general linear model (glm) with Poisson errors. The main effect results are presented with or without interactions throughout the modeling simplification. The Akaike's information criterion (AIC) was used to find the minimal adequate model, which is credited to the smallest AIC observed after all simulation steps.

AIC for each step	Model complexity (main effects kept in the model)	DF	Deviance	AIC
318.22	Team*Grade*Sex*Likert	16	10.275	296.50
296.50	Team*Grade*Likert	16	20.419	274.64
	Grade*Sex*Likert	8	14.271	284.49
	Team*Sex*Likert	8	15.696	285.92
	Team*Grade*Sex	4	13.965	292.19
274.64	Grade*Sex*Likert	8	22.444	260.67
	Team*Sex*Likert	8	23.845	262.07
	Team*Grade*Sex	4	23.837	270.06
260.67	Team*Sex*Likert	8	25.843	248.06
	Grade*Likert	8	32.995	255.22
	Team*Grade*Sex	4	25.683	255.91
248.07	Team*Likert	8	30.854	237.08
	Sex*Likert	4	26.864	241.09
	Grade*Likert	8	36.392	242.61
	Team*Grade*Sex	4	29.084	243.31
237.08	Sex*Likert	4	32.135	230.36
	Grade*Likert	8	41.439	231.66
	Team*Grade*Sex	4	34.130	232.35
230.36	Grade*Likert	8	42.809	225.03
	Team*Grade*Sex	4	35.410	225.63
225.03	Team*Grade*Sex	4	46.085	220.31
	Likert	4	184.424	358.65
220.31	Team*Grade	4	47.182	213.40
	Grade*Sex	2	46.711	216.93
	Team*Sex	2	48.404	218.63

	Likert	4	187.700	353.92
213.40	Grade*Sex	2	47.850	210.07
	Team*Sex	2	49.542	211.76
	Likert	4	188.797	347.02
210.07	Team*Sex	2	50.210	208.43
	Grade	2	55.528	213.75
	Likert	4	189.464	343.69
208.43	Team	2	51.375	205.60
	Sex	1	50.280	206.50
	Grade	2	57.888	212.11
	Likert	4	191.824	342.05
205.60	Sex	1	51.445	203.67
	Grade	2	59.054	209.28
	Likert	4	192.990	339.21
203.67	Grade	2	59.123	207.35
	Likert	4	193.059	337.28

Table 11S. Results from the modeling process using the step function, considering frequencies of the type of answers provided by students for question nine, according to the team, grade, sex and likert. Each model was fitted using a general linear model (glm) with Poisson errors. The main effect results are presented with or without interactions throughout the modeling simplification. The Akaike's information criterion (AIC) was used to find the minimal adequate model, which is credited to the smallest AIC observed after all simulation steps.

AIC for each step	Model complexity (main effects kept in the model)	DF	Deviance	AIC
284.27	Team*Grade*Sex*Likert	16	1.190	253.46
253.46	Team*Grade*Likert	16	11.715	231.99
	Team*Sex*Likert	8	2.091	238.36
	Grade*Sex*Likert	8	13.504	249.77
	Team*Grade*Sex	4	12.252	256.52
231.98	Team*Sex*Likert	8	14.798	219.07
	Grade*Sex*Likert	8	23.236	227.51
	Team*Grade*Sex	4	17.636	229.91
219.07	Grade*Sex*Likert	8	26.241	214.51
	Team*Grade*Sex	4	20.538	216.81
	Team*Likert	8	29.082	217.35
214.51	Grade*Likert	8	35.766	208.04
	Team*Grade*Sex	4	30.654	210.92
	Team*Likert	8	39.202	211.47
	Sex*Likert	4	40.303	220.57
208.04	Team*Grade*Sex	4	39.042	203.31

	Team*Likert	8	48.106	204.38
	Sex*Likert	4	48.442	212.71
203.31	Team*Grade	4	40.139	196.41
	Team*Likert	8	51.382	199.65
	Grade*Sex	2	39.668	199.94
	Team*Sex	2	41.389	201.66
	Sex*Likert	4	51.718	207.99
196.41	Team*Likert	8	52.479	192.75
	Grade*Sex	2	40.807	193.08
	Team*Sex	2	42.543	194.81
	Sex*Likert	4	52.815	201.09
192.75	Grade*Sex	2	53.147	189.42
	Team*Sex	2	54.839	191.11
	Sex*Likert	4	65.111	197.38
189.42	Team*Sex	2	55.507	187.78
	Grade	2	60.825	193.09
	Sex*Likert	4	65.779	194.05
187.78	Team	2	56.672	184.94
	Grade	2	63.185	191.46
	Sex*Likert	4	68.139	192.41
184.94	Grade	2	64.351	188.62
	Sex*Likert	4	69.304	189.57

Table 12S. Results (coefficients) from the general linear models (glm) with Poisson errors considering only those models with significant results. Results are shown for questions one, two, three and eight, which had Grade + Likert as the minimal adequate model.

Questions	Source of variation*	Estimate	Std. error	z-value	*P*†
One	Intercept	2.010	0.142	14.111	*< 0.001*
	Fundamental 2	−0.120	0.200	−0.600	0.549
	Médio	−0.603	0.231	−2.611	*0.009*
	Likert (2)	−22.088	3633.992	−0.006	0.995
	Likert (4)	−2.389	0.330	−7.230	*< 0.001*
	Likert (3)	−2.389	0.330	−7.230	*< 0.001*
	Likert (1)	−22.088	3633.992	−0.006	0.995
Two	Intercept	1.862	0.148	12.624	*< 0.001*
	Fundamental 2	−0.120	0.200	−0.600	0.549
	Médio	−0.603	0.231	−2.611	*0.009*
	Likert (2)	−4.543	1.005	−4.519	*< 0.001*
	Likert (4)	−1.142	0.210	−5.446	*< 0.001*
	Likert (3)	−3.157	0.511	−6.184	*< 0.001*

	Likert (1)	−20.930	2193.735	−0.010	0.992
Three	Intercept	1.773	0.151	11.757	*< 0.001*
	Fundamental 2	−0.120	0.200	−0.600	0.549
	Médio	−0.603	0.231	−2.611	*0.009*
	Likert (2)	−21.840	3614.100	−0.006	0.995
	Likert (4)	−0.817	0.195	−4.193	*< 0.001*
	Likert (3)	−2.845	0.460	−6.184	*< 0.001*
	Likert (1)	−21.840	3614.100	−0.006	0.995
Eight	Intercept	1.650	0.156	10.587	*< 0.001*
	Fundamental 2	−0.120	0.200	−0.600	0.549
	Médio	−0.603	0.231	−2.611	*0.009*
	Likert (2)	−2.721	0.462	−5.894	*< 0.001*
	Likert (4)	−1.035	0.224	−4.619	*< 0.001*
	Likert (3)	−1.286	0.247	−5.217	*< 0.001*
	Likert (1)	−20.710	2185.946	−0.009	0.992

*For contrasts (estimate), the grade Fundamental 1 was compared with Fundamental 2 and Médio, and the Likert (5) was compared with Likerts (1), (2), (3) and (4).

[†] Significant effects are shown at $P < 0.05$.

Table 13S. Results (coefficients) from the general linear models (glm) with Poisson errors considering only those models with significant results. Results are shown for questions four and five, which had the interaction Grade*Likert as the minimal adequate model.

Questions	Source of variation*	Estimate	Std. error	z-value	*P*[†]
Four	Intercept	2.100	0.143	14.700	*< 0.001*
	Fundamental 2	−0.458	0.230	−1.995	*0.046*
	Médio	−1.059	0.282	−3.761	*< 0.001*
	Likert (2)	−22.403	6344.939	−0.004	0.997
	Likert (4)	−2.506	0.520	−4.818	*< 0.001*
	Likert (3)	−22.403	6344.939	−0.004	0.997
	Likert (1)	−22.403	6344.939	−0.004	0.997
	Fundamental 2*Likert (2)	0.458	8973.099	0.000	0.999
	Médio*Likert (2)	1.059	8973.099	0.000	0.999
	Fundamental 2*Likert (4)	1.844	0.604	3.052	*0.002*
	Médio*Likert (4)	1.752	0.674	2.599	*0.009*
	Fundamental 2*Likert (3)	0.458	8973.099	0.000	0.999
	Médio*Likert (3)	20.956	6344.939	0.003	0.997
	Fundamental 2*Likert (1)	0.458	8973.099	0.000	0.999
	Médio*Likert (1)	1.059	8973.099	0.000	0.999
Five	Intercept	1.992	$1.508\text{x}10^{-1}$	13.216	*< 0.001*
	Fundamental 2	$-1.206\text{x}10^{-1}$	$2.199\text{x}10^{-1}$	−0.548	0.583

	Médio	−9.510x10^{-1}	2.856x10^{-1}	−3.330	*0.001*
	Likert (2)	−22.300	6345.000	−0.004	0.997
	Likert (4)	−22.300	6345.000	−0.004	0.997
	Likert (3)	−1.587	3.658x10^{-1}	−4.338	*< 0.001*
	Likert (1)	−22.300	6345.000	−0.004	0.997
	Fundamental 2*Likert (2)	1.206x10^{-1}	8973.000	0.000	0.999
	Médio*Likert (2)	19.460	6345.000	0.003	0.997
	Fundamental 2*Likert (4)	1.206x10^{-1}	8973.000	0.000	0.999
	Médio*Likert (4)	19.460	6345.000	0.003	0.997
	Fundamental 2*Likert (3)	2.845x10^{-3}	5.334x10^{-1}	0.005	0.996
	Médio*Likert (3)	6.997x10^{-1}	5.792x10^{-1}	1.208	0.227
	Fundamental 2*Likert (1)	1.206x10^{-1}	8973.000	0.000	0.999
	Médio*Likert (1)	20.560	6345.000	0.003	0.997

*For contrasts (estimate), the grade Fundamental 1 was compared with Fundamental 2 and Médio, and the Likert (5) was compared with Likerts (1), (2), (3) and (4).

[†] Significant effects are shown at $P < 0.05$.

Table 14S. Results (coefficients) from the general linear models (glm) with Poisson errors considering only those models with significant results. Results are shown for questions six and seven, which had the structures “Sex*Likert + Grade*Likert” and “Grade*Likert” as the minimal adequate models, respectively.

Questions	Source of variation*	Estimate	Std. error	z-value	$P^{\dagger}$
Six	Intercept	1.797	1.959x10^{-1}	9.174	*< 0.001*
	Sex (Boys)	4.349x10-2	2.086x10^{-1}	0.208	0.835
	Likert (2)	−40.710	1.122x10^{4}	−0.004	0.997
	Likert (4)	−22.000	8520.000	−0.003	0.998
	Likert (3)	−8.164x10^{-1}	3.664x10^{-1}	−2.228	*0.026*
	Likert (1)	−22.000	8520.000	−0.003	0.998
	Fundamental 2	1.027x10^{-1}	2.268x10^{-1}	0.453	0.651
	Médio	−9.719x10^{-1}	3.138x10^{-1}	−3.097	*0.002*
	Boys*Likert (2)	18.540	6575.000	0.003	0.998
	Boys*Likert (4)	−19.410	5608.000	−0.003	0.997
	Boys*Likert (3)	−4.349x10^{-2}	4.205x10^{-1}	−0.103	0.918
	Boys*Likert (1)	−19.410	5608.000	−0.003	0.997
	Likert (2)*Fundamental 2	−1.027x10^{-1}	1.286x10^{4}	0.000	0.999
	Likert (4)*Fundamental 2	−1.027x10^{-1}	1.205x10^{4}	0.000	0.999
	Likert (3)*Fundamental 2	−1.083	5.297x10^{-1}	−2.045	*0.041*
	Likert (1)*Fundamental 2	−1.027x10^{-1}	1.205x10^{4}	0.000	0.999
	Likert (2)*Médio	20.200	9090.000	0.002	0.998
	Likert (4)*Médio	21.170	8520.000	0.002	0.998
	Likert (3)*Médio	2.787x10^{-1}	5.347x10^{-1}	0.521	0.602

	Likert (1)*Médio	21.170	8520.000	0.002	0.998
Seven	Intercept	1.992	0.151	13.216	*< 0.001*
	Fundamental 2	−0.147	0.222	−0.662	0.508
	Médio	−1.012	0.292	−3.465	*0.001*
	Likert (2)	−21.295	3848.400	−0.006	0.996
	Likert (4)	−3.784	1.011	−3.742	*< 0.001*
	Likert (3)	−1.705	0.384	−4.435	*< 0.001*
	Likert (1)	−21.295	3848.400	−0.006	0.996
	Fundamental 2*Likert (2)	0.147	5442.4598	0.000	0.999
	Médio*Likert (2)	18.522	3848.400	0.005	0.996
	Fundamental 2*Likert (4)	−17.364	3848.400	−0.005	0.996
	Médio*Likert (4)	1.705	1.259	1.354	0.176
	Fundamental 2*Likert (3)	0.264	0.534	0.495	0.621
	Médio*Likert (3)	0.724	0.614	1.179	0.238
	Fundamental 2*Likert (1)	0.147	5442.460	0.000	0.999
	Médio*Likert (1)	19.909	3848.400	0.005	0.996

*For contrasts (estimate), the grade Fundamental 1 was compared with Fundamental 2 and Médio, the sex “girls” was compared with “boys”, and the Likert (5) was compared with Likerts (1), (2), (3) and (4). †Significant effects are shown at $P < 0.05$.

Table 15S. Results (coefficients) from the general linear model (glm) with Poisson errors considering the model nine, which had Grade + Sex*Likert as the minimal adequate model.

Question	Source of variation*	Estimate	Std. error	z-value	$P^{\dagger}$
Nine	Intercept	2.019	0.171	11.797	*< 0.001*
	Fundamental 2	−0.120	0.200	−0.600	0.549
	Médio	−0.603	0.231	−2.611	*0.009*
	Sex (Boys)	−0.136	0.198	−0.689	0.491
	Likert (2)	−4.007	1.009	−3.971	*< 0.001*
	Likert (4)	−2.909	0.593	−4.906	*< 0.001*
	Likert (3)	−2.398	0.467	−5.134	*< 0.001*
	Likert (1)	−3.314	0.720	−4.604	*< 0.001*
	Boys*Likert (2)	−16.950	3112.204	−0.005	0.996
	Boys*Likert (4)	1.603	0.670	2.391	*0.017*
	Boys*Likert (3)	−0.780	0.860	−0.908	0.364
	Boys*Likert (1)	−17.643	3112.204	−0.006	0.995

*For contrasts (estimate), the grade Fundamental 1 was compared with Fundamental 2 and Médio, the sex “girls” was compared with “boys”, and the Likert (5) was compared with Likerts (1), (2), (3) and (4). †Significant effects are shown at $P < 0.05$.