Máster en Minería de Datos e Inteligencia de Negocio

Facultad de Estudios Estadísticos

Universidad Complutense de Madrid

Applying data mining techniques to Game Learning Analytics

Cristina Alonso Fernández

Trabajo de Fin de Máster

Director: Rafael Caballero Roldán
Codirector: Baltasar Fernández Manjón

UNIVERSIDAD
COMPLUTENSE
MADRID

AUTORIZACIÓN PARA LA DIFUSIÓN DEL TRABAJO DE FIN DE MÁSTER (TFM) Y SU DEPÓSITO EN EL REPOSITORIO INSTITUCIONAL E-PRINTS COMPLUTENSE DE ACCESO ABIERTO A LA DOCUMENTACIÓN CIENTÍFICA

Los abajo firmantes, estudiante y tutor/es del trabajo fin de máster (TFM) en el Máster en MINERIA DE DATOS E INTELIGENCIA DE NEGOCIO De la Facultad de ESTADÍSTICA, autorizan a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor el trabajo de fin de máster (TFM) cuyos datos se detallan a continuación. Así mismo autorizan a la Universidad Complutense de Madrid a que sea depositado en acceso abierto en el repositorio institucional con el objeto de incrementar la difusión, uso e impacto del TFM en Internet y garantizar su preservación y acceso a largo plazo.

Periodo de embargo (opcional):

- ☐ 6 meses
- ☐ 12meses

TÍTULO del TFM: APPLYING DATA MINING TECHNIQUES TO GAME LEARNING ANALYTICS

Curso académico: 2016 / 2017

Nombre del Estudiante:
CRISTINA ALONSO FERNÁNDEZ

Tutor/es del TFM y departamento al que pertenece/n:
RAFAEL CABALLERO ROLDÁN
DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN

Fecha de aprobación por el Tribunal:

....

....

Calificación ....

Firma del estudiante	Firma del tutor/es	Firma de la Institución Colaboradora (en su caso)
Cristina	R. Caballero	

Agradecimientos

En primer lugar, quiero agradecer a mis directores Rafael y Baltasar su dedicación y consejos en estos meses de trabajo.

Gracias a mis compañeros del Aula 16 por hacer que ir a trabajar llegue a ser incluso divertido. Sin vosotros habría terminado el proyecto mucho antes.

Gracias a mi familia y amigos por su apoyo constante.

Este trabajo está especialmente dedicado a mis padres.

Acknowledgment

This work has been partially funded by the European Commission (BEACONING H2020-ICT-2015-687676).

Table of contents

Table of figures

Table of tables

1. Introduction

This project aims to provide a full exploratory analysis of a dataset obtained from students before, during and after their interaction with an educational game or serious game. The main goal of this project is to stablish students' knowledge after the gameplay, predicted from their game interactions and, if needed, their previous knowledge, to determine the suitability of the serious game as a learning tool.

1.1. Motivation

Educational games or serious games (SGs) are videogames whose purpose is not only to entertain but also to teach, to change an attitude or behavior or to create awareness of a certain issue [1]-[2]. Throughout the years, many SGs have achieved their goals, for instance: the adventure videogame *Aislados* [3] has received several awards for its help in changing its players' attitude towards drug dependency, sexist behaviors and other risk attitudes for teenagers; the SG *Darfur is Dying* [4] contributed to shed a light on the ongoing war in the Darfur region of Sudan in 2006 attracting hundreds of thousands of players in just a few months; and the online puzzle SG *Foldit* [5] helped decipher the crystal structure of an important protease to antiretroviral drug development.

SGs have been successfully applied in many fields such as mathematics [6], physics [7], engineering [8], medicine [9] and literature [10]. In particular, in the field of education, SGs have proven to be effective due to their goal-oriented nature and their capacity to engage players in the game and to encourage them to improve their outcomes and outdo themselves. For instance, games have been used to teach children about asthma [11] or about the cardiopulmonary resuscitation protocol (CPR) [12].

However, some unresolved issues are behind the still low application rate of serious games in education: high development costs, lack of understanding of how students interact with games and lack of understanding of the actual impact games have on students [1]. These reasons make games a complementary educational resource that is usually not reflected in students' marks as educators lack from tools that allow them to control what is happening while students are playing. In most cases, it is clear that students love to play, but we still have to ensure that they learn while playing [13].

In the field of entertaining videogames, data analysis has been used to improve user experience and improve as a last resort the profits. This analysis is called Game Analytics (GA). Different tools are available that allow to monitor what players do to ideally understand their behavior [14].

Data analysis has also been used to prove the efficacy and efficiency of different educational methods trying to obtain information about their usefulness to improve the knowledge that players gain. This analysis is called Learning Analytics (LA) [15].

1.2. Game Learning Analytics

We define Game Learning Analytics (GLA) as the process of capturing, storing, analyzing and obtaining information from players' interactions with a SG, combining the technologies already used in the entertaining videogame industry (GA) with the educational goals pursued by the analysis of the knowledge gained by players / students (LA) [1]-[16]. The process of GLA (Figure 1) usually consists of the following steps:

1. **Data tracking:** while students play with the SG, collect data in a non-disruptive manner.
2. **Analysis of the data captured**: seeking useful information such as places where players make most mistakes, scores obtained, levels that take longer time to complete, and so on.
3. **Visualization of the results**: in a visual manner to ease its communication with the different stakeholders interested on it (students, teachers, game developers or designers, game managers, researchers).

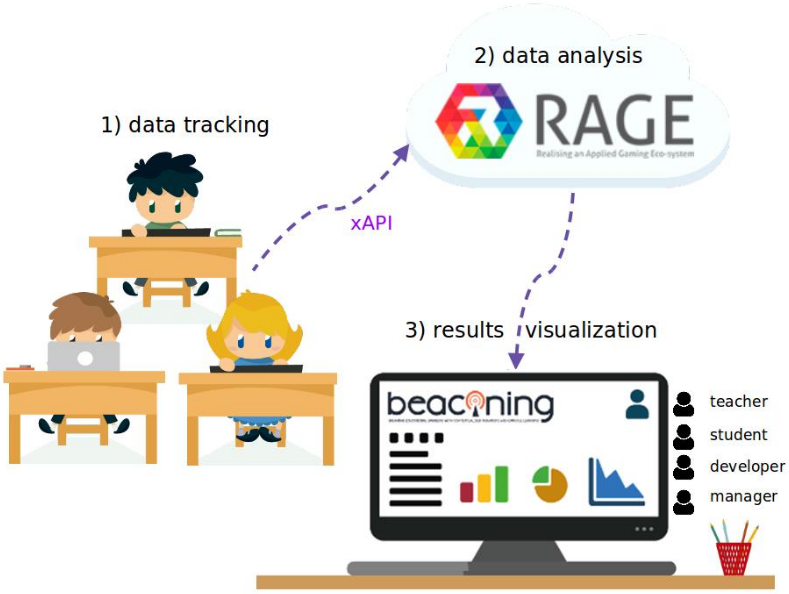


Figure 1. Overview of the process of Game Learning Analytics

SGs can be formally evaluated through several methods such as talk aloud and self-reports or media comparison [2], but still the most common method consists of carrying through a pre-test and a post-test and compare results [17]. In this project, we collected data both from pre and post tests as well as from the SG interactions.

1.3. Structure of the work

This document is structured as follows: Chapter 2 describes project goals; Chapter 3 explains the methods used including experiments description, the serious game used and its validation, population sample and software used; Chapter 4 explains the data capture process: questionnaires used, game generated traces, classes and surveys management, real time information for teachers and other considerations; Chapter 5 explains the data analysis to be applied; Chapter 6 describes the project development and results; Chapter 7 summarizes the conclusions; Chapter 8 contains the main bibliography; and finally, Appendix 1 contains the questionnaires used to collect data, Appendix 2 contains a report about students' game habits given to the school and Appendix 3 contains additional figures and table results.

2. Project goals

With this project we aim to collect and later analyze the information obtained from interactions of several students with a SG, as well as the information obtained from those students of their knowledge before and after their interactions with the game.

With the information collected before, during and after the gameplay, we have conducted an analytic study, at first exploratory, about the data collected. Particularly, the main project goals are the following:

- G1. Determine the influence of previous knowledge in game results.
- G2. Determine the influence of game habits in game results.
- G3. Determine the capability of game interactions to predict post test results when combined with the pre test.
- G4. Compare the previous capability to that of game interactions on their own to predict post test results.

As secondary goals, we plan to establish a classification of the types of players that appear among the students according to their game habits as well as to their interactions and results obtained in the game. With this classification, we aim to be able to predict students' results and the acquired knowledge of a student after the gameplay according to their game habits and their initial knowledge.

To sum up, we intend to analyze the three following areas:

1. knowledge acquisition
2. game habits
3. attitude towards the game

With the GLA architecture developed by the e-UCM research group as part of the H2020 European projects RAGE [18] and BEACONING [19], we are going to capture, collect, analyze and visualize data about the game and players' interactions with it. We have also used previous experience in similar experiments [20].

As the SG that we are going to use for the experiment has already been validated [21], in this project we can focus on analyzing the information obtained from game interactions together with the information collected from the pre-post questionnaires. This way, the project can focus on seeking and finding the possible relations that may appear between the variables obtained from these different data sources.

The ultimate goal is to predict players' results and to establish whether we can predict players' results, that is, their knowledge after playing the game, merely from in-game interactions or if, otherwise, we need more information (collected prior playing the game) to be able to predict players' results.

3. Experiments methodology

In this project, we analyze real data from students so it can be split into two main stages:

I. Data collection from experiments in high-schools.
II. Analysis of the data captured with different techniques.

This chapter describes all the features of the experiments where data was collected: a general description, the serious game used in the experiments and its previous validation, the classes and surveys management and the real time information shown to teachers. Chapter 4 provides further details on the data captured in game generated traces and in the questionnaires.

3.1. Experiments description

The first step consisted on carrying out experiments with high school students playing a SG, collecting information of their interactions with the game. The goal was to obtain at least around a hundred students. Different sessions were carried out in a classroom with several students and, optionally, a teacher.

Before each experiment, a brief presentation was displayed explaining the experiment goal, the goals of this SG and its characteristics (e.g. the three levels students need to complete and other particularities that may surprise them such as the fact that is not possible to go back in the game). After the presentation, the teacher or the session manager gave students their unique identification code that allowed them to access the game and to have all the information of a single student together but anonymized.

Each student / player completed, in this order:

1. A questionnaire before the game (pre-test).
2. A complete game play in the selected SG.
3. A questionnaire after the game (post-test).

These three steps provide information to be analyzed: both questionnaires directly with the answers and the gameplay through all the information tracked and collected. Each player is related to the three data sources via the unique player identification code.

Figure 2 shows a flow chart with the experiments description: students completed a pre questionnaire, a complete game session and a post questionnaire. These three data sources of each student are related by the unique code provided by the teacher or session manager.

During the experiments, teachers also received information of what students were doing (game statistics or mistakes made) as simple visualizations.

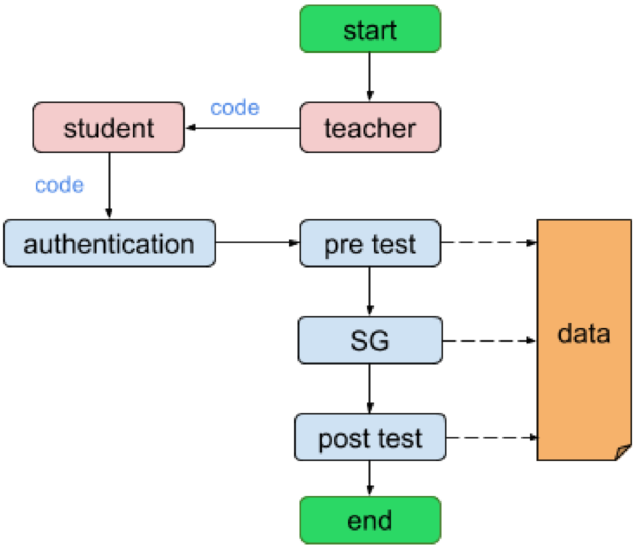


Figure 2. Experiments structure overview

3.2. Serious game

The SG used for the experiments is *First Aid Game*, developed and previously tested by the e-UCM research group [22]. This game has been adapted and updated by Iván Pérez Colado to the videogame engine Unity 3D using the new editor uAdventure [23].

The game, designed for players from 12 years old, contains three levels or initial situations to instruct basic life support maneuvers in situations of:

- chest pain
- unconsciousness
- choking

Figure 3 shows the initial screen of the game *First Aid Game* with its three possible initial situations: chest pain, unconsciousness and choking.

In each situation, the game presents elements the player may interact with: the main character suffering one of the previous situations, or a mobile phone available at the bottom right corner of the screen (Figure 4).

When an inappropriate action is selected, the game shows a brief message and allows the player to try again.

The game also includes a little randomness: for instance, a semi-automatic external defibrillator (SAED) is not always available.

Figure 3. Initial screen of the serious game First Aid Game

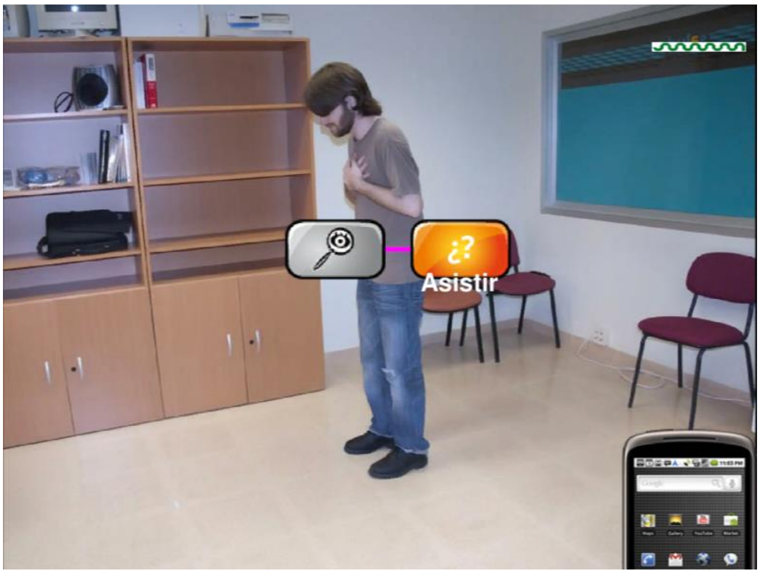


Figure 4. Possible interactions in the serious game First Aid Game

When a suitable action is selected, the game offers different types of multiple-choice questions to be answered containing the specific first aid knowledge to be tested and learned through the game. With their answers in these multiple-choice questions, players learn if their decisions are appropriate or not: if they choose an incorrect answer, the game reports the error and lets them try again until they choose the correct answer.

The questions are shown with several options the player has to choose from: either as simple text (such as in Figure 5, asking for the emergency number), as different pictures with possible actions to perform (such as in Figure 6, asking for the correct position to place the patient in), or as positions inside the game context (such as in Figure 7, where options are presented as different arrows inside the game context, asking for the correct are to place our head for checking the patient's breathing).

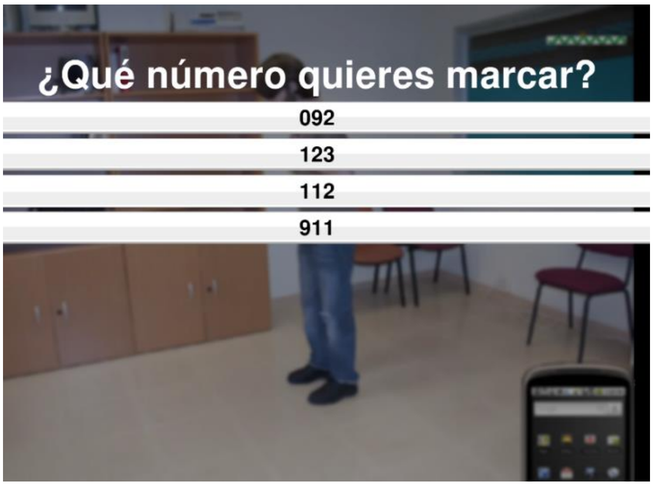


Figure 5. Example question with different textual options

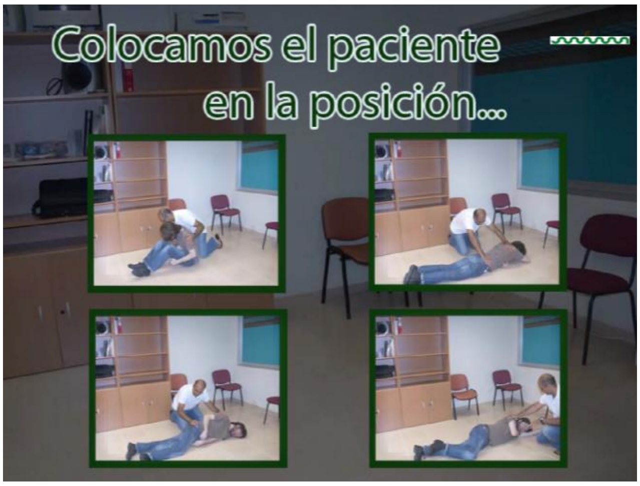


Figure 6. Example question with different visual options

Figure 7. Example question with options inside the game context

After completing each of the three levels, a score from 1 to 10 is shown based on the options taken in the situation, the number of mistakes made and their importance (Figure 8 shows a score of 2 in "pain chest" mode and 9 in "unconsciousness" mode). Players may repeat each situation as many times as they want to improve their scores.

Figure 8. Main screen of the serious game First Aid Game showing scores

3.3. Game validation

The SG *First Aid Game* was validated through an experiment with more than 300 students between 12 and 14 years old in four secondary schools of the Autonomous Community of Aragon (Spain). In each school, the students were randomly split into a group control, who assisted a practical demonstration by an instructor about basic life support maneuvers, and an experimental group who used the SG with no further supervision or additional intervention. Both groups completed an initial questionnaire about basic life support maneuvers and a final questionnaire about the same topics. The results of both questionnaires were analyzed and compared with different statistics methods (Student's t test and two-way ANOVA) to establish their differences.

The experiments, conducted during 2011, showed statistically significant differences in both groups. Although the differences were slightly smaller in the experimental group, the cost per session of this method is much cheaper, proving both the game validity to show basic life support maneuvers and the adequateness of using SGs rather than traditional methods for several reasons such as: possible of access them repeatedly and through the time, or the possibility of using them without presence of specialized instructors [21].

As the game remains structurally and content-wise the same, the current project can focus not on the SG validation, but in establish the influence of previous knowledge and game interactions for knowledge after gameplay (given by post test score) as stated in Chapter 2.

3.4. Classes and surveys management

To keep all the information of a single student together, the teacher or session manager provides each student with a unique identifier. To control those identifiers for the different classes, we have designed and implemented a classes and surveys management system for teachers.

To start with, teachers can register in the system or log-in with their user and password if they are previously registered. Once the teacher has accessed the system, classes and surveys management are available, either to visualize the already-created ones or to create new ones:

- In the case of creating a new class, the teacher is asked to specify the number N of students. Then, the N random codes are generated, each one being a set of four capital letters that the player has to introduce in the game a single time at the beginning of the experiment, before the first survey.
- In the case of registering a new survey, the teacher is asked to add the two files with the previous and after surveys (if they have been generated with LimeSurvey as in our experiment they would be .lss files). By default, the questionnaires are a pair of pre-post surveys, although it is possible to specify only one of the two questionnaires in case that the experiment only needs to capture information before or after the game session.

Lastly, from the complete view of the surveys registered, the teacher can assign a previously-created class to a previously-created survey. With this assignment, we keep track of the classes in which certain survey has been used.

A general diagram of the implemented system to manage classes, codes and survey can be seen in Figure 9: teachers can register or log-in if previously registered to view classes and surveys and create new classes with a number of students and new surveys with pre-test and post-test files. Classes can be assigned to previously created surveys.

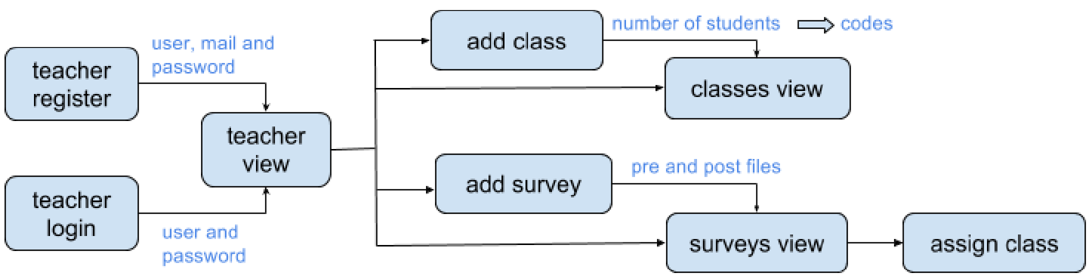


Figure 9. Diagram of the system for classes and surveys management

3.5. Real time information for teachers

During the experiments, while students are completing the three possible initial situations of the game, teachers can obtain useful information to know what is happening in the class. This information is displayed via:

- Visualizations showing information of players' interactions with the game, such as level of completion of errors made.
- Warnings and alerts for special situations: a warning appears in situations that teachers need to know (e.g. a student has not interacted with the game in the last two minutes), while alerts appear in situations that require the teacher's immediate action (e.g. a student has made too many mistakes or has made some serious mistake).

The amount of information that teachers can receive is limited, as this may difficult its correct understanding decreasing its usefulness. For that reason, the next following questions have been identified as questions teachers may want to answer while students play, and that we answers them with visualizations and simple warnings and alerts:

1. How many students are playing? (Answered in Figure 10).
2. What is the students' progress? Which students have finished? (Answered in Figure 11).
3. Which errors are making students the most? (Answered in Figure 12)
4. Has any student stopped playing?

For each of the three first questions, the system answers with the information provided in the following visualizations shown at (near) real time for teachers during the experiments:

Figure 10.Visualization showing the number of active sessions

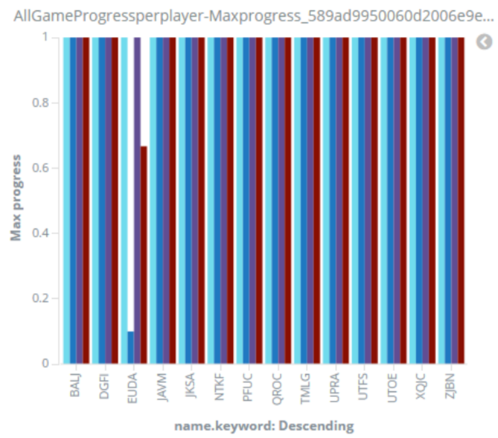


Figure 11. Visualization showing for each player their progress (from 0 to 1) in the three levels and in the game

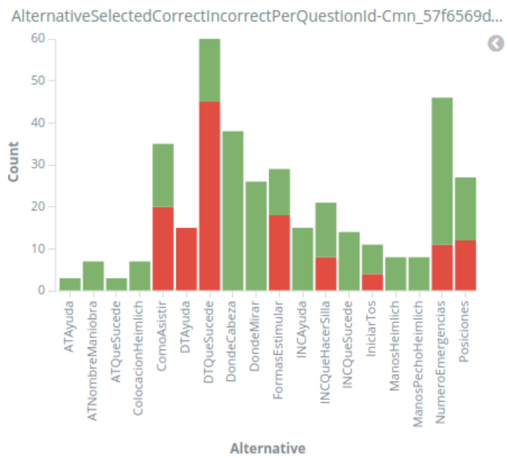


Figure 12. Visualization showing correct answers and mistakes in each game alternative

It is important to notice that the previous visualizations (shown in Figure 10, Figure 11 and Figure 12) can be developed with no additional game knowledge, but simply from the information obtained in the tracked traces [24], as detailed in next chapter.

To answer the question "*Has any student stopped playing?*" we use an alert or warning that is activated when the player has not interacted with the game in the last two minutes. A complete view of all the alerts and warnings triggered by the students during a session is shown for teachers on top of the available visualizations (Figure 13).

Alerts & Warnings

User	Alerts	Warnings
LNDN	0	2
IZLH	0	1
LCJU	0	1
BCGR	0	1
IDLV	0	3
UFAY	0	1
XTUB	0	1
BCGR	0	0
LCJU	0	2

Figure 13. General view of alerts and warnings

By clicking on a specific user, teacher can access the detail of the warnings that the specific learner has triggered. Warnings examples include "*the user has failed, at least, once the question about the emergency number (112)*" or "*the user has failed chest pain game mode*" (Figure 14). In this case, the user given by code EZTP has triggered four warnings including failing the emergency number and failing the "Chest pain" game mode.

Alerts & Warnings

×

USER: EZTP

ALERTS (0) WARNINGS (4)

2 - Has failed, at least, once the question about the emergency number (112)

5 - Has completed Chest Pain or Unconscious and has not used the defibrillator

6 - He completed Chest Pain or Unconscious and never performed cardiopulmonar reanimation (CPR)

7 - This user has failed Chest Pain game mode

Figure 14. Detailed view of warnings a specific user has triggered

In each of the previous visualizations, the username corresponds to the unique code of the student, so teachers can easily relate the information provided by the visualizations and alerts or warnings with the corresponding student, keeping the information anonymized in the system.

4. Data capture

4.1. Pre and post questionnaires

The previous questionnaire (pre-test) and the after questionnaire (post-test) have been prepared with LimeSurvey, an open-source free tool to manage surveys. LimeSurvey allows to create, import and export surveys, add participants to a survey (with a unique access code) and obtain the results of the surveys [25].

Both questionnaires share fifteen multiple-choice questions (e.g. "*Cuando una persona se atraganta lo primero es*"), being just one of the answers correct, that aim to obtain the knowledge of basic life support maneuvers of which the game is about. This set of questions was developed following the contents that appear in the game according to the suitable standards and was used in the original experiment to validate the game [21].

The previous questionnaire starts by asking the students' age and gender. It also contains eleven questions about game habits (e.g. *"¿Con qué frecuencia juegas a videojuegos?"*) to be answered with a Likert scale of 5 levels being 1 "never" and 5 "daily". These questions about game habits were obtained from the article [20] and slightly adapted for the current experiment.

The after questionnaire includes five statements to obtain the players opinion about the game (e.g. "*He aprendido con el juego.*") with a Likert scale of 5 levels being 1 "strongly disagree" and 5 "strongly agree". A sixth optional question allows students to leave any additional comments on the game.

Both complete questionnaires can be seen in Appendix 1: Questionnaires.

4.2. Game generated traces

The information of each game session is obtained from game-generated traces during the player's interactions with the SG. These traces are anonymized and solely identified by the unique player code; the same code is used for the pre-test and post-test so the system can relate all the information of a single user by the unique code. The key step to ensure anonymization is that the system does not collect any further information about the particular student the code belongs to and, therefore, it cannot identify particular students in the experiment as they remain completely anonymous.

However, for the teacher it is relevant to identify which particular student corresponds to a given code in several situations: to help a student having trouble, to provide more tasks to an advanced player or to evaluate players based on their scores or interactions with the game. For these reasons, teacher manage the codes, giving them to the students and allowing the teacher to keep a personal register of the correspondence between students and codes. This function could also be made by a session manager if no teacher is available.

The traces are automatically generated by the game in Experience API (xAPI) format, a widely-used e-learning specification to capture traces (called *statements* in the xAPI world) [26]. The statements contain learning activities and have three main fields: an actor, a verb and an object (*who* did *what action* on *which object*). The statements may also include additional fields such as a *timestamp* to register the moment when the trace was generated. Concretely, we are using a standard interactions model for SGs that has been defined and implemented on xAPI [27].

The main actions registered during each game session are:

- Start a level or the complete game (*initialized*)
- Progress in a level or in the complete game *(progressed)*
- End a level or the complete game (*completed*)
- Select an alternative, a question or a menu (*selected*)
- Access a screen, cut scene or game area (*accessed*)
- Skip a cut scene (*skipped*)
- Interact with an item on an non playable character (NPC) (*interacted*)

Sample statements in xAPI format generated during the game for *initialized* and *completed* verbs can be seen in Figure 15 (the object "JuegoCompleto" was initialized) and Figure 16 (the object "Atragantamiento" was completed in 15.819951 seconds with score 8) respectively.


```
{
    "timestamp":"2017-01-17T11:50:05.339Z",
    "verb":
        {
            "id":"https://w3id.org/xapi/adb/verbs/initialized"
        },
    "object":
        {
            "id":"JuegoCompleto",
            "definition":
                {
                    "type":"https://w3id.org/xapi/seriousgames/activity-types/serious-game"
                }
        }
}
```


Figure 15. Experience API trace generated when the game starts


```
{
    "timestamp":"2017-01-17T15:12:15.105Z",
    "verb":
        {
            "id":"http://adlnet.gov/expapi/verbs/completed"
        },
    "object":
        {
            "id":"Atragantamiento",
            "definition":
                {
                    "type":"https://w3id.org/xapi/seriousgames/activity-types/level"
                }
        },
    "result":
        {
            "extensions":
                {
                    "time":"15.819951",
                    "success":"True",
                    "score":"0.8"
                }
        }
}
```


Figure 16. Experience API (xAPI) trace generated when a level is completed

4.3. Population sample

The experiment was conducted with a total of 227 students of between 12 and 16 years old of La Inmaculada Escolapias School of the Community of Madrid (Spain). These students comprise all the students of the School of the academic years First, Second, Third and Fourth Year of Secondary Education (ESO) and the First Year of Bachillerato.

The experiment comprised 16 sessions with 14 or 15 students per session. The two first sessions, conducted in late January 2017, with a total of 28 students were considered as a training evaluation, so the final number of observations for the study is of **199 students**, from the 14 sessions conducted in the first fortnight of February 2017.

The final sample has the distribution of gender and academic year shown in Table 1. Due to an error with one of the previous questionnaires, gender was not collected for 15 students of First Year of ESO.

Table 1. Gender and academic year distribution of the final population

Gender / Year	1 ESO	2 ESO	3 ESO	4 ESO	1 BACH	Total
Female	23	21	27	22	5	98
Male	16	25	16	26	3	86
Unspecified	15	0	0	0	0	15
Total	54	46	43	48	8	199

We can see graphically in Figure 17 how the distribution per gender is almost equal for each academic year.

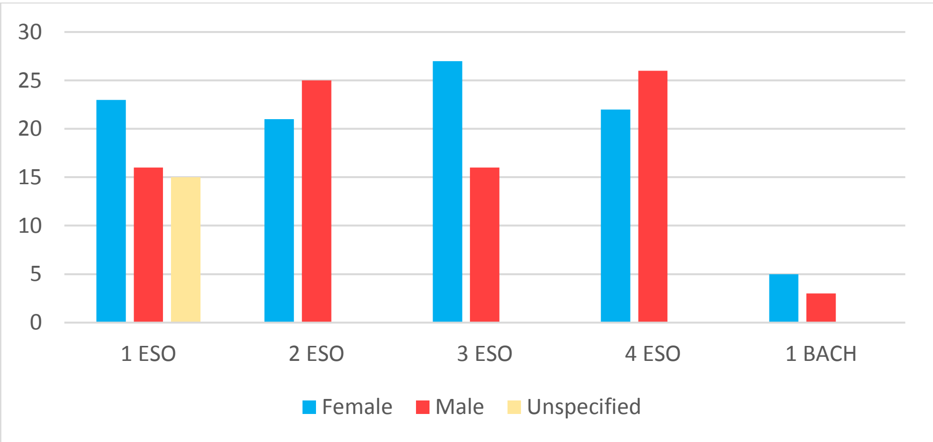


Figure 17. Gender distribution per academic year of the final population

Each student completed, in order, the previous questionnaire, the complete game session (which consists in completing, whichever the order, the three initial situations of the game – pain chest, unconsciousness and choking -), and the after questionnaire.

4.4. Additional recall experiment

In March 2017, it came out the opportunity to return to the school to perfom an additional session with students for a TV programme piece of news. This additional session was used to collect information from another 14 students of 3rd and 4th years of ESO, 10 out of which had participated in the previous experiments. Therefore, these students were considered as an independent set for the analysis, due to the different time they were perfomed and the different conditions (e.g. television camera filming during the experiment).

However, the data from the 10 students who were able to repeat the experiment is interesting to be analyzed to determine whether students recall what they may learn in the experiment after some time or not. In this case, the lapsus of time was just 2-3 weeks.

4.5. Software to use

To collect the data, the GLA architecture developed for the e-UCM research group as part of the H2020 European projects RAGE [18] and BEACONING [19] was used.

The data analysis was performed with R, a free software environment for statistical computing [28], and the integrated development environment (IDE) for R, RStudio [29].
The complete project code can be found at: https://github.com/crisal24/data-mining-gla

4.6. Other considerations

To ensure the data are correctly captured and that no data are lost, we have established a dual capture mode that both sends the data to the RAGE server and also stores it locally in each computer where the game is playing. For the surveys, we also ensured their data capture taking some printed copies in case that there was some problem with the survey management system.

In the experiments, the local storage of game sessions xAPI traces was required twice in two different sessions were internet connection was lost for a few minutes at the end of the gameplays, resulting in lost traces in the server for players with codes "EUDA" and "RADJ", that were correctly stored locally and could be copied to obtain that information. The printed copies for questionnaires were also required in one session were internet was lost at the time of completing the post-test. The answers were then manually transcribed following the other post-test format.

To verify that the data were correctly being received, we can directly access the server to verify that xAPI traces were being sent during the experiment. Also, the visualizations are automatically updated when new traces are received.

Annonimization is ensured using unique codes, as explained previously. Still, indirect identification of students may be possible derived from personal data collected (age and sex) but this is not the case as for each pair of age and sex, there is more than one student. If this was not the case, we could report results aggregated per year intervals to ensure that no student can be uniquely identified.

The school asked for some information about students' game habits. Therefore, after the experiment conclusion, a complete report was handed back to the school with game habits information (always annonymized and aggregated) about students. The report can be found in Appendix 2: Game habits report for school.

5. Data analysis

The data collected from the questionnaires contains:

- 15 variables with the answers of first aid knowledge in the pre-test.
- 11 variables with the answers of the game habits questions, obtained in the pre-test.
- 15 variables with the answers of first aid knowledge in the post-test.
- 5 variables with the answers of the game opinion, obtained in the post-test.

From the xAPI statements collected during the students' interactions with the game, we obtain, at least, directly:

- 3 variables with the scores in the three game levels.
- A session-dependent number of variables with the answers in each multiple-choice questions (where options can be either images or text).
- A session-dependent number of variables with the player's choices in each multiple-choice situation (e.g. arrow to act).
- 3 variables with the interactions of the player with the possible elements in the situation: mobile phone, character who needs help and defibrillator.
- A session-dependent number of variables indicating whether the player has accessed each video or not, and in the first case, whether the video has been skipped or fully watched.

We may also generate other variables derived from the information obtained from the xAPI statements as:

- 4 variables with the time needed to complete each game level and the complete game (only 3 of these variables are independent, as time to complete the game is the sum of the other 3 times).
- 4 variables with the number of errors made by the player in each level and in the complete game (only 3 of these variables are independent, as total number of errors is the sum of the other 3).

The variables obtained from the questionnaires are qualitative having as possible values all the questionnaire answers for that question (four answers in the game contents questions; scale from 1 to 5 in the player's game habits questions and the players' game opinion). From the game interactions, we obtain both quantitative (e.g. amount of time spent on each level or in the complete game) and qualitative variables (e.g. selected option in each game question).

After the data collection, the analysis stage starts with a descriptive analysis of the variables. By making different visual graphics, we try to obtain some early information of each variable, see if we can apply some specific statistical methods (linear regression or other kind), look for possible errors in the data, and so on.

In case that we find any error, we carry out a data cleaning step that may include: treating missing values, correct incorrect values, and so on [30]-[31]. With the cleaned data we can apply different data mining techniques, identified in the data description step.

A brief summary of the techniques that are going to be applied in this project, whether they are supervised or not, as well as the goal of their application, can be seen in Table 2.

Table 2. Data mining and anlysis techniques to be applied in the project and specific goal of their application

Data mining / analysis technique	**Supervised**	**Application goal**
Descriptive analysis / Analysis of correlation	No	Obtain information about variables and about their relations
Cluster analysis	No	Classification of players based on game habits and game results.
Principal component analysis	No	Discovery of the variables of greater influence on the game results, and on the knowledge acquired, as well as discovery of variables relations.
Factor analysis	No	
Regression trees	Yes	Prediction of a player's results in the game levels based on game habits and initial knowledge.
Linear regression	Yes	Predict value of score based on values of other independent variables.
Classification trees	Yes	Predict value of score (as a binary category pass / fail) based on other independent variables.
Logistic regression	Yes	Predict value of score (as a binary category pass / fail) based on other independent variables.
Naïve Bayes classification	Yes	Predict value of score (as a binary category pass / fail) based on other independent variables.
Statistical tests (dependent t-Test / Wilconxon Singed-rank test)	No	Measure significancy of players' learning and players' recall

Other interesting analysis techniques, such as neural networks, are not suitable for this project as the number of observations is not big enough.

Note that as the goal of this project is to conduct an exploratory study, no starting hypotheses have been establish for the study.

6. Project development and main results

After the experiments, in the data obtained from pre-test, post-test and xAPI traces of game sessions, we have identified the following 104 variables to be analyzed in this project, with a total of 199 observations. Of these 104 variables, 84 are directly obtained from pre and post tests (Table 3) and 20 are inferred from the xAPI traces (Table 4). Therefore, the magnitude of data that we are working with in this project is around ~20000 data points.

Table 3. Name, type and description of the variables identified in the pre-test and post-test

Variable name	Type	Description
Code	Identifier	Unique learner code
PREQi, i in [1,15]	Categorical, 4 levels	Answer about first aid (pre)
POSTQi, i in [1,15]	Categorical, 4 levels	Answer about first aid (post)
Hi, i in [1,11]	Categorical, 5 levels	Answer about game habits
Ji, i in [1,5]	Categorical, 5 levels	Answer about game opinion
Class	Categorical, 5 levels	Academic year of the student
Sex	Binary: F, M	Learners' gender
Age	Categorical	Learners' age
PREQiRIGTH, i in [1,15]	Binary: T, F	T if answered correctly in pre
POSTQiRIGTH, i in [1,15]	Binary: T, F	T if answered correctly in post
PREQSCORE	Interval in [0,15]	Number of correct answers in pre
POSTQSCORE	Interval in [0,15]	Number of correct answers in post
GAIN	Interval in [0,15]	Difference of correct answers in post and pre
PASS	Binary: T, F	T if learner answered more than 7 questions correctly in post

Table 4. Name, type and description of the variables inferred from game xAPI traces

Variable name	Type	Description
gameCompleted	Binary: T, F	T if learner completed game
Score	Interval in [0, 10]	Total score obtained
maxScoreCP	Interval in [0, 10]	Max score in chest pain
maxScoreU	Interval in [0, 10]	Max score in unconsciousness
maxScoreCH	Interval in [0, 10]	Max score in choking
firstScoreCP	Interval in [0, 10]	First score in chest pain
firstScoreU	Interval in [0, 10]	First score in unconsciousness
firstScoreCH	Interval in [0, 10]	First score in choking
timesCP	Interval (integer)	Times completed chest pain
timesU	Interval (integer)	Times completed unconsciousness
timesCH	Interval (integer)	Times completed choking
mostRepeatedSituation	Categorical, 3 levels	Situation learner repeated the most
int_patient	Interval (integer)	Number of interactions with patient
int_phone	Interval (integer)	Number of interactions with phone
int_saed	Interval (integer)	Number of interactions with defibrillator
failedEmergency	Binary: T, F	T if learner failed, at least once, the question about emergency number
failedThrusts	Binary: T, F	T if learner failed, at least once, the number of abdominal thrusts per minute
failedHName	Binary: T, F	T if learner failed, at least once, the name of Heimlich maneuver
failedHPosition	Binary: T, F	T if learner failed, at least once, the initial position for Heimlich maneuver
failedHHands	Binary: T, F	T if learner failed, at least once, the hands position for Heimlich maneuver

Variable J6 containing optional comments about the game is treated in a different analysis to identify students' opinion.

It is also important to notice for the following anlaysis that variables measure different areas:

1. Knowledge: variables from pre and post tests about first aid techniques.
2. Perception: variables of players' answers about game habits and game opinion.
3. Interaction: variables directly observed in the game.

Correct answers for the pre-test and post-test questions have been obtained from the European Resuscitation Council Guidelines for Resuscitation [32].

In the following analysis, we describe the main goal of the analysis, the subset of the previous set of variables used and the results obtained in the analysis.

6.1. Data cleaning

After a first descriptive analysis, the following issues in the data were identified:

- 15 observations have missing values in sex, age and H11, due to an old pre-test being handed in that session. These observations are included or discarded depending on the importance of these variables for the specific analysis.
- The observation with code "ZMZV" has missing values in POSTQi for i in [10, 15] and Jj for j in [1, 5]. We only consider this observation when these variables are not required for the specific analysis (e.g. for classification according to game habits).
- The observation with code "JUNF" has an outsider value of age 92. As we have the additional knowledge of the relation between age and class, we check its class and see is "1BACH" and the most common value of age for that class is 16, therefore we update the value of age for this observation to 16 (*imputation*).

6.2. Descriptive analysis

The first step is to obtain some information of the variables graphically; for instance: bar chart of the age distribution of players which varies from 12 to 17 years old, or boxplot of gender and age distribution (Figure 18) where we see that age variability is similar for both genders.

Game play frequency and its relatedness with other variables can be intuited from some graphic visualizations. First, the bar chart in Figure 19 shows general game play frequency of all students. The boxplot of game frequency per academic year (Figure 20) already shows some differences between lower and higher courses. Also, the boxplot of game play frequency per sex (Figure 21) also seems to show a relation between those two variables: female learners seem to play less in general than male learners.

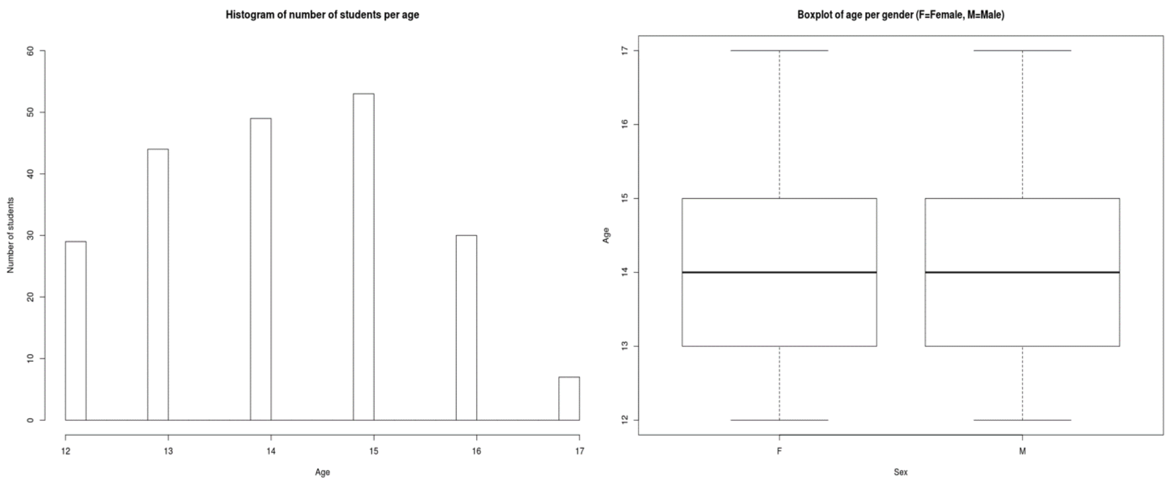


Figure 18. Number of players per age and distribution of age per gender

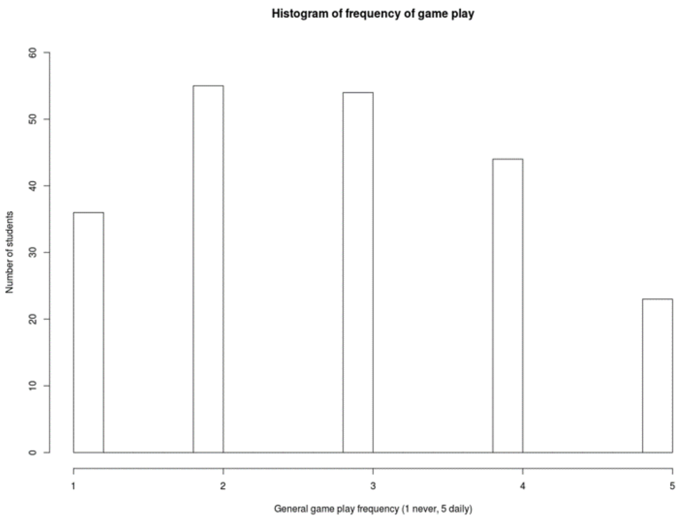


Figure 19. General game play frequency

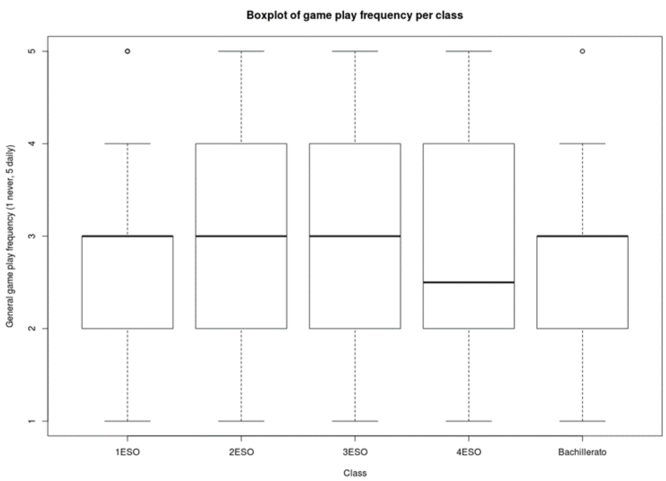


Figure 20. General game play frequency per class

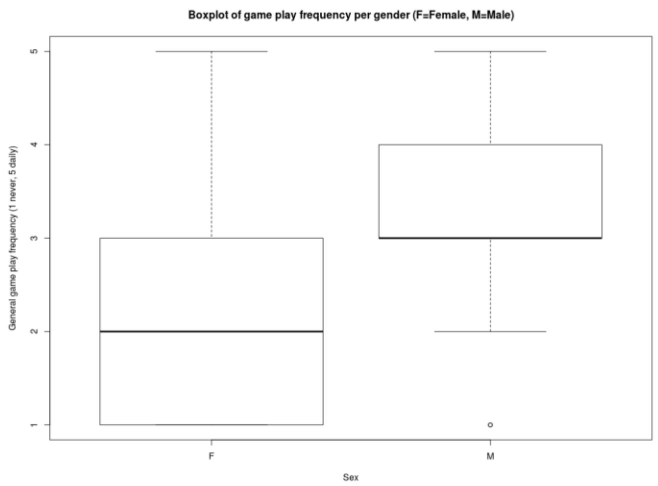


Figure 21. General game play frequency per sex

At this first descriptive analysis, we can also get a glimpse at the distribution of the interval variables containing scores, times and number of interactions. Table 5 summarizes the main features for the 16 interval variables considered in our project. Note that the complete score variable is calculated as the mean of the scores obtained the first time players complete all three levels, and it is not updated with subsequent scores.

Table 5 . Minimum, maximum, mean, standard deviation, skewness and kurtosis of continuous variables

Variable	**Minimum**	**Mean**	**Maximum**	**Standard deviation**	**Skewness**	**Kurtosis**
PREQSCORE	2	8.0555	14	2.0481	-0.1390	3.0198
POSTQSCORE	2	9.8282	14	2.3824	-0.6288	3.3798
GAIN	-6	1.7727	9	2.5616	286	3.5233
score	0	5.8040	9.333333	1.9635	-0.5591	2.9500
maxScoreCP	0	7.5879	10	2.1036	-0.9359	3.7840
maxScoreU	0	7.4271	10	2.7530	-1.1675	3.5359
maxScoreCH	0	8.0502	10	1.8333	-1.4645	5.9662
firstScoreCP	0	7.5879	10	2.1676	-0.2487	2.5860
firstScoreU	0	7.4271	10	2.7991	-0.1056	2.0409
firstScoreCH	0	8.0502	10	1.9956	-0.7236	3.4293
timesCP	0	3.5778	8	1.5251	0.6482	3.1939
timesU	2	5.3065	20	2.7600	1.4594	7.0045
timesCH	0	3.0050	14	1.7964	2.4028	11.3847
int_patient	11	32.8492	126	14.4144	2.0191	11.4441
int_phone	3	10.1658	39	6.0916	1.5606	6.1265
int_saed	0	3.6080	13	2.8457	0.7569	3.2098

6.3. Variables correlation

Besides variables derived from others (such as GAIN from PREQSCORE and POSTQSCORE), there are other correlations between variables studied in the following.

Correlation between interval variables

There is a positive significant correlation (>0.5) for variables: score with maxScoreU, as score is influenced by score in different levels; maxScoreCP with timesCP, maxScoreU with timesU, and maxScoreCH with timesCH, it may indicate that the more times a student repeates a level, the higher the score gets; timesCP with timesU, it may indicate that students who repeat one level tend to repeat the other too; timesCP and timesU with int_patient, they may be related as the number of interactions increases if you repeat the levels.

Correlation between categorical variables

There is a positive significant correlation (>0.5) for variables: sex with H1, H2, H8 which may show a relation between sex and game habits, studied later; PREQ6 and POSTQ6, it may show a bigger influence on this question from pre-test than from the game; H1 with H2 and H3, H2 with H3, H5 and H10, H3 with H5 and H10, H5 with H10, they also may show different related game categories, studied later in detail; J1 with J2 and J3, J2 with J3, it may show that students who find the game interesting also found it fun and learnt with it.

There is also a negative significant correlation (-0.508) found for variables sex and H4 which may indicate that music games are more played by women than men (sex codified as F=1, M=2).

Correlation between interval and categorical variables

There is an expected positive significant correlation (0.749) found for variables age with class.

6.4. Classification according to game habits

- **Goal**: establish a learners' classification according to their game habits.
- **Variables used**: age; sex; class; Hi for i in [1, 11].
- **Number of observations**: 184.
- **Analysis method**: cluster analysis.

The variables considered have mixed data types (continuous and nominal), so we need to define a suitable distance other than the Euclidean distance to measure how similar observations are. For this classification, we are going to use the Gower distance [33]-[34], that selects a distance metric that works for each variable type and scales to fall between 0 and 1. In particular, nominal variables with k categories are first converted into k binary columns and then Dice coefficient is used for each pair of columns to compute their distance. The Dice coefficient is equal to $\frac{2a}{2a+b+c}$ where a is the number of dummies 1 for both individuals, b the number of dummies 1 for this individual, 0 for the other, and c is the number of dummies 0 for this individual, 1 for the other.

After calculating the Gower distance using the *daisy* function [35], we choose the clustering algorithm partitioning around medoids (PAM), similar to the k-means algorithm [36], but whose cluster centers are medoids (i.e. in the observations) instead of centroids.

Finally, we select the number of clusters using silhouette width, a metric that compares how similar an observation is to its own cluster compared to its closets neighboring cluster [37]-[38]. After running the algorithm, we obtain a graph (see Figure 36 in Appendix 3: Additional figures and table results) that shows that the adequate number of clusters is two (the higher plot silhouette).

After obtaining the clusters, we can analyze statistics about the data obtained. We observe that most females fall into cluster 1, while most males fall into cluster 2. We also observe that higher frequency of game play corresponds to cluster 2. With PAM algorithm, medoids provide information of clusters. For the two clusters obtained, their medoids can be seen in Table 6, providing information about the clusters.

Table 6 . Medoids values for the two clusters

cluster	class	sex	AGE	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11
1	2ESO	F	13	2	1	1	2	1	2	1	1	2	1	3
2	4ESO	M	16	4	3	2	1	1	2	2	4	1	1	4

Cluster 1 seems to correspond to younger students, mostly female and with lower game play frequency, while cluster 2 seems to comprise older students, mostly male, and with higher game play frequency.

In particular, observations in cluster 2 play more frequently all kind of videogames (H1), first person shooters (FPS) games (H2), adventure or thriller games (H3), strategy games (H7), sports, racing or simulation games (H8) and mobile or tablet games (H11). On the other side, observations in cluster 1 play more often singing, dancing or playing instruments games (H4) and social games (H9). No difference appears between two clusters in the other three types of games: fighting games (H5), intelligence and quiz/trivia games (H6) and internet collaborative games (H10), all of which are hardly ever play by observations in both clusters. We decide to represent both clusters grouped by gender in Figure 22 to see the possible correspondence between cluster 1 (in red) and female (circular shape); and cluster 2 (in blue) and male (triangular shape). The graph has been personalized based on examples in [39].

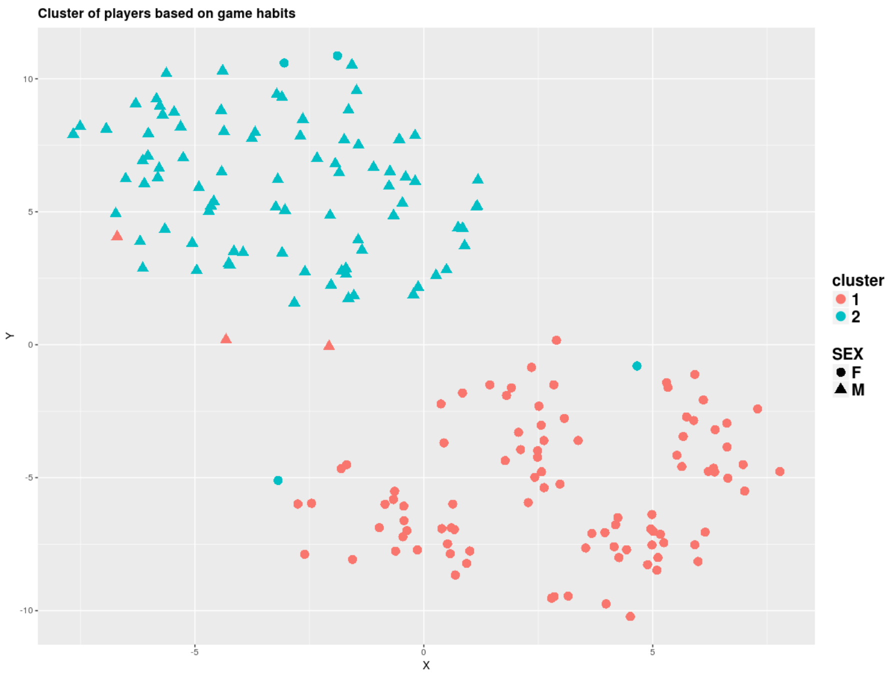


Figure 22. Clusters for learners based on their game habits

Although most women do seem to fall in cluster 1 there are 4 women (a 4% of the total of 98 female) who have been classified in cluster 2, that is, with higher gameplay habits.

Similarly, there are 3 male students assigned to cluster 1 (a 3% of the total of 87 male) with lower gameplay habits.

The visualization in Figure 22 has been created using the dimension reduction technique of t-SNE (t-distributed stochastic neighborhood embedding) which tries to preserve local structure making clusters visible in a 2D visualizatin handling the custom distance metric created.

We can conclude that, although sex is an influential factor, game habits also seem to have a great impact on this classificacion (as seen in Table 6). In fact, if we perfom a cluster analysis solely based on sex, results greatly differ from the previous and no clear separation on two clusters appears.

6.5. Players' learning with the game

- **Goal**: establish whether learners' first aid techniques knowledge has improved with the serious game.
- **Variables used**: PREQi, POSTQi, PREQiRIGTH, POSTQiRIGTH for i in [1, 15], PREQSCORE, POSTQSCORE, GAIN.
- **Number of observations**: 198.
- **Analysis method**: Wilcoxon Signed-rank test.

The number of observations is 198 and not 199 as the variables POSTQi, for i in [1, 15] had a big number of missing values for one observation, so the full observation was deleted. Analyzing each question in the pre-test and post-test, there are four possibilities studied in Table 24 in Appendix 3: Additional figures and table results.

Mean of variable POSTQSCORE is 9.83 and of variable PREQSCORE is 8.06 giving a mean of variable GAIN of 1.77, so we can already intuit that there is a difference between pre-test and post-test scores. The boxplots in Figure 23 of the variables also show that higher sccores appear in post-test. We can also compare the results in a plot (Figure 23) where the diagonal line means no difference in scores in both tests, points above the line have higher score in post-test than in pre-test and points below the line have less scores in post-test than in pre-test. We see that more students are above the line as wanted.

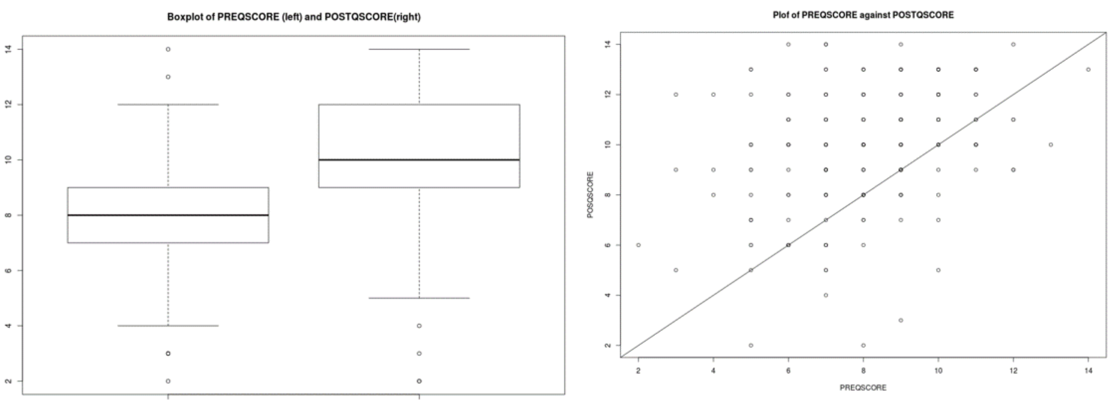


Figure 23. Boxplot of pre-test and post-test scores and plot of students' scores in pre-test (x-axis) compared to scores in post-test (y-axis)

Measurement of pre-test and post-test relations is done through a paired sample t-Test. This test is used to determine if the mean difference of two sets of observations is zero. In our case, we consider the variable GAIN = POSTQSCORE – PREQSCORE. The null hypothesis assumes that GAIN is zero, while the upper tailed alternative hypothesis H1 assumes that GAIN is not zero, that is, that is greater than zero:

H0: GAIN = 0 (*null hypothesis*) , that is, PREQSCORE = POSTQSCORE
H1: GAIN ≠ 0 (*alternative hypothesis*), that is, PREQSCORE ≠ POSTQSCORE

We try to use paired sample t-Test with α=0.05 to compare pre and post scores. To perform the test, we have to verify the four assumptions:

- Variable GAIN is continuous. It is continuous as it is the difference of two continuous variables.
- All observations are independent as each comes from a different student.
- Variable GAIN does not contain any outliers. In boxplot in Figure 24 we can see that variable may have some outliers.
- Variable GAIN has to be approximately normally distributed. The Q-Q plot in Figure 24 does not clarify if the distribution of the variable is normal or not.

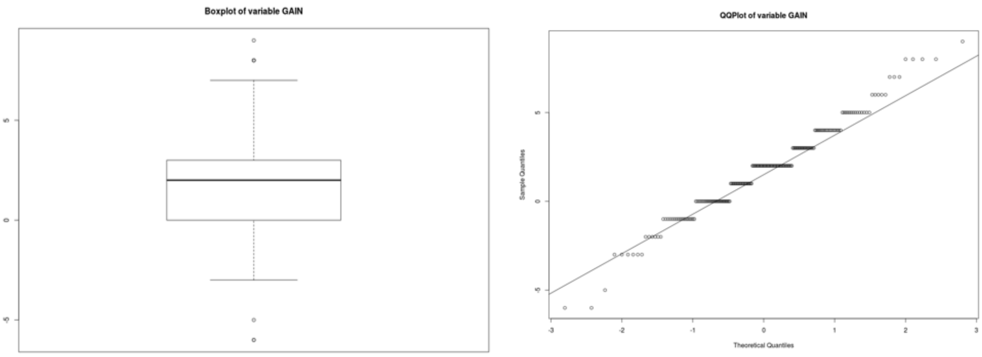


Figure 24. Boxplot of variable GAIN

Displaying histograms for the variables PREQSCORE and POSTQSCORE (Figure 25) we already intuit that they do not have a normal distribution.

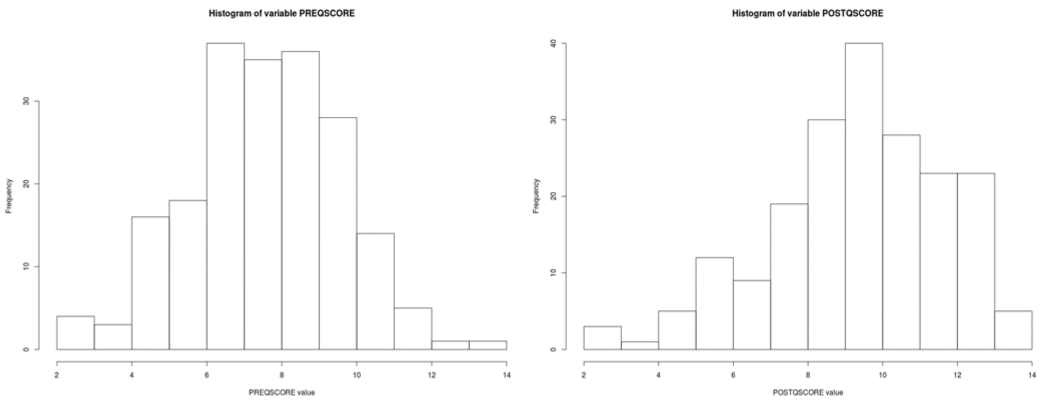


Figure 25. Histograms of variables PREQSCORE and POSTQSCORE

The Shapiro-Wilk normality test for both variables PREQSCORE and POSTQSCORE results in p-values < 0.01, confirming that variables do not have a normal distribution.

Therefore, GAIN does not follow a normal distribution either and we can not perform the paired sample t-Test which assumes normality of variables. Instead, we can perform its non-parametric equivalent, the Wilcoxon Signed-Rank Test, which does not make assumptions on the variables. This test returns a p-value < 0.05, therefore, we can reject the null hypothesis H0 and assume the alternative hypothesis: true difference in means is greater than 0. This means that there is a significant difference in scores in pre-test and post-test. As seen in Figure 23, this differences means that players obtained a significative higher score after playing the game.

Calculating the effect size, as the z-score value for the test was z=-8.3456 we obtain a Pearson correlation coefficient *r* of:

$$r = \frac{-8.3456}{\sqrt{199 * 2}} = -0.41$$

This represents a large effect as it is close to Cohen's benchmack of .5 ([40] as cited in [41]). Therefore, not only the test statistic is significant, but also its effect is meaningful. This could mean that students improved their first-aid techniques knowledge after playing the game.

6.5.1. Comparative with original validation experiment

The original experiment showed the following scores [21]:

- For the 187 students in the experimental group (playing the game), the score increased from 5.41 before playing the game to 7.48 afterwards, out of a maximum score of 10.
- For the 144 students in the control group (attending a theorical and practical demonstration), the mean score was 4.95 before playing the game and 8.56 afterwards, again out of a maximum score of 10.

For the 198 in this experiment, the mean score before playing was 8.06 out of 15 (5.37 out of 10) and after playing was of 9.83 out of 15 (6.55 out of 10). These scores can be seen graphically in Figure 26 where we can see that pre scores of current experiment align with those of the original experiment while post score is lower than in the original experiment.

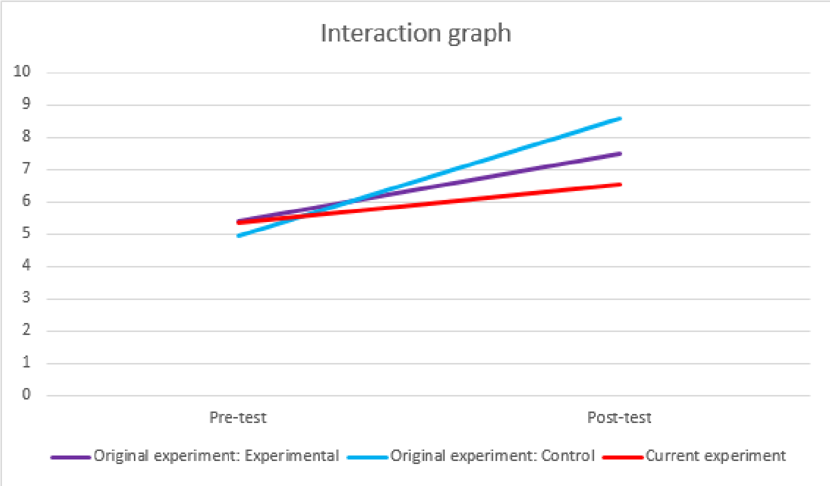


Figure 26. Interaction graph of pre and post scores in original experiment (experimental and control group) and current experiment

6.6. Relations between variables and groups of variables (I)

- **Goal**: establish relations among variables.
- **Variables used**: AGE, Hi for i in [1,11], PREQSCORE, POSTQSCORE, score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, int_patient, int_phone, int_saed.
- **Number of observations**: 184.
- **Analysis method**: principal component analysis.

The principal component analysis [42] reveals the standard deviations, proportions of variances explained and cumulative proportion for the first eight components shown in Table 25 in Appendix 3: Additional figures and table results.

To determine the number of components to retain, we display a scree plot (Figure 37 in Appendix 3: Additional figures and table results) that reveals that five components may be adequate as they explained 52% of variance.

We can analyze the coefficients of each variable in each of the five selected components (see Table 26 in Appendix 3: Additional figures and table results). We observe that components are mainly related to:

PC1. Majority of variables of game habits (H1, H2, H3, H5, H7, H8, H10 and H11) and int_saed.

PC2. Majority of variables from game play, in particular it is the component that best collects the information of: all three variables containing max scores, two of the three variables containing first scores (firstScoreCP and firstScoreU), all three variables of times of completion and two of the three variables containing interactions (int_patient and int_phone).

PC3. The third component describes best the variables AGE, the variable firstScoreCH and two essential variables: POSTQSCORE and score.

PC4. Describes the remaining variables of game habits (not so well described in PC1), that is, H4, H6 and H9.

PC5. It best describes the variable PREQSCORE.

If we plot the two first components (Figure 27) we observe graphically how component 1 contains information about most variables of game habits and int_saed (number of game interactions with the defibrillator), while component 2 describes best variables about scores and times of completion. We may wonder about the relation between components and the class variable: we display this grouping in Figure 38 in Appendix 3: Additional figures and table results. We can see that components do not seem to be related to class.

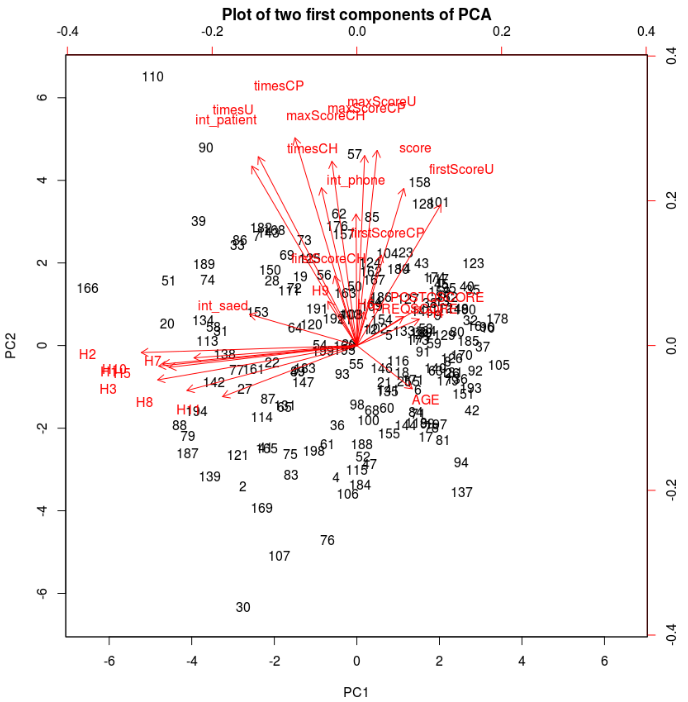


Figure 27. Plot of two first principal components

6.7. Relations between variables and groups of variables (II)

- **Goal**: establish relations among variables.
- **Variables used**: age, Hi for i in [1,11], PREQSCORE, POSTQSCORE, score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, int_patient, int_phone, int_saed.
- **Number of observations**: 184.
- **Analysis method**: factor analysis.

Through a factor analysis, we aim to find groups of variables explained by a smaller group of unobserved variables (factors). First, we need to determine how many factors we are going to keep. Using the nFactors package, Cattell's proposed scree test is improved with the following methods [43]:

- Parallel analysis which modifies Kaiser-Guttman rule - which stops retaining components when they explain less variance than the original standardized variables – to work in the non-asymptotic case.

- Scree test optimal coordinates which compares actual eigenvalues with the estimated predicted ones.
- Scree test acceleration factor which measures abrupt changes of the slope of the curve.

Results can be seen in Figure 39 in Appendix 3: Additional figures and table results.

The recommended number of factors to retain, by two of the three methods, is six. However, the p-value obtained for six factors is < 0.01 so we reject the null hypothesis that 6 factors are sufficient. It is not up to ten factors that we obtain a p-value of 0.0257 > 0.01, so we cannot reject the null hypothesis, that is, accept that the model with ten factors fits the data. The cumulative variance explained by those ten factors is of 60%. The results with ten factors can be seen in Table 27 in Appendix 3: Additional figures and table results. We identify that:

- Factor 1 most relates to game habits for all pre-test questions except H4, H6 and H9.
- Factor 2 relates to game interactions (with patient, phone and saed) and partially with time of completion.
- Factor 3 relates to scores, especially global score and scores of "unconsciousness" mode.
- Factor 4 also relates to scores, especially for "chocking" mode.
- Factor 5 relates to game habits not explained in Factor 1 (H4, H6 and H9).
- Factor 6 contains the information of scores in pre-test and post-test.
- Factor 7 also relates to scores, especially for "chest pain" mode.
- Factor 8 relates to times of completion in "choking" mode and partially to its max score.
- Factor 9 relates to time of completion in "chest pain" mode and its max score.
- Factor 10 relates partially to interactions with phone and score.

The representation of the two main factors can be seen in Figure 28: first one contains most information about game habits; the second one about interactions and times of completion.

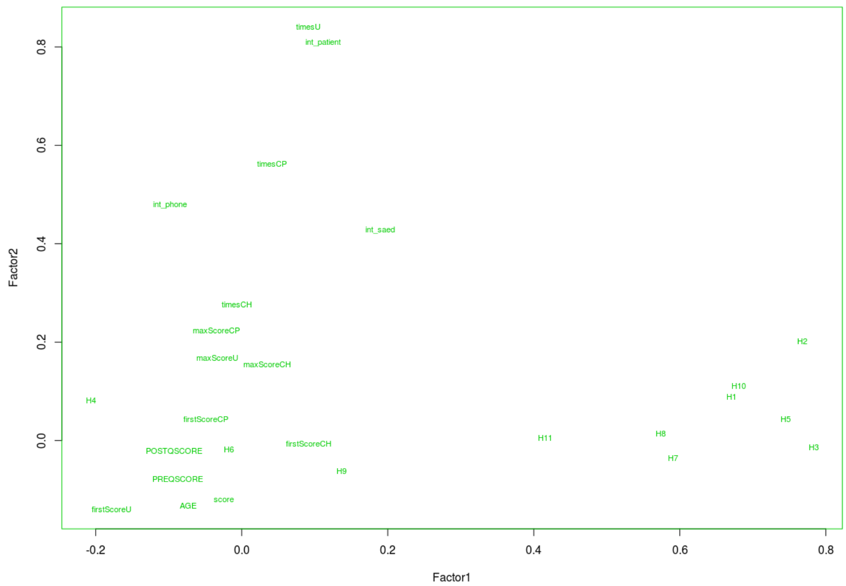


Figure 28. Representation of two main factors.

6.8. Prediction of players' post-test score (I)

To predict players' score in the post-test (as a measure of players' knowledge after playing the game), we can use two approaches:

- Include the pre-test information, to predict the score based both on their previous knowledge and their in-game actions.
- Exclude the pre-test information and try to predict the post-test result merely with the information obtained from the game interactions. With this approach we expect to obtain less accurate results but it has a more interesting outcome as the ideal long-term aim is to be able to predict players' learning based purely on what they do in the game, without requiring pre-post tests.

Both these approaches are studied in the following two sections (6.8 and 6.9), in this case to predict the post-test score as a interval value. Both approaches would also be studied in following sections (6.10 and 6.11) to predict a binary pass / fail result instead.

6.8.1. Using pre-test information

- **Goal**: predict players' score in post-test.
- **Variables used**: class, sex, age, PREQi, PREQiRIGTH for i in [1, 15], Hi for i in [1, 11], Jj for j in [1, 5], score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, int_patient, int_phone, int_saed, POSTQSCORE (target variable).
- **Number of observations**: 198 (missing included).
- **Analysis method**: regression trees.

We want to predict the value of the interval variable POSTQSCORE. The original 198 observations have an average POSTQSCORE value of 9.83.

The decision trees [44] created contain, in each leaf, the number of observations and prediction of POSTQSCORE, given by the average target variable value of the observations that fall in that leaf. New observations could then have a predictions based on the values of the variables of relevance included in the tree.

We create a first sample tree with the following stops criteria:

1. The decrease in the deviance goes below a threshold of 0.01.
2. The number of samples in the node is less than the threshold of 20.
3. The tree depth exceeds the value of 30.

The tree can be seen in Figure 40 in Appendix 3: Additional figures and table results. The lowest cross-validated error for this tree is 0.96177 obtained for *cp*=0.05. The error obtained doing cross validation is 5,4314, calculated as the root node error (5.6473) times the cross validation error of the tree (0.96177). In this first example, we notice that variables of relevance are: PREQSCORE, int_patient, PREQ15, int_saed, maxScoreCP, PREQ2, class, AGE, firstScoreU, H6, PREQ3, firstScoreCH, PREQ7, H11.

Using cross validation with K=10 groups for this tree, we obtain a mean absolute error of 0.25%, shown in Figure 41 in Appendix 3: Additional figures and table results. The variables used in the 10 trees constructed can be seen in Table 7, where the left column (N) specifies the number of trees in which those variables were used.

Table 7. Variables selected in cross validation iterations for score prediction with a regression tree

N	Variables selected
10 trees	int_patient, PREQSCORE
9 trees	PREQ15
8 trees	-
7 trees	PREQ7
6 trees	maxScoreCP, firstScoreCH
5 trees	maxScoreCH, PREQ9
4 trees	PREQ2, PREQ10, timesCP
3 trees	H4, H6, H8, PREQ13, int_phone, maxScoreU, mostRepeatedSituation, score, firstScoreU, J5
2 trees	H7, H11, PREQ3, int_saed, firstScoreCP, failedHPosition, failedHName, J2, J3
1 tree	AGE, class, H10, PREQ2RIGHT, PREQ4, PREQ5, PREQ5RIGHT, PREQ8RIGHT, J1

Tree variation 1

We fix the variable *minsplit* that specifies the minimum number of observations that must exist in a node in order for a split to be attempted, to 30 (default=20), and the variable *minbucket* that specifies the minimum number of observations in any terminal (leaf) node to 15 (default=7). The tree obtained with these parameters can be seen in Figure 42 in Appendix 3: Additional figures and table results. The lowest cross-validated error for this tree is 0.97878 obtained for *cp*=0.05. The error obtained doing cross validation is 5,2287, calculated as the root node error times the cross validation error of the tree.

Tree variation 2

We maintain the previous values of *minsplit* (30) and *minbucket* (15) and fix the value of *cp*, the complexity parameter, specifying that any split that does not decrease the overall lack of fit by a factor of *cp* is not attempted, to 0.02 (default=0.01). We obtain the tree in Figure 43 in Appendix 3: Additional figures and table results. The lowest cross-validated error for this tree is 0.95628 obtained for *cp*=0.05. The error obtained doing cross validation is 5,1085.

Tree variation 3

With all previous parameters with their original values, we fix the *manxval* variable that specifies the number of cross-validations to 20 (default=10) and obtain the tree in Figure 29.

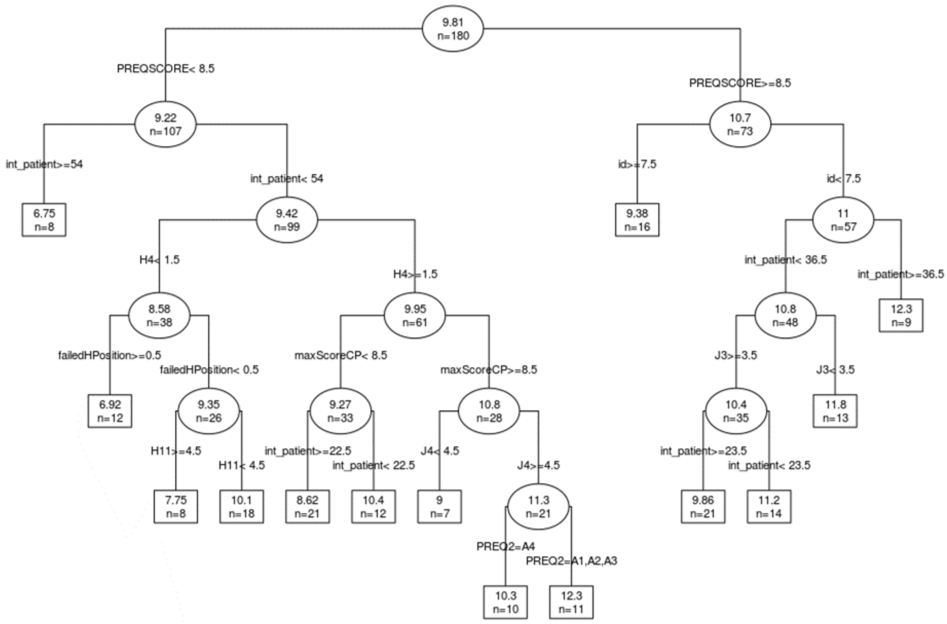


Figure 29. Tree for score prediction with 20 cross-validations

The lowest cross-validated error for this tree is 0.92262 obtained for *cp*=0.05. The error obtained doing cross validation is 4,9287, which improves that of all previous trees.

6.8.2. Using only game interactions information

- **Goal**: predict players' score in post-test.
- **Variables used**: gameCompleted, score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, mostRepeatedSituation, int_patient, int_phone, int_saed, failedEmergency, failedThrusts, failedHName, failedHPosition, failedHHands, POSTQSCORE (target variable).
- **Number of observations**: 198 (missing included).
- **Analysis method**: regression trees.

The first constructed tree, with default parameters, can be seen in Figure 30. The lowest cross-validated error for this tree is 1.0067 obtained for *cp*=0.06. The error obtained doing cross validation is 5,6851, calculated as the root node error (5,6473) times the cross validated error of the tree. In this first example, we notice that variables of relevance are: failedEmergency, failedHPosition, firstScoreU, int_patient, int_saed, maxScoreCP, maxScoreU, score and timesU.

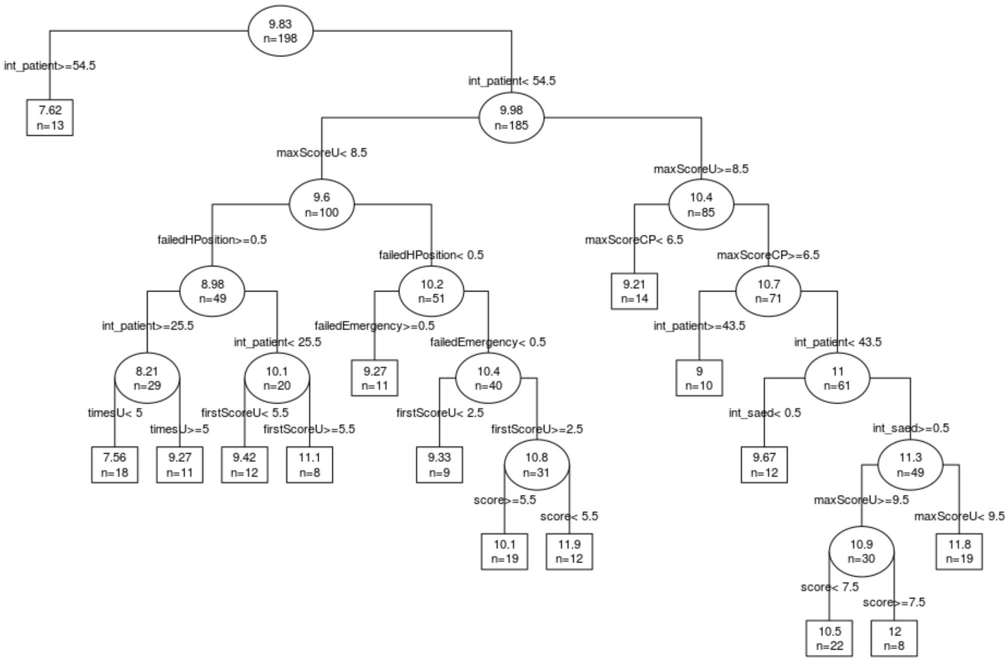


Figure 30. Tree for score prediction based only in game interactions variables

Tree variation 1

Fixing *minsplit*=30 and *minbucket*=15, we get the lowest cross-validated error for the tree in Figure 44 in Appendix 3: Additional figures and table results is 1.0082 obtained for *cp*=0.05. The error obtained doing cross validation is 5,6936.

Tree variation 2

Fixing *cp*=0.02, we obtain the tree in Figure 45 in Appendix 3: Additional figures and table results, which has a lowest cross-validated error of 1.077 obtained for *cp*=0.05. The error obtained doing cross validation is 5,6907.

Fixing the number of cross validation groups to 20, we obtain the same tree of Figure 30. Using cross validation with K=10 groups, we obtain a mean absoulte error of 0.23%, shown in Figure 46 in Appendix 3: Additional figures and table results. The variables used in the 10 trees constructed can be seen in Table 8.

Table 8. Variables selected in cross validation iteration for score prediction with a regression tree based only in game interactions variables

N	Variables selected
10 trees	int_patient
9 trees	int_saed, maxScoreCP
8 trees	firstScoreU, failedEmergency, maxScoreU
7 trees	failedHPosition, timesU
6 trees	-
5 trees	firstScoreCP, firstScoreCH, score
4 trees	-
3 trees	int_phone, failedThrusts
2 trees	-
1 tree	failedHName, maxScoreCH, timesCP

6.9. Prediction of players' post-test score (II)

6.9.1. Using pre-test information

- **Goal**: predict players' score in post-test.
- **Variables used**: class, sex, age, PREQi, PREQiRIGTH for i in [1, 15], Hi for i in [1, 11], Jj for j in [1, 5], score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, int_patient, int_phone, int_saed, POSTQSCORE (target variable).
- **Number of observations**: 183 (missing excluded).
- **Analysis method**: linear regression.

Different regression models were tested. We can compare their values of residual standard error (on some degrees of freedom), r^2 (which always increases adding more variables), their values of adjusted r^2. N is the number of variables included in the model.

In each iteration, we consider the next model based on the variables of greater significancy in the previous models. Then, we include interaction between class variables and interactions between class and interval variables. The results can be seen in Table 9.

Table 9. Different linear regression models and parameters obtained

Model	**N**	**RSE**	**Df**	r^2	**Adj** r^2
All variables	72	1.909	94	0.6524	0.3269
PREQ12 + PREQ13 + maxScoreCP + maxScoreCH + timesCP + PREQ8 + PREQ9 + PREQ15 + J3 + gameCompleted	10	2.032	162	0.3211	0.2373
PREQ12 + PREQ13 + maxScoreCP + maxScoreCH + timesCP	5	2.186	173	0.1612	0.1176
PREQ12 + PREQ13 + maxScoreCP + PREQ8	4	2.159	172	0.1868	0.1395
PREQ12 + PREQ13 + maxScoreCP	3	2.202	175	0.1388	0.1043
PREQ12 + PREQ13 + maxScoreCP + PREQ12*PREQ13	4	2.226	169	0.1507	0.0853
PREQ12 + PREQ13 + maxScoreCP + PREQ12*maxScoreCP + PREQ13*maxScoreCP	5	2.185	169	0.1816	0.1187
PREQ12 + PREQ13 + maxScoreCP + PREQ12*PREQ13 + PREQ12*maxScoreCP + PREQ13*maxScoreCP	6	2.205	163	0.1956	0.1019

The best model seems to be PREQ12 + PREQ13 + maxScoreCP + PREQ8, with only four variables obtains a 0.1395 of adjusted r^2. Comparing this model with the following model [45], which excludes variable PREQ8, we obtain that the latest is better.

We may wonder if including interactions improves the results significantly or not. The last model, which includes interaction PREQ13*maxScoreCP, decreases the previous results. So we try the variable selection methods with the previous model that includes two interactions and compare the obtained AIC values. The results can be seen in Table 10.

Table 10. Variable selection methods for linear regression model for score prediction

Selection method	**Model obtained**	**AIC**
Forward	PREQ12 + PREQ13 + maxScoreCP + PREQ12*maxScoreCP + PREQ13*maxScoreCP	299.46
Backward	PREQ12 + PREQ13 + maxScoreCP + PREQ13*maxScoreCP	293.90
Step by step	PREQ12 + PREQ13 + maxScoreCP + PREQ13*maxScoreCP	293.90

With both backward and step by step we obtain that the best model is given by variables: PREQ12 + PREQ13 + maxScoreCP + PREQ13*maxScoreCP. Comparing the model to the one with all variables, we obtain that this model provides a better result.

Finally, we can perfom cross validation to have a better evaluation of the model. Performing cross validation with K=10 cross validation groups, we obtain a total mean square error of 5,81.

6.9.2. Using only game interactions information

- **Goal**: predict players' score in post-test.
- **Variables used**: gameCompleted, score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, mostRepeatedSituation, int_patient, int_phone, int_saed, failedEmergency, failedThrusts, failedHName, failedHPosition, failedHHands, POSTQSCORE (target variable).
- **Number of observations**: 198 (missing included).
- **Analysis method**: linear regression.

Different regression models were tested. We can compare their values of residual standard error (on some degrees of freedom), r^2 (which always increases adding more variables), their values of adjusted r^2. N is the number of variables included in the model.

In each iteration, we consider the next model based on the variables of greater significancy in the previous models. Then, we include interaction between class variables and interactions between class and interval variables. The results can be seen in Table 11.

Table 11. Different linear regression models and parameters obtained for score prediction based only in game interactions variables

Model	N	RSE	Df	r^2	Adj r^2
All variables	20	2.232	176	0.2157	0.1221
int_patient + maxScoreCP + firstScoreCP + mostRepeatedSituation + failedHPosition	5	2.231	191	0.1495	0.1228
int_patient + maxScoreCP + firstScoreCP + failedHPosition	4	2.243	193	0.1319	0.1139
int_patient + maxScoreCP + failedHPosition	3	2.249	194	0.1226	0.1090
int_patient + maxScoreCP	2	2.277	195	0.0954	0.0861
int_patient + maxScoreCP + firstScoreCP + mostRepeatedSituation + failedHPosition + mostRepeatedSituation*failedHPosition	6	2.237	190	0.1499	0.1186
int_patient + maxScoreCP + firstScoreCP + mostRepeatedSituation + failedHPosition + int_patient*maxScoreCP + int_patient*firstScoreCP + maxScoreCP*firstScoreCP	8	2.237	188	0.1588	0.1185
int_patient + maxScoreCP + firstScoreCP + mostRepeatedSituation + failedHPosition + int_patient*mostRepeatedSituation + int_patient*failedHPosition + maxScoreCP*mostRepeatedSituation + maxScoreCP*failedHPosition + firstScoreCP*mostRepeatedSituation + firstScoreCP*failedHPosition	11	2.227	182	0.1931	0.1266
int_patient + maxScoreCP + firstScoreCP +	15	2.239	178	0.2017	0.1165

mostRepeatedSituation + failedHPosition + mostRepeatedSituation*failedHPosition + int_patient*maxScoreCP + int_patient*firstScoreCP + maxScoreCP*firstScoreCP + int_patient*mostRepeatedSituation + int_patient*failedHPosition + maxScoreCP*mostRepeatedSituation + maxScoreCP*failedHPosition + firstScoreCP*mostRepeatedSituation + firstScoreCP*failedHPosition					

Different variable selection methods can be applied to the previous best model to improve the value of AIC with less effects. Results can be seen in Table 12.

Table 12. Variable selection methods for linear regression model for score prediction based only in game interactions variables

Selection method	Model obtained	AIC
Forward	int_patient + maxScoreCP + firstScoreCP + mostRepeatedSituation + failedHPosition + mostRepeatedSituation*failedHPosition + int_patient*maxScoreCP + int_patient*firstScoreCP + maxScoreCP*firstScoreCP + int_patient*mostRepeatedSituation + int_patient*failedHPosition + maxScoreCP*mostRepeatedSituation + maxScoreCP*failedHPosition +firstScoreCP*mostRepeatedSituation + firstScoreCP*failedHPosition	338.17
Backward	int_patient + maxScoreCP + firstScoreCP + mostRepeatedSituation + failedHPosition + int_patient*mostRepeatedSituation	322.33
Step by step	int_patient + maxScoreCP + firstScoreCP + mostRepeatedSituation + failedHPosition + int_patient*mostRepeatedSituation	322.33

Performing cross validation with K=10 on the model with all interactions, we obtain an average square error of 5.92. Perfoming it on the model only with interactions between interval and categorial variables, the error decreases to 5.71. However, comparing both models with the *anova* function [46], the models are not significantly different at the level of α=0.05. Neither are significantly different the original model obtained with all variables and the model with interval and categorical interactions.

6.10. Prediction of players' pass/fail result (I)

6.10.1. Using pre-test information

- **Goal**: predict players' result as binary category pass/fail.
- **Variables used**: class, sex, age, PREQi, PREQiRIGTH for i in [1, 15], Hi for i in [1, 11], Jj for j in [1, 5], score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, int_patient, int_phone, int_saed, PASS (target variable).
- **Number of observations**: 198 (missing included).
- **Analysis method**: classification trees.

To transform the original score into a binary (pass / fail) category, as the test contains 15 questions, we establish the threshold in 7.5 (that is, with up to 7 questions correctly answered the category assigned is "fail"; with more than 7 questions correctly answered, the category is "pass").

From the original 198 observations, 168 students passed and 30 failed. This makes us deal with imbalanced classes (85%-15%) [47]. Therefore, we define a new binary variable taking 40% of it as fail class and the rest 60% as pass, that is, 47 observations were randomly selected from pass set. The final balanced set contains 78 observations.

The first classification tree [48] derived from the balanced set can be seen in Figure 31, where relevant variables are PREQSCORE, score and PREQ2.

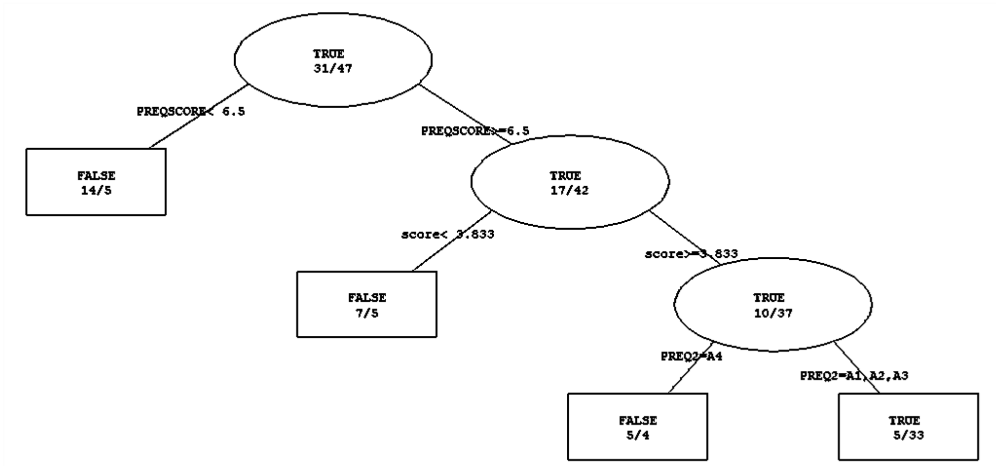


Figure 31. First tree for pass/fail prediction

The lowest cross-validated error for this tree is 0.96774 obtained for *cp*=0.06. The error obtained doing cross validation is 0,2435 (or a 24.35%), calculated as the root node error times the relative error of the tree.

Pruning the tree

Pruning the tree at *cp*=0.06, we obtain the tree in Figure 32. The error obtained for this tree is 0,3870 (or a 38.7%), calculated as the root node error times the relative error of the tree.

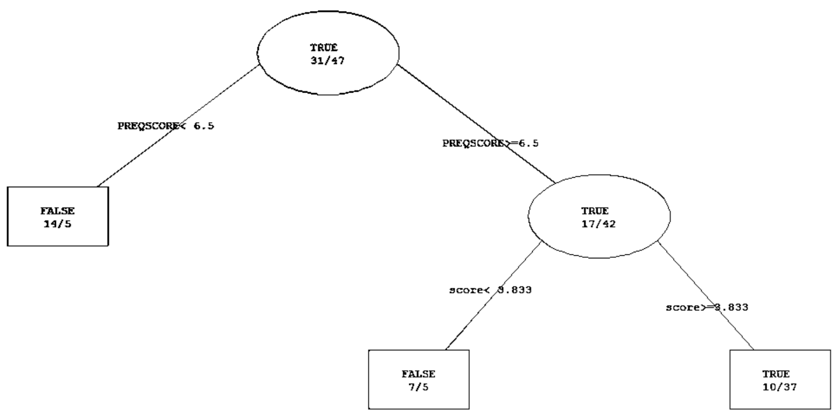


Figure 32. Tree pruned at cp=0.06

Cross validation

Applying cross validation with K=10 groups we obtain the error for prediction, precision and recall in each iteration can be seen in Figure 47 in Appendix 3: Additional figures and table results. The average error is of 3.4, the average precision is 0.6504 and the average recall is 0.8110.

Cross validation on original imbalanced data

We can wonder how accurate our predictions would have been using the original imbalanced dataset. Again with cross validation of K=10 groups, we obtain the number of errors, precision and recall for the original imbalanced dataset seen in Figure 48 in Appendix 3: Additional figures and table results. The average error is of 3.5, the average precision is 0.8157 and the average recall is 0.9424.

6.10.2. Using only game interactions information

- **Goal**: predict players' result as binary category pass/fail.
- **Variables used**: gameCompleted, score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, mostRepeatedSituation, int_patient, int_phone, int_saed, failedEmergency, failedThrusts, failedHName, failedHPosition, failedHHands, PASS (target variable).
- **Number of observations**: 198 (missing included).
- **Analysis method**: classification trees.

Balanced data

Again, we take the 30 observation in fail class as the 40% of new balanced data, the other 60% (45 observations) are taking randomly from passed set. The result balanced set has therefore 75 observations.

The first tree can be seen in Figure 33; variables of relevance are int_patient and firstScoreCH.

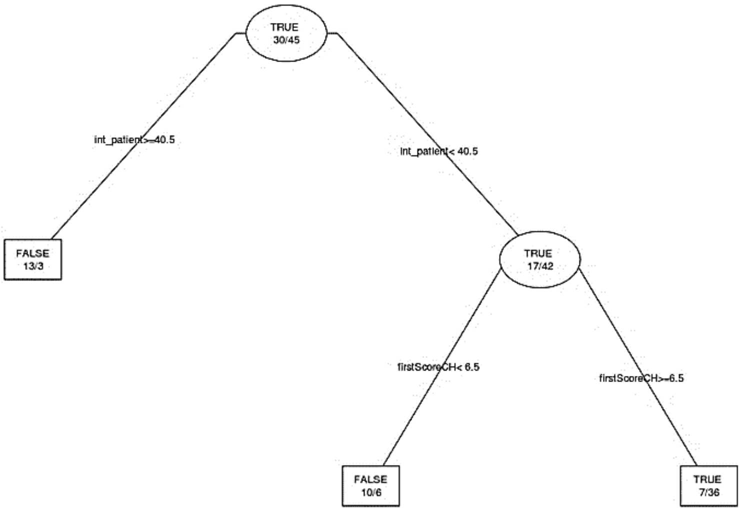


Figure 33. Tree for pass fail prediction based only in game interactions variables

Performing cross validation with K=10 groups we obtain a mean number of prediction errors of 1.7, a mean precision of 0.8949 and a mean recall of 0.8032 (see Figure 49 in Appendix 3: Additional figures and table results).

On the original imbalanced data, the number of errors raises to 2.5, the precision is 0.8859 and the recall is 0.9237 (and Figure 50 in Appendix 3: Additional figures and table results).

6.11. Prediction of players' pass/fail result (II)

6.11.1. Using pre-test information

- **Goal**: predict players' result as binary category pass/fail.
- **Variables used**: class, sex, age, PREQi, PREQiRIGTH for i in [1, 15], Hi for i in [1, 11], Jj for j in [1, 5], score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, int_patient, int_phone, int_saed, PASS (target variable).
- **Number of observations**: 183 (missing excluded).
- **Analysis method**: logistic regression.

Balanced data

We can obtain a balanced dataset taking the 25 observations in fail class as the 40% of this new set; the other 37 observations (60%) randomly selected from pass class. The balanced set has therefore 62 observations.

The models tested are summarized in Table 13, where N is the number of effects of the model and the value of AIC.

Table 13. Logistic regression models for pass fail prediction

Model	N	AIC
class + PREQ1 + PREQ2 + PREQ3 + PREQ4 + PREQ5 + PREQ6 + PREQ7 + PREQ8 + PREQ9 + PREQ10 + PREQ11 + PREQ12 + PREQ13 + PREQ14 + H1 + H2 + H3 + H4 + H5 + H6 + H7 + H8 + H9 + H10 + J1 + J2 + PREQSCORE	28	46.92

PREQ12 + PREQ15 + mostRepeatedSituation	3	84.34
PREQ12 + PREQ15	2	82.92
PREQ12 + PREQ15 + mostRepeatedSituation + H1 + J1	5	87.52
PREQ12 + PREQ15 + mostRepeatedSituation + PREQ12*PREQ15 + PREQ12*mostRepeatedSituation + PREQ15*mostRepeatedSituation	6	93.41
PREQ12 + PREQ15 + PREQ12*PREQ15	3	89.60

From the first model, we perform a backward variable selection method to reduce the number of effects. Results can be seen in Table 14.

Table 14. Variable selection methods for pass fail prediction

Selection method	Model obtained	AIC
Backward	class + PREQ1 + PREQ2 + PREQ3 + PREQ4 + PREQ5 + PREQ6 + PREQ7 + PREQ8 + PREQ9 + PREQ10 + PREQ11 + PREQ12 + PREQ14 + H1 + H3 + H4 + H5 + H6 + H7 + H8 + H9 + H10 + J1 + J2 + PREQSCORE	41.05

Cross validation

We can the following models using cross validation with K=10 groups:

1. PASS ~ class + PREQ1 + PREQ2 + PREQ3 + PREQ4 + PREQ5 + PREQ6 + PREQ7 + PREQ8 + PREQ9 + PREQ10 + PREQ11 + PREQ12 + PREQ13 + PREQ14 + H1 + H2 + H3 + H4 + H5 + H6 + H7 + H8 + H9 + H10 + J1 + J2 + PREQSCORE
2. PASS ~ class + PREQ1 + PREQ2 + PREQ3 + PREQ4 + PREQ5 + PREQ6 + PREQ7 + PREQ8 + PREQ9 + PREQ10 + PREQ11 + PREQ12 + PREQ13 + PREQ14 + H1 + H2 + H3 + H4 + H5 + H6 + H7 + H8 + H9 + H10 + J1 + J2 + J3 + J4 + J5 + PREQSCORE
3. PASS ~ PREQ12 + PREQ15

Comparing four metrics on these models (misclassfication rate, area under ROC curve, precision and recall), which can be seen in Figure 51 in Appendix 3: Additional figures and table results, we obtain that the two first models provide almost identical results, while the third simpler model turns out to provide better results in all four metrics: with better means in missclassification rate (0.1072), area under ROC curve (0.6723) and precision (0.8943), it is even more significant its better performance in recall (0.9856).

Original imbalanced data

We can test now different models on the original imbalanced dataset, including interactions between class variables and number of effects. Comparation of their values of AIC can be seen in Table 15.

Table 15. Different models for logistic regression

Model	N	AIC
All variables	72	140.73
PREQ12 + PREQ15 + mostRepeatedSituation	3	147.98
PREQ12 + PREQ15 + mostRepeatedSituation + PREQ9	4	144.46
PREQ12 + PREQ15 + mostRepeatedSituation + PREQ12*PREQ15 + PREQ12*mostRepeatedSituation + PREQ15*mostRepeatedSituation	6	146.62
PREQ12 + PREQ15 + mostRepeatedSituation + PREQ9 + PREQ12*PREQ15 + PREQ12*mostRepeatedSituation + PREQ15*mostRepeatedSituation + PREQ9*PREQ12 + PREQ9*PREQ15 + PREQ9*mostRepeatedSituation	10	106.11

The best model seems to be the latest with 10 variables greatly decreases the value of AIC. We can test different variable selection methods on this model in Table 16.

Table 16. Variable selection methods for logistic regression model for pass fail prediction

Selection method	Model obtained	AIC
Forward	PREQ12 + PREQ15 + mostRepeatedSituation + PREQ9 + PREQ12*PREQ15 + PREQ12*mostRepeatedSituation + PREQ15*mostRepeatedSituation + PREQ9*PREQ12 + PREQ9*PREQ15 + PREQ9*mostRepeatedSituation	106.11
Backward	PREQ12 + PREQ15 + mostRepeatedSituation + PREQ9 + PREQ15*mostRepeatedSituation	84.01
Step by step	PREQ12 + PREQ15 + mostRepeatedSituation + PREQ9 + PREQ15*mostRepeatedSituation	84.01

Both backward and step by step obtain the best model with 5 variables and AIC of 84.

Cross validation

We use cross validation to compare the previous best model to some of the previous, simpler and more complex, and verify if this is the best model indeed. Concretely, we compare the following four models:

1. PASS ~ PREQ12 + PREQ15 + mostRepeatedSituation
2. PASS ~ PREQ12 + PREQ15 + mostRepeatedSituation + PREQ9
3. PASS ~ PREQ12 + PREQ15 + mostRepeatedSituation + PREQ9 + PREQ15*mostRepeatedSituation
4. PASS ~ PREQ12 + PREQ15 + mostRepeatedSituation + PREQ9 + PREQ12*PREQ15 + PREQ12*mostRepeatedSituation + PREQ15*mostRepeatedSituation +PREQ9*PREQ12 + PREQ9*PREQ15 + PREQ9*mostRepeatedSituation

We use K=10 groups of cross validation and 100 repetitions. The results for the four models and the evaluation parameters: missclasification rate, area under ROC curve, precision and recall can be seen in Figure 52 in Appendix 3: Additional figures and table results. Model 2 has the best area under curve ROC, model 3 has the best precision and model 1 has the best results in missclassification rate (0.1049) and recall (0.9833) with still aceptable area under ROC (0.6817) and precision (0.8983), so we could take model 1.

We can change the threshold and see that the one that minimizes the missclasification rate is 0.28 and the one that maximizes the Youden index is 0.89.

6.11.2. Using only game interactions information

- **Goal**: predict players' result as binary category pass/fail.
- **Variables used**: gameCompleted, score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, mostRepeatedSituation, int_patient, int_phone, int_saed, failedEmergency, failedThrusts, failedHName, failedHPosition, failedHHands, PASS (target variable).
- **Number of observations**: 183 (missing excluded).
- **Analysis method**: logistic regression.

Balanced data

Different models were tested, including interactions between class and interval variables. Results for AIC and number of effects (N) are shown in Table 17.

Table 17. Different models for logistic regression with balanced data based only in game interaction information

Model	**N**	**AIC**
All variables	20	120.83
mostRepeatedSituation + int_patient + maxScoreCP	3	99.21
int_patient + maxScoreCP	2	100.05
mostRepeatedSituation + int_patient + maxScoreCP + mostRepeatedSituation*failedHPosition + mostRepeatedSituation*failedThrusts	5	99.87
mostRepeatedSituation + int_patient + maxScoreCP + mostRepeatedSituation*int_patient + mostRepeatedSituation*maxScoreCP	5	104.33
mostRepeatedSituation + int_patient + maxScoreCP + int_patient*maxScoreCP	4	101.21

Cross validation

We use cross validation to compare the best model to some of the previous, simpler and more complex, and verify if this is the best model indeed. Concretely, we compare the following four models:

1. PASS ~ mostRepeatedSituation + int_patient + maxScoreCP
2. PASS ~ int_patient + maxScoreCP
3. PASS ~ mostRepeatedSituation + int_patient + maxScoreCP + mostRepeatedSituation*int_patient + mostRepeatedSituation*maxScoreCP
4. PASS ~ mostRepeatedSituation + int_patient + maxScoreCP + int_patient*maxScoreCP

We use K=10 groups of cross validation and 100 repetitions. The results for the four models and the evaluation parameters: missclasification rate, area under ROC curve, precision and recall can be seen in Figure 53 in Appendix 3: Additional figures and table results.

On balanced data, we see more clearly how model 3 is worse than the others which are still similar. Now model 1 provides the best area under ROC curve. Model 2 has the lowest misclassication rate mean (0.2580), still aceptable area (0.7639), slighlty better precision (0.7272) and recall (0.8312), so in this case we could keep model 2 which is better in three of the metrics considered with the lowest number of effects. We can change the threshold and see that the threshold that minimizes missclasification rate is 0.59; the same threshold maximizes Youden index.

Original imbalanced data

We can apply those models to the original dataset and compare the values of AIC and their number of effects (N) in Table 18. We can see that the error AIC greatly increases on the original data.

Table 18. Different models for logistic regression based only in game interaction information

Model	**N**	**AIC**
All variables	20	149.44
mostRepeatedSituation + int_patient + maxScoreCP	3	128.69
int_patient + maxScoreCP	2	134.00

mostRepeatedSituation + int_patient + maxScoreCP + mostRepeatedSituation*failedHPosition + mostRepeatedSituation*failedThrusts	5	133.66
mostRepeatedSituation + int_patient + maxScoreCP + mostRepeatedSituation*int_patient + mostRepeatedSituation*maxScoreCP	5	129.20
mostRepeatedSituation + int_patient + maxScoreCP + int_patient*maxScoreCP	4	130.29

The best model seems to be the second one with only three variables has the lowest value of AIC. Different variable selection methods leave this model as it is.

Cross validation

We use cross validation to compare the previous best model to some of the previous, simpler and more complex, and verify if this is the best model indeed. Concretely, we compare the following four models:

1. PASS ~ mostRepeatedSituation + int_patient + maxScoreCP
2. PASS ~ int_patient + maxScoreCP
3. PASS ~ mostRepeatedSituation + int_patient + maxScoreCP + mostRepeatedSituation*int_patient + mostRepeatedSituation*maxScoreCP
4. PASS ~ mostRepeatedSituation + int_patient + maxScoreCP + int_patient*maxScoreCP

We use K=10 groups of cross validation and 100 repetitions. The results for the four models and the evaluation parameters: missclasification rate, area under ROC curve, precision and recall can be seen in Figure 54 in Appendix 3: Additional figures and table results.

The second model has the highest recall while the first model is slighly better in misclasification rate (0.1271), area under ROC (0.7327) and precision (0.8716), with a still aceptable recall (0.9881) containing only 3 effects, so we would take this as the best model.

We can change the threshold and the one that minimizes the missclasification rate is 0.51 and the one that maximizes the Youden index is 0.88.

6.12. Prediction of players' pass/fail result (III)

6.12.1. Using pre-test information

- **Goal**: predict players' result as binary category pass/fail.
- **Variables used**: class, sex, age, PREQi, PREQiRIGTH for i in [1, 15], Hi for i in [1, 11], Jj for j in [1, 5], score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, int_patient, int_phone, int_saed, PASS.
- **Number of observations**: 198 (missing included).
- **Analysis method**: Naïve Bayes.

Finally, we decided to perform Naïve Bayes as they are a commonly used method for classification with good results [49]. Splitting the dataset randomly into 70% train and 30% test 100 iterations, we obtain the results shown in Table 19.

Table 19. Results of Naïve Bayes classificator with pre test information

Measure	Min	Median	Mean	Max	Sd
Fraction correct	0.7647	0.8441	0.8465	0.9565	0.0390
Precision	0.8478	0.9252	0.9262	1.0000	0.0318
Recall	0.8113	0.8924	0.8907	0.9767	0.0353

With a low standard deviation, we obtain quite good results with Naïve Bayes.

6.12.2. Using only game interactions information

- **Goal**: predict players' result as binary category pass/fail.
- **Variables used**: score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, int_patient, int_phone, int_saed, PASS.
- **Number of observations**: 198 (missing included).
- **Analysis method**: Naïve Bayes.

Splitting the dataset randomly into 70% train and 30% test 100 iterations, we obtain the results shown in Table 20.

Table 20. Results of Naïve Bayes classificator only with game interactions information

Measure	Min	Median	Mean	Max	Sd
Fraction correct	0.7119	0.8305	0.8282	0.9062	0.0409
Precision	0.8182	0.8889	0.8895	0.9630	0.0345
Recall	0.8043	0.9099	0.9089	0.9804	0.0365

Slighlty worse than results with pre-test information, except for recall, still this method provides good prediction results.

6.13. Prediction of previous knowledge from game interactions

- **Goal**: predict players' PREQSCORE as binary category pass/fail with the information from game variables.
- **Variables used**: score, maxScoreCP, maxScoreU, maxScoreCH, firstScoreCP, firstScoreU, firstScoreCH, timesCP, timesU, timesCH, int_patient, int_phone, int_saed, PREQPASS.
- **Number of observations**: 198 (missing included).
- **Analysis method**: Naïve Bayes.

After obtaining good results for both Naïve Bayes methods, with pre-test and without it, we wonder if we can infer previous knowledge (given by pre-test score) from game interactions. For this purpose, we transform our continious variable PREQSCORE to a binary T/F variable and predict it with game variables. Results are shown in Table 21.

Table 21. Results of Naïve Bayes classificator to predict previous knowledg with game interactions information

Measure	Min	Median	Mean	Max	Sd
Fraction correct	0.5593	0.6767	0.6781	0.7931	0.0447
Precision	0.5625	0.6978	0.6924	0.8250	0.0548
Recall	0.7333	0.8485	0.8459	0.9630	00449

6.14. Analysis of knowledge recall

- **Goal**: determine if players retained knowledge after 2-3 weeks.
- **Variables used**: PREQi, POSTQi, PREQiRIGTH, POSTQiRIGTH for i in [1, 15], PREQSCORE, POSTQSCORE, GAIN. All these variables duplicated for first original experiment and second recall experiment.
- **Number of observations**: 10 observations.
- **Analysis method**: dependent t-Test.

First, we can plot the scores in both pre and post-test for both experiments (Figure 34). Analyzing this graph, we can see that there is a bigger variance in the pre-test scores of the second experiment that in the original. Comparing post-test scores, they seem to be slightly lower in the second experiment. To analyze what students recall, we focus on the post-test score of the first experiment and the pre-test score of the second experiment. In the graph, it seems that the latest is lower and has a bigger variance.

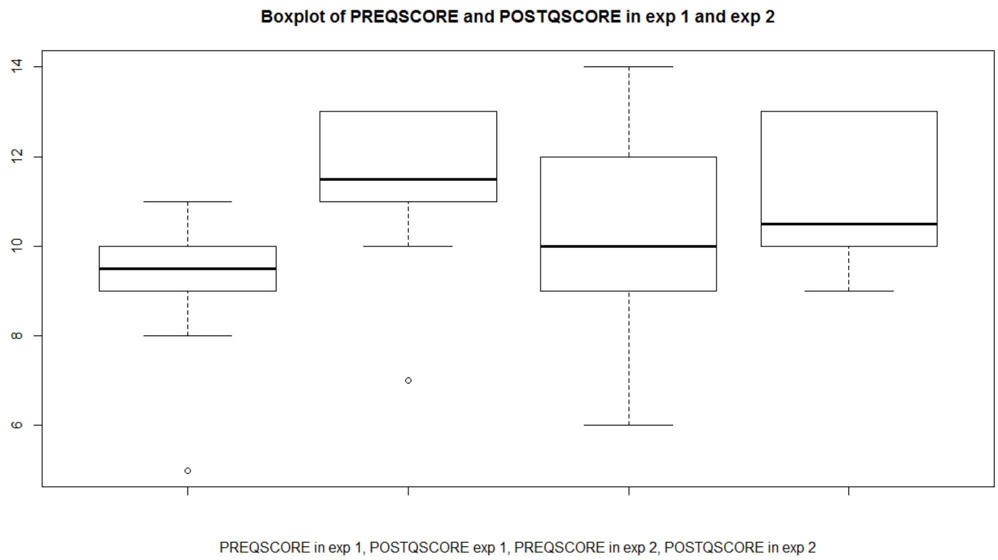


Figure 34. Boxplot of PREQSCORE and POSTQSCORE in first and second experiment

Variables are continuous, do not contain outliers and observations are independent. Testing the normality of both variables with Shapiro-Wilk normality test, we obtain both p-values > 0.01, so we can accept that variables follow a normal distribution. Therefore, we can apply a paired t-Test that returns a p-value > 0.05, so we reject the null hyphotesis of means difference, and accept the alternative hyphotesis that means are equal. This would lead us to think that there is a relation indeed between post-test score in the first experiment and pre-test score in the second experiment, that is, that students do recall what they learnt.

We may also analyze simply if students learnt in this experiment or not, and compare their learning to that of the original experiment analyzing its effect size. In this new experiment, out of the 15 students which participated, the mean PREQSCORE was 9.93 and the mean POSTQSCORE was 10.46, that is the mean in GAIN was 0.53. As POSTQSCORE does not have a normal distribution, we apply Wilcoxon signed-rank test to evaluate the equality between PREQSCORE and POSTQSCORE at level $\alpha=0.05$. The test returns a p-value > 0.05, so we cannot reject the null hyphotesis, therefore no significant difference appears between both scores. But we can still measure the effect size of this difference, from the z-score of -1.2613 we obtain a Pearson correlation coefficient *r* of:

$$r = \frac{-1.2613}{\sqrt{15 * 2}} = -0.23$$

This represents a medium effect as it is close to Cohen's benchmack of 0.30 ([40] as cited in [41]). This value is lower than in the original experiment (note that we can compare these values as effect size are independent of sample size), therefore we can conclude that the effect of this new experiment was less significant than the original one.

6.15. Analysis of number of game interactions

- **Goal**: determine the possible relation between game interactions and game habits.
- **Variables used**: Hi for i in [1, 11], int_patient, int_phone, int_saed.
- **Number of observations**: 199 observations.
- **Analysis method**: correlation and descriptive analysis.

Several of the previous analysis have made clear the relevance of game interactions: the variable int_patient appears as one of the variables of greater relevance in all previous models that do not use information from pre-test (that is, models for both score and pass / fail prediction with both trees and regression) and in the best model of regression tree even with information from pre-test. We may wonder if these interactions are related to game habits, that is, do gamers perform more clicks in the game than non gamers?

Correlations studied at the beginning of the analysis section were not significant (at 0.5 level) for game habits variables and game interactions, however, we may still find some relations between these variables. Figure 55 in Appendix 3: Additional figures and table results shows total interactions for each of the 5 levels of general game play frequency (as given in varible H1). Although the mean seams to be practically the same, there is a higher variance in number of iterations for players who play more frequently.

We can analyze this relation for particular types of games as well: in Figure 56 (Appendix 3: Additional figures and table results) we see that players who play daily music related games (value 5 – daily – in variable H4) perfom more interactions in mean (and with less variance) than players who play music related games less frequently. On the contrary, frequency of play for sport games (Figure 57) is not direcly related to number of game interactions: the mean is very similar for players with all frequencies and variance does not show a clear correspondence between frequency values and number of interactions either. As for the other categories, Mario games have a similar relation as music games with the number of game interactions, while there is no clear relation for the rest of categories.

We can analyze this relation between game interactions and game habits per sex. In terms of general game play frequency, which showed no relation with game interactions, considering both sexs separetely, there is a clear difference in variance (Figure 58 in Appendix 3: Additional figures and table results): females who play less vary a lot in the number of interactions they perfom while male players who play less perfom less interactions with a much smaller variance. For players who play more frequently, females have a bigger mean than males, who vary more in their game interations.

If we consider the different game types, now sport games do show a difference: mean for females is more variant, and variance is greater for less frequent players. Males show less relation between sport game play frequency and game interactions (Figure 59 in Appendix 3: Additional figures and table results).

6.16. Word frequency analysis of comments

- **Goal**: establish if students coments show a positive perception of the experiment.
- **Variables used**: J6.
- **Number of observations**: 94 observations.
- **Analysis method**: frequency analysis.

Out of the total players, only 94 made comments (47.24% of total) as this was an optional field in the post-test. Figure 35 shows the most commonly used words in the comments, after stopwords have been eliminated. As we observe, most frequent words are all positive: "game", "liked", "good", "fun", "learn" and combinations of words as "liked a lot" and "learnt a lot". Smiles were also frequent in the comments.

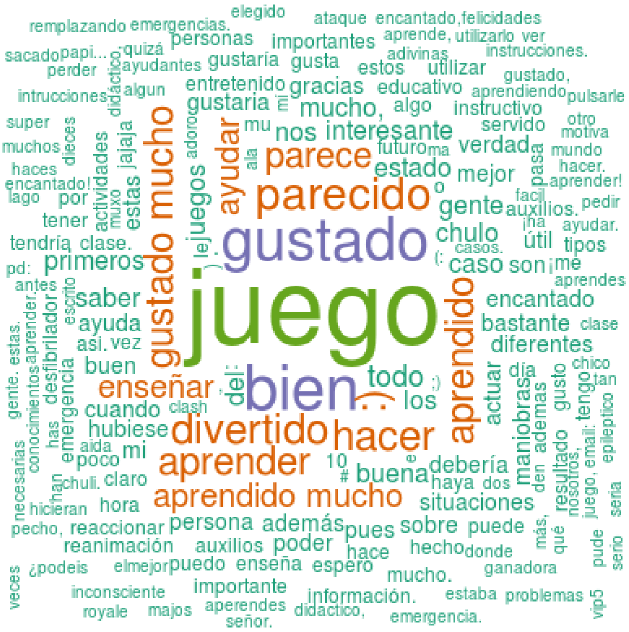


Figure 35. Word cloud of most used words in comments

7. Conclusions

In the previous sections we have stated the description of the experiments to collect the data, the several analysis carried out in the data and the results obtained in those analysis. At the beginning of the project, we established the following four main goals. Now, in the light of the results obtained in this project, we can conclude that:

G1. Determine the influence of previous knowledge in game results.

It has been established the influence of previous knowledge in game results by obtaining several prediction models and also by comparing all prediction models with and without pre-test information. Results show that previous knowledge influence has a great relevance in predictions and models without this previous knowledge generally provide worse results than with it.

G2. Determine the influence of game habits in game results.

The influence of game habits in game results has also been established as game habits variables appear to be significant in several of the prediction models containing previous information. Results also show the influence of frequeny of gameplay for different types of games in the way students play: specifically, in the number of interactions they perform in the game.

G3. Determine the capability of game interactions to predict post test results when combined with the pre test.

The prediction capability of pre test together with game interactions to predict post test results differs from different models; however an adequate average square error of 4.92 points is obtained for the best linear model.

When predicting simpler pass / fail results, it is clearer the good performance of the models: the best precision obtained is of 92% and the best recall is of 98%.

G4. Compare the previous capability to that of game interactions on their own to predict post test results.

The prediction capability of game interactions on their own to predict post test results, without any pre test information has also been established. With generally worse results than with pre-test information, still good prediciton models have been obtained only from game interactions. An average square error of 5.68 is obtained for the best score prediction model.

When predicting simpler pass / fail results, the best precision obtained is of 89% and the best recall is of 98%, similar to that of the best model with pre test information.

Specific prediction results for best models (both for score prediction and for pass / fail prediction) are summarized in subsection 7.1. Main conclusions derived from other analysis are summarized in subsection 7.2.

7.1. Summary of prediction results

Table 22 summarizes the results obtained for the best prediction models for post test score. Similar results are obtained for the different models, being the best the regression tree with pre-test information.

Table 22. Result of best prediction models for score

Method	**Use pre-test data**	**ASE**
Regression trees	Yes	4.9287
	No	5.6851
Linear regression	Yes	5.8100
	No	5.7100

Table 23 summarizes the results for the best prediction models for pass / fail prediction. Note that Naïve Bayes has not been used for regression as studies have shown that its good performance is restricted to classification [50]. Logistic regression provides the best recall results while Naïve Bayes has similarly good results both with pre-test information and without it.

Table 23. Results of best prediction models for pass / fail

Method	**Use pre-test data**	**Balanced data**	**Precision**	**Recall**
Decision trees	Yes	Yes	0.6504	0.8110
		No	0.8147	0.9424
	No	Yes	0.8949	0.8032
		No	0.8859	0.9237
Logistic regression	Yes	Yes	0.8943	0.9856
		No	0.8983	0.9833
	No	Yes	0.7272	0.8312
		No	0.8716	0.9881
Naïve Bayes	Yes	-	0.9257	0.8967
	No	-	0.8975	0.9057

The generally good prediction results could lead to predict post-test results without the need of conducting the post-test itself. This will greatly reduce the costs of performing experiments both in time and effort. This way, games could be played by bigger samples of students whose results could be predicted by current models.

Comparing models with and without pre-test information, no great differences were found. In fact, we also tried to predict pre-test scores from game interactions obtaining an aceptable predictions (mean of 0.67 fraction correct; mean of 0.69 precision and mean of 0.84 recall, to predict pre pass / fail result). This result could lead to another essential difference: games could not only be used to teach an issue but also to measure students' knowledge on that issue. If previous knowledge can be derived from game interactions, games could be used as an assessment tool by themselves.

7.2. Summary of analysis results

The several analysis conducted for the data collected in our experiment, for learners from ages 12 to 17, also derive some general conclusions:

- Male students play videogames more often than female students.
- Students from higher years (Fourth Year of ESO and Bachillerato) play less often than younger students (First, Second and Third Year of ESO).
- Principal components analysis reveals five main components containing, information of: (1) most game habits, (2) game situations' times and scores, (3) age, final score and post-test score, (4) rest of game habits and (5) pre-test score.
- Factor analysis reveals relations between: (1) most game habits questions, (2) in-game interactions with different elements, (3) scores in pre-test and post-test, and (4) max score and time of completion in the three game levels.
- The analysis of the comments reveales that those players who left comments enjoyed the game. Most frequent words include: "game", "liked", "learnt", "fun" or "learn". A smile also appeared quite frequently in the comments.

7.2.1. Summary of learning results

- The scores obtained in the questionnaire after playing the game were significantly higher than scores before playing the game ($p < .05$, $r=-.41$), that is, students did improve their knowledge (as far as it is measured in the questionnaire) playing the game. Compared to the original validation experiment, their knowledge improved less in this new experiment. Still results show significance so this proves the replicablity of the experiment.
- The scores obtained in the pre-test for students who repeated the experiment, showed no significant difference to scores obtained in the post-test the first time they played ($p > .05$, $r=.23$). This could mean that students do recall what they learnt in the first experiment – although notice that the small sample $N=10$ does not ensure results.

7.2.2. Summary of game habits results

- There are two main group of leaners based on game habits: one is mainly related to male students, who play more often all types of games and in particular, some specific types of games (e.g. first person shooter games, adventure or thriller games, sports, racing or simulation games and mobile or tablet games), and the other is more related to female.
- Both PCA and factor analysis found two main groups of game habits questions: H4, H6 and H9 in one group, the rest in the other group. Analyzing the content of those three questions we see they are related to: music – singing, dancing or playing instruments games - (H4), thinking – intelligence and quiz/trivia games - (H6) and social - Super Mario, Mario Kart or Wii Sports - (H9). We can find a meaning for this group of questions as they are related to more lightweight games, compared to first person shooters, fighting or sports games.
- No clear relation appears between number of game interactions and players' general game habits. For particular game categories, a relation appears; for instance: player who play music games daily interact with the game more than players who play music games less frequently. If we study this relation for both sexs separately, we obtain that less frequent female players perform more variant number of intreactions than male players, who interact less if they play less frequently. For specific game categories, a similar relation appears as women who play less frequently usually are more variant in their number of game interactions than male players.

8. Bibliography

[1] M. Freire, Á. Serrano-Laguna, B.M. Iglesias, I. Martínez-Ortiz, P. Moreno-Ger, B. Fernández-Manjón, Game Learning Analytics: Learning Analytics for Serious Games, in: Learn. Des. Technol., Springer International Publishing, Cham, 2016: pp. 1–29. doi:10.1007/978-3-319-17727-4_21-1.

[2] C.S. Loh, Y. Sheng, D. Ifenthaler, Serious Games Analytics, Springer International Publishing, Cham, 2015. doi:10.1007/978-3-319-05834-4.

[3] Asociación Servicio Interdisciplinar de Atención a las Drogodependencias (SIAD), Aislados, (2014). http://www.aislados.es/zona-educadores/ (accessed November 13, 2016).

[4] L. interFUEL, Darfur is Dying, (2006). http://www.gamesforchange.org/play/darfur-is-dying/.

[5] Center for Game Science at University of Washington in collaboration with UW Department of Biochemistry., Foldit: Solve Puzzles for Science, (2008). http://fold.it/portal/ (accessed November 4, 2016).

[6] Center for Game Science at the University of Washington, Treefrog Treasure, (2016). http://centerforgamescience.org/blog/portfolio/treefrog-treasure/ (accessed November 15, 2016).

[7] K. Squire, M. Barnett, J.M. Grant, T. Higginbotham, Electromagnetism supercharged!:learning physics with digital simulation games, Int. Conf. Learn. Sci. (2004) 513–520.

[8] M.J. Mayo, Video Games: A Route to Large-Scale STEM Education?, Science (80-.). 323 (2009) 79–82. doi:10.1126/science.1166900.

[9] K.H. Evans, W. Daines, J. Tsui, M. Strehlow, P. Maggio, L. Shieh, Septris: a novel, mobile, online, simulation game that improves sepsis recognition and management., Acad. Med. 90 (2015) 180–4.

[10] B.M. Iglesias, C. Fernandez-Vara, B. Fernandez-Manjon, E-Learning Takes the Stage: From La Dama Boba to a Serious Game, IEEE Rev. Iberoam. Tecnol. Del Aprendiz. 8 (2013) 197–204. doi:10.1109/RITA.2013.2285023.

[11] M. Greer, L. Lin, R.K. Atkinson, Using a computer game to teach school-aged children about asthma, Interact. Learn. Environ. 25 (2017) 431–438. doi:10.1080/10494820.2015.1135469.

[12] I. Boada, A. Rodriguez-Benitez, J.M. Garcia-Gonzalez, S. Thió-Henestrosa, M. Sbert, 30 : 2: A Game Designed to Promote the Cardiopulmonary Resuscitation Protocol, Int. J. Comput. Games Technol. 2016 (2016) 1–14. doi:10.1155/2016/8251461.

[13] K. McConachie, They're having fun but are they learning?, (2017). https://www.media.mit.edu/posts/they-re-having-fun-but-are-they-learning/.

[14] GameAnalytics, GameAnalytics, (2015). http://www.gameanalytics.com/ (accessed June 3, 2016).

[15] T. Elias, Learning Analytics: Definitions, Processes and Potential, (2011). http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf.

[16] C. Alonso-Fernández, A. Calvo Morata, M. Freire, I. Martínez-Ortiz, B. Fernandez-Manjon, Systematizing game learning analytics for serious games, in: IEEE Glob. Eng. Educ. Conf., 2017: pp. 1106–1113. http://www.e-ucm.es/drafts/e-UCM_draft_303.pdf.

[17] A.C. and M. Ruiz, A systematic literature review on serious games evaluation: An application to software project management, Comput. Educ. 87 (2015) 396–422.

[18] The Open University of the Netherlands, University of Trento, Hull College of Further Education, Universidad Complutense de Madrid, T.U. of Bolton, R.

INMARK Instituto de Engenhariade Sistemas e Computadores, Investigacao e Desenvolvimiento im Lis, H2020 RAGE (Realising an Applied Gaming Eco-system) Project, (2015). http://rageproject.eu/ (accessed November 30, 2016).

[19] H2020 EU, BEACONING Breaking Educational Barriers with Contextualised, Pervasive and Gameful Learning, (n.d.). http://beaconing.eu/ (accessed November 4, 2016).

[20] B. Manero, J. Torrente, M. Freire, B. Fernández-Manjón, An instrument to build a gamer clustering framework according to gaming preferences and habits, Comput. Human Behav. 62 (2016) 353–363. doi:10.1016/j.chb.2016.03.085.

[21] E.J. Marchiori, G. Ferrer, B. Fernandez-Manjon, J. Povar-Marco, J.F. Suberviola, A. Gimenez-Valverde, Video-game instruction in basic life support maneuvers, Emergencias. 24 (2012) 433–437.

[22] e-ucm, First Aid Game, (n.d.). http://first-aid-game.e-ucm.es/.

[23] I. Perez Colado, uAdventure: desarrollo del intérprete y de un emulador de videojuegos de -Adventure sobre Unity3D, 2016.

[24] C. Alonso-Fernández, Gaming Learning Analytics for Serious Games, 2016.

[25] The LimeSurvey Project Team, LimeSurvey, (2013). https://www.limesurvey.org/.

[26] ADL, Experience API, (n.d.). https://www.adlnet.gov/adl-research/performance-tracking-analysis/experience-api/ (accessed March 20, 2016).

[27] Á. Serrano-Laguna, I. Martínez-Ortiz, J. Haag, D. Regan, A. Johnson, B. Fernández-Manjón, Applying standards to systematize learning analytics in serious games, Comput. Stand. Interfaces. (2016). doi:10.1016/j.csi.2016.09.014.

[28] The R Foundation, R, (n.d.). https://www.r-project.org/.

[29] RStudio, (n.d.).

[30] L.L. Pipino, Y.W. Lee, R.Y. Wang, Data quality assessment, Commun. ACM. 45 (2002) 211. doi:10.1145/505248.506010.

[31] J. Osborne, Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data, SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States, 2013. doi:10.4135/9781452269948.

[32] European Resuscitation Council Guidelines Writing Group, European Resuscitation Council Guidelines for resuscitation, (2015). http://ercguidelines.elsevierresource.com/european-resuscitation-council-guidelines-resuscitation-2015-section-1-executive-summary/fulltext (accessed March 6, 2017).

[33] J.C. Gower, A General Coefficient of Similarity and Some of Its Properties, Biometrics. 27 (1971) 857. doi:10.2307/2528823.

[34] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis (Wiley Series in Probability and Statistics), 1990.

[35] Department of Mathematics ETH Zürich, Dissimilarity Matrix Calculation, (n.d.). https://stat.ethz.ch/R-manual/R-devel/library/cluster/html/daisy.html (accessed February 21, 2017).

[36] U.B. Department of Computer Science, College of Science and Mathematics, The PAM Clustering Algorithm, (n.d.). http://www.cs.umb.edu/cs738/pam1.pdf (accessed February 21, 2017).

[37] P.J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65. doi:10.1016/0377-0427(87)90125-7.

[38] Wicked Good Data - r, Clustering mixed data types in R, (2016). https://www.r-bloggers.com/clustering-mixed-data-types-in-r/ (accessed February 21, 2017).

[39] W. Chang, Cookbook R, (n.d.). http://www.cookbook-r.com/.
[40] J. Cohen, A power primer., Psychol. Bull. 112 (1992) 155–159. doi:10.1037/0033-2909.112.1.155.
[41] A.P. Field, G. Hole, How to design and report experiments, 2003. doi:10.1016/B978-0-12-384727-0.00002-1.
[42] P. Gerrard, R.M. Johnson, Mastering Scientific Computing with R, 2015.
[43] G. Raîche, M. Riopel, J.-G. Blais, Non Graphical Solutions for the Cattell's Scree Test, Int. Meet. Psychom. Soc. (2006) 12. http://www.empowerstats.com/manuals/paper/scree.pdf#1 (accessed March 21, 2017).
[44] Y. Zhao, R and Data Mining: Examples and Case Studies, 2012. doi:10.1016/B978-0-12-396963-7.00001-5.
[45] Department of Mathematics ETH Zürich, ANOVA for Linear Model Fits, (2017). https://stat.ethz.ch/R-manual/R-devel/library/stats/html/anova.lm.html (accessed May 2, 2017).
[46] L. Torgo, Data Mining with R : learning by case studies, Technometrics. 45 (2003) 112–113. doi:10.1198/tech.2003.s39.
[47] T. Fawcett, Learning from Imbalanced Classes, (2016). https://svds.com/learning-imbalanced-classes/#ref5 (accessed April 7, 2017).
[48] G. Daróczi, Mastering Data Analysis with R, 2015. doi:10.1017/CBO9781107415324.004.
[49] H. Zhang, The Optimality of Naive Bayes, Proc. Seventeenth Int. Florida Artif. Intell. Res. Soc. Conf. FLAIRS 2004. 1 (2004) 1–6. doi:10.1016/j.patrec.2005.12.001.
[50] E. Frank, L. Trigg, G. Holmes, I.H. Witten, Naive Bayes for Regression, Mach. Learn. 41 (2000) 5–25. doi:10.1023/A:1007670802811.

9. Appendix 1: Questionnaires

9.1. Pre and post questionnaires: common questions

Conocimientos sobre reanimación cardiopulmonar básica
(En las siguientes preguntas sólo una respuesta es correcta)

1.- Tras estimular a una persona inconsciente y ésta no responder, la actitud correcta sería:
  a. Llamar pidiendo ayuda a 061 o 112
  b. Iniciar respiraciones boca a boca
  c. Estimular al paciente hasta que despierte
  d. Darle un azucarillo por si es una bajada de azúcar

2.- Ante una persona inconsciente de las siguientes opciones cuál le parece la más correcta:
  a. Levantarle las piernas
  b. Echarle un vaso de agua por la cara para que despierte
  c. Tratar de ver, oír y sentir para comprobar si respira
  d. Poner la persona tumbada de lado, en posición de seguridad

3.- Ante una persona inconsciente también sería una actitud correcta:
  a. Iniciar compresiones torácicas
  b. Incorporarle para que recupere la consciencia
  c. Hacer boca a boca
  d. Dar golpes en la espalda, por si está atragantado

4.- Cuando se hacen compresiones torácicas el número a realizar por minuto será:
  a. 200 por minuto
  b. 50 por minuto
  c. 100 por minuto
  d. 150 por minuto

5.- Si encontramos una persona inconsciente o con dolor torácico a la vez que pedimos ayuda podemos pedir:
  a. Un vaso de agua
  b. Una manta para tapar al enfermo y que no se enfríe
  c. Que venga la policía
  d. Un desfibrilador automático (DEA)

6.- Si una persona inconsciente está respirando, la actitud a tomar será:
  a. Poner la persona tumbada de lado, en posición de seguridad
  b. Sentarle en una silla
  c. Abrigarle una vez que esté boca arriba
  d. Ponerle una almohada en la cabeza

7.- Cuando una persona se atraganta lo primero es
  a. Realizar compresiones abdominales rápidas (Maniobra de Heimlich)
  b. Animarle a que tosa
  c. Dar golpes en la espalda
  d. Sujetarle la frente para que no se canse

8.- Si una persona atragantada no puede respirar, la actitud correcta sería:
  a. Realizar compresiones abdominales rápidas (Maniobra de Heimlich)
  b. Meterle los dedos en la boca por si encontramos algo
  c. Esperar a que pase alguien
  d. Dar un vaso de agua

9.- Si dispones de un desfibrilador automático y una persona queda inconsciente
  a. Conectar y aplicar los electrodos, separarse mientras analiza el ritmo, y seguir las instrucciones que nos del aparato
  b. Tras aplicar los electrodos, realizar descarga
  c. Esperar a poner los electrodos por si recupera el nivel de consciencia
  d. Leer el manual de instrucciones antes de utilizarlo

10.- Si vamos a realiza una descarga con un desfibrilador automático, lo mejor será:
  a. Sujetarle para que no se mueva
  b. Pedir a alguien que le sujete mientras se realiza la descarga
  c. No tocar a la persona ya que podríamos recibir una descarga
  d. Esperar a que vengan los médicos de urgencias para administrar la descarga

11.- Si una persona está mojada y tenemos un desfibrilador automático:
  a. No se puede utilizar un desfibrilador automático
  b. Secarle y aplicar los electrodos del desfibrilador automático
  c. Poner los electrodos del desfibrilador automático encima la ropa mojada
  d. Esperar a que vengan los médicos de urgencias

12.- Si el desfibrilador automático no funciona porque no tiene batería:
  a. Pedir otro y esperar a que llegue
  b. Iniciar maniobras de reanimación con compresiones en el pecho
  c. Hacer boca a boda, hasta que llegue otro desfibrilador automático
  d. Poner en posición lateral de seguridad hasta que se tenga un desfibrilador automático

13.- En una persona con dolor torácico hay que:
  a. Ayudar a que se siente y pedir ayuda urgente
  b. Dar un vaso de agua con azúcar para mejorar el dolor
  c. Animarle a hacer movimientos para que mejore el dolor
  d. Un dolor torácico suele ser irrelevante, no tiene por qué ser grave

14.- Si una persona con dolor torácico queda inconsciente
  a. Iniciar compresiones torácicas hasta que llegue un desfibrilador automático o un médico de urgencias
  b. Probablemente haya quedado dormido tras el dolor
  c. Debemos ponerle tumbado en posición lateral de seguridad
  d. Debemos iniciar boca a boca

15.- La posición lateral de seguridad consiste en:
  a. Levantar las piernas
  b. Levantar las piernas y brazos
  c. Poner a la persona de lado con la mano alejada del suelo debajo de la cara y la pierna del mismo lado que la mano cruzada a 45º sirviendo de apoyo
  d. Elevar la cabeza

9.2. Pre questionnaire: questions about game habits

Contesta del 1 (nunca) al 5 (a diario) a las siguientes preguntas sobre tus hábitos de juego.

	Nunca (1) A diario (5)				
	1	2	3	4	5
H1 - ¿Con qué frecuencia juegas a videojuegos?					
H2 - ¿Con qué frecuencia juegas a videojuegos de disparos en primera persona (Call of Duty, Black Ops, Borderlands, Halo, Bioshock)?	1	2	3	4	5
	1	2	3	4	5
H3 - ¿Con qué frecuencia juegas a videojuegos de aventura (Uncharted, Heavy Rain, Resident Evil, Assassin's Creed)?					
H4 - ¿Con qué frecuencia juegas a videojuegos de cantar, bailar o tocar instrumentos (Guitar Hero, Sing Star, Just Dance)?	1	2	3	4	5
H5 - ¿Con qué frecuencia juegas a videojuegos de peleas (Tekken, Mortal Kombat, Street Fighter)?	1	2	3	4	5
H6 - ¿Con qué frecuencia juegas a videojuegos de inteligencia, preguntas/respuestas (Preguntados, Trivial, Brain Training)?	1	2	3	4	5
H7 - ¿Con qué frecuencia juegas a videojuegos de estrategia (Civilization, Age of Empires, Starcraft)?	1	2	3	4	5
H8 - ¿Con qué frecuencia juegas a videojuegos de deportes, carreras o simulación (FIFA, PES, NBA Live, Gran Turismo, Need for Speed)?	1	2	3	4	5
H9 - ¿Con qué frecuencia juegas a videojuegos como Super Mario, Mario Kart o Wii Sports?	1	2	3	4	5
H10 - ¿Con qué frecuencia juegas a videojuegos de internet multijugador masivos (World of Warcraft, RuneScape, League of Legends)?	1	2	3	4	5
H11 - ¿Con qué frecuencia juegas a videojuegos en el móvil o tablet (Clash Royale, Clash of Clans, Candy Crush)?	1	2	3	4	5

9.3. Post questionnaire: questions about game opinion

Valora del 1 (nada de acuerdo) al 5 (muy de acuerdo) los siguientes aspectos del juego, según tu opinión.

	Nada de acuerdo(1)			Muy de acuerdo(5)	
J1 - Usar el juego es interesante	1	2	3	4	5
J2 - He aprendido con el juego	1	2	3	4	5
J3 - El juego ha sido divertido	1	2	3	4	5
J4 - Me gustaría tener más juegos en clase	1	2	3	4	5
J5 - El juego es fácil de usar	1	2	3	4	5
J6 - Si quieres hacer algún comentario más sobre el juego, puedes hacerlo aquí.					

10. Appendix 2: Game habits report for school

Cátedra Telefónica
Educación Digital y Juegos Serios
Cátedra Telefónica-Complutense

Enseñando técnicas de primeros auxilios a alumnos de instituto mediante un juego serio:

Informe sobre el experimento y los hábitos de juego de los alumnos

La Inmaculada Escolapias – Puerta de Hierro

Elaborado por el grupo e-UCM

www.e-ucm.es

Con la colaboración de:

Objetivo del experimento

El objetivo del experimento es comprobar la eficacia de un videojuego como herramienta de aprendizaje para los alumnos. Para el experimento se utiliza un juego educativo o *juego serio*, es decir, un videojuego cuyo objetivo principal no es entretener sino enseñar, cambiar alguna actitud o comportamiento o crear conciencia sobre algún asunto. En este caso, el videojuego utilizado es *First Aid Game*, un juego que pretende enseñar diferentes técnicas de primeros auxilios.

En el experimento, cada estudiante/jugador realizó:

1. Un cuestionario previo a la partida sobre conocimiento de primeros auxilios, así como unas preguntas sobre sus hábitos de juego.
2. Una partida completa al juego serio seleccionado.
3. Un cuestionario posterior a la partida sobre conocimiento de primeros auxilios para comprobar lo que han aprendido con el juego, así como unas preguntas sobre su opinión del mismo.

Toda la información de los estudiantes se recoge de manera anónima: los alumnos acceden al juego mediante un código (compuesto por cuatro letras) que deben introducir una sola vez al comienzo del experimento. Este código es proporcionado por el profesor, que mantiene la lista con la correspondencia entre alumnos y códigos, siendo el único que puede relacionar la información recogida con un alumno concreto.

El juego First Aid Game

El juego *First Aid Game* fue desarrollado por el grupo e-UCM en 2011 financiado por el Centro Aragonés de Tecnologías para la Educación (CATEDU). El objetivo del juego es enseñar conocimientos teóricos sobre soporte vital básico a alumnos a partir de 12 años en tres situaciones: dolor torácico, inconsciencia y atragantamiento. El juego fue validado mediante la realización de un experimento con más de 300 alumnos de entre 12 y 14 años de cuatro institutos de la Comunidad Autónoma de Aragón (España).

Ilustración 1. Pantalla inicial del juego First Aid Game con las tres situaciones: dolor torácico, inconsciencia y atragantamiento.

Entre los conocimientos específicos que se enseñan en el juego se incluye: número de emergencias en España, realización de maniobra de Heimlich, realización de compresiones torácicas o la utilización de un desfibrilador automático.

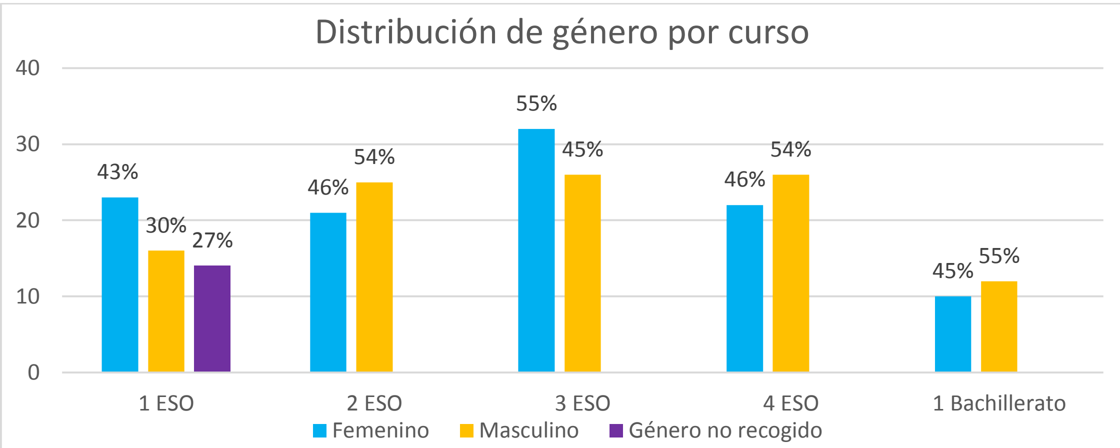
Alumnos que participaron en el experimento

Los experimentos se realizaron entre Enero y Febrero de 2017 en 16 sesiones con un número total de 227 alumnos de los cursos de 1º, 2º, 3º y 4º de ESO y 1º de Bachillerato del Colegio La Inmaculada Escolapias Puerta de Hierro.
La distribución concreta de los alumnos por cursos y sesiones fue la siguiente:

Sesión	Fecha	Curso	Número de alumnos
1	30/Enero/2017	1 Bach.	14
2	30/Enero/2017	3 ESO	14
3	7/Febrero/2017	1 ESO	14
4	7/Febrero/2017	2 ESO	14
5	7/Febrero/2017	3 ESO	14
6	7/Febrero/2017	4 ESO	14
7	8/Febrero/2017	4 ESO	14
8	8/Febrero/2017	1 ESO	14
9	8/Febrero/2017	2 ESO	14
10	8/Febrero/2017	3 ESO	14
11	14/febrero/2017	3 ESO	14
12	14/febrero/2017	1 ESO	14
13	14/febrero/2017	2 ESO	14
14	14/febrero/2017	4 ESO	15
15	15/febrero/2017	1 ESO / 2 ESO	15 (11 / 4)
16	15/febrero/2017	3 ESO / 4 ESO / 1 Bach.	15 (2 / 5 / 8)

El reparto total así como la distribución por curso y género se muestra en la siguiente tabla. Por un fallo con uno de los cuestionarios iniciales, no se recogió el género de 14 alumnos de 1º de la ESO como se refleja en la tabla. Como observamos, la cantidad de chicos y chicas en cada curso y en el total está aproximadamente proporcionada.

Curso	1 ESO	2 ESO	3 ESO	4 ESO	1 Bachillerato	Total
Femenino	23	21	32	22	10	108
Masculino	16	25	26	26	12	105
Género no recogido	14	0	0	0	0	14
Número de alumnos	53	46	58	48	22	227



Cuestionario sobre hábitos de juego

Como parte del cuestionario inicial del experimento, se incluían las siguientes preguntas sobre hábitos de juego que completaron todos los alumnos. De las respuestas a estas preguntas se obtiene la información analizada en los siguientes apartados del informe.

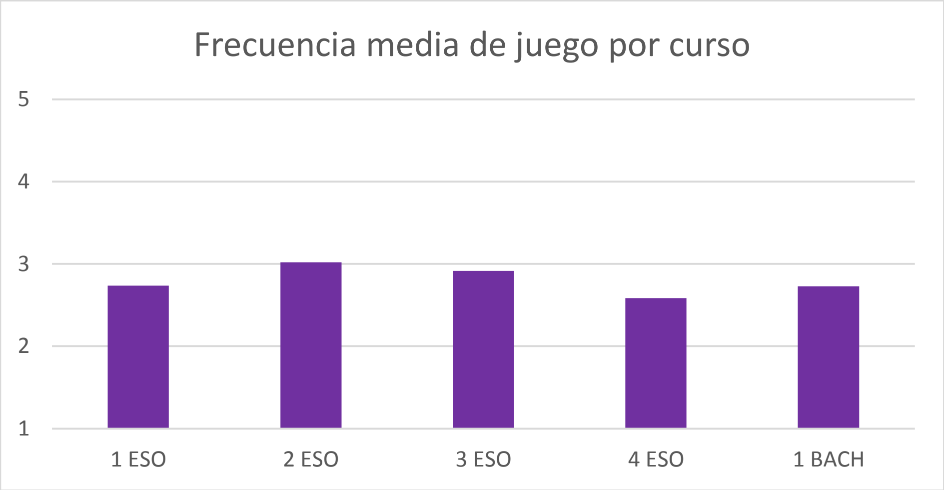
Contesta del 1 (nunca) al 5 (a diario) a las siguientes preguntas sobre tus hábitos de juego.

	Nunca (1)				A diario (5)
	1	2	3	4	5
H1 - ¿Con qué frecuencia juegas a videojuegos?					
H2 - ¿Con qué frecuencia juegas a videojuegos de disparos en primera persona (Call of Duty, Black Ops, Borderlands, Halo, Bioshock)?	1	2	3	4	5
H3 - ¿Con qué frecuencia juegas a videojuegos de aventura (Uncharted, Heavy Rain, Resident Evil, Assassin's Creed)?	1	2	3	4	5
H4 - ¿Con qué frecuencia juegas a videojuegos de cantar, bailar o tocar instrumentos (Guitar Hero, Sing Star, Just Dance)?	1	2	3	4	5
H5 - ¿Con qué frecuencia juegas a videojuegos de peleas (Tekken, Mortal Kombat, Street Fighter)?	1	2	3	4	5
H6 - ¿Con qué frecuencia juegas a videojuegos de inteligencia, preguntas/respuestas (Preguntados, Trivial, Brain Training)?	1	2	3	4	5
H7 - ¿Con qué frecuencia juegas a videojuegos de estrategia (Civilization, Age of Empires, Starcraft)?	1	2	3	4	5
H8 - ¿Con qué frecuencia juegas a videojuegos de deportes, carreras o simulación (FIFA, PES, NBA Live, Gran Turismo, Need for Speed)?	1	2	3	4	5
H9 - ¿Con qué frecuencia juegas a videojuegos como Super Mario, Mario Kart o Wii Sports?	1	2	3	4	5
H10 - ¿Con qué frecuencia juegas a videojuegos de internet multijugador masivos (World of Warcraft, RuneScape, League of Legends)?	1	2	3	4	5
H11 - ¿Con qué frecuencia juegas a videojuegos en el móvil o tablet (Clash Royale, Clash of Clans, Candy Crush)?	1	2	3	4	5

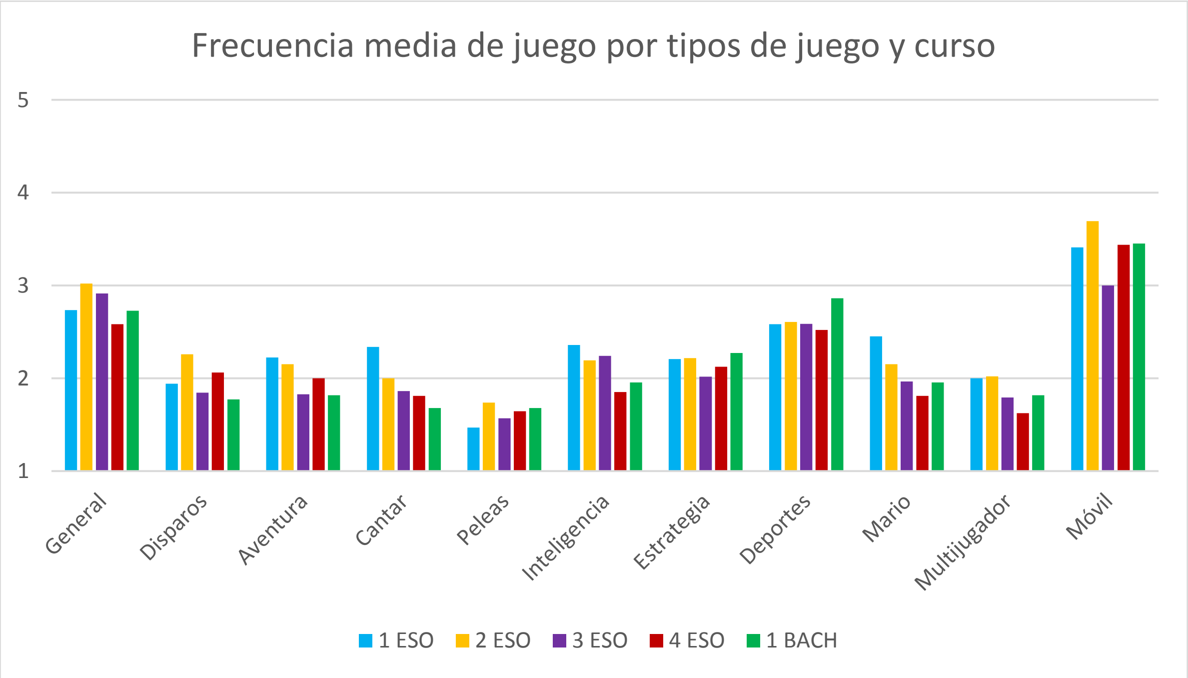
Hábitos de juego de los alumnos

En las gráficas siguientes, se muestran las frecuencias de juego en la escala seguida por el cuestionario, siendo 1 "nunca", 5 "a diario".

En primer lugar, obtenemos la frecuencia media de juego por cada curso, donde observamos que el curso en el que más juegan es Segundo de la ESO, reduciéndose esta frecuencia de juego en los dos cursos siguientes, y volviendo a subir ligeramente en Bachillerato:

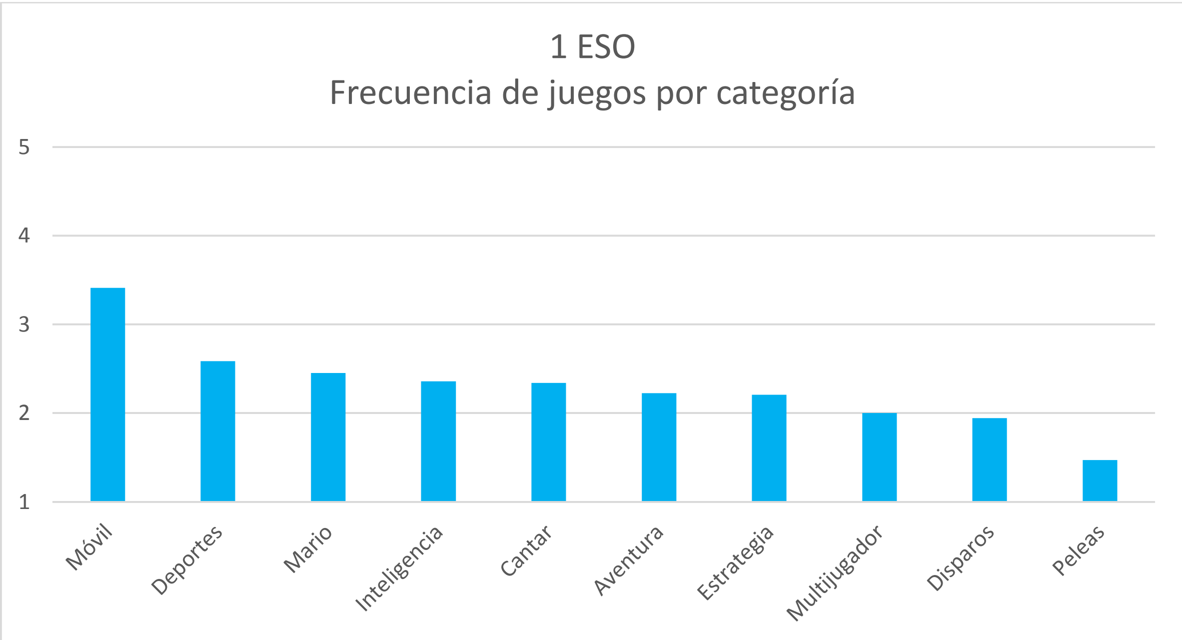


Si distinguimos por las distintas categorías de juegos, podemos observar la frecuencia de juego de cada categoría en cada clase, donde ya observamos que los juegos de móvil son jugados con mayor frecuencia en todos los cursos. En los siguientes apartados se analizan estas categorías de manera independiente para cada curso.



Primero de la ESO

En Primero de la ESO, la frecuencia de juegos de 1 "nunca" a 5 "a diario" para cada categoría, ordenadas de mayor a menor, es la siguiente:



Categoría de juegos de mayor a menor frecuencia	Ejemplos de juegos
Móvil o tablet	Clash Royale, Clash of Clans, Candy Crush
Deportes, carreras o simulación	FIFA, PES, NBA Live, Gran Turismo, Need for Speed
Mario	Super Mario, Mario Kart o Wii Sports
Inteligencia, preguntas/respuestas	Preguntados, Trivial, Brain Training
Cantar, bailar o tocar instrumentos	Guitar Hero, Sing Star, Just Dance
Aventura	Uncharted, Heavy Rain, Resident Evil, Assassin's Creed
Estrategia	Civilization, Age of Empires, Starcraft
Internet multijugador masivos	World of Warcraft, RuneScape, League of Legends
Disparos en primera persona	Call of Duty, Black Ops, Borderlands, Halo, Bioshock
Peleas	Tekken, Mortal Kombat, Street Fighter

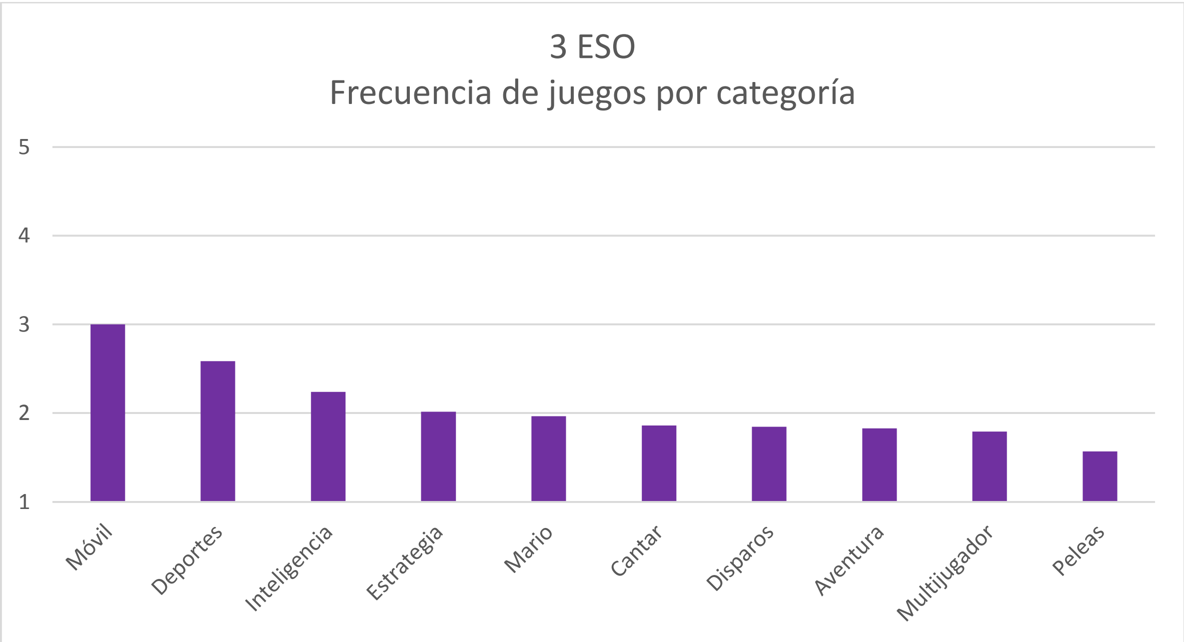
Segundo de la ESO

En Segundo de la ESO, la frecuencia de juegos de 1 "nunca" a 5 "a diario" para cada categoría, ordenadas de mayor a menor, es la siguiente:

Categoría de juegos de mayor a menor frecuencia	Ejemplos de juegos
Móvil o tablet	Clash Royale, Clash of Clans, Candy Crush
Deportes, carreras o simulación	FIFA, PES, NBA Live, Gran Turismo, Need for Speed
Disparos en primera persona	Call of Duty, Black Ops, Borderlands, Halo, Bioshock
Estrategia	Civilization, Age of Empires, Starcraft
Inteligencia, preguntas/respuestas	Preguntados, Trivial, Brain Training
Aventura	Uncharted, Heavy Rain, Resident Evil, Assassin's Creed
Mario	Super Mario, Mario Kart o Wii Sports
Internet multijugador masivos	World of Warcraft, RuneScape, League of Legends
Cantar, bailar o tocar instrumentos	Guitar Hero, Sing Star, Just Dance
Peleas	Tekken, Mortal Kombat, Street Fighter

Tercero de la ESO

En Tercero de la ESO, la frecuencia de juegos de 1 "nunca" a 5 "a diario" para cada categoría, ordenadas de mayor a menor, es la siguiente:



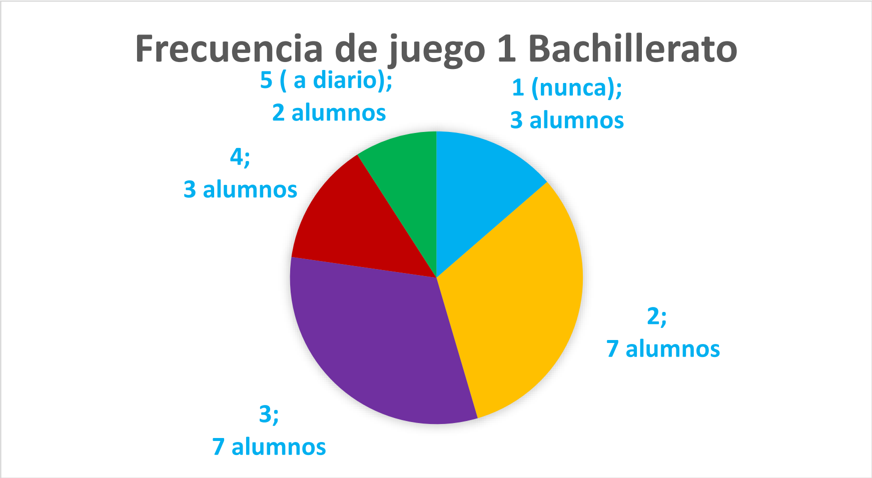
Categoría de juegos de mayor a menor frecuencia	Ejemplos de juegos
Móvil o tablet	Clash Royale, Clash of Clans, Candy Crush
Deportes, carreras o simulación	FIFA, PES, NBA Live, Gran Turismo, Need for Speed
Inteligencia, preguntas/respuestas	Preguntados, Trivial, Brain Training
Estrategia	Civilization, Age of Empires, Starcraft
Mario	Super Mario, Mario Kart o Wii Sports
Cantar, bailar o tocar instrumentos	Guitar Hero, Sing Star, Just Dance
Disparos en primera persona	Call of Duty, Black Ops, Borderlands, Halo, Bioshock
Aventura	Uncharted, Heavy Rain, Resident Evil, Assassin's Creed
Internet multijugador masivos	World of Warcraft, RuneScape, League of Legends
Peleas	Tekken, Mortal Kombat, Street Fighter

Cuarto de la ESO

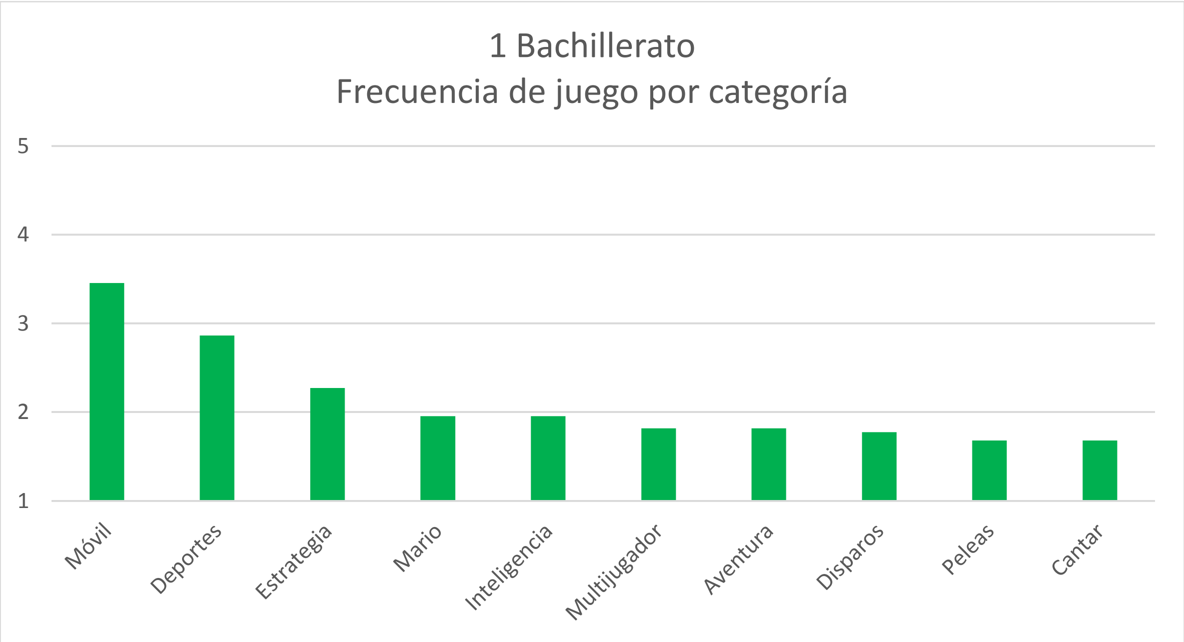
En Cuarto de la ESO, la frecuencia de juegos de 1 "nunca" a 5 "a diario" para cada categoría, ordenadas de mayor a menor, es la siguiente:

Categoría de juegos de mayor a menor frecuencia	**Ejemplos de juegos**
Móvil o tablet	Clash Royale, Clash of Clans, Candy Crush
Deportes, carreras o simulación	FIFA, PES, NBA Live, Gran Turismo, Need for Speed
Estrategia	Civilization, Age of Empires, Starcraft
Disparos en primera persona	Call of Duty, Black Ops, Borderlands, Halo, Bioshock
Aventura	Uncharted, Heavy Rain, Resident Evil, Assassin's Creed
Inteligencia, preguntas/respuestas	Preguntados, Trivial, Brain Training
Cantar, bailar o tocar instrumentos	Guitar Hero, Sing Star, Just Dance
Mario	Super Mario, Mario Kart o Wii Sports
Peleas	Tekken, Mortal Kombat, Street Fighter
Internet multijugador masivos	World of Warcraft, RuneScape, League of Legends

Primero de Bachillerato



En Primero de Bachillerato, la frecuencia de juegos de 1 "nunca" a 5 "a diario" para cada categoría, ordenadas de mayor a menor, es la siguiente:



Categoría de juegos de mayor a menor frecuencia	Ejemplos de juegos
Móvil o tablet	Clash Royale, Clash of Clans, Candy Crush
Deportes, carreras o simulación	FIFA, PES, NBA Live, Gran Turismo, Need for Speed
Estrategia	Civilization, Age of Empires, Starcraft
Mario	Super Mario, Mario Kart o Wii Sports
Inteligencia, preguntas/respuestas	Preguntados, Trivial, Brain Training
Internet multijugador masivos	World of Warcraft, RuneScape, League of Legends
Aventura	Uncharted, Heavy Rain, Resident Evil, Assassin's Creed
Disparos en primera persona	Call of Duty, Black Ops, Borderlands, Halo, Bioshock
Peleas	Tekken, Mortal Kombat, Street Fighter
Cantar, bailar o tocar instrumentos	Guitar Hero, Sing Star, Just Dance

Conclusiones sobre los hábitos de juego

- El curso en el que juegan con mayor frecuencia es Segundo de la ESO. Esta frecuencia de juego va disminuyendo hasta Bachillerato donde aumenta ligeramente.

- En todos los cursos hay más alumnos que no juegan nunca a videojuegos que alumnos que juegan a diario.

- El número de alumnos que juegan a diario se mantiene en Primero y Segundo de la ESO (6 alumnos), aumenta en Tercero de la ESO (8 alumnos) y disminuye en Cuarto de la ESO y Bachillerato (tan sólo 2 alumnos).

- En todos los cursos, la categoría a la que más juegan es *Móvil o tablet*, que incluye ejemplos como *Clash Royale, Clash of Clans* o *Candy Crush*.

- En todos los cursos, la segunda categoría a la que más juegan es *Deportes, carreras o simulación*, que incluye ejemplos como *FIFA, PES, NBA Live, Gran Turismo, Need for Speed*.

- Cuanto mayores son los alumnos, con mayor frecuencia juegan a juegos de *Estrategia* como *Civilization, Age of Empires* o *Starcraft* (siendo la séptima categoría más jugada para Primero de la ESO hasta llegar a ser la tercera más jugada para Primero de Bachillerato).

- Los juegos de *Aventura, Inteligencia* y *Mario* son jugados con frecuencia media en todos los cursos, no observándose una relación clara entre la edad y la frecuencia con la que los juegan.

- Los juegos de *Disparos* se juegan con mucha frecuencia en Segundo de la ESO y también en Cuarto de la ESO, siendo menos frecuentes en el resto de los cursos.

- En todos los cursos, las categorías de *Peleas, Multijugador* o *Cantar* se juegan con una frecuencia baja.

Recepción por parte de los alumnos

El cuestionario posterior al juego incluía una pregunta opcional para que los alumnos pudieran dejar los comentarios que quisieran sobre el juego. A continuación se incluye una selección de los comentarios de los alumnos, señalando en cada caso el curso al que pertenecen:

- "El juego me ha parecido divertido y he aprendido cosas." (1º ESO)
- "A mí me ha gustado mucho, y espero que haya ese del cyberbulling, para que aprendamos más sobre él." (1º ESO)
- "Me ha gustado, me gustaría tener otro juego parecido." (1º ESO)
- "Me ha gustado mucho, es una buena forma de enseñar esto." (2º ESO)
- "Me ha parecido un juego muy bueno y parece una dinámica muy buena para aprender." (2º ESO)
- "Me ha servido mucho porque antes casi no tenía conocimientos de este tema así que gracias a esta actividad pues puedo en algún futuro salvar a gente." (2º ESO)
- "¡Me ha gustado mucho, porque es una forma fácil de aprender!" (3º ESO)
- "Muy buen recurso para poder aprender cómo actuar de primeros auxilios remplazando una charla, se aprende mucho más." (3º ESO)
- "La verdad es que me gusta el juego y me parece muy bien que los colegios den esta oportunidad a alumnos para instruirnos. " (4º ESO)
- "Estas actividades deberían de hacerse más" (4º ESO)
- "Me gustaría hacer más actividades de estas." (4º ESO)
- "Es fácil y simple, además de didáctico, ojalá hicieran más cosas de estas." (1º Bachillerato)

11. Appendix 3: Additional figures and table results

11.1. Cluster analysis

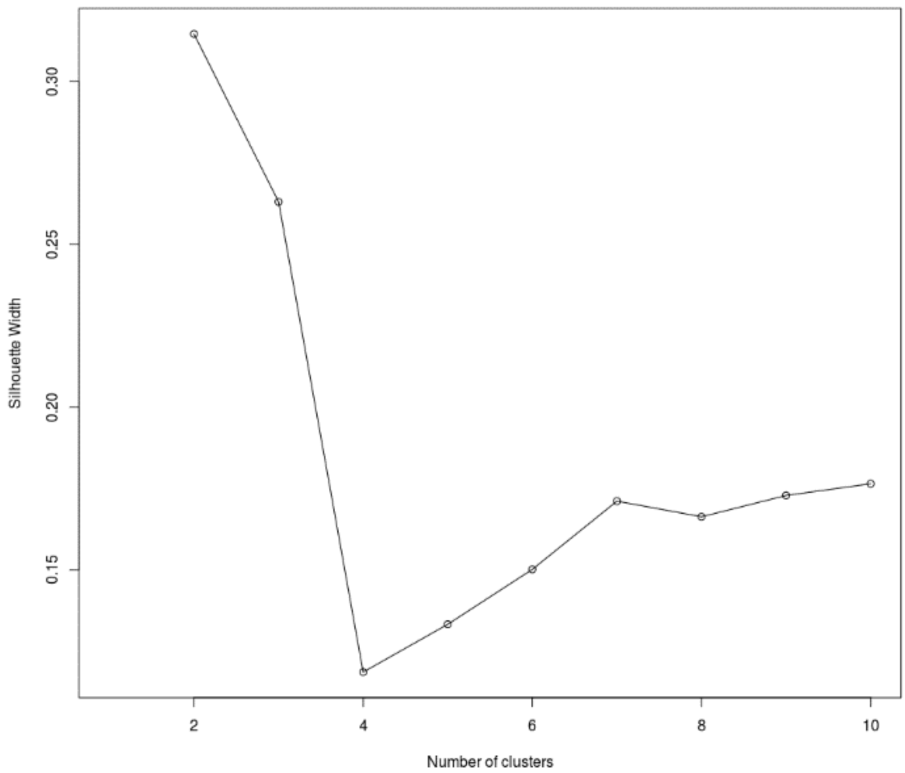


Figure 36. Adequate number of clusters for classification based on game habits

11.2. Players' learning with the game

Analyzing each question in the pre-test and post-test, there are four possibilities:

1. Players answered correctly in pre-test and post-test [Masters]
2. Players answered wrongly in pre-test and correctly in post-test [Learners]
3. Players answered correctly in pre-test and wrongly in post-test [Non-learners]
4. Players answered wrongly in pre-test and post-test [Unlearners]

Table 24 summarizes the number of masters, learners, non-learners and unlearners, as defined above, for each of the 15 first aid techniques questions in the questionnaires. High values of non-learners appear in questions 12 and 14.

Table 24 . Distribution of players based on correct and incorrect answers in pre-test and post-test

Question	Masters	Learners	Non-learners	Unlearners
1	137	30	16	15
2	55	50	30	63
3	60	48	37	53
4	50	99	8	41

5	110	58	14	16
6	64	39	18	77
7	10	55	11	122
8	133	45	6	14
9	80	50	24	44
10	78	58	15	47
11	77	48	20	53
12	104	24	46	24
13	103	38	25	32
14	54	18	64	62
15	133	38	13	14

11.3. PCA

Table 25. Standard deviation, proportion of variance explained and cumulative proportion of the first eight principal components.

	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8
Standard deviation	2.1313	1.9495	1.5628	1.3626	1.2601	1.1489	1.0983	1.0252
Proportion of Variance	0.1682	0.1407	0.0904	0.0687	0.0588	0.0488	0.0446	0.0389
Cumulative Proportion	0.1682	0.3090	0.3994	0.4682	0.5270	0.5759	0.6206	0.6595

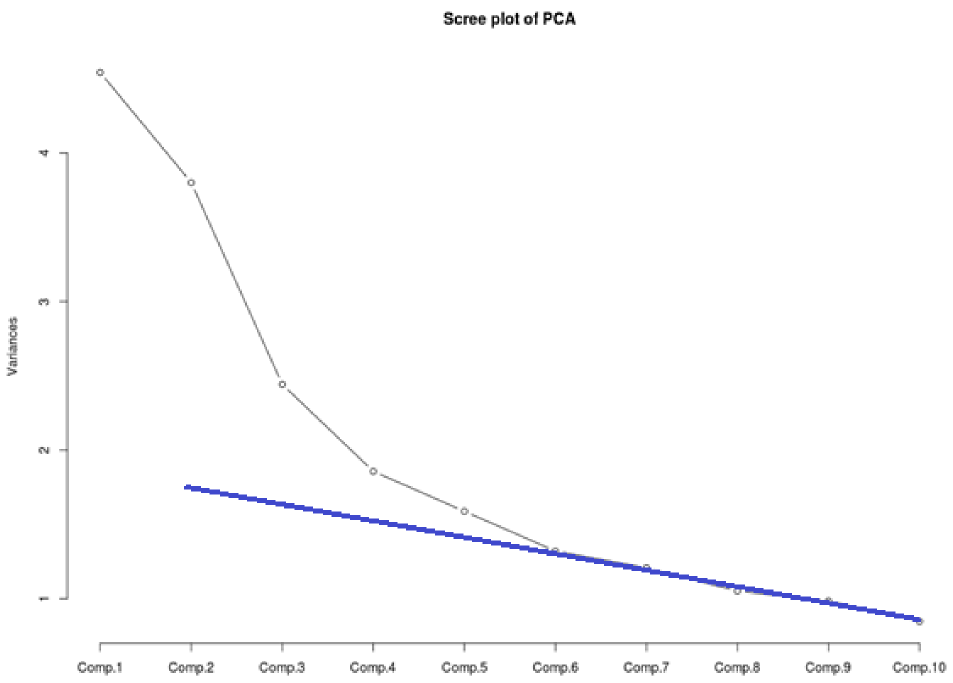


Figure 37. Scree plot of Principal Component Analysis that shows that five components may be adequate

Table 26. Coefficients of the variables in the five principal components

Variable	PC1	PC2	PC3	PC4	PC5
AGE	-0.20	-0.15	0.29	-0.36	-0.01
H1	0.73	-0.07	0.15	-0.01	0.05
H2	0.79	-0.02	0.10	-0.24	-0.17
H3	0.73	-0.12	0.20	0.06	-0.26
H4	-0.23	0.09	-0.28	0.52	-0.09
H5	0.69	-0.07	0.25	-0.12	-0.25
H6	-0.03	0.11	-0.05	0.68	0.00
H7	0.60	-0.04	0.24	0.25	0.01
H8	0.63	-0.15	0.16	0.14	0.06
H9	0.11	0.15	-0.01	0.73	-0.24
H10	0.71	-0.06	0.07	-0.06	-0.17
H11	0.49	-0.17	0.08	0.21	0.28
PREQSCORE	-0.17	0.10	0.37	0.00	0.44
POSTQSCORE	-0.24	0.13	0.28	0.05	0.14
score	-0.17	0.53	0.62	-0.09	-0.10
maxScoreCP	-0.03	0.64	0.04	0.01	-0.13
maxScoreU	-0.07	0.66	0.32	-0.05	-0.28
maxScoreCH	0.09	0.62	0.35	0.24	0.30
firstScoreCP	-0.09	0.30	0.28	-0.08	0.28
firstScoreU	-0.31	0.47	0.30	-0.10	-0.56
firstScoreCH	0.08	0.23	0.58	0.08	0.44
timesCP	0.23	0.70	-0.30	-0.08	-0.13
timesU	0.36	0.64	-0.37	-0.08	0.27
timesCH	0.13	0.53	-0.23	0.18	0.00
int_patient	0.39	0.60	-0.48	-0.19	0.13
int_phone	0.00	0.44	-0.32	-0.24	-0.01
int_saed	0.40	0.11	-0.33	-0.14	0.37

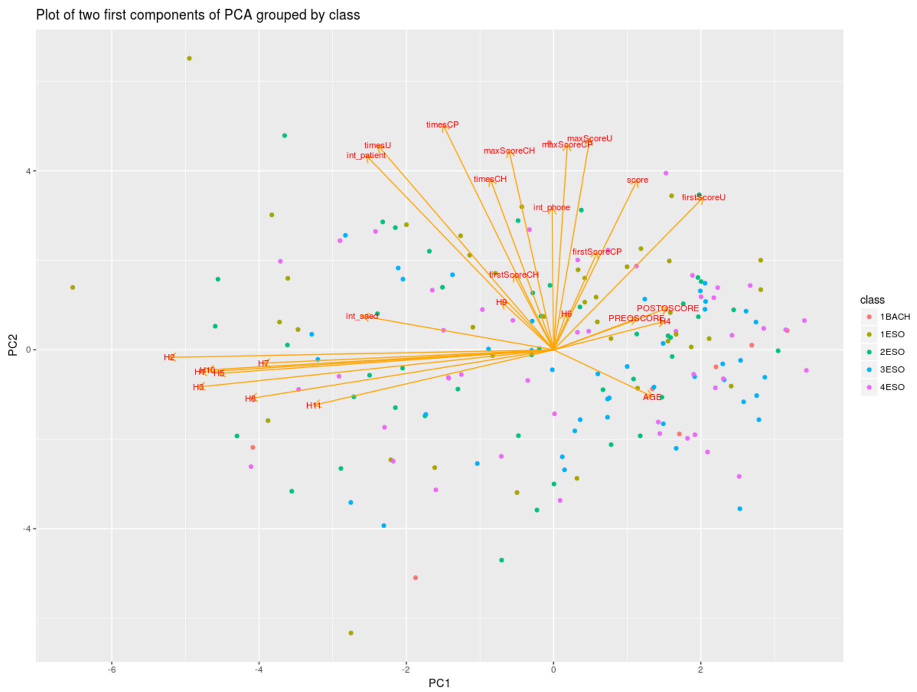


Figure 38. Plot of two first principal components grouped by class

11.4. Factor analysis

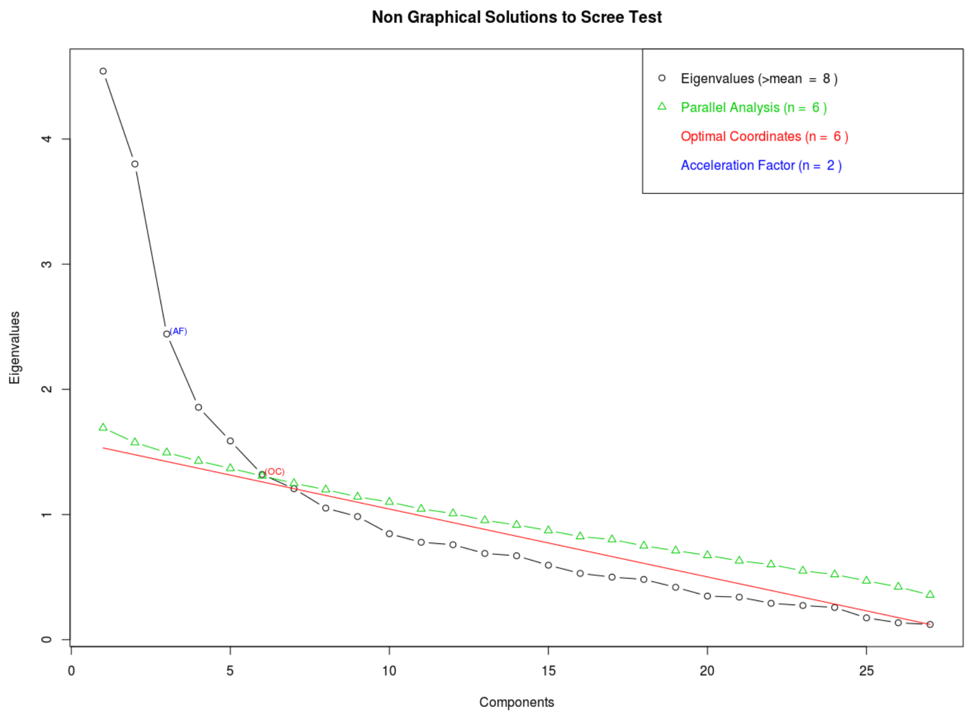


Figure 39. Results of parallel analysis, optimal coordinates and acceleration factor for the number of factors to retain

Table 27. Results of factor analysis for the numeric variables in the ten factors retained

Variable	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	**F10**
age										
PREQSCORE						0.35				
POSTQSCORE						0.98				
H1	0.67									
H2	0.77									
H3	0.78									
H4					0.53					
H5	0.74									
H6					0.55					
H7	0.59									
H8	0.57									
H9					0.63					
H10	0.68									
H11	0.42									
score			0.50	0.39			0.33			0.35
maxScoreCP			0.31				0.39		0.51	
maxScoreU			0.95							
maxScoreCH				0.68				0.41		
firstScoreCP							0.79			
firstScoreU			0.67							
firstScoreCH				0.97						
timesCP		0.56							0.77	
timesU		0.84						0.31		
timesCH								0.79		
int_patient		0.81								

int_phone		0.48								0.37
int_saed		0.43								

11.5. Regression trees

11.5.1. Using pre-test information

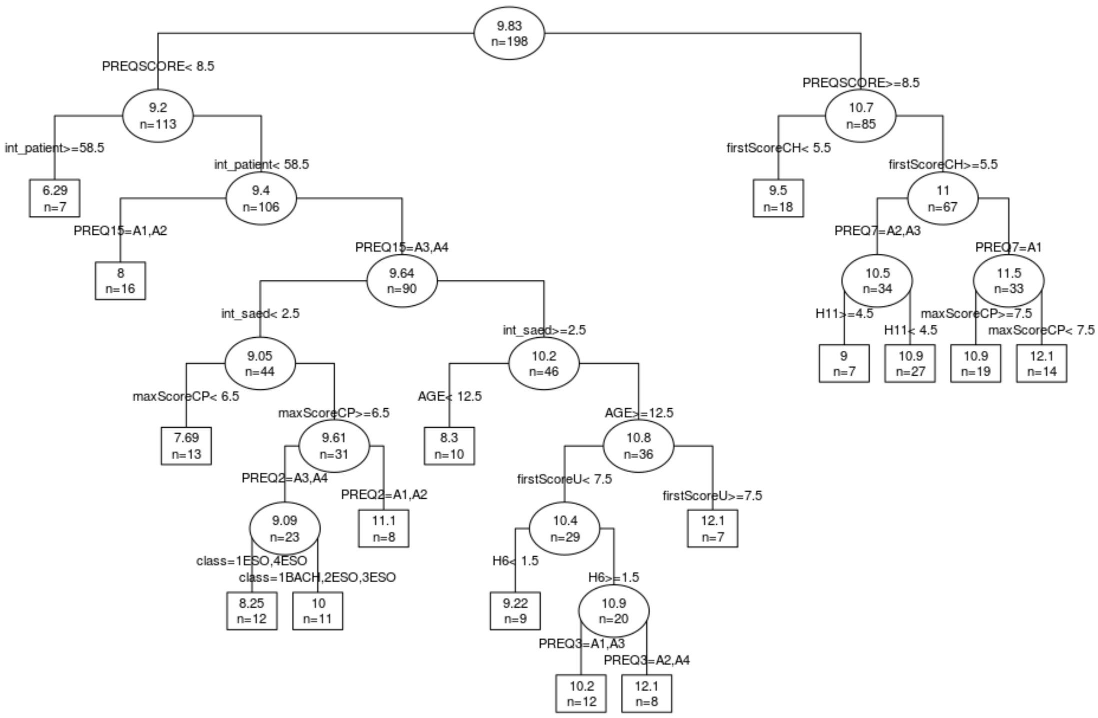


Figure 40. Tree for score prediction

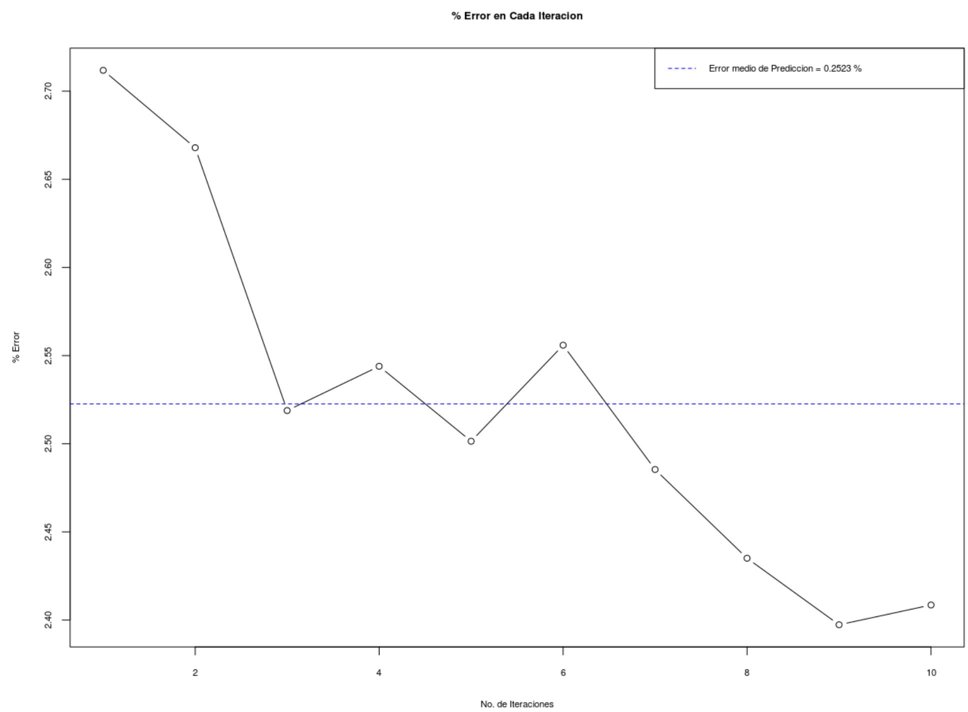


Figure 41. Error in cross validation iterations for score prediction tree

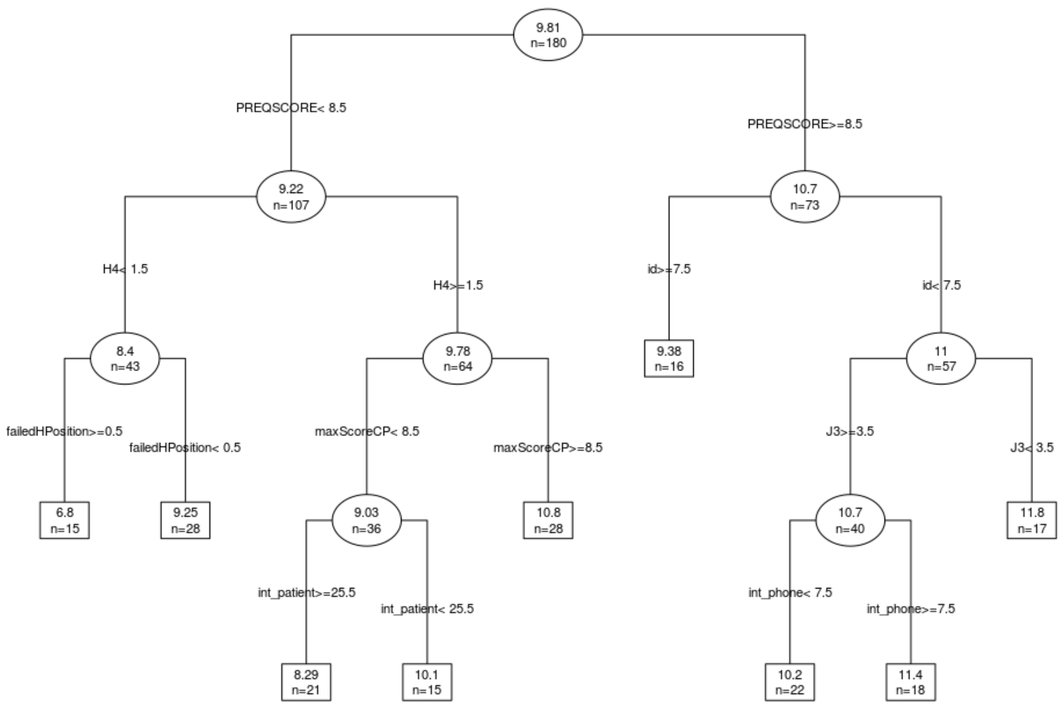


Figure 42. Tree for score prediction with min 30 observations per node and min 15 observations per terminal

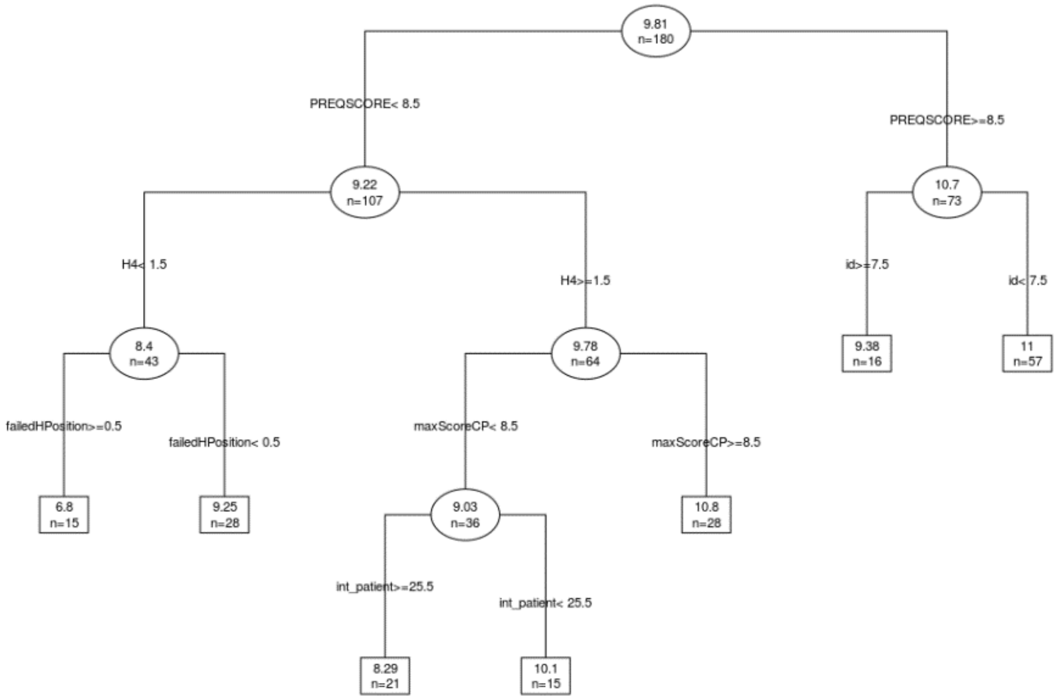


Figure 43. Tree for score prediction with cp factor of 0.02

11.5.2. Only with game interactions

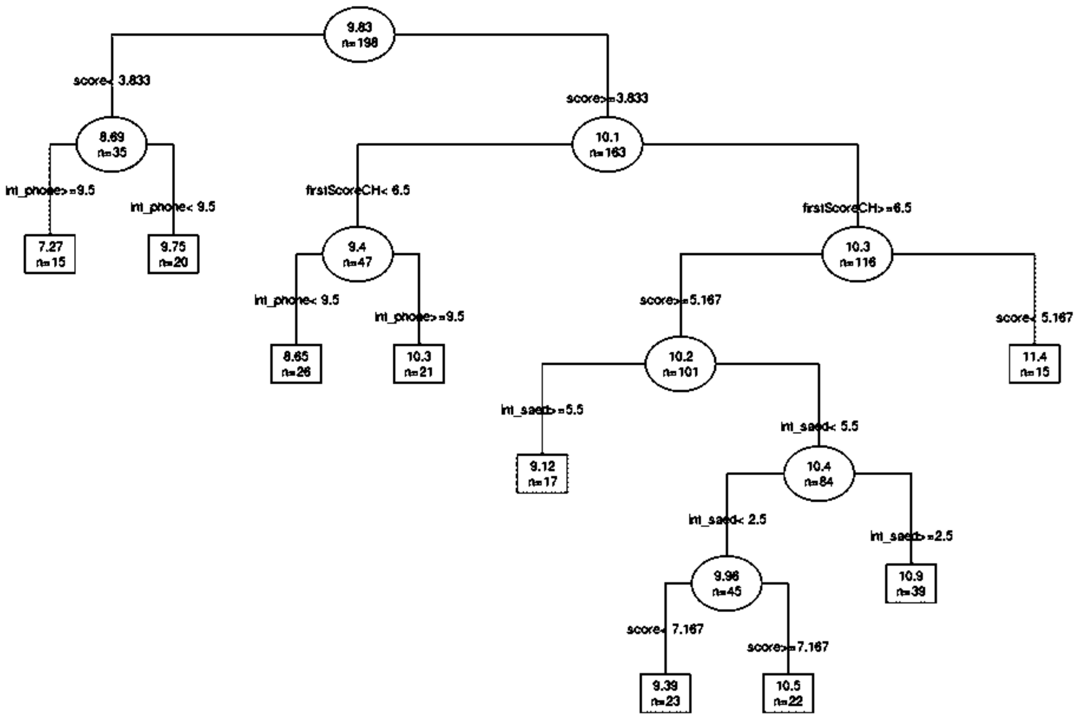


Figure 44. Tree for score prediction based only in game interactions variables with min 30 observations per node and min 15 observations per terminal

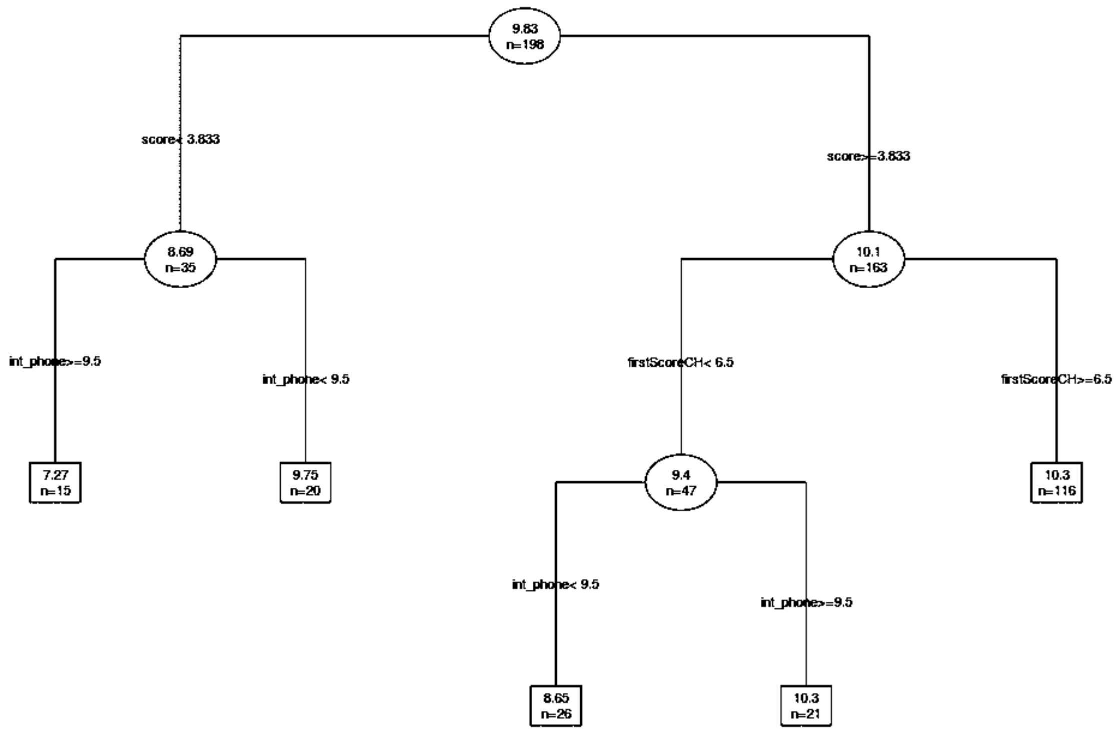


Figure 45. Tree for score prediction based only in game interactions variables with cp factor of 0.02

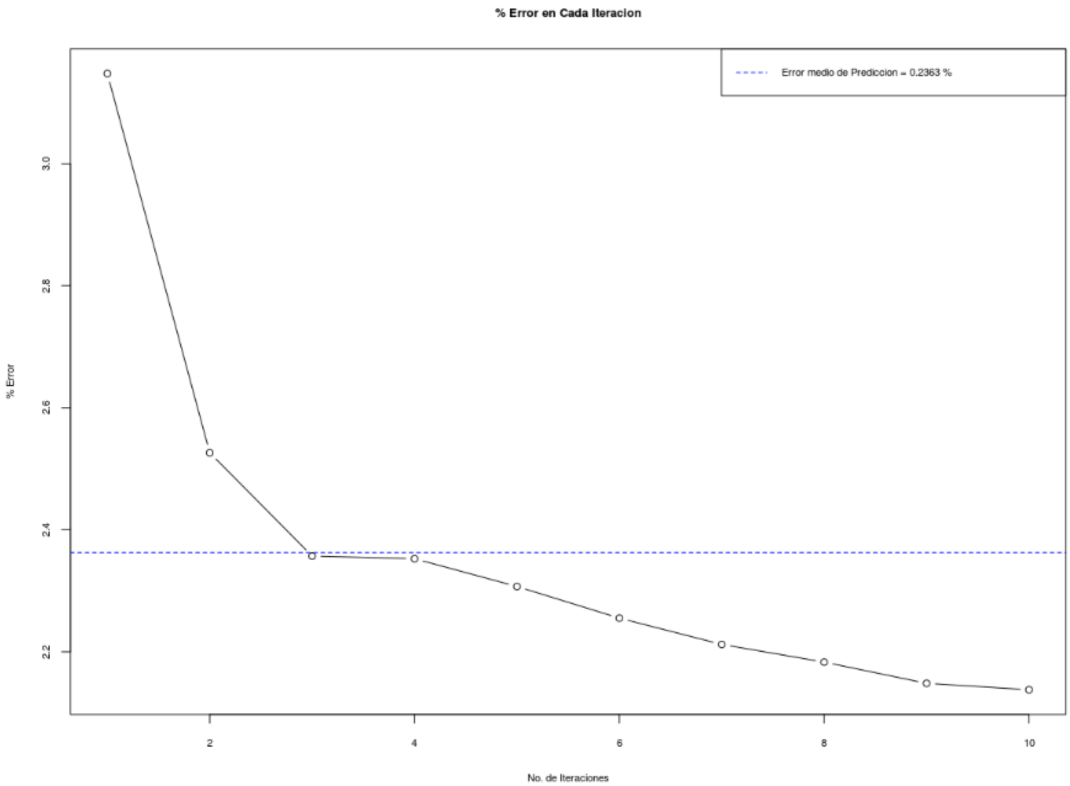


Figure 46. Error in cross validation iterations for score prediction tree based only in game interactions variables

11.6. Classification trees

11.6.1. With pre test information

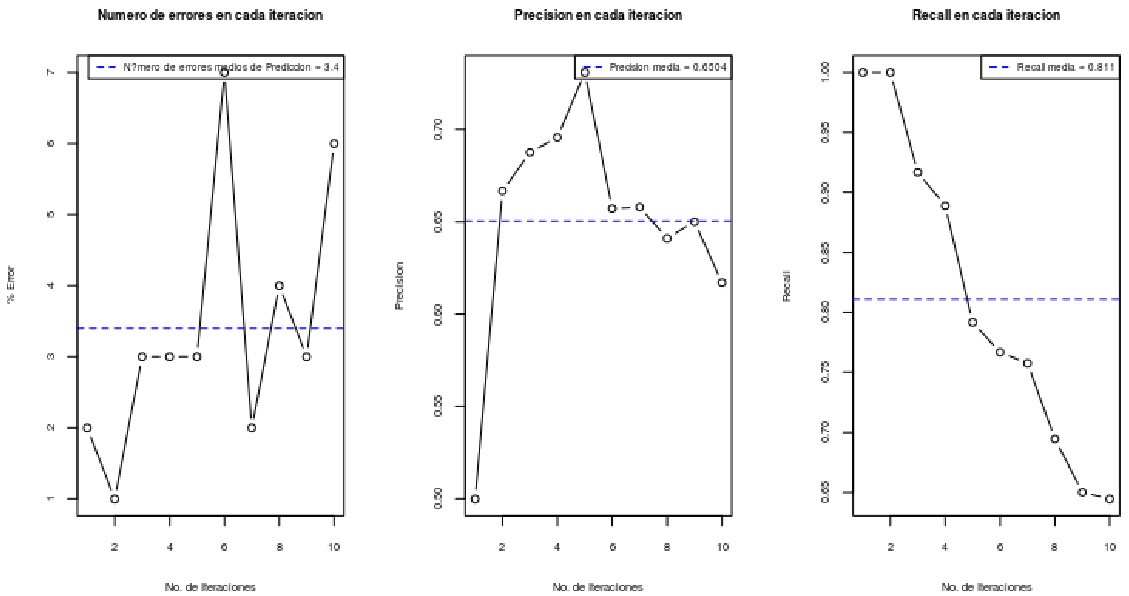


Figure 47. Number of errors, precision and recall in each cross validation iteration on balanced data

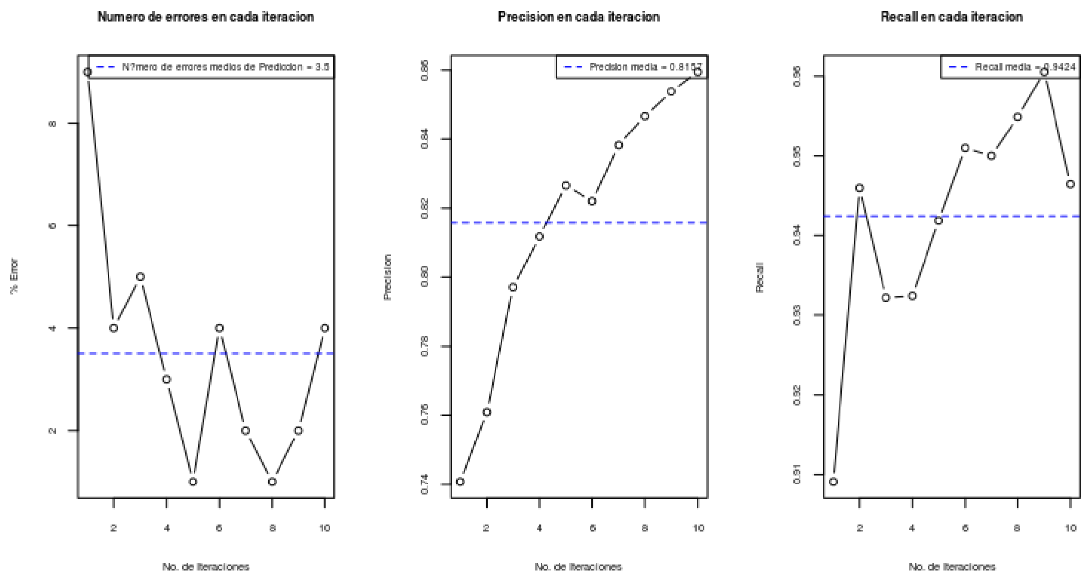


Figure 48. Number of errors, precision and recall in each cross validation iteration in original imbalanced data

11.6.2. Using only game interactions information

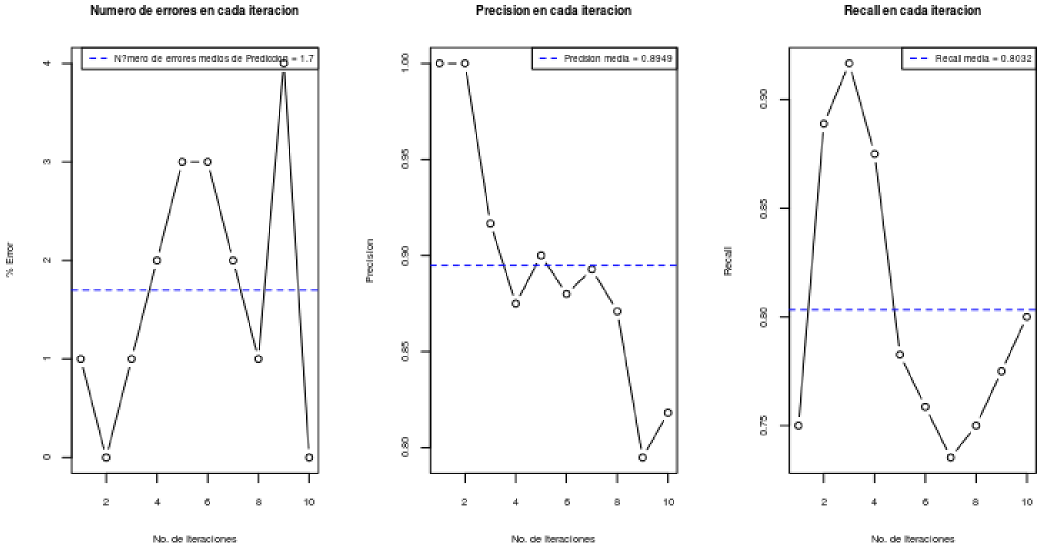


Figure 49. Number of errors, precision and recall in each cross validation iteration on balanced data for pass fail prediction based only in game interaction variables

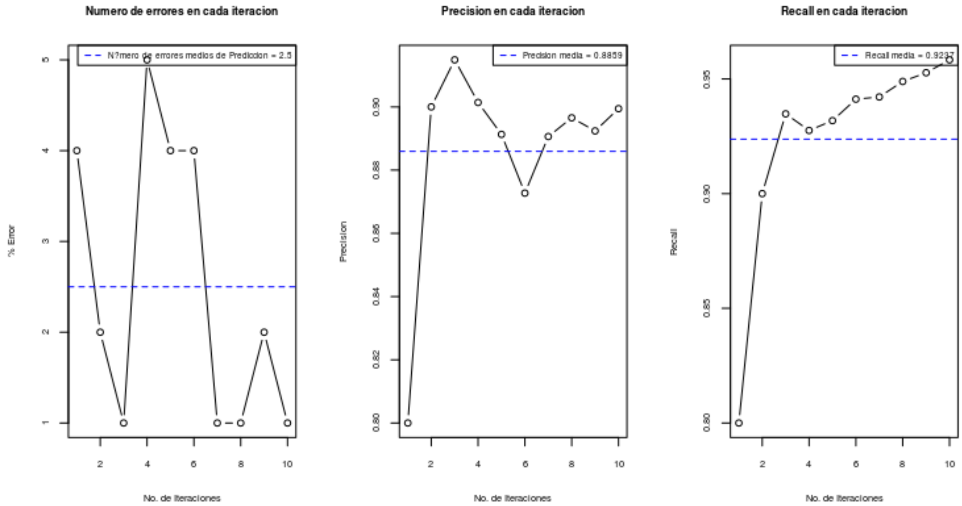


Figure 50. Number of errors, precision and recall in each cross validation iteration on balanced data for pass fail prediction based only in game interaction variables

11.7. Logistic regression

11.7.1. Using pre test information

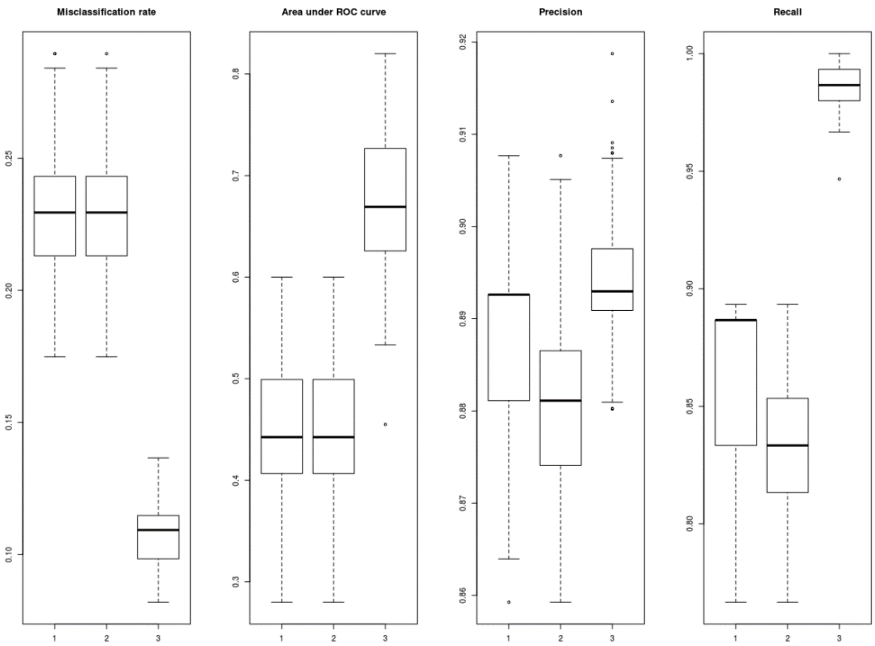


Figure 51. Missclassification rate, area under ROC curve, precision and recall for the logistic regression models with cross validation on balanced data

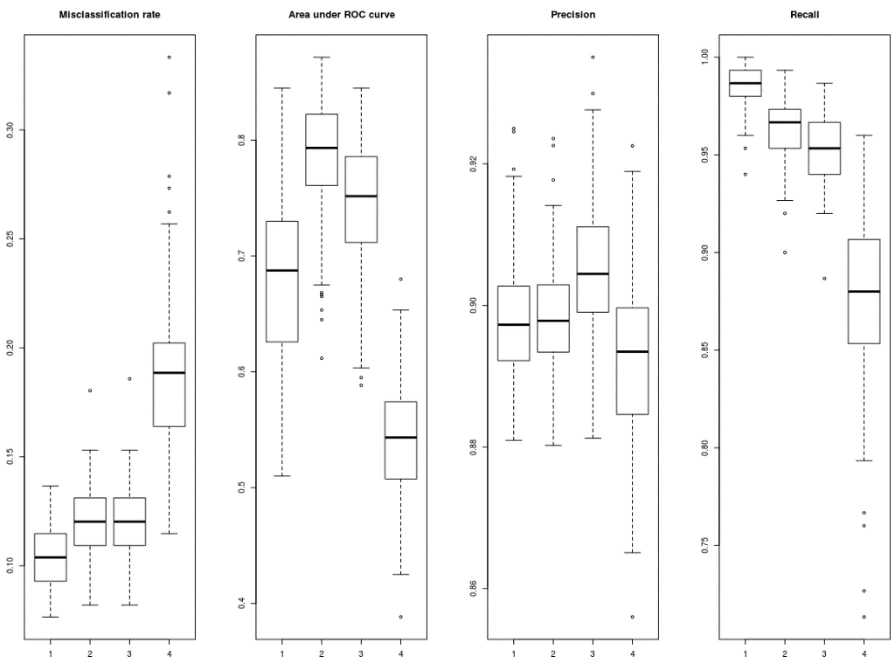


Figure 52. Missclassification rate, area under ROC curve, precision and recall for the logistic regression models with cross validation on original imbalanced data

11.7.2. Using only game interactions information

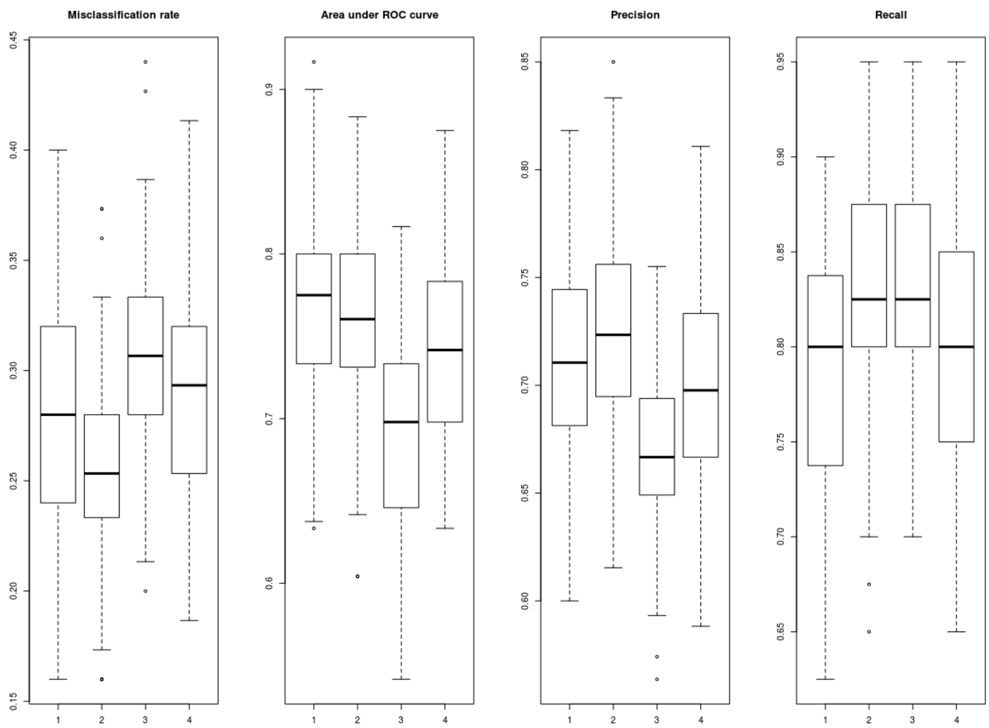


Figure 53. Results for the four logistic regression models with cross validation on balanced data based only in game interactions variables

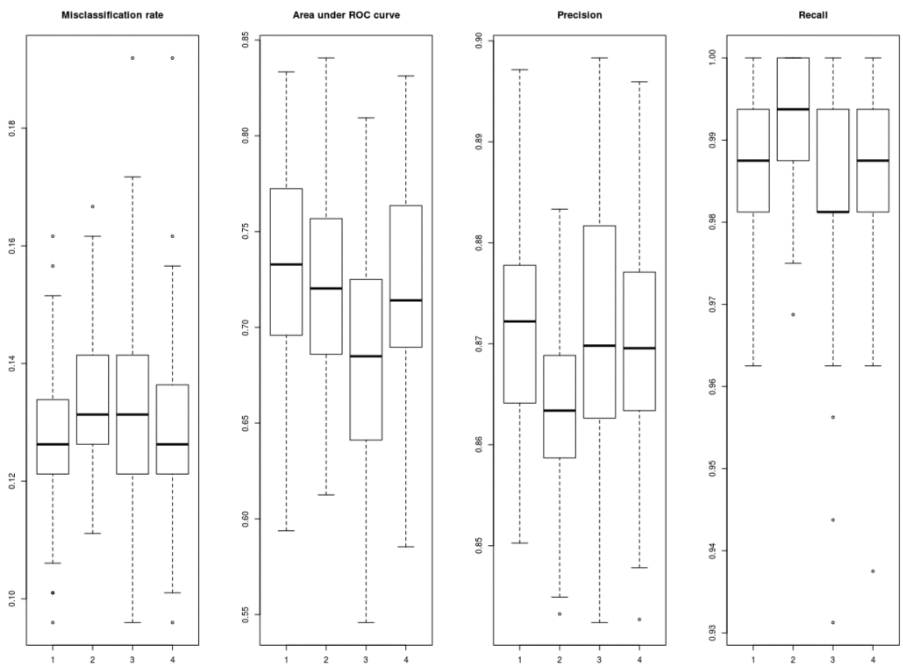


Figure 54. Results for the four logistic regression models with cross validation on original imbalanced data based only in game interactions variables

11.8. Analysis of number of game interactions

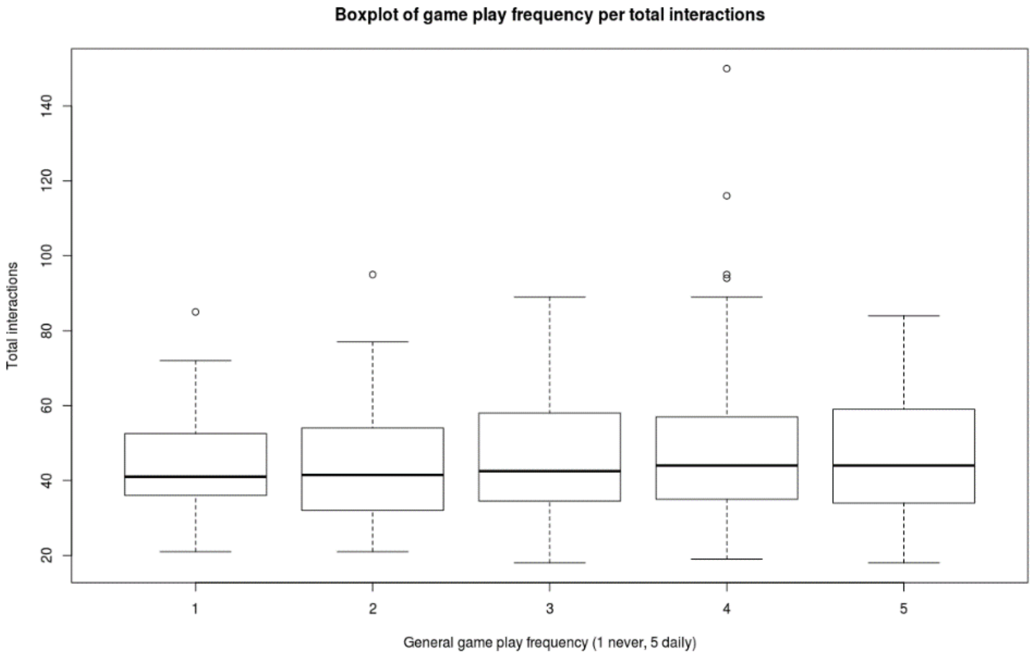


Figure 55. Total interactions per game play frequency

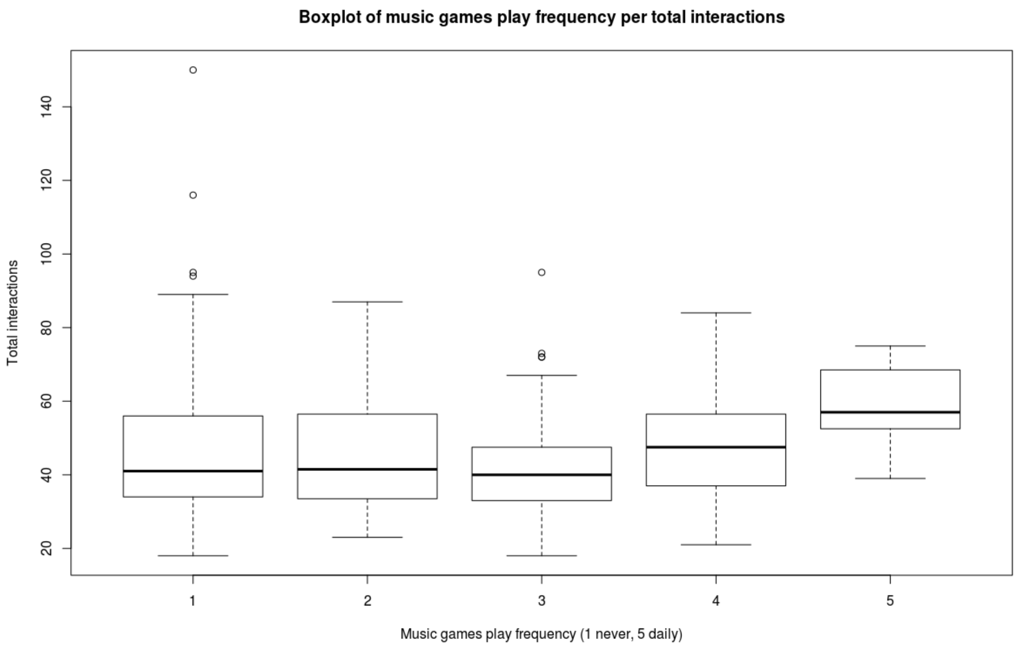


Figure 56. Total interactions per music games play frequency

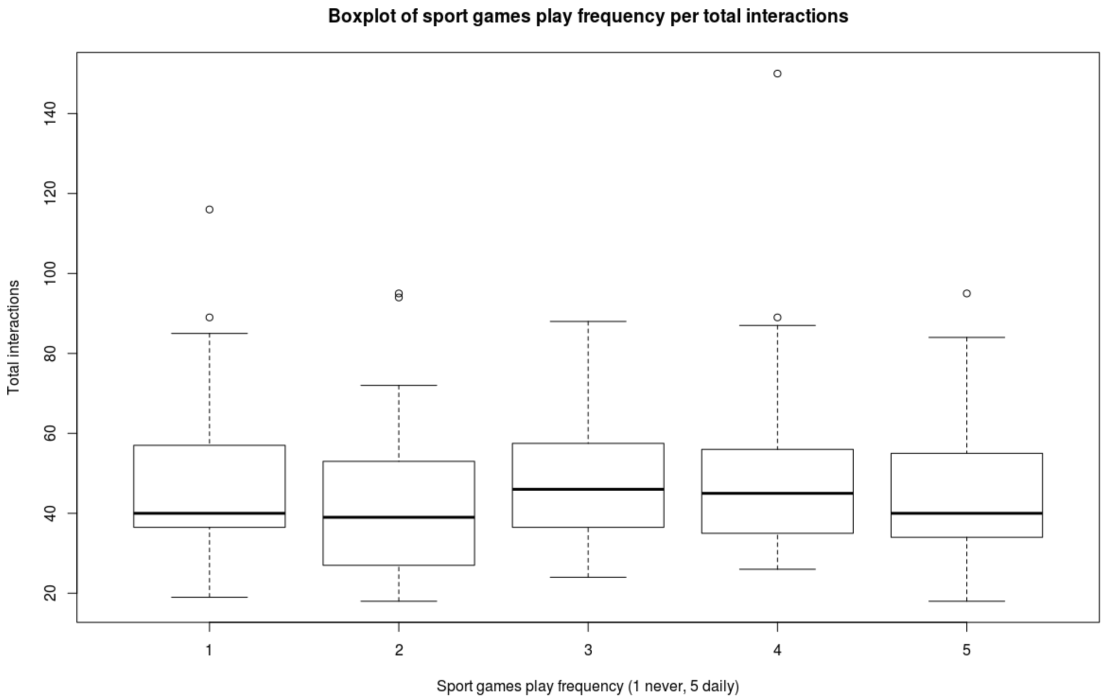


Figure 57. Total interactions per sport games play frequency

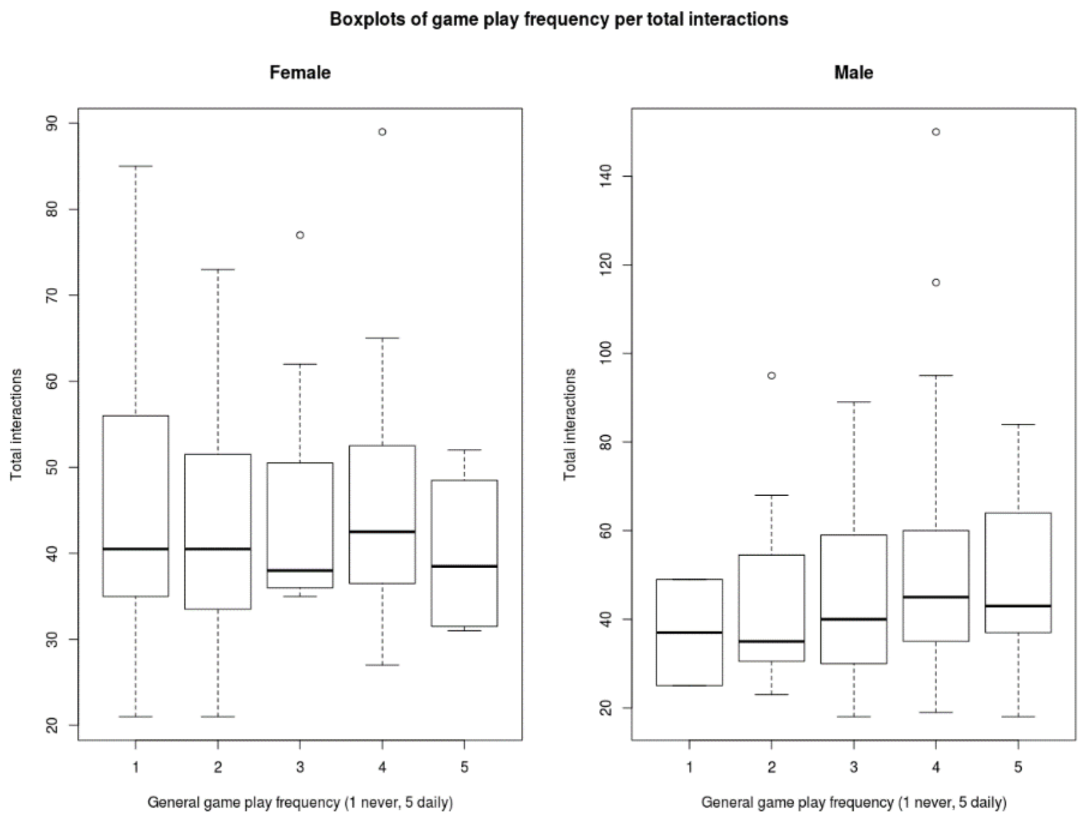


Figure 58. Total interactions per games play frequency for female (left) and male (right) players

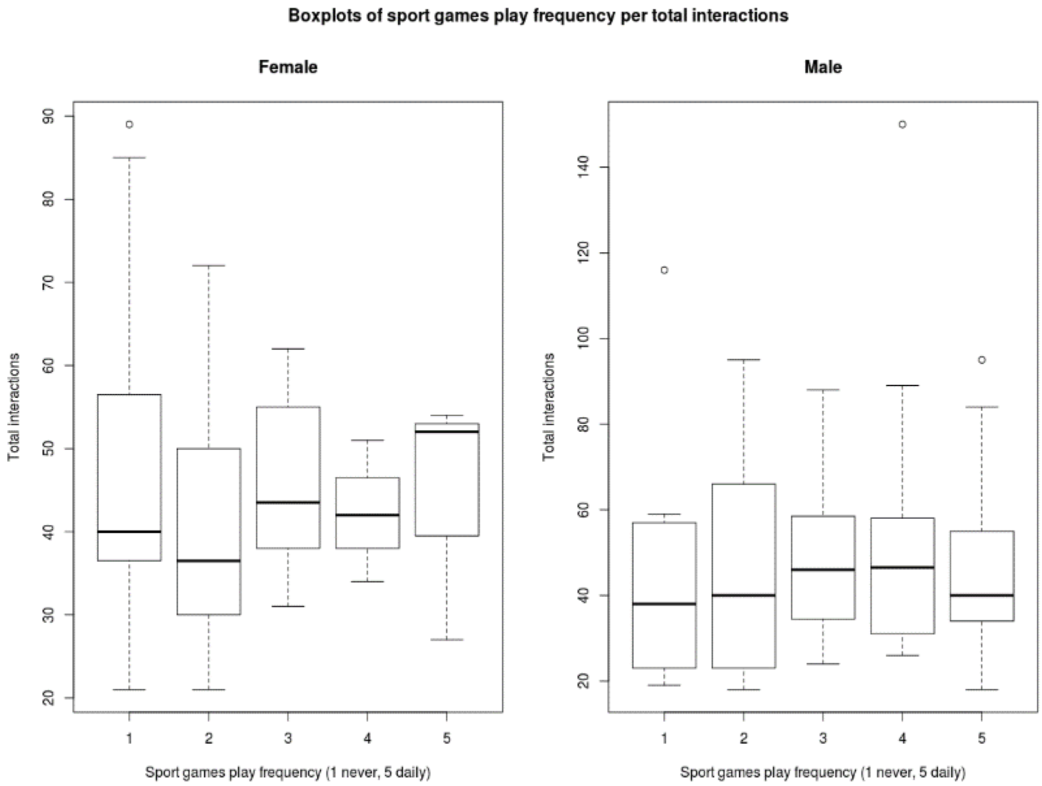


Figure 59. Total interactions per sport games play frequency for female (left) and male (right) players