The original article has been published in Telematics and Informatics with DOI: https://doi.org/10.1016/j.tele.2021.101568

Creating awareness on bullying and cyberbullying among young people: validating the effectiveness and design of the serious game *Conectado*

Abstract

Serious games have proven to be educational tools with numerous positive effects, capable of promoting learning, changing behavior, or improving training and skills development, among other effects. Serious games can also be applied to address and prevent social problems. This study describes the experiments and analyses we have carried out to test the effectiveness of *Conectado*, a serious game created to prevent bullying and cyberbullying in schools by increasing awareness of the problem in players. We have used the results from these experiments to answer different research questions regarding the game's acceptance and effectiveness, validating several characteristics of the game design. We have also studied the influence of players' characteristics on the effect of the game, as well as the relationship between the in-game behaviors of players and their degree of awareness. We have verified a positive effect of the game in terms of an increase of awareness regarding bullying and cyberbullying for all target users, and have provided further insight into how this increase is related to different players' characteristics and behaviors.

Keywords: Serious games; Improving classroom teaching; Secundary Education; Learning Analytics; Bullying

1. Introduction

Bullying is a social problem that is present all over the world, regardless of culture or country. With the extensive use of the internet and technology in daily life, cyberbullying has also become increasingly prevalent and relevant. Many studies have been carried out to understand the characteristics of (cyber)bullying and find methods to prevent it (Kowalski, Giumetti, Schroeder, & Lattanner, 2014; Menesini & Salmivalli, 2017; Zych, Ortega-Ruiz, & Del Rey, 2015). Both bullying and cyberbullying are highly prevalent among young people and can seriously affect them in many ways: causing psychological and physical effects, and manifesting themselves through changes in behavior, difficulties in adapting or in maintaining attention. They can also trigger psychological disorders that accompany the victim throughout their adult life and, in the most severe cases, lead to suicidal ideation and even documented suicides (Beran & Li, 2007; Carr-Gregg & Manocha, 2011).

Because of this high prevalence and the devastating effects (cyber)bullying can have on victims, it is crucial to find, develop and test tools that help society to address these problems at all phases, including detection, prevention, and fighting the aggression when it has already started. Among the resources available, some are established at school level such as prevention campaigns (Gaffney, Farrington, Espelage, & Ttofi, 2019) or training talks, while other resources are geared towards families and teachers, including online resources such as best-practices guides or initiatives such as pantallasamigas.net and stopbullying.gov. There are also other approaches, such as chats and telephone numbers, to help victims and families (Ballesteros, Santiago Pérez de Viñaspre, Díaz, & Toledano, 2018).

We consider that serious games are one of the most promising multimedia resources, both as (cyber)bullying prevention and detection tools. Serious games have been applied for educational purposes in multiple fields such as medicine, education, and research (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Wattanasoontorn, Boada, García, & Sbert, 2013). Video games have certain characteristics that make them useful tools for education, and which are also especially useful to improve the prevention or detection of (cyber)bullying. Games allow players to safely experience risky situations through simulations and controlled environments, providing instant feedback to in-game actions. The combination of storytelling and feedback, in turn, keeps players engaged and immersed, breaking their usual barrier of 10 minutes of attention (Weinschenk, 2011).

This study presents *Conectado* a serious game to raise awareness of bullying and cyberbullying, and the experiments carried out to validate it. The complete game validation has been performed using pre-post questionnaires to analyse the effectiveness of the game at increasing the awareness of (cyber)bullying, and collecting player-game interaction data to provide additional evidence-based validation. Game learning analytics techniques have been used to further validate some aspects of the

Cite: Calvo-Morata, A., Alonso-Fernández, C., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2021). Creating awareness on bullying and cyberbullying among young people: validating the effectiveness and design of the serious game The Videogame. *Telematics and Informatics*, *60*(January), 101568. https://doi.org/10.1016/j.tele.2021.101568

game design, such as gameplay time.

2. Related Work

Studies have shown the effectiveness of serious games to prevent bullying, sometimes as part of prevention programs (Garaigordobil & Martínez-Valderrey, 2015), and in other cases as support tools for teachers or students (Bradley & Kendall, 2019; Hall, Jones, Paiva, & Aylett, 2009). These studies, however, have several limitations, and it is necessary to continue studying game effectiveness. Among these limitations, we find that many experiments have been carried out with a low number of users, and most of the serious games studied are no longer available. Hence, it is challenging to replicate or compare their results with those of new studies. Additionally, most of those video games focused specifically on either bullying or cyberbullying, although it is common for both to occur simultaneously at schools. Moreover, most studies use a particular approach for (cyber)bullying prevention, focusing on different aspects such as empathy, awareness, change of behavior, or teaching how to detect the situations. (Álvarez-Bermejo, Belmonte-Ureña, Martos-Martínez, Barragán-Martín, & del Mar Simón-Marquez, 2016; DeSmet et al., 2016; Lievense, Vacaru, Liber, Bonnet, & Sterkenburg, 2019; McEvoy, Oyekoya, Ivory, & Ivory, 2016; Raminhos et al., 2016). This variety of approaches lead us to think that there is little evidence of the effectiveness of each specific approach. For example, one experiment may focus on empathy and be targeted towards 12-14 year olds, but this leaves a gap regarding other strategies in this age range, or the same approach on other age ranges. The review (Calvo-Morata, Alonso-Fernández, Freire, Martínez-Ortiz, & Fernández-Manjón, 2020) presents other serious games with scientific publications for the prevention and detection of bullying and cyberbullying.

Regarding how these studies are carried out, a majority of them use questionnaires to get the experimental data, and very few collect in-game player information to analyse how users interact with the game and whether this affects the effectiveness of the tool. This broader use of questionnaires is common in all serious game studies (Calderón & Ruiz, 2015), although recent research is introducing use of Game Analytics (El-Nasr, Drachen, & Canossa, 2013) or Game Learning Analytics to provide a richer insight into players' actions in the game (Alonso-Fernández, Calvo-Morata, Freire, Martínez-Ortiz, & Fernández-Manjón, 2019). The use of these techniques can help not only to evaluate the effectiveness of the game, but also to provide contrasting data to those obtained by questionnaires, thus helping to validate the design and ensure that the hypotheses of how the game is played and its effects are correct.

3. Materials and methods

We have developed and validated a serious game to be used as a tool for teachers to raise awareness on (cyber)bullying at school and to create empathy towards the victims. The evaluation has been performed through multiple experiments focused on validating different characteristics of the game and its effectiveness. During these experiments, the developed video game has been applied in various high schools from different regions of Spain.

3.1. Research questions

The experiments carried out as well as the analyses performed aim to answer the following research questions:

- RQ1. Is *Conectado* well accepted among students?
- RQ2. How long does it take for players to complete the game? In particular, are they able to complete the game in the average time available in one school session?
- RQ3. Does playing *Conectado* produce a positive effect on players' awareness of bullying and cyberbullying?
- RQ4. Does playing *Conectado* produce a different effect depending on players' previous use of social networks?
- RQ5. Does playing *Coenctado* produce a different effect depending on players' previous experience with cyberbullying?
- RQ6. Does playing *Conectado* produce a different effect depending on players' age or gender?
- RQ7. Does players' previous awareness affect the way players interact with the characters of *Conectado*?
- RQ8. Are there gender differences in the way players interact with *Conectado*?
- RQ9. Are players' interactions, decisions and behaviors in *Conectado* related to their change in awareness?

All these questions are answered in Section 4, the results section.

3.2. Conectado, the game

Conectado is a first-person point&click 2D graphic adventure where players take the role of a (cyber)bullying victim. Note that students are not warned beforehand that they are going to play this role. To increase the identification of players with their in-game characters, players are not graphically represented as avatars in the game. The player starts in a new school and has to make friends. However, every day more and more in-game characters will begin to (cyber)bully the player in different ways. Players can make different decisions when interacting with other game avatars to address the (cyber)bullying. For example, having a good relationship with their parents by talking with them about what happens at school, asking for help from the teacher when they have problems, avoiding confrontation with bullies, not sharing their social network password, or being nice to other classmates score positively. However, no choice will lead to effectively solving the problem before the fifth and last day of the story. The decisions and freedom of movement through the different scenes are designed to make players feel in control of the game. Only a few of those decisions have a significant impact on the story (making players reach one of the three possible game endings), since we want to combine user choices with the educational requirement that all players should have similar in-game experiences that the teacher can work with and elaborate into a class-wide discussion afterward. Apart from decision-making and dialogue, *Conectado* has four simple mini-challenges at the end of each of the first four days. These mini-games are represented as nightmares, and are designed to continuously increase in difficulty so that players cannot overcome them. The mini-challenges are designed to elicit feelings of helplessness and frustration in the players, and to convey that cyberbullying aggression is 24/7. The game is designed to

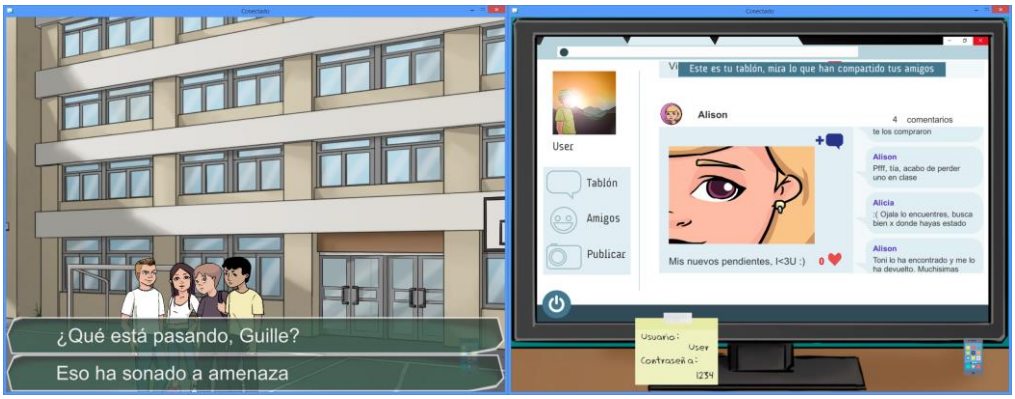

Figure 1. Conectado screenshots with the school and the in-game simulated social network.

show (cyber)bullying at school realistically and to bring the player closer to the feelings of frustration, helplessness, sadness, and anger that victims of such (cyber)bullying may experience. It also conveys the key idea to the player that bullying can only be overcome by asking for external help (e.g., family, teachers) and never by remaining silent.

Conectado has been developed to serve as a tool for teachers to use in class, and designed for students from 12 to 17 years old. Once the students have completed the game, their teacher can conduct a discussion and reflection session based on their shared experience in the game. *Conectado* is a free and open-source project available on GitHub. The video game is designed to last about 35 minutes and can be downloaded at *(URL Blinded for Review)*. A teachers' guide is also available, explaining the events in the game and giving ideas on how to use this information to conduct a discussion and reflection session. Although currently available only in Spanish, we are working on publishing the corresponding English version. Our main goal with this research, including both the video game and its user guide (both free and openly available online) is to provide a free resource, versatile, easy to use, and which can be complemented with other tools used by schools.

3.3. Data acquisition

In order to measure the effectiveness of *Conectado*, its acceptance in the classroom, and ultimately answer the research questions, different data have been collected during the experiments. The experiments were conducted in those schools that contacted the group after promoting the study at various

teacher events and activities related to technological innovations in education. Apart from the age of the students, the type and age of the computers and the mobility limitations of the research group, no other inclusion or exclusion criteria were applied. The two main sources were questionnaires and players' interactions within the game, which were sent as traces to a game analytics server. Before carrying out the experiments, we informed each of the schools about the experiment, and their management team reviewed and approved the use of the game as a school educational activity, signing an informed consent document which described the data we intended to collect, including plans for anonymization, retrieval, and later use. Both questionnaires and interaction data were anonymized, and linked to each other only by a unique identifier that was randomly assigned to each student at the beginning of the session. Once the sessions are over, particular students can no longer be traced back to their data again, as no student-to-identifier mapping is retained. The questionnaires used during the experiments included:

- Demographic questions, eliciting players' age and gender. This information of the players that have participated in the different experiments is useful to analyse later if these variables affect the previous awareness of the users and the effect that the game has on them.
- Bullying and cyberbullying awareness test, an adaptation and combination of four different questionnaires: CUVE3 (Álvarez-García, Nuñez-Pérez, & González, 2013), ECIP-Q, EBIP-Q (Ortega-Ruiz, Del Rey, & Casas, 2016) and Cyberbullying test (Garaigordobil & Aliri, 2013). This adaptation presents a high internal consistency with a Cronbach's Alpha of 0.95. The questionnaire is composed of 18 items that are evaluated on a Likert-type scale from 1 (Strongly Disagree) to 7 (Strongly Agree). 11 of the items assess awareness of cyberbullying, and the other 7 assess awareness of bullying. We interpret the average questionnaire score of each player as that player's level of awareness. The players fill in this test before and after playing, and the comparison of both scores shows the effect of the game on the player.
- Cyberbullying test, a questionnaire to measure the level of cyberbullying in the classroom for the three roles (observer, victim and aggressor). The test is made up of a total of 45 items, 15 for each of the roles, rated on a Likert-type scale from 1 (Never) to 4 (Daily). Its objective is to study, in the schools where the experiments were conducted, the degree to which these roles were present.
- Use of social networks, a questionnaire to measure the frequency of use of 10 different applications and social network on a Likert-type scale from 1 (I do not know it) to 5 (I use it every day).
- Acceptance of the video game, consisting of two free text questions: "What did you think of the game's look, and what would you change?" and "Do you think you have learned something?".

The following process was used to evaluate the answers to the two qualitative questions:

- A set of categories were defined to assign to the answers of each of the questions. In total 25 binary categories were created, some are common to both questions while others are specific. Examples of these categories and their meaning are:
    - Positive: positive comment about the game or about learning with the game
    - Negative: negative commentary about the game or about learning with the game
    - Awareness: comments that mention awareness and/or putting themselves in the victim's shoes
    - TooLong: comments about the length of the game being too long
    - Know: comments that mention that they already know what the game shows
- Two researchers, independently, read and analysed the responses and assigned each response to a set of categories.
- The categories assigned to each answer by each researcher were compared. If a similarity of more than 95% in the complete set of answers was found, the classification of the researcher 1 was accepted. For the cases where the difference remains, both researchers met to review those answers on which they had not agreed. The final set of categories per response was reached.
- The appearance of each category in each of the questions was counted.

Moreover, the in-game player interaction data is sent as traces following the Experience API for Serious Games (xAPI-SG) standard format (Serrano-Laguna et al., 2017). These xAPI-SG traces include all the game relevant interaction data such as the timestamp of the interaction, the choices in the player's dialogues, the interactions with objects or characters, and the changes of day and place (home, classroom, hallway, etc.).

3.4. First pilot, a formative evaluation

Once the first version of the game was developed, it was tested in three different high schools (2 in different cities and one rural school). The goal was to formally validate that the game fulfilled its goal of raising bullying and cyberbullying awareness and if the students liked it (answering RQ1, RQ2, RQ3, RQ4, and RQ5). This evaluation involved 257 users from three different high schools. Each session of the experiment lasted between 50 and 60 minutes and consisted of three parts in the following order:

- Pre-test questionnaire: including the demographics questions (age and gender) and the bullying and cyberbullying awareness test.
- *Conectado* session: complete gameplay of the video game, up to the maximum available time.
- Post-test questionnaire: including the bullying and cyberbullying awareness test, the cyberbullying test, the use of social networks, and the acceptance of the video game.

64 of the 257 participated in a first set of sessions. From their results, a shortened version of the game was created, which was used in the following sessions with the remaining 193 players. Also, to verify if the game produced any change in the players, we wanted to examine in a real setting and with actual final users if there were any usability problems and if the gameplay time was adequate, we used the data collected from the interactions made by each player during the session.

3.5. Experiment with control group

This experiment sought to verify, using a control group, that differences between the initial (pre-test), and subsequent (post-test) questionnaires were caused by playing *Conectado* and not by the context of use (e.g., extra motivation from the use of a videogame) and the questionnaires themselves (providing further evidence to answer RQ3). To this end, an experiment was carried out with 40 users aged 15 and 16 years old from 15 different high schools, randomly divided into two groups of 20 students. The sessions of both groups lasted an hour and a half, the last 30-40 minutes for play a second video game, and consisted of four phases in the following order:

- Pre-test questionnaire: including the demographics questions (age and gender) and the bullying and cyberbullying awareness test.
- Game play session using *Conectado* in the experimental group and using a different serious game (*First Aid Game*, unrelated to bullying) in the control group.
- Post-test questionnaire: including the bullying and the cyberbullying awareness test.
- Game-play session using *First Aid Game* in the experimental group and *Videogame* in the control group.

In the control group, we used *First Aid Game*, a serious game for learning basic resuscitation procedures (Marchiori et al., 2012). We chose it because it covered a topic of interest for the schools entirely unrelated to bullying. The experiment was carried out in the laboratories of the *Faculty of Computer Science of the Complutense University of Madrid* during the career orientation week. This is the only experiment that could be carried out with a control group. In all other experiments, participating educational centers had several restrictions that made use of control groups impractical. For example, schools were reluctant to have only a part of the students receiving the intervention, had limited numbers of computer rooms and computers available, or had scheduling constraints that made extended use of these resources and classrooms difficult to justify for, from the point of view of these schools, little educational benefit.

It is also important to note that *Conectado* is a tool for teachers to address the issue of bullying and cyberbullying after the students play. Teachers can use the shared experience that players acquire while playing the game (use of social networks, exclusion, nicknames, and insults, etc.) to adapt it to their specific case, going deeper into these topics and carrying out reflection sessions or other complementary activities. In this experiment, the game is compared with another game that is not about bullying or cyberbullying. We do not want to study if *Conectado* is better than another tool, but to corroborate that the game has a positive effect on the player and that this effect is not caused by the experimental design, the questionnaires or the possible memory effect (Lavrakas et al., 2019).

3.6. Summative evaluation

This large-scale evaluation was carried out to measure with more data the effect of the game and to assess whether different variables such as gender or age lead to a distinct increase in (cyber)bullying

awareness (providing further evidence to answer RQ3, and answering RQ6, RQ7, RQ8 and RQ9). A total of 1004 students from 12 to 17 years old from 8 different high schools participated in this evaluation. Five of the schools are public, and the other three are charter/private schools. This set of schools is a representative sample of the schools that exist in Spain (Instituto Nacional de Evaluación Educativa, 2019). The phases, time and questionnaires used in this experiment were the same as those used in the pilot test: the Pre-test questionnaire including the demographics questions (age and gender) and the bullying and cyberbullying awareness test; the *Conectado* playing session; and the Post-test questionnaire including the bullying and cyberbullying awareness test, the cyberbullying test, the use of social networks and the acceptance of the video game. In this experiment, as in the pilot, all in-game player interactions were collected. With this data, we also wanted to validate the game design and analyse if the decisions taken in its development have the desired effects and if there is any kind of relationship between the questionnaires and the way of interacting with the game.

4. Results

With the results obtained in each of the experiments through the analysis of the completed questionnaires and the players' interactions during the different sessions, we answer the research questions and also compare the various experiments in order to discuss the validation process carried out.

4.1. Results of the first pilot

This experiment was divided into two groups of sessions. 64 of the 257 total users participated in the first set of sessions. From the traces collected, we found that actual game completion time was longer than expected. Therefore, we decided to shorten the game by removing some reiterative events. In the first version, players had to repeat, although with some changes, the second in-game day; and had the option to repeat the fourth, depending on certain decisions made on previous days. These repetitions were intended to frustrate players and examine if they took different actions when repeating days, but had no effect on the final outcome of their games. The 193 players in the second set of sessions who participated in the first pilot's experiment used the shortened version without repetitions. When comparing the results between the shorter and longer version of the game, using the Mann-Whitney test, the results do not show a statistically significant difference (Statistics=3312, p-value=0.367).

From a total of 257 users, 223 players (45 from the first version and 178 from the second one) completed both questionnaires. Addressing RQ1, "Is *Conectado* well accepted among the students?", the general acceptance of the game was high, as 65% of the players liked the game and left positive comments. Only 2% said they did not like it. 18% did not fill in the question, and the remaining 14% left neutral comments that did not indicate whether they liked the game or not. Besides, 63% considered that they had learned while playing *Conectado*, and 19% did not answer. On the other hand, 10% believed that they had not learned anything, while an additional 9% thought that they already knew what the game covered.

Table 1. First pilot pre-post awareness results. Awareness ranges from 1 (lowest) to 7 (highest).

	N	Mean age	Pre-Test	Post-Test	p-value
All	223	14.20	M = 5.72 SD = 1.26	M = 6.38 SD = 1.11	<0.001
Initial version	45	14.93	M =5.81 SD = 0.97	M = 6.22 SD = 1.29	<0.001
Final, shortened version	178	14.01	M = 5.68 SD = 1.30	M = 6.42 SD = 0.79	<0.001

Concerning RQ2, "How long does it take for players to complete the game? In particular, are they able to complete the game in the average time available in one school session?". First, this pilot evaluation allowed us to find and fix minor errors and to discover that it was challenging to complete the game in the assigned time. With the improvements made to shorten the game, the time to complete the game was reduced by 8 minutes on average (from M=41.07 and SD=5.68 in the original version to M=32.97 and SD=6.23 in the modified version). Only 62% of players in the extended version completed the game, while the short version was completed by 86%. Therefore, our modifications made the game easier to complete within a school session, although a small percentage of users could still not finish it within the allotted time. Notice that this is in part due to shorter-than-usual sessions, as they had to

complete the questionnaires before and after playing, which in a classroom setting can take longer than

Table 2. Awareness and game effect by users cyberbullying profile.

	Victim (N=17)	Bystander (N=127)	Bully (N=8)	None (N=73)	p-value
Awareness	M=5.93 SD=0.94	M=5.73 SD=1.11	M=4.18 SD=1.67	M=5.84 SD=1.18	0.030
Effect	M=0.36 SD=0.83	M=0.66 SD=0.83	M=0.44 SD=1.70	M=0.56 SD=1.10	0.059

expected.

Regarding RQ3, "Does playing *Conectado* produce a positive effect on players' awareness of bullying and cyberbullying?", both the free-text responses of the players and the comparison between the pre- and post-game awareness test show that the game has a positive effect, increasing the awareness of its players (Table 1). We analysed the pre-post awareness results for each age and compared it using the Mann-Whitney test, discovering that the awareness increase was significantly higher in students aged 12-15 (N=184, M=0.72, SD=1.10) than in those aged 16-17 (N=39, M=0.40, SD=0.59) with p-value=0.006.

Looking into RQ4, "Does playing *Conectado* produce a different effect depending on players' previous use of social networks?", and contrary to what we expected, the data did not show an influence

Table 3. First Pilot summary results.

Data collected and analysed	Results

Interactions game data	• Initial game often took too long to complete (>50 minutes)
	  ○ 62% of users completed the game in under 50 minutes
	  ○ Time (N=28): M=41.07 min, SD=5.68
	• After game modifications for a shorter version
	  ○ 86% of users completed the game in under 50 minutes
	  ○ Time (N=153): M=32.97 min, SD=6.23

Pre-Post Awareness questionnaire (Likert-7)	• Positive effect in awareness
	  ○ Pre-test score (N=223): M=5.72, SD=1.26
	  ○ Post-test score (N=223): M=6.38, SD=1.11
	  ○ Pre-post difference M=0.66
	  ○ Wilcoxon test p-value < 0.001
	• Better results for students aged 12-15
	  ○ 12-15 aged (N=184): M=0.72, SD=1.10
	  ○ 16-17 aged (N=39): M=0.40, SD=0.59
	  ○ Mann-Whitney test p-value=0.006

Pre-Post Awareness questionnaire (Likert-7) and Post Social Network Questionnaire (Likert-5)	• The frequency of use of social networks (SN) does not influence the game effect
	  ○ High use of SN pre-post difference (N=46): M=0.89, SD=1.18
	  ○ Low use of SN pre-post difference (N=177): M=0.60, SD=0.98
	  ○ Mann-Whitney test p-value=0.171

Pre-Post Awareness questionnaire (Likert-7) and Post Test Cyberbullying (Likert-5)	• Game effect is not affected by previous cyberbullying experience
	  ○ Victim profile (N=17): M=0.36, SD=0.83
	  ○ Bystander profile (N=127): M=0.66, SD=0.83
	  ○ Bully profile (N=8): M=0.44, SD=1.70
	  ○ No profile (N=73): M=0.56, SD=1.10
	  ○ Kruskal-Wallis test p-value=0.059
	• Pre-test awareness is affected by previous cyberbullying experience
	  ○ Victim profile (N=17): M=5.93, SD=0.94
	  ○ Bystander profile (N=127): M=5.73, SD=1.11
	  ○ Bully profile (N=8): M=4.18, SD=1.67
	  ○ No profile (N=73): M=5.84, SD=1.18
	  ○ Kruskal-Wallis test p-value=0.030

of the use of social networks on the effect of the game or the level of awareness of the previous questionnaire. There appeared to be a small, statistically not-significant difference in initial awareness, of 0.33 points in a 7-point scale, linked to the number of social networks that participants reported as frequently-used. We have applied the Mann-Whitney test by comparing the initial awareness (Statistical=3737, p-value=0.196) and the effect of the game (Statistical=3700, p-value=0.171) of both groups: players who do not use any social network frequently, or only use WhatsApp frequently, and the group of players who do use other social networks frequently.

Finally, with regard to RQ5, "Does playing *Conectado* produce a different effect depending on players' previous experience with cyberbullying?", while differences in pre-game awareness have been observed, being for example somewhat lower in those students who have reported acting as bullies before, we observed no differences in the effect of *Conectado* depending on the roles of players (Table 2).

4.2. Results of experiment with control group

Of the 40 students who agreed to participate in the experiment, four missed the session, and one did not submit the post-game questionnaire, leaving a valid data set of 19 students from the control group, which played *First Aid Game*, and 16 from the intervention group, playing *Conectado*.

Table 4. Changes in awareness in control and intervention groups. Awareness ranges from 1 (lowest) to 7 (highest).

	N	Mean age	Male% Female%	Pre-Test	Post-Test	p-value
Control	19	15.6	57.89% 42.11%	M = 5.37 SD = 1.31	M = 5.21 SD = 1.16	0.0585
Intervention	16	15.5	56.25% 43.75%	M = 5.54 SD = 0.82	M = 6.15 SD = 0.57	0.0016

As can be seen in Table 4, the intervention group showed statistically significant changes when conducting a Wilcoxon Test while the control group did not. Although the number of users in both groups is small, the results are very positive as this change is seen in the intervention group but not in the control group (which even decreases its awareness results from pre- to post-test). If we compare the distribution of the results of the initial questionnaire of both groups with a Mann-Whitney U Test, we do not find a statistically significant difference. However, comparing the post-test questionnaires using another Mann-Whitney U Test, the difference between both groups is statistically significant with a p-value of 0.0014: *Conectado* produced a change in the awareness and perception of bullying and cyberbullying in users while *First Aid Game*, as expected, did not. These results reinforce the affirmative response to RQ3, "Does playing *Conectado* produce a positive effect on players' awareness of bullying and cyberbullying?", obtained by the pilot experiment.

Table 5. Experiment with control group summary results.

Data collected and analysed	Results

Pre-Post Awareness questionnaire (Likert-7)	• Positive effect in intervention group
	  ○ Pre-test score (N=16): M=5.54, SD=0.82
	  ○ Post-test score (N=16): M=6.14, SD=0.57
	  ○ Wilcoxon test p-value = 0.0016
	• No effect in control group
	  ○ Pre-test score (N=19): M=5.37, SD=1.31
	  ○ Post-test score (N=19): M=5.21, SD=1.16
	  ○ Wilcoxon test p-value = 0.0585

4.3. Results of summative evaluation

This experiment was conducted with 1004 participants aged 12-17, 902 of which completed both questionnaires, of which 860 had a corresponding game traces file which contained their main in-game interactions. Apart from a larger scale than previous experiments, the availability of in-game interaction data from participants sets the summative evaluation apart from other experiments, as interactions provide a richer insight than the simple increase in awareness obtained with the pre-post questionnaires.

The original article has been published in Telematics and Informatics with DOI: https://doi.org/10.1016/j.tele.2021.101568

Table 6. Summative evaluation pre-post awareness results by gender and bullying/cyberbullying questions. Awareness ranges from 1 (lowest) to 7 (highest).

	Pre-Test	Post-Test	p-value
All users (N=860 Mean age=13.79 SD=1.49)			
Total	M=5.78 SD=1.28	M=6.29 SD=1.30	<0.001
Cyberbullying	M=5.86 SD=1.33	M=6.34 SD=1.30	<0.001
Bullying	M=5.65 SD=1.42	M=6.22 SD=1.37	<0.001
Male (N=454 Mean Age=13.82 SD=1.50)			
Total	M=5.51 SD=1.43	M=5.98 SD=1.55	<0.001
Cyberbullying	M=5.60 SD=1.49	M=6.02 SD=1.56	<0.001
Bullying	M=5.36 SD=1.57	M=5.87 SD=1.62	<0.001
Female (N=406 Mean Age=13.78 SD=1.48)			
Total	M=6.08 SD=1.01	M=6.66 SD=0.80	<0.001
Cyberbullying	M=6.15 SD=1.07	M=6.69 SD=0.79	<0.001
Bullying	M=5.98 SD=1.14	M=6.60 SD=0.87	<0.001

Table 7. Summative evaluation pre-post awareness results by age and gender. Awareness ranges from 1 (lowest) to 7 (highest).

	Pre-Test	Post-Test	p-value
Age 12 (N=205)			
All users	M=5.77 SD=1.25	M=6.33 SD=1.18	<0.001
Male (55%)	M=5.74 SD=1.18	M=6.17 SD=1.24	<0.001
Female (45%)	M=5.80 SD=1.32	M=6.51 SD=1.07	<0.001
Age 13 (N=217)			
All users	M=5.61 SD=1.42	M=6.14 SD=1.52	<0.001
Male (49%)	M=5.28 SD=1.69	M=5.73 SD=1.85	<0.001
Female (51%)	M=5.93 SD=1.01	M=6.53 SD=0.96	<0.001
Age 14 (N=167)			
All users	M=5.83 SD=1.30	M=6.27 SD=1.36	<0.001
Male (50%)	M=5.38 SD=1.53	M=5.77 SD=1.72	<0.001
Female (50%)	M=6.29 SD=0.77	M=6.78 SD=0.46	<0.001
Age 15 (N=138)			
All users	M=5.80 SD=1.30	M=6.37 SD=1.27	<0.001
Male (60%)	M=5.49 SD=1.46	M=6.10 SD=1.52	<0.001
Female (40%)	M=6.26 SD=0.84	M=6.79 SD=0.51	<0.001
Age 16 (N=85)			
All users	M=6.12 SD=0.83	M=6.48 SD=0.99	<0.001
Male (52%)	M=5.86 SD=0.95	M=6.18 SD=1.23	<0.001
Female (48%)	M=6.40 SD=0.58	M=6.80 SD=0.47	<0.001
Age 17 (N=50)			
All users	M=5.75 SD=1.18	M=6.29 SD=1.11	<0.001
Male (52%)	M=5.34 SD=1.27	M=5.75 SD=1.30	<0.001
Female (48%)	M=6.20 SD=0.90	M=6.87 SD=0.30	<0.001

The original article has been published in Telematics and Informatics with DOI: https://doi.org/10.1016/j.tele.2021.101568

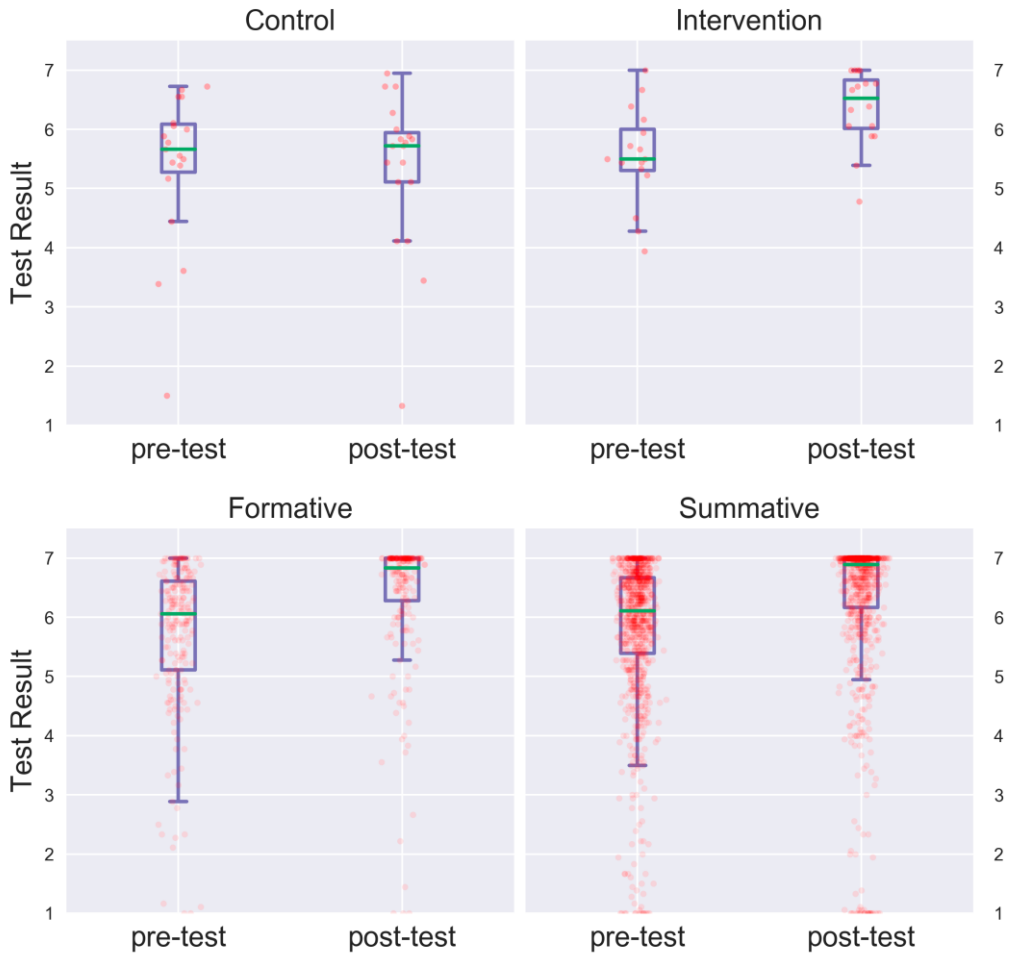

Figure 2. Pre-post test awareness results in experiment with control (upper-left) and intervention (upper-right) groups and in the formative (lower-left) and summative evaluations (lower-right).

The results of the awareness test before and after playing *Conectado* can be seen in Table 6. Revisiting RQ3, "Does playing *Conectado* produce a positive effect on players' awareness of bullying and cyberbullying?", as in the pilot experiment and the intervention group of the control group experiment, we found positive results. Again, we find a statistically significant change in (cyber)bullying awareness, with a p-value < 0.001, applying the Wilcoxon Test. The mean difference in the score obtained between both questionnaires is 0.51 (SD=0.97, Q1=0.06, Q2=0.33, Q3=0.78), which, due to already high pre-test awareness scores, amounts to ~42% of the possible score improvement on our 7-point scale. The positive change in awareness occurs in both bullying and cyberbullying questions. This positive change is demonstrated in all items, with statistically significant values, the average value of increase in these items ranging from 0.15 to 1.34. The highest increase occurs in those items where the pre-test shows lower awareness values. The mean of awareness for each item in the pre-test ranges from 4.74 to 6.28, while those in the post-test range from 6.08 to 6.46. These data show that the video game has a positive effect on the awareness of players by making them reflect on acts that they previously considered as minor, as compared to those which they already considered hurtful.

For example, in the pre-test questions related to exclusion, discrimination and avoidance, many participants initially considered these behaviors, when carried out over social networks and chat applications, as only weakly constituting bullying or cyberbullying (M=5.74 and M=4.74, respectively). In the post-test, players were much more aware of the effects that these behaviors can have on the victims when repeated over time (M=6.08 and M=6.21).

Regarding RQ6, "Does playing *Conectado* produce a different effect depending on players' age or gender?" as we see in Table 6, the game is effective in all ages and for both genders. However, boys show a lower average awareness than girls both before and after playing. When analysing the questionnaires from the point of view of age, for girls, the older they are, the higher the initial awareness. Also, in girls, the awareness gain among the 12 and 13-year-old groups is higher than in the 14, 15 and 16-year-old groups with a p<0.01. Although after playing the awareness of the 12 and 13-year-old group is still lower,

the difference in awareness between the two age groups decreases. This difference in awareness depending on age is not observed in boys. Another result that stands out is that 17 years old players, particularly boys, show less awareness. This result can be due to lower diversity and smaller sample sizes for this age group: 89% of the 17 year-old players were from the same school, which had already carried out previous prevention campaigns (Table 7).

By comparing the results obtained from pre- and post-game questionnaires, the game has a positive effect on players in all experiments and groups – except in the control group, which played a different game. This effect can be seen in Figure 2, which depicts the results of both awareness questionnaires of the different experimental groups. The same happens if we only consider 14- and 15-year-old students, the age of the players who participated in the experiment with the control group.

4.3.1. Completeness and gameplay time

With the data collected, we can answer RQ2, "How long does it take for players to complete the game *Conectado*?", and we can further explore if the game has an effect on those players who do not complete it or if there are differences in the time it takes for players according to their gender or age.

The game is designed to be completed in about 35 minutes, allowing time for a 5-minute pre-test and a 10-minute post-test. These times add to 50 of the 60 minutes of a typical class in participating high schools. Therefore, we expected the percentage of completion to be high. 82.79% of the players successfully completed the game, with a higher percentage of boys (88.10%) than girls (76.85%). All players completed day 1 of the game, while only 9 players (1.04% of the total) stayed on day 2 of the game, 36 players (4.19%) did not advance past day 3, 65 (7.56%) remained in day 4 and 38 (4.42%) arrived at day 5 but did not complete it. It should be noted that all schools had computer rooms in dedicated classrooms, including some in different buildings from where the students had their regular classes. In those schools, participating students had less time to play, because they first had to reach the computer room. In other schools, players had to leave as soon as their sessions were over, as opposed to being allowed to stay a little longer until they finished. When using the game as a tool in a regular school session, without doing the questionnaires, there would be additional time to complete the game. In some cases, it may be necessary to do the reflection part in a later session.

Analysing the results in the questionnaires of those players who have not completed the game, we can see that the game also had an effect on them, although this effect was a little lower (N=148 M=0.42, SD=0.41, MD=0.28) than that of players who completed the game (N=712, M=0.52, SD=1.01, MD=0.36). Besides, for those players who did not complete the game, the results show that the more they have progressed in the game, the higher the effect is. This result is consistent with the design of *Conectado* as the bullying and cyberbullying that appear is progressive in the game: the game shows new forms of bullying and cyberbullying and more game characters opposing the player as the player progresses through the game. Data from players who have completed the game show that girls generally take longer to complete the game (Figure 3). The average difference is only 1 minute and 37 seconds, but is statistically significant and may be related to a higher percentage of girls not finishing the game. But this difference is reduced and even disappears after 45 minutes of play due to the experimental design. At this point, players were forced to leave the game and fill out the questionnaire if they had to go to another class.

All players take the longest to complete the first game-day (Figure 4). This result aligns with our expectations because, at the beginning of the game, players are learning how the game works, how to move, which doors they can go through, and what they can interact with. In terms of age, players aged 12-14 take more time to complete the game than those aged 15-17. This finding is true for both genders, although girls take longer than boys of the same age. Since one of the main mechanics is that of

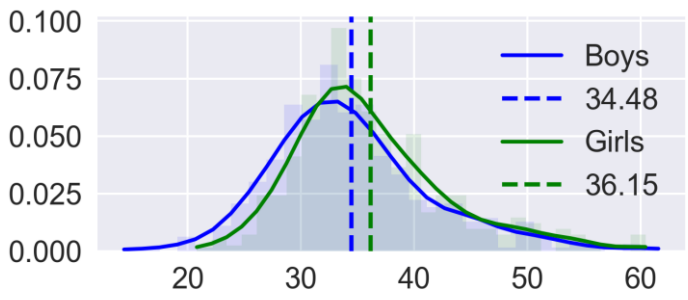

Figure 3. Time to complete the game by gender

The original article has been published in Telematics and Informatics with DOI: https://doi.org/10.1016/j.tele.2021.101568

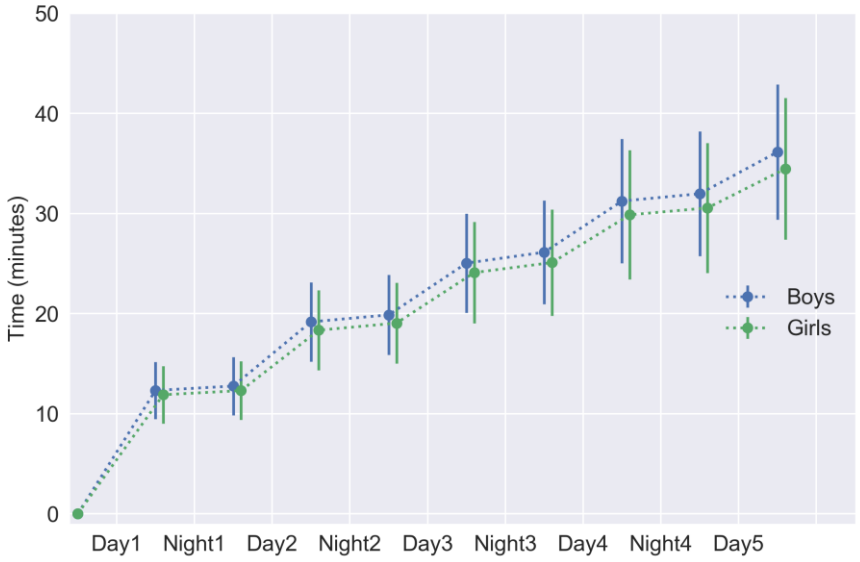

Figure 4. Time (accumulated) to complete each part of the game by gender

dialogues, it is normal that younger players take longer to complete the game, as their reading speed is generally lower.

4.3.2. In-game choices

During the game, players have the option to choose from several replies in the dialogues. Some of these choices influence the game story. The data collected from players' interactions allows us to study if these choices also change the effect of the game on the players. Additionally, it allows us to analyse if the decisions made by players depend on their pre-game awareness. To analyse player behaviors, we have classified the players according to their in-game choices. Choices that decrease the risk of victimization score positively, while choices that increase the risk score negatively. We have classified players according to their choice-scores into 4 different groups, ranging from Wrong (most decisions increased victimization) to Perfect (all decisions minimized victimization). Table 8 shows the classification of the 4 groups and their pre-game awareness scores.

With regard to RQ7, "How does players' previous awareness affect the way players interact with the characters of *Conectado*?" and RQ8, "Are there gender differences in the way players interact with *Conectado*?", results show that there is a higher percentage of girls who have a higher score than boys during the game (Table 8). In particular, a higher percentage of girls are ranked in the group with highest score. Also, the group of players with the highest score in both genders show a greater awareness pre-

Table 8. Differences between in-game behavior of players, by gender.
Wrong = choices increased likelihood of victimization; Perfect = all choices decreased likelihood of victimization.

Behavior	N	Pre-Test	p-value comparing with			
			Wrong	Neutral	Good	Perfect
Male						
Wrong	99 (24.75%)	5.09	-	0.030	0.055	<0.001
Neutral	175 (43.75%)	5.43	0.030	-	0.407	0.028
Good	44 (11.00%)	5.63	0.055	0.407	-	0.109
Perfect	82 (20.50%)	5.95	<0.001	0.28	0.109	-
Female						
Wrong	47 (15.06%)	5.94	-	0.289	0.335	0.018
Neutral	115 (36.86%)	5.99	0.289	-	0.136	0.027
Good	54 (17.31%)	5.82	0.335	0.136	-	0.003
Perfect	96 (30.77%)	6.31	0.018	0.027	0.003	-

game with some statistically significant differences. Regardless of their prior awareness, players appear to make any of the possible decisions in the game. Still, the more awareness they have, the more likely they are to make choices that decrease the risk of victimization during the game.

In terms of the effect of decisions and behavior on the increased awareness caused by the game, to answer RQ9, "Are players' interactions, decisions and behaviors in *Conectado* related to their change in awareness?" different effects are observed for girls and boys. On the one hand, the increase for girls is inversely proportional to the previous awareness, being lower in the group of girls with the highest score. This effect can be explained by the high level of initial awareness, which limits the extent of possible improvements. However, the opposite is observed in boys. Those boys with greater previous awareness have a greater increase, to statistically significant degree when comparing the group with lowest scores and those with highest. This last effect on boys is unexpected, and therefore some additional analyses have been carried out, which are detailed in the following sub-section 4.3.3.

While players cannot see their own avatars, they can, at the start of the game, choose their avatar's gender, which changes some of the game dialogues and some of the graphic resources of the game (e.g. the player's social network, and the school bathroom the player can enter). These small changes help the immersion of the player in the story. But do players choose their actual gender? To verify this, we have compared the gender stated in the questionnaire before playing with the gender selected at the beginning of *Conectado*. 98% of the players selected characters of the same gender. Only 5 girls and 8 boys chose to play with a character of the opposite gender. Comparing the results of the awareness test of the group who select their gender with those that select a different gender, a small but significant difference in awareness ($p=0.028$) was found in those girls who selected the opposite gender in the game: they tended to score less. Notice that the sample of players who did not select their gender to play is tiny. In future experiments, it would be interesting, given the gender properties of Spanish language, to study whether the immersion and change produced by the game are affected by the gender used by the player. The data collected from the boys make us think that there may be several players who did not take the session seriously enough to affect some of the analyses made. When observing their game-interaction traces, we have not found evidence of groups of players that played randomly or without paying attention. However, in this analysis, we have realized that it would have been useful to know when the dialogues of the game finished, and not only when they started, to identify which players could have skipped through conversations without actually reading them. This unexpected outcome, together with the good results observed in a previous study in which a reflection session guided by a counsellor was carried out after the students played *Conectado* (Calvo-Morata, García-Diego, Freire, Martínez-Ortiz, & Fernández‐Manjón, 2019), may indicate that perhaps some students answered their questionnaires randomly. We explore this in the next section.

4.3.3. Analysis of divergences between game effect and questionnaires results

While the literature has pointed out some gender differences in video games and bullying and cyber-bullying attitudes, the decisions made while the users are playing *Conectado* are not designed to create different levels of awareness. The difference in awareness increase among boys depending on their

Table 10. Differences between the in-game behavior of filtered players.
Wrong = choices increased likelihood of victimization; Perfect = all choices decreased likelihood of victimization.

Behaviour	N	Pre-Test	p-value comparing with			
			Wrong	Neutral	Good	Perfect
Male						
Wrong	64 (21.12%)	5.89	-	0.227	0.499	0.030
Neutral	133 (43.89%)	5.96	0.227	-	0.273	0.106
Good	38 (12.54%)	5.81	0.499	0.273	-	0.073
Perfect	68 (22.44%)	6.20	0.030	0.106	0.073	-
Female						
Wrong	43 (15.03%)	6.08	-	0.254	0.443	0.036
Neutral	105 (36.71%)	6.15	0.254	-	0.278	0.072
Good	46 (16.08%)	6.16	0.443	0.278	-	0.034
Perfect	92 (32.17%)	6.36	0.036	0.072	0.034	-

choices, together with a higher standard deviation in the questionnaires results, led us to believe that some students, especially boys, filled in the questionnaires randomly or paying little attention, and yet they had been engaged while playing the game. Because of this, we have performed a more in-depth analysis to attempt to detect such students. To filter them out, we have identified the following patterns and discarded those players who, in any of the questionnaires, showed one or more of the following:

- Answering in a continuous cascade, where answers to Likert questions are always the previous value plus or minus one (e.g.: 3,4,5,6,7,6,5,4,3).
- Answering all questions with the same value.

Table 9. Summative evaluation pre-post awareness results after filtering out users with random-like questionnaires. Awareness ranges from 1 (lowest) to 7 (highest).

	Pre-Test	Post-Test	p-value
All users (N=719 Mean age=13.81 SD=1.51)			
Total	M=6.12 SD=0.83	M=6.62 SD=0.67	<0.001
Male (N=346 Mean Age=13.79 SD=1.52)			
Total	M=5.98 SD=0.90	M=6.50 SD=0.74	<0.001
Female (N=373 Mean Age=13.82 SD=1.49)			
Total	M=6.24 SD=0.75	M=6.74 SD=0.58	<0.001

Extreme positive and negative values have also been filtered out. A total of 141 users were discarded (33 girls and 108 boys), leaving a valid sample for this analysis of 719 players. 589 of these users complete the game, still with a higher percentage of boys completing the game (77% of girls and 88% of boys). Analysing both questionnaires for the 719 players, girls still have higher values of pre- and post-game awareness; however, similar values in the increase of awareness were found in both genders (Table 9). The difference between the two genders also remains statistically significant, with a p-value < 0.001. The game had a positive effect on 81% of boys and 82% of girls.

We have also re-analysed the effect of awareness on in-game behavior, and the impact of these decisions on awareness increase (Table 10). In this case, there is still a higher percentage of girls who have a higher score than boys in the game choices. Besides, statistically significant differences in pre-game awareness remain between the groups with the highest score and those with the lowest. But in this case, the differences in the increase in awareness between the groups of boys disappear. Discarded users had average gameplay times as well as number and type of interactions. Therefore, we consider that these users paid attention to the game but filled out questionnaires without reading them.

5. Conclusions and discussion

In this study, the evaluation of the *Conectado* videogame has been carried out in actual schools and with a large number of students, to validate both its effectiveness and some key elements of its design. Nowadays, questionnaires are the most widely accepted and used method for validating serious games, but we consider that questionnaires combined with game learning analytics (based on in-game player interaction data) provide a more powerful insight for validating both the design and effectiveness of a serious game. We believe that these two methods are complementary and very useful when it comes to getting a better understanding of how a serious game affects its players.

The different experiments carried out and the data collected have allowed us to answer nine research questions about the game's acceptance and effectiveness, and validate different parts of the game design. First, the game has been well accepted by the players, and the results show that *Conectado* can be a useful tool when it comes to raising awareness on its own. However, *Conectado* is designed to be used as a classroom tool that allows teachers to initiate reflection sessions afterwards on the shared experience obtained by the players. The pre-post questionnaires and the small control group experiment showed us that the game had positive effects on awareness in 80% of the sample, as well as gender differences. Girls show increased awareness before and after playing, although the game has a positive effect on both boys and girls. The results of the questionnaires also shows age differences, although these are only present in girls, where the group of 12 and 13-year-old users showed less pre-awareness than the group of 14 to 17-year-old players. On the other hand, analysis of interactions collected during the gaming sessions has

The original article has been published in Telematics and Informatics with DOI: https://doi.org/10.1016/j.tele.2021.101568

Table 11. Summative evaluation summary results.

Data collected and analysed	Results

Pre-Post Awareness questionnaire (Likert-7)	• Positive effect in awareness
	  ○ Pre-test score (N=860): M=5.78, SD=1.28
	  ○ Post-test score (N=860): M=6.29, SD=1.30
	  ○ Pre-post difference M=0.51
	  ○ Wilcoxon test p-value < 0.001
	• Higher initial awareness of girls
	  ○ Girls pre-test score (N=406): M=6.08, SD=1.01
	  ○ Boys pre-test score (N=454): M=5.51, SD=1.43
	  ○ Mann-Whitney test p-value < 0.001
	• Higher game effect in girls
	  ○ Girls pre-test score (N=406): M=0.57, SD=0.84
	  ○ Boys pre-test score (N=454): M=0.45, SD=1.08
	  ○ Mann-Whitney test p-value = 0.012

Interactions game data	• 82.79% of users completed the game
	• Girls spend more time than boys to complete the game
	  ○ Girls M=36.15 min
	  ○ Boys M=34.48 min

Pre-Post Awareness questionnaire (Likert-7) and Interactions game data	• More days completed → More game effect
	• Girls have better in-game behavior
	  ○ Girls: 32.17% have perfect behavior and 16.08% have good behavior
	  ○ Boys: 22.44% have perfect behavior and 12.54% have good behavior
	• Boys with perfect behavior have higher pre-test awareness than those with wrong behavior
	  ○ Boys wrong → pre-test score = 5.89
	  ○ Boys perfect → pre-test score = 6.20
	  ○ Mann-Whitney test p-value = 0.030
	• Girls with perfect behavior have higher pre-test awareness than those with wrong behavior
	  ○ Girls wrong → pre-score = 6.08
	  ○ Girls perfect → pre-score = 6.36
	  ○ Mann-Whitney test p-value = 0.036

provided insights and allowed us to validate several design decisions: (1) The game is easy to complete in the established time, and although some players could not finish it, they show an increase in their awareness proportional to the percentage of game completed. This is consistent with the design of the game, where the intensity of bullying and cyberbullying increases as the game progresses. (2) We found differences between groups of players. Regarding genders, girls take longer to complete the game, and there is a higher percentage of girls that take correct, victimization-minimizing actions within the game. On the other hand, the group of players who choose better actions shows a higher previous awareness, although this does not condition their choices. Finally, (3) the different choices that players can make that change some dialogues and parts of the game do not affect the perception and experience of those players, at least from the point of view of awareness increase; that is, the effect of the game does not seem to depend on the choices taken by the player.

All these results showed differences mainly in gender, regarding completion time, game choices and awareness of bullying and cyberbullying. This gender effect has been previously pointed out by other authors, who have shown differences in the preferences and way of playing (Lucas, Sherry, & Sherry, 2004) as well as in the prevalence of bullying and cyberbullying, with girls showing more knowledge and awareness than boys (Waasdorp & Bradshaw, 2015; Zych et al., 2015). Other video games studied as prevention tools have also shown positive results in preventing bullying with different strategies. At Risk for Middle School Educators trains teachers on how to identify bullying and how to deal with it (Bradley & Kendall, 2019). Quest for the Golden Rule is a set of 3 games that demonstrated increases in knowledge and social skills, teaching prevention strategies to students from 6 to 11 years old (Rubin-Vaughan, Pepler, Brown, & Craig, 2011). FearNot! Increased empathy teaching bullying-prevention skills to 8-12 year olds (Paiva et al., 2004; Watson et al., 2010). However, many of these validated games are not available for

public use or require a license fee, restricting their possible impact and outreach.

Also, the differences found in the results could make the design of games that can accurately detect the likelihood of being a victim or a bully more difficult. Although we have seen that players tend to behave better in the game if they are more aware, results also show that there is no strict relationship between the type of player and their in-game choices. The games found do not study the differences of the players according to their profile – neither in cases of bullying in which they may have been involved nor based on their knowledge. It would be interesting to study the existence of this relationship in other video games.

Finally, during the validation and analysis steps, the authors have carried out some actions that have proven to be especially useful in our experience. Considering that they may help other researchers who perform serious game validations by combining questionnaires and analysis of interactions, we have gathered them in the following recommendations:

- Improving and filtering the sample data: although the set of users who fill in the questionnaires arbitrarily and without taking questionnaires seriously may not affect the analysis of the entire sample, it could have an effect when we want to analyse and compare groups of users, as in our case by gender. In our analysis, the majority of users who did not take the questionnaires seriously were boys, which affected the comparison of interactions with the results of the questionnaires. Although there are studies that show that the set of users that fill in the questionnaires wrongly in general obtain worse results (Osborne et al., 2011), it is convenient to carry out this filter despite losing sample size. Note that players may take the game seriously but not the questionnaires, likely because they find the game activity more enjoyable. In our case, we could also study the questionnaires and partial interaction data of the players who did not complete the game.
- Collecting and analysing player interactions: while the use of questionnaires is most common and widespread when validating serious games (Calderón & Ruiz, 2015), the analysis of interactions has proven to be a very useful complement. Analysis of interactions allows for filters and groupings that would otherwise not be possible with questionnaires alone. In this case, filtering by game behavior or examining the effect of *Conectado* on players according to the percentage of completion.
- Collecting more detailed and fine-grained analytics: for instance, in the case of games with dialogues, it is advisable to capture the beginning and ending of these conversations. During the analysis of the interactions, we realized that when analysing if the players had paid attention to the game, it would have been useful to collect also when they reached the end of dialogues. With this information, we could have analysed the time players spent reading, and therefore found when they advanced too fast to actually read the dialogue. Together with the number of interactions and scene changes, this would have given us better filters when identifying which players are more thorough in interacting, and gaining a proxy value for engagement.

We consider that this approach combining both questionnaires and game learning analytics information will contribute to obtain scientifically validated serious games. For example, if changing behavior or increasing player awareness, use of Learning Analytics by itself to measure changes in players would be challenging, but relevant information can be missed if only using questionnaires. We consider that proving efficacy of games with this combined approach can contribute to the generalization of serious games in mainstream education, and that it is key for games to contribute to evidence-based educational approaches.

5.1. Limitations

One of the limitations of this research is that the effect of playing *Conectado* has been measured by the awareness increase in players, instead of measuring if playing decreases the prevalence of bullying in the classes participating in the research. To verify if the use of the game results in a reduction of bullying in the school, a follow-up questionnaire could be conducted after enough time has elapsed since the game was played. This has several associated problems, including difficulty to determine whether results are only caused by the game or if they are related to other activities or interventions carried out by each school. On the other hand, bullying is a global problem, and although the number of users in the experiments is large (compared to other research with serious games), the video game has only been used in one country. Furthermore, although we have proven a positive effect on the use of the game, we have not been able to compare its effect with that of other tools and video games, nor have we found other video games for the prevention of bullying that make such comparisons. Many of the video games with

published research regarding their effects are also not freely available, and/or would require several sessions to be applied for comparison purposes. We are currently working on translating the video game into more languages and we encourage other researchers to compare their video games and resources with *Conectado*, which has been designed to be applied in a single one-hour session, as the game is both free and open-source. Finally, while the short duration of the game makes it easy to use in any classroom session, it also means that there is less variety in the interactions and paths that the player can take. Therefore, the lack of differences when playing between different user profiles may be due to the lack of variability in the game story.

Acknowledgements

This work has been partially funded by Regional Government of Madrid (eMadrid S2018/TCS-4307, co-funded by the European Structural Funds FSE and FEDER), by the Ministry of Education (TIN2017-89238-R), by the European Commission (Erasmus+ IMPRESS 2017-1-NL01-KA203-035259), by MIT-La Caixa ((MISTI program, LCF/PR/MIT19/5184001) and by Telefonica-Complutense Chair on Digital Education and Serious Games. Also thanks to Julio Santilario Berthilier and Ana Rus Cano for their contribution to Conectado

Bibliography

Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers and Education*, *141*. https://doi.org/10.1016/j.compedu.2019.103612

Álvarez-Bermejo, J. A., Belmonte-Ureña, L. J., Martos-Martínez, A., Barragán-Martín, A. B., & del Mar Simón-Marquez, M. (2016). System to Detect Racial-Based Bullying through Gamification. *Frontiers in Psychology*, *7*(November), 1–13. https://doi.org/10.3389/fpsyg.2016.01791

Álvarez-García, D., Nuñez-Pérez, J. C., & González, A. D. (2013). Cuestionarios para evaluar la violencia escolar en Educación Primaria y en Educación Secundaria: CUVE3-EP y CUVE3-ESO. *Apuntes de Psicología*, *31*, 191–202.

Ballesteros, B., Santiago Pérez de Viñaspre, Díaz, D., & Toledano, E. (2018). *III Estudio sobre acoso escolar y cyberbullying según los afectados. Informe del Teléfono ANAR*. Retrieved from https://www.anar.org/wp-content/uploads/2018/09/III-Estudio-sobre-acoso-escolar-y-ciberbullying-según-los-afectados.pdf

Beran, T., & Li, Q. (2007). The Relationship between Cyberbullying and School Bullying. *Journal of Student Wellbeing*, *1*(2), 15–33. https://doi.org/10.1007/s10212-010-0050-5

Bradley, E. G., & Kendall, B. (2019). Training Teachers to Identify and Refer At-Risk Students Through Computer Simulation. *Journal of Technology in Behavioral Science*, (Cdc 2018). https://doi.org/10.1007/s41347-019-00106-w

Calderón, A., & Ruiz, M. (2015). A systematic literature review on serious games evaluation: An application to software project management. *Computers & Education*, *87*, 396–422. https://doi.org/10.1016/j.compedu.2015.07.011

Calvo-Morata, A., Alonso-Fernández, C., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2020). Serious games to prevent and detect bullying and cyberbullying: a systematic serious games and literature review. *Computers and Education*, *157*(November). https://doi.org/10.1016/j.compedu.2020.103958

Calvo-Morata, A., García-Diego, C., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Conectado in the classroom : applying a video game for the prevention of bullying at the school. In *12th annual International Conference of Education, Research and Innovation* (pp. 1484–1493).

Carr-Gregg, M., & Manocha, R. (2011). Bullying: Effects, prevalence and strategies for detection. *Australian Family Physician*, *40*(3), 98–102.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, *59*(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

DeSmet, A., Van Cleemput, K., Bastiaensens, S., Poels, K., Vandebosch, H., Malliet, S., … De Bourdeaudhuij, I. (2016). Bridging behavior science and gaming theory: Using the Intervention Mapping Protocol to design a serious game against cyberbullying. *Computers in Human Behavior*, *56*(January), 337–351. https://doi.org/10.1016/j.chb.2015.11.039

El-Nasr, M. S., Drachen, A., & Canossa, A. (2013). *Game Analytics*. (M. Seif El-Nasr, A. Drachen, & A.

Canossa, Eds.). London: Springer London. https://doi.org/10.1007/978-1-4471-4769-5

Gaffney, H., Farrington, D. P., Espelage, D. L., & Ttofi, M. M. (2019). Are cyberbullying intervention and prevention programs effective? A systematic and meta-analytical review. *Aggression and Violent Behavior*, *45*(March), 134–153. https://doi.org/10.1016/j.avb.2018.07.002

Garaigordobil, M., & Aliri, J. (2013). Ciberacoso ("Cyberbullying") en el País Vasco: Diferencias de sexo en víctimas, agresores y observadores. *Behavioral Psychology*, *21*(3), 461–474.

Garaigordobil, M., & Martínez-Valderrey, V. (2015). Effects of Cyberprogram 2.0 on "face-to-face" bullying, cyberbullying, and empathy. *Psicothema*, *27*(1), 45–51. https://doi.org/10.7334/psicothema2014.78

Hall, L., Jones, S., Paiva, a, & Aylett, R. (2009). FearNot! providing children with strategies to cope with bullying. *Proceedings of IDC 2009 The 8th International Conference on Interaction Design and Children*, 276–277. https://doi.org/10.1145/1551788.1551854

Instituto Nacional de Evaluación Educativa. (2019). *Sistema estatal de indicadores de la educación 2019*.

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, *140*(4), 1073–1137. https://doi.org/10.1037/a0035618

Lavrakas, P. J., Traugott, M. W., Kennedy, C., Holbrook, A. L., Leeuw, E. D. de, & West, B. T. (2019). *Experimental Methods in Survey Research: Techniques That Combine Random Sampling with Random Assignment*. John Wiley.

Lievense, P., Vacaru, V. S., Liber, J., Bonnet, M., & Sterkenburg, P. S. (2019). "Stop bullying now!" Investigating the effectiveness of a serious game for teachers in promoting autonomy-supporting strategies for disabled adults: A randomized controlled trial. *Disability and Health Journal*, *12*(2), 310–317. https://doi.org/10.1016/j.dhjo.2018.11.013

Lucas, K., Sherry, J. L., & Sherry, J. L. (2004). Sex Differences in Video Game Play : A Communication-Based Explanation. *Communication Research*, *31*(5). https://doi.org/10.1177/0093650204267930

Marchiori, E. J., Ferrer, G., Fernández-Manjón, B., Povar-Marco, J., Suberviola, J. F., & Giménez-Valverde, A. (2012). Video-game instruction in basic life support maneuvers. *EMERGENCIAS*, *24*(6), 433–437.

McEvoy, K. A., Oyekoya, O., Ivory, A. H., & Ivory, J. D. (2016). Through the eyes of a bystander: The promise and challenges of VR as a bullying prevention tool. *Proceedings - IEEE Virtual Reality*, *2016-July*, 229–230. https://doi.org/10.1109/VR.2016.7504737

Menesini, E., & Salmivalli, C. (2017). Bullying in schools: the state of knowledge and effective interventions. *Psychology, Health and Medicine*, *22*, 240–253. https://doi.org/10.1080/13548506.2017.1279740

Ortega-Ruiz, R., Del Rey, R., & Casas, J. A. (2016). Evaluar el bullying y el cyberbullying validación española del EBIP-Q y del ECIP-Q. *Psicología Educativa*, *22*(1), 71–79. https://doi.org/10.1016/j.pse.2016.01.004

Osborne, J. W., Blanchard, M. R., Bartlett, J., State, N. C., Bartlett, M. E., Carolina, N., & Brookshire, R. G. (2011). Random responding from participants is a threat to the validity of social science research results. *Frontiers in Psychology*, *1*(January), 1–7. https://doi.org/10.3389/fpsyg.2010.00220

Paiva, A., Dias, J., Sobral, D., Woods, S., Aylett, R., Sobreperez, P., … Hall, L. (2004). Caring for agents and agents that care: Building empathic relations with synthetic agents. *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS 2004*, *1*, 194–201.

Raminhos, C., Cláudio, A. P., Carmo, M. B., Gaspar, A., Carvalhosa, S., & Candeias, M. de J. (2016). A serious game-based solution to prevent bullying. *International Journal of Pervasive Computing and Communications*, *12*(2), 194–215. https://doi.org/10.1108/IJPCC-04-2016-0022

Rubin-Vaughan, A., Pepler, D., Brown, S., & Craig, W. (2011). Quest for the Golden Rule: An effective social skills promotion and bullying prevention program. *Computers and Education*, *56*(1), 166–175. https://doi.org/10.1016/j.compedu.2010.08.009

Serrano-Laguna, Á., Martínez-Ortiz, I., Haag, J., Regan, D., Johnson, A., & Fernández-Manjón, B. (2017). Applying standards to systematize learning analytics in serious games. *Computer Standards & Interfaces*, *50*(September 2016), 116–123. https://doi.org/10.1016/j.csi.2016.09.014

Waasdorp, T. E., & Bradshaw, C. P. (2015). The overlap between cyberbullying and traditional bullying. *Journal of Adolescent Health*, *56*(5), 483–488. https://doi.org/10.1016/j.jadohealth.2014.12.002

Watson, S. E. J., Vannini, N., Woods, S., Dautenhahn, K., Sapouna, M., Enz, S., … Aylett, R. (2010). Inter-cultural differences in response to a computer-based anti-bullying intervention. *Educational Research*, *52*(1), 61–80. https://doi.org/10.1080/00131881003588261

Wattanasoontorn, V., Boada, I., García, R., & Sbert, M. (2013). Serious games for health. *Entertainment Computing*, *4*(4), 231–247. https://doi.org/10.1016/j.entcom.2013.09.002

Weinschenk, S. (2011). *100 Things Every Designer Needs To Know About People*. (New Riders Publishing, Ed.), *Voices that matter*. Berkeley, CA 94710: Pearson Education.

Zych, I., Ortega-Ruiz, R., & Del Rey, R. (2015). Systematic review of theoretical studies on bullying and cyberbullying: Facts, knowledge, prevention, and intervention. *Aggression and Violent Behavior*, *23*, 1–21. https://doi.org/10.1016/j.avb.2015.10.001

Cite: Calvo-Morata, A., Alonso-Fernández, C., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2021). Creating awareness on bullying and cyberbullying among young people: validating the effectiveness and design of the serious game The Videogame. *Telematics and Informatics*, *60*(January), 101568. https://doi.org/10.1016/j.tele.2021.101568