Evaluation of an Intervention Program Based on Mobile Apps to Learn Sexism Prevention in Teenagers

PEDRO MORILLO, JOSÉ J. NAVARRO-PÉREZ, JUAN M. ORDUÑA, and MARCOS FERNÁNDEZ, Universidad de Valencia

The fight against sexism is nowadays one of the flagship social movements in western countries. Adolescence is a crucial period, and some empirical studies have focused on the socialization of teenagers, proving that the socialization with the surrounding environment prevent sexist practices. In a previous work, we developed and tested the effectiveness of a mobile app, called *Liad@s*, with the goals of helping teenagers to prevent sexism and build healthy couple relationships. In this article, we carry out a study where (using a real situation) we compare the effectiveness of the Liad@s app in front of traditional interventions like a workshop about sexism for teenagers. Also, we evaluate the usability of the app and the user satisfaction with this application. In this study, our primary hypothesis is that the effectiveness of using our mobile application, in terms of knowledge acquired about sexism, would be at least as good as attending the workshop. Our secondary hypothesis is that the user satisfaction with the mobile application would be higher than the one with the workshop, causing a preference for the app. The results of this study show significant differences in learning appeared between gender and between the two different procedures when separately evaluating the data collected from both hostile sexism (HS) and benevolent sexism (BS) questionnaires. These results validate our primary hypothesis. Also, most of the population under study preferred the mobile app in front of the traditional workshop, validating also our secondary hypothesis.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI;**

Additional Key Words and Phrases: Ambivalent sexism, mobile apps, real user evaluation, usability outcomes

ACM Reference format:
Pedro Morillo, José J. Navarro-Pérez, Juan M. Orduña, and Marcos Fernández. 2022. Evaluation of an Intervention Program Based on Mobile Apps to Learn Sexism Prevention in Teenagers. *ACM Trans. Multimedia Comput. Commun. Appl.* xx, x, Article x (2022), xx-xx
https://doi.org/10.1145/3471139

1 INTRODUCTION

Sexism was defined in the mid-fifties as a prejudiced attitude toward women [3]. Many studies have appeared last years about different aspects of sexism [16, 26, 38, 43]. Ambivalent Sexism (SA) was defined as the integration in a same level both positive and negative attitudes toward the gender

This work has been supported by Spanish MCIU and EU ERDF programs under grant RTI2018-098156-B-C55. It has been also supported by the Valencian Regional government under grant GV2017/208.
Authors' address: P. Morillo, J. J. Navarro-Pérez, J. M. Orduña, and M. Fernández, Universidad de Valencia, Avda. Universidad, s/n, Burjassot, Valencia 46100, Spain; emails: {Pedro.Morillo, J.Javier.Navarro, Juan.Orduna, Marcos.Fernandez}@uv.es.
 1551-6857/2022/02-ART45 $15.00
https://doi.org/10.1145/3471139

construction [15]. SA is configured starting from opposed affective connections, and it is represented in two dimensions: **Hostile Sexism** (**HS**) and **Benevolent Sexism** (**BS**). An HS refers to negative attitudes toward individuals who violate traditional gender stereotypes [29]. The hostile representation shares a negative emotional burden with traditional sexism, which diagnoses the role of women as a weak human being, inferior to men [7]. Hostility toward women segments the gender duality, splitting genders into strong and weak ones. This negative perception is associated to other representation with a pseudo-positive affective connection, the BS, which reproduces the female figure as a fragile being [21]. These two forms of expression of sexism are relevant to the purpose of this work, as we will explain below.

Sexism is tightly linked to violence against women [27]. Gender violence is a widespread practice not only in underdeveloped societies, but also in the first world societies. Adolescence is a crucial period, because identity patterns and citizenship constructions start to consolidate in these years. Thus, some empirical studies have focused on the socialization of teenagers [23], and they have proven that the socialization with the surrounding environment prevent sexist practices, and they deliver gratitude, safety, emotional welfare, and positive self-esteem [12]. In order to help in these tasks, we developed a mobile app, called *Liad@s*, with the goals of helping teenagers to identify sexism and inequality, discouraging them from violent behaviors, and informing teenagers about how to build healthy couple relationships. In a previous work [37], we described the Liad@s 1.0 mobile app and studied its effectiveness from a psycho-educational point of view through a classic experiment design [10], analyzing a control group who did not use the app in front of an experimental group who used the app.

In this article, we carry out a study based on a experiment conducted with a counterbalanced measures design [39], which allows to determine whether using one of the two alternatives first has some effect on the scores for the alternative system, given that both alternatives are not exclusive and can coexist in real life [17]. Instead of a quantitative research strategy, more suitable for exploratory research, we have carried out a quantitative research using structured interviews, models, and questionnaires, which have been widely validated previously. In particular, we compare the effectiveness of the updated app Liad@s 2.0 (defined in this context as the ability of a user to complete a task and gain knowledge in the context of sexism prevention), in front of traditional interventions at high schools like a workshop about sexism for teenagers. The comparison is made for both types of sexism, hostile and benevolent. Also, we evaluate and discuss the usability of the mobile app and the user satisfaction with this application (defined as the level of comfort and pleasantness afforded to the user through the use of the mobile app [19]), using a population of 42 users. Following the typical guidelines of an analysis in quantitative research, this article uses questionnaires as evaluation instrument. These questionnaires are used for obtaining measurements of both the acquired knowledge about ambivalent sexism [15] and usability/satisfaction with software apps [48]. In this study, our primary hypothesis is that the effectiveness of using our mobile application, in terms of knowledge acquired about sexism, would be at least as good as attending the workshop. Our secondary hypothesis is that the user satisfaction with the mobile application would be higher than the one with the workshop, causing a preference for the app.

2 RELATED WORK

Sexist beliefs are associated with different disorders, in the form of stereotypes, reprehensive attitudes, and assumption of false beliefs. A recent study has focused on sexism and risky sexual behaviors in adolescents, depending on gender [40]. This study concludes that men have more hostile and sexist beliefs than women. Another recent study carried out on young university and pre-university Norwegian students has found that men showed more sexist attitudes and

hostility than women [6]. Also, this study has correlated some variables of HS in pre-university male students to attitudes of rape.

Regarding the effects on new technologies on adolescents, a study has detected that those adolescents who spend the longest time playing video games show more sexist attitudes [5]. In this sense, an interesting research on sexism in digital games show that user experience of sexist games is not predicted by player gender, but by the player's pre-existing beliefs about gender [8]. However, there are some works studying how apps change attitudes in different aspects (although not in the particular field of sexism prevention) [14, 47]. Related to the access mechanism to digital information, smartphones have been identified as one of the items with the greatest influence on adolescents [35], since they use this device for auto-determination, self-empowering, and establishing virtual relationships. Similar results are confirmed by qualitative research described in [33], which results indicate that teenagers currently use their phones to access sexual health information due to ease of access and privacy.

Also, the downloading of mobile apps represents another attraction for adolescents [24]. The fact that the mobile phone is the best friend of adolescents is interesting, because it allows the educators to access their interests in a socializing way, but adding an educative aspect [45]. In the same line of playful apps with educational connotations, a study [18] incentivizes the use of apps for increasing the pro-social conscience, since they are complementary with developmental psychology, and they provide teenagers with positive behavior patterns.

Thus, although some initiatives have been developed to provide teenagers with educative apps in this context, only few of them have been assessed properly. Girl Talk [9] is a free and stand-alone smartphone application containing comprehensive sexual health information including mental health, body image, breast health, healthy lifestyles, and reproductive health. Detectamor [13], and some other mobile apps available at the Apple/Play Store such as Sheboard [1], which is a standalone mobile app with multiple option games, which explore knowledge and skills to find a partner within what is considered as true love (without sexism). Myplan [2] was developed for helping college women who are victims of hostile sexist behaviors, including a section of family resources, which helps in the decision-making as a social support.

The latest work proposes and evaluates the effectiveness of an app (called EMOTIC) designed as a tool for psychological intervention on teenagers [11]. Nevertheless, to the best of our knowledge, there are no contributions neither in the specific topic of effectiveness of apps as intervention tools for sexism prevention in teenagers, nor in the topic of the usability evaluation of these apps. Thus, this is the first work where a quantitative research approach compares the electiveness of a custom mobile application with a more traditional intervention like a workshop about sexism for teenagers. Authors determined the application's desirability and appeal among teenage girls, and the results show that the app potentially connects teenage girls to more information about sexual health than traditional methods. Moreover, the participants recommended the application as a valuable resource to learn about comprehensive sexual health.

3 METHOD

The purpose of the Liad@s app (in terms of playability) is to complete four targets (identified by colors) in the player board, achieving the maximum number of points as possible. The colors of the targets are randomly assigned at the beginning of each game from a set of eight possibilities, where repetitions can occur. In order to complete these targets, the player should play a roulette wheel located at the center of the screen. The roulette is composed of twelve positions, eight of which can be identified by the colors of the targets. If the player falls in any of these positions, he/she will have to pass a small test or mini-game about skill, ability, and reflexes, which will force the player to reflect and take a stand on situations and experiences related to intimate relations and sexism

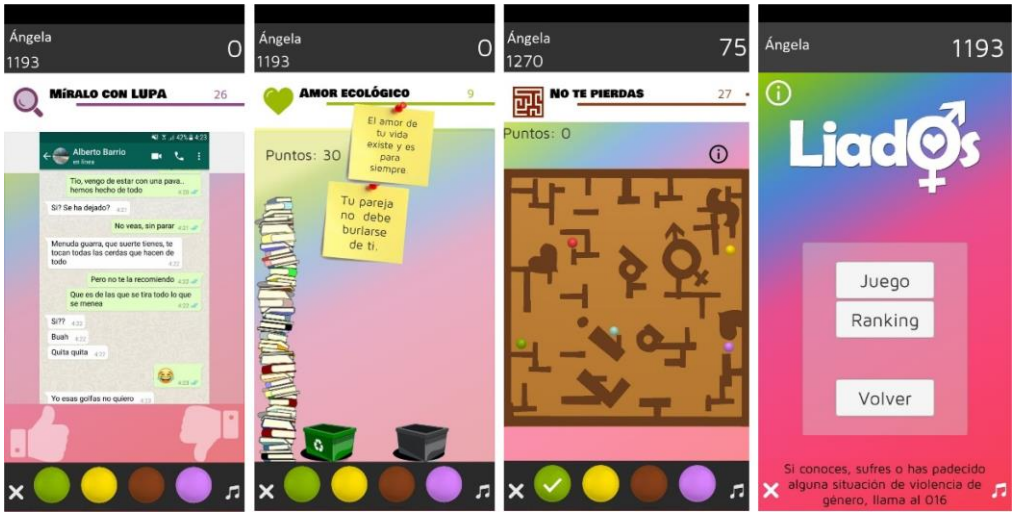

Fig. 1. Example snapshots of some mini-games of the app, as well as the screen showing the general score.

in adolescence. If the player pass the test, he/she completes the corresponding color in the player board. Depending on how fast the player pass the test, the player will get some points, which will be added to his/her score (which is reset at the beginning of the game). The other four positions in the roulette wheel are positions where the players can loose, duplicate, and add their score or be randomly re-directed to any of the eight mini-games to go on with the game. In order to play Liad@s app, it is necessary to complete a previous registering process, which allows that every player is authenticated in the system. Thus, when a player successfully finishes a game (he/she has completed the four targets in his/her player board) the score in that game is added to the general score of the player (an action named consolidation). This mechanism allows to significantly increase the gameplay of the app, since the players can play an unlimited number of games to increase their score, consult their score in local, national or international rankings, challenging another players or comparing their scores. In our experiment, all the general scores were reset and the payers were given a maximum time of 25 minutes to increase their general score. Figure 1 shows some snapshots of the mini-games, which can be played in the app, as well as the screen showing the general score.

Next, we describe the experiments carried out by the participants to validate the hypotheses stated in the introduction. This study with minors was approved by Dirección General de Infancia y Adolescencia, Conselleria de Igualdad y Políticas Inclusivas de la Generalitat Valenciana (Childhood and Adolescence Directorate, Valencian Regional Government, Spain) in a prospective way, several months before the study was carried out. Also, we obtained consent from parents or guardians before the study, which was conducted in June 2018.

The study contains the following limitations: first, the sample size is limited, which has prevented the authors to make comparisons with wider samples with different levels of gameplay. Second, is the ignorance of the playability in children of younger age. Third, we neither assess the benefits in gender violence derived from the use of the app during a long time in boys (as possible aggressors) nor in girls (as possible victims). Another important limitation is the rigid rules about the use of mobile phones in official teaching institutions, which prevents that this study can be extended to a longer usability within teaching institutions. Finally, another important limitation is the absence of playful apps for reducing sexism, which limits the discussion of the existing literature on this topic.

We have used *probability sampling* to ensure that all members of the group of interest have an equal chance of being selected to participate in the study. To that end, one random day a week of the month of February 2018 our research team presented the experiment in each of the five groups (morning and afternoon shifts, each of them including 25–30 students) at the Gonzalo Anaya Secondary School, located in Xirivella (Valencia), Spain. In this presentation, the students were invited to register in a web-based event organizer platform. This *simple random sampling process* [25] was done without replacement, and it provided fourth grade students to the experiments. The sampling process finished when almost 60 participants registered in our web-based event organizer platform. In addition, we rejected some random participants and some others who had experience in using similar mobile apps, such as Girl Talk, Detectamor, or Sheboard, to ensure an unbiased sample population. Later, the final group of participants were randomly assigned to two groups, denoted, respectively, as A and B in the experiment.

The initial group was composed of 45 students (35 boys and 10 girls) with average age of 15.01 years, with a standard deviation of 1.41 years, corresponding to the fourth year of the Spanish secondary school. Although it appears to be gender representative bias (there are more males (32) than females (10) in the selected sample) our intervention was carried out within an official training cycle of matrices and moulds, where the Gender Parity Index is only 10%. A detailed 10 minute explanation was given to this group, explaining each stage of the whole experiment. After this talk, the students were asked to sign an express consent to participate in the experiment.

At this point, the participants filled in a PreTest questionnaire, and they were randomly aligned to two groups (denoted as A and B, respectively), balancing the number of girls per group. Group A participants started the procedure using the "Liad@s 2.0" app [37]. An explanatory YouTube video-clip was shown on the participants smart-phone, and after this video a session was carried out to solve any question related to the app functionality. The participants then had 5 minutes of free game with the app to register in the game and get familiar with each of the associated mini-games included in the app. Once this warm-up stage was ended, the participants were given a maximum time of 25 minutes to get as many points as possible with the mini-games as the previously registered user. Then, they filled in a PostTest1 questionnaire (which among other items it recorded the punctuation reached by the participant), as well as the satisfaction questionnaire.

After a 10 minute break, Group A participants received a 30-minute educative talk about sexism. As part of this talk, the concept of ambivalent sexism was explained and related to BS and HS. This talk also highlighted in which consisted the expressive feeling related with couple relationships myths, the attitudes and believes influencing healthy couple relationships in front of risky relationships, specially highlighting the control against the partner, jealousy, manipulation, and so on. After this talk, Group A participants filled in the PostTest2 questionnaire, finishing the experiment.

Group B participants inverted the order of the activities. They first attended the educative 30-minute talk about sexism, then they filled in the PostTest1 questionnaire, then they played with the Liad@s app as described above, completing the experiment by filling the PostTest2 and user satisfaction questionnaires.

It must be noted that during the experiment 5 of the participants (4 of them in Group A and 1 in Group B, all of them boys) either they did not sign the consent form or the technical staff detected that they were arbitrarily answering the questions in the questionnaires. Due to these reasons, the answers of these 5 participants were discarded, finally obtaining 20 valid answers in Group A and 17 valid answers in Group B for each of the questionnaires.

The threats both the internal and external validity of the experiments [41] were checked and minimized in the following terms: in order to minimize the threats of selection and diffusion of interventions, we randomly selected Groups A and B of the counterbalanced design through a

random sample of ESO fourth course students of a public high school randomly selected from the network of high schools belonging to Generalitat Valenciana, the regional government with education competences. In order to minimize the threat of maturation, both Groups A and B were randomly populated with students whose age, gender, maturity, and education level were checked as similar. Our research uses questionnaires as evaluation instrument. The literature shows many contributions proving the validity and reliability of both the questionnaires used for obtaining measurements of the knowledge acquired about ambivalent sexism [15] and the usability/satisfaction of the application [48]. In order to minimize the threat of the instrument, all the PreTest and PostTest in the counterbalance design were delivered with adequate instructions and deadlines. The history and test threats were minimized by custodying both the PreTest and PostTest during eight weeks. Finally, in order to limit the threat of inconsistent data collection, the teacher taught the talks about sexism to both groups of the counterbalance design, and the same three technicians were used to solve any technical doubt during the exposition and the training period with the mobile app. Regarding the threats corresponding to the external validity of the experiments, the threat of selection bias was mitigated by assigning the participants to the groups of the experiments randomly. Moreover, we used the counterbalanced design in order to minimize the effect of having taken the same test previously and to control for the testing effect. The threat of the Hawthorne effect was controlled by including the experiments as a part of students' classes in their academic scheduling, conducting our study in an unobtrusive way. Additionally, the threat of experimenter effect was diminished by including a technical staff composed of four experts in intervention with adolescents, who conducted the experiments and interacted with the participants. Finally, the threat of aptitude-treatment was softened by selecting participants from the same class (fourth year of the Spanish secondary school).

4 EXPOSITION

The objective of the study was to compare the effectiveness of the updated app Liad@s 2.0 in front of traditional interventions at high schools like a workshop about sexism for teenagers. Also, we evaluate the usability of the app and the user satisfaction with this application, using an initial population of 45 users.

4.1 Participants

The initial group was formed by 45 students, in order to obtain statistically significant results [20]. It was formed by 35 boys and 10 girls, with average age of 15.01 years and a standard deviation of 1.41 years, corresponding to the fourth year of the Spanish secondary school. All the students were enrolled in the Gonzalo Anaya Secondary School, located in Xirivella (Valencia), Spain. A detailed 10 minute explanation was given to this group, explaining each stage of the whole experiment. After this talk, the students were asked to sign an express consent to participate in the experiment. Five of the participants refused to sign or were discarded, because they were caught completing the questionnaires arbitrarily. Thus, they could not finally participate, being 40 people the initial group, which started the experiments. As stated before, three of the participants were further discarded due to their non-collaborative behavior in the experiment or their experience in similar apps. They were randomly assigned to two groups (denoted as A and B, respectively), balancing the number of girls per group.

The initial group of participants consisted of 45 participants, where 35 of them were boys (77.78%) and 10 of them were girls (22.22%). Although the ratio of girls and boys in Spanish secondary schools (term denoted as GPI, standing for Gender Parity Index) is roughly equal, as stated by Institute of Statistics of the United Nations Educational, Scientific, and Cultural Organization (UNESCO) (https://data.worldbank.org/indicator/se.enr.prsc.fm.zs), the difference between

the ratio of boys and girls in the group of participants has two reasons. First, the participants who where discarded from the group due to randomness or having experience in similar applications (such as Girl Talk, Detectamor, or Sheboard) were all girls. Second, the experiments were developed during the classes of an optional subject in the Spanish study plan called "technical drawing". This is a subject enrolling a high ratio of male students. Nevertheless, it should be noted that the results of the experiments carried out show that the gender of the participants does not have any effect neither in the knowledge acquired regarding ambivalent sexism, nor with the usability of the mobile app and the user satisfaction with this application. This type of conclusions, where there is no relationship between the results obtained with respect to the gender of the participants, are similar to those obtained in previous interventions completed by the authors in previous works [36, 37].

4.2 Measurements

Different metrics were measured during and after the activities described in the Method Section. The measurements came from the participants and from the staff (observers) through questionnaires that they should fill out at different moments, as described below. In particular, we used four questionnaires:

- PreTest: This questionnaire allows to determine the degree of the participant's ambivalent sexism before carrying out any intervention, and it directly derived from well-known questionnaires on this subject [15]. In particular, this questionnaire has been validated with Spanish population [30]. It contains some general statements about girls/boys, asking the user to determine, within a Likert scale ranging from 0 to 5 [42], his/her level of agreement with each statement.
- PostTest1: The participants filled in this form after using the first method. The questionnaire contains the same questions appearing in the PreTest questionnaire.
- PostTest2: The participants filled in this form after using the second method. The questionnaire contains the same questions appearing in the PreTest questionnaire and an additional free text field where the participants could express their personal opinion about the overall experiment and the mobile application.
- Usability and Satisfaction: The participants filled in this questionnaire after using the app. In this questionnaire, the participant expressed his/her opinion (determined as the level of agreement within a Likert scale ranging from 0 to 6) about different aspects of the application, like comfort factors, and so on.

Since the three questionnaires PreTest, PostTest1, and PostTest2 measure the level of ambivalent sexism, they contain the same questions. The only difference among them is that the participants filled each of them at different moments of the procedure. Table 1 shows these questionnaires. Table 2 shows the usability questionnaire.

4.3 Procedure

As described above, the participants in the study are split into two groups, denoted as A and B. The reason behind this separation is to check if the order in which the interventions are carried out has an effect on both the knowledge acquired or the perception of the app analyzed. The procedure followed by the participants is illustrated in Figure 2.

Concretely, the protocol followed by the participants is the following one: first, all participants complete the PreTest. At this point, Group A participants use first the Liad@s app, obtaining their score. Then they perform the PostTest1 and fill the satisfaction and usability survey about the app. Next, they attend the school-based education intervention and perform again the PostTest2.

Table 1. Questions about Ambivalent Sexism in PreTest, PostTest1, and PostTest2 Questionnaires

Q1	Boys are physically superior to girls.
Q2	Boys should control with whom their girlfriends have relationships.
Q3	Girls should help their mothers at home more than boys.
Q4	Girls do better in domestic tasks, while boys are better skilled in repairing things.
Q5	Girls know how to get what they want from boys.
Q6	Sometimes girls use the fact of being female for being treated differently from men.
Q7	When girls are beaten by boys in a fair competition, they usually complain that they have been discriminated.
Q8	Girls are very easily offended.
Q9	Girls usually interpret innocent comments as sexist.
Q10	Girls usually exaggerate their problems.
Q11	With the excuse of equality, girls expect to gain more power than boys.
Q12	Boys should accompany girls to their home at night, to avoid that something happens to them.
Q13	Girls should be loved and protected by boys.
Q14	Boys should take care of girls.
Q15	A good boyfriend should be ready to sacrifice things he likes in order to please his girlfriend.
Q16	In case of catastrophe girls should be saved before boys.
Q17	Girls are more sensitive to other people's feelings than boys.
Q18	Dating a girl is important for boys.
Q19	Couple relationships are essential for reaching true happiness.
Q20	A boy may feel himself as incomplete if he is not dating a girl.

The procedure followed by Group B participants is the opposite one: they attend the school-based education intervention and perform the PostTest1, and after that they use the Liad@s app, obtaining their score and performing afterwards the PostTest2 and filling the satisfaction and usability survey.

This procedure was successfully completed by 37 participants. It should be remembered that, although 45 participants initially started the experiments, 3 of them were randomly removed (or due to their previous experience in similar apps) and 5 of them did not sign the consent form or the technical staff detected that they were arbitrarily answering the questions in the questionnaires.

5 RESULTS AND DISCUSSION

In this Section, we analyze and discuss the data obtained about the participants learning as well as the participant's satisfaction with the mobile app. All the data have been collected using questionnaires. We have used the statistics program "Statistical Package for the Social Sciences" (SPSS version 24.0) (https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software). For all of the results shown below, all significance tests were two-tailed and conducted at the 0.05 significance level. Also, we have included in the results metrics related to the effect size of the experiments performed, in terms of Cohen's d and Eta-squared (η^2). These type of metrics have yielded valuable information about the size of the samples when comparing the results of our experiments. This property is not evaluated by the p-values. All the underlined values in the tables shown below indicate statistically significant differences.

Table 2. Questions about the Liad@s APP in Usability Questionnaire

Q1	The use of the application did not require a great mental effort.
Q2	The information displayed in the screen was adequate.
Q3	The information displayed in the screen was easy to read.
Q4	The information displayed in the screen was clear.
Q5	The use of the smart-phone did not require a great effort of the arms.
Q6	The use of the smart-phone was comfortable for my hands and arms.
Q7	The handling of the smart-phone was easy.
Q8	I haven't felt any dizziness during the experiment.
Q9	My hands and arms did not get tired.
Q10	It was easy to control the application.
Q11	At no time did it seem to me that the smart-phone was going to fall out.
Q12	The handling of the application was uncomplicated and simple.
Q13	The application reacted properly to my actions.
Q14	The handling of the application was natural.
Q15	I did not notice delays between may actions and the expected results.
Q16	I got used to the application at once.
Q17	The application was easy to use.
Q18	I think that I've learned concepts and ideas about sexism with this app.
Q19	I would like to use a mobile game like this to learn other things.
Q20	I liked this experience.
Q21	I have been concentrated on the tasks to be done, not in the smart-phone.
Q22	I have felt motivated during the experience.
Q23	I have felt as expert on the application at the end of the experience.
Q24	Mark the application.

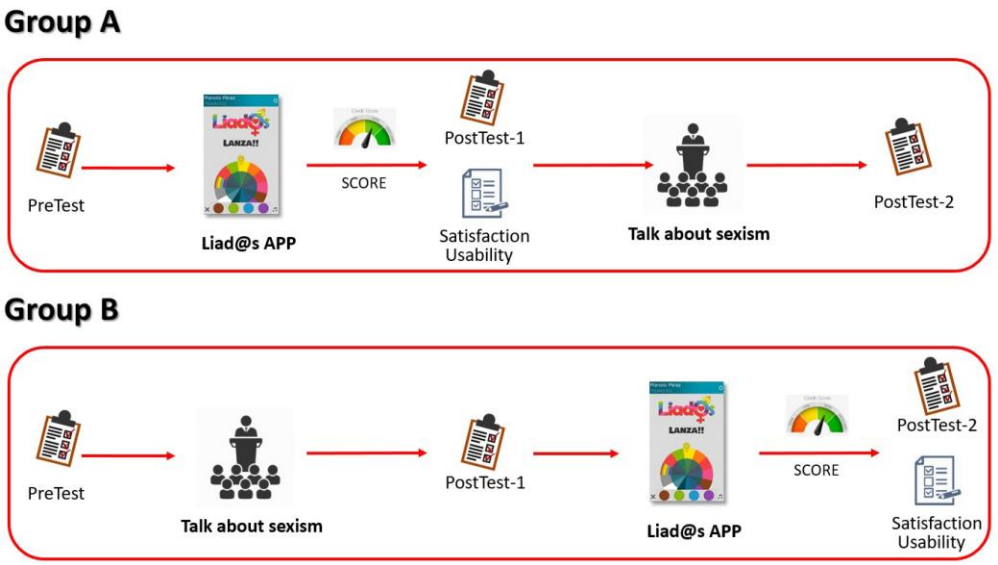

Fig. 2. Procedures carried out by each of the groups.

Table 3. Comparison of the Learning Achieved by the Participants

	Avg. ± Dev.		Avg. ± Dev.	T	p	d
EA	1.77 ± 1.44	EB	1.71 ± 1.55	0.26	0.71	0.07
EA	1.77 ± 1.44	O1A	1.43 ± 1.22	−1.16	0.002	2.37
EA	1.77 ± 1.44	O2A	1.19 ± 1.08	2.01	0.003	0.34
O1A	1.44 ± 1.28	O2A	1.19 ± 1.08	1.07	0.001	0.21
EB	1.72 ± 1.55	O1B	1.50 ± 1.54	−0.29	0.008	0.35
EB	1.72 ± 1.55	O2B	1.26 ± 1.44	1.16	0.002	0.33
O1B	1.50 ± 1.54	O2B	1.26 ± 1.44	1.70	0.006	0.14
O1A	1.44 ± 1.28	O1B	1.50 ± 1.54	0.10	0.301	−0.05
O2A	1.19 ± 1.08	O2B	1.27 ± 1.44	−2.11	0.24	0.14

First, we have analyzed if the collected data follow a normal distribution. All data collected in the experiments follow a normal distribution, and therefore we have used parametric tests: the t-test and the Cohen's test for paired and unpaired data, as well as multifactorial ANOVA for analyzing relationships among different parameters involved in the experiment. As a representative example, the Kolmogorov–Smirnov test [32] ($D = 0.2587$ and p-value = 0.1998), Anderson–Darling test [4] ($A = 0.5471$ and p-value = 0.2580), and Shapiro–Wilk test [44] ($W = 0.3874$ and p-value = 0.3698) confirmed that the data collected from the Post1A questionnaire follow a normal distribution. The same happened in the rest of the datasets.

We analyzed the data in questionnaires PreTest, PostTest1, and PostTest2 in order to evaluate the participants' learning about ambivalent sexism when attending the talk and using the mobile app. For each of the questionnaires, we created a "knowledge" variable summarizing the answers given for each test, denoted as ASI score [15]. An exhaustive analysis of the tests results is required (comparing the average±standard deviation values obtained in PreTest, PostTest1, and PostTest2 results of Group A to the ones of Group B) to study the effectiveness of each learning method (talk or mobile app). Table 3 shows the learning results using both tests, as well as the significance level, for all the datasets for both participants groups. The underlined values mean statistically significant differences. Due to space limitations, we have coded the labels denoting each test, so that "PreTest" is coded as an E, and "PostTest" is coded using the prefix O. Thus, the label denoted as EA means PreTest A, and O1A means PostTest1 A. In this Table, the column labeled as "T" denotes Student's T parameter, and the column labeled as "d" denotes Cohen's d values.

First row in Table 3 shows that the starting level of knowledge of both groups (A and B) is similar, since there are no significant differences in the average values ($T = 0.258$, $p = 0.714$). Also, this table shows that both groups significantly improved their knowledge when using the first learning method, since the second and fifth row show statistically significant differences. However, the level of knowledge acquired by both groups can be considered as similar, since there is no a significant difference when comparing PostTest1 A with PostTest1 B results (last but one row). Similar results are obtained for the second learning stage with a second learning tool. Both groups significantly improve their knowledge. However, the last row in the table shows that there is not a significant difference in the knowledge acquired after having used both learning tools. These initial results suggest that both intervention instruments (APP or talk about sexism) produce a significant improvement in the learning achieved by the participants. Both instruments attract the attention of the participants, who add new information about sexism in teenagers to their knowledge. Also, once the first procedure stage has finished (01 A vs 01B) this new knowledge is similar, in spite it has been achieved through different intervention tools. Therefore, we can conclude that both methods can be considered as valid for teenager ambivalent sexism learning.

Table 4. Mixed Design ANOVA for the HS Scores of the Pre-Test and Post-Test (1 and 2) Questionnaires

Variable	Sum sqr	df	F-value	p-value	η^2
Gender	125.369	1	2.301	0.002	0.236
Age	25.154	2	3.369	0.343	0.157
SE status	39.254	5	1.205	0.156	0.359
Family status	87.239	5	2.364	0.436	0.157
Number of B/S	69.145	6	0.216	0.536	0.894
Other interactions	41.687	1	0.369	0.096	0.354
Procedure	234.361	1	0.581	0.007	0.287

Table 5. Mean and Standard Deviation Values for the Gender and Procedure Variables in Mixed Design ANOVA for the HS

	Gender	
	Mean	Std. dev.
Boys	1.34	0.43
Girls	0.45	0.25
	Procedure	
	Mean	Std. dev.
App	0.48	0.29
Talk	1.53	0.67

Next, we evaluated separately the data collected from both HS and BS questionnaires. Table 4 shows the mixed design ANOVA for the HS scores of the Pre-Test and Post-Test (1 and 2) questionnaires. This Table shows two variables where significant differences appear: gender and the order of questionnaires (this latter one labeled as "Procedure"). That is, significant differences in learning have occurred for different gender groups (boys and girls) and for groups A and B.

It must be noted that the number of data subsets in each variable is the value in the "df" column (which stands for *degree of freedom*) plus one (e.g., the df value for the "Gender" row is 1, since there are two genders, male and female). That is, there is a high number of subsets. Although showing the mean and standard deviations of each level (subset) in the differentials would facilitate a more meaningful understanding of the levels of differences observed, it would also add a huge amount of values, making the results difficult to be analyzed. Instead, we shown in Table 5 the mean and standard deviations only for those variables where significant differences appear (the underlined variables), gender and procedure.

Table 5 shows that the interventions related to HS score shows a higher average effectiveness on the female population of the study, in front of the one for the male population. Later we will verify if both groups started from similar levels of knowledge about HS. Also, Table 5 shows that the average effectiveness using the app is higher than the one achieved with the talks. The intervention effectiveness of the mobile app (in terms of HS score) with respect to the talks about sexism could be explained by the power of seduction that these mobile applications have over this population segment. This attraction power achieves that the intervention sessions take longer than the talks about sexism, which makes the participants to better work and integrate the new contents related to HS, showing higher values of HS score.

Table 6. Mixed Design ANOVA for the BS Scores of the Pre-Test and Post-Test (1 and 2) Questionnaires

Variable	Sum sqr	df	F-value	p-value	η^2
Gender	167.365	1	1.987	0.008	0.548
Age	36.664	2	2.369	0.037	0.698
SE status	54.236	5	0.265	0.356	0.541
Family status	97.105	5	0.419	0.214	0.369
Number of B/S	25.691	6	0.335	0.198	0.254
Other interactions	74.215	1	0.878	0.597	0.397
Procedure	199.587	1	0.558	0.003	0.167

Table 7. Mean and Standard Deviation Values for the Gender, Age, and Procedure Variables in Mixed Design ANOVA for the BS

	Gender	
	Mean	Std. dev.
Boys	1.37	0.51
Girls	0.52	0.30
	Procedure	
	Mean	Std. dev.
App	0.45	0.31
Talk	1.63	0.59
	Age	
	Mean	Std. dev.
12–13	1.74	0.47
14–15	1.16	0.33
16–17	0.61	0.28

Table 6 shows the mixed design ANOVA for the BS scores of the Pre-Test and Post-Test (1 and 2) questionnaires. In this case, three variables show significant differences: gender and learning procedure (Group A or B). This Table shows that there are three variables where significant differences appear: gender and the order of questionnaires, like in the case of HS, plus age. That is, in this case significant differences appear also in the learning acquired by certain age segments.

As in the case for the HS, Table 7 shows the mean and standard deviation values for those variables where significant differences have appeared: gender, age, and procedure.

The data shown in Table 7 follow a very similar pattern to the data shown in Table 5 about HS. Nevertheless, it must be noted the higher average effectiveness of the interventions on female teenagers, and also in the utilization of the mobile app as intervention tool. Also, Table 7 shows that the interventions seem to show a higher effectiveness as the age of the population increases. The cause of this trend could be that the the concepts related to BS are usually more difficult to be identified and understood than the ones of the HS, which are much more evident in the current society [15]. It could also explain why the BS score increases with the age of participants.

Since Tables 4 and 6 show that there is a significant correlation in the procedure and Table 3 shows that both groups significantly improve their knowledge, we can conclude that these results validate our primary hypothesis. However, both tables above show that there is a significant

gender correlation in both HS and BS, and we have studied the scores obtained by these groups. Figure 3 graphically shows, using a box-plot representation, the scores obtained with the app, grouped by gender. The authors have intentionally decided to keep the extreme data of the box plots (outliers and extreme outliers) shown in Figures 3 and 4. In the former case, the outliers show the exceptional score obtained by three female participants in the experiment. Although the objective of the game is to complete a set of challenges (represented as mini-games) until completing all the portions of a cheese (with a game mechanic similar to Trivial Pursuit), the experiment allowed the participants to accumulate points indefinitely, by playing again to the mini-games, until the game time (25 minutes) ended. The addictive features of some mini-games and the skill of these participants resulted in remarkable scores. In the latter case, we found some participants who obtained excellent evaluations in the PreTest stage, in terms of ASI score. Consequently, the valuation of this knowledge after the PostTest1 and PostTest2 stages was also much higher than the average, in general terms, and were highlighted in the boxplot as an outlier. Finally, we analyze the origin of the extreme outlier caused by a low score (in terms of ASI score) obtained by a female participant in the PreTest stage and shown in this figure. This participant was part of a group of girls of gypsy origin who takes classes at the high school where we completed the intervention/experiment. In general terms, female participants have been very interested in participating and collaborating during all the interventions carried out, unlike male participants. This interest in these age ranges could be due to, among other causes, the different education campaigns about sexism that most of Spanish secondary schools have carried out inside integral programs for teenager student women . These campaigns focus on women, encouraging them to acquire knowledge about the social problems surrounding sexism in teenagers.

Figure 3 shows that women (mean$\pm$ std_dev) played much better than men (mean$\pm$ std_dev), obtaining much higher scores. Also, three singular points arise. The singular points depicted as an asterisk (*) indicate extreme cases where the value is located further than three times the height of the 75-percentile box. The singular point depicted as a circle (o) indicates outlier values, which are located further than 1.5 times the height of the 75 percentile box. These results suggest that the interventions are being carried out in the correct moment, since they start from male and female populations with similar levels of ASI-score. Nevertheless, due to the reasons explained above, female participants acquire a higher level of knowledge than the male ones for both type of interventions.

Next, Figure 4 shows the global knowledge acquired by men and women along the undertaken procedure. This Figure shows that the different starting level of ambivalent sexism between men and women was significant, but it is significantly reduced along the procedure. It is also worth mention the different results in the Post1 questionnaire for men and women. Regarding other prior studies, similar differences between gender appeared in some studies, which conclude that the majority of players of video games and apps with aggressive connotations are male [28, 46].

Next, we focus on the results obtained for the user preferences. Figure 5 results (separated by groups) for the first additional question. Group A was formed by 20 users (54% of the population), and 3 of these users (15% of Group A) preferred the talk about sexism as the learning method, while 17 of them (85% of Group A) preferred the mobile app. Group B was formed by 17 users (46% of the total population), and from these 17 users, 2 (11.76% of Group B) preferred the talk and 15 (82.24% of Group B) preferred the app as the learning method. If we study these results by the selected system, from the 37 users, 32 (86.49%) preferred the mobile app. From these 32 users, 17 were in Group A and 15 were in Group B. From the 5 users (13.51%) who preferred the talk, 3 of them were in Group A and the other 2 were in Group B. The preference of the participants for the mobile app in front of the talk about sexism can be explained in the light of the comments the participants

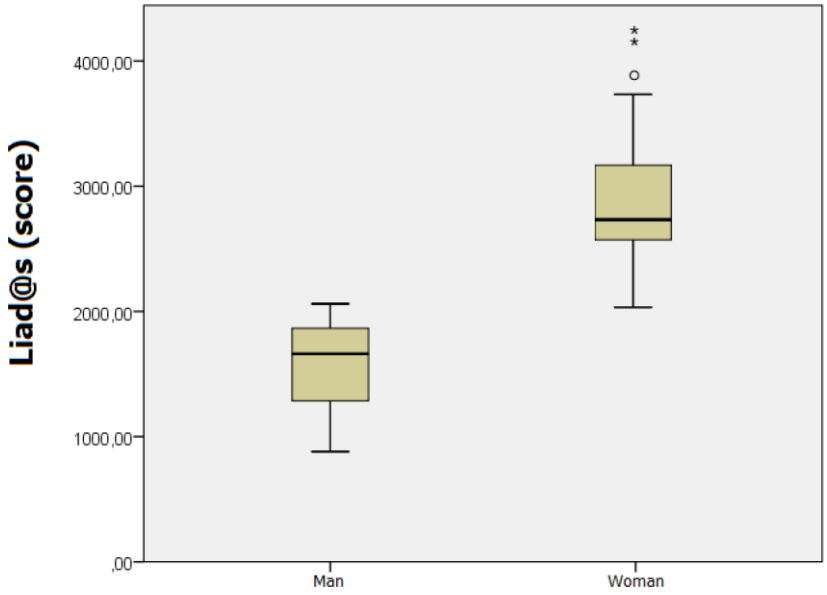

Fig. 3. Box-plot representation (by gender) of the global scores obtained in the app.

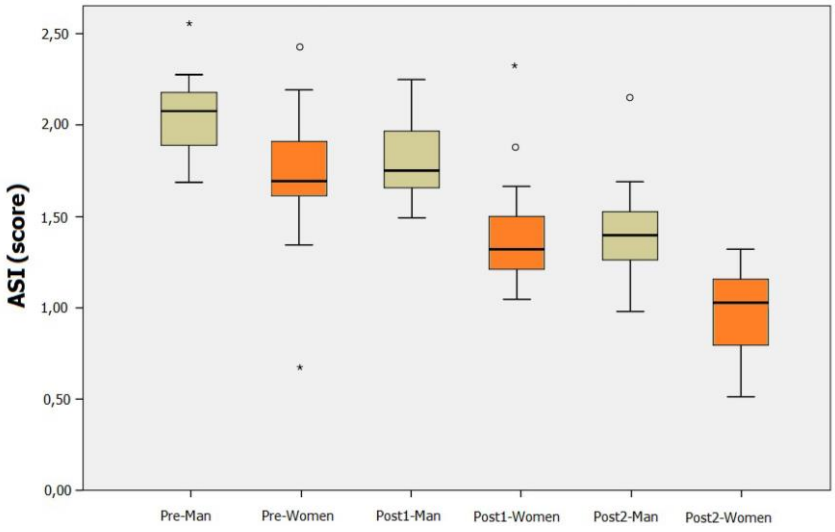

Fig. 4. Box-plot representation of the global knowledge acquired by gender.

themselves made to the staff, and also in the comments written in the space for free text comments in the questionnaires. These comments reflected the attraction and seduction these devices project over the target population. Although the participation of the students was encouraged in the talks, allowing questions to the staff about any kind of question, the interventions through the use of the Liad@s APP showed that the participants preferred to autonomously discover concepts related to sexism among peers through the use of the mobile app. These results strongly validate our secondary hypothesis.

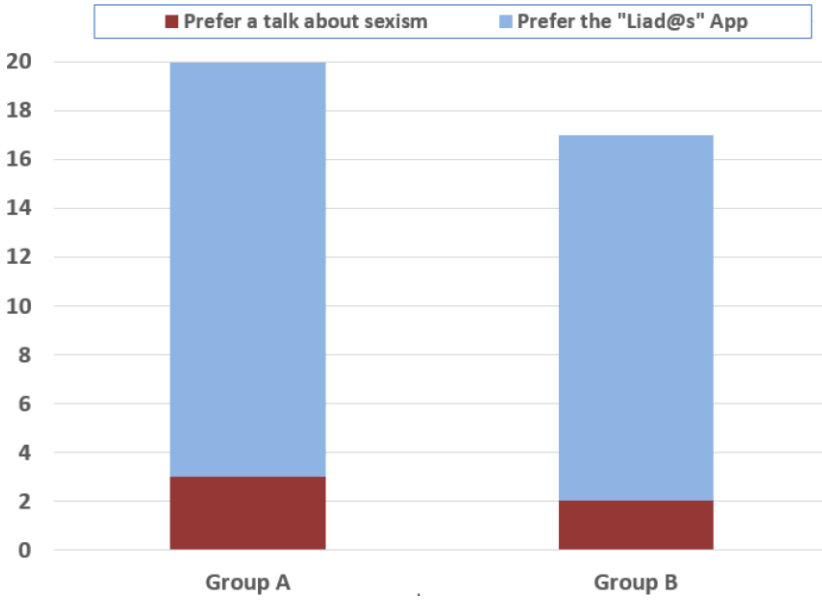

Fig. 5. User preferences.

Apart from the knowledge acquired by teenagers about sexism prevention with the app, and given that they have used the app, we also want to study different aspects of the usability of the app, and check if the user satisfaction with the app is at least equal than with the traditional workshop. Thus, we analyzed the usability and satisfaction of the participants with different aspects of the mobile app developed, starting from the answers in the usability questionnaire. However, it must be noted that this study does not provide theoretical (usability evaluation studies) or practical (industry) general significance.

Although the satisfaction questionnaire is not shown here for the sake of shortness, the factors considered in the usability questionnaire (as an adaptation of the WSPQ questionnaire [49] and containing twenty-three questions) were the following ones: Eight questions related to **Control Factors** (**CF**): Degree of control, immediacy of control, anticipation of events, mode of control, and physical environment modifiability; Four questions related to **Sensory Factors** (**SF**): Sensory modality, environmental richness, multi-modal presentation, consistency of multi-modal information, degree of movement perception, and active search; We have also added five questions related to ergonomics (EF): devices comfort when using, effort, and so on, and six other questions (OF), scoring different aspects related to the experiment: perception, usefulness, and so on.

The Kolmogorov–Smirnov test, ($D = 0.12547$, p-value = 0.3525), Anderson–Darling test ($A = 0.43522$, p-value = 0.7541), and Shapiro–Wilk test ($W = 0.8708$, p-value = 0.3399) performed on the usability questionnaire indicated that these data follow a normal distribution, and therefore we have used parametric tests (the t-test and the Cohen's test for paired data and the ANOVA test).

Table 8 shows the mean values and the standard deviation obtained for each of the four factors considered from the answers to the 23 questions. These answers showed that the user satisfaction with the app was high, being the average value above 4 (in a Likert scale from 0 to 6, since all sexism questionnaires starts from 0) in all the considered factors. These results were compared to similar experiments for other applications using the same kind of questionnaires [31, 34]. The results of this comparison, due to the high values shown in Table 8, seem to indicate that both the design

Table 8. Scores Obtained in the Usability and Satisfaction Questionnaire

Variable	Avg.	Dev.
Global Score	4.26	1.25
SF	4.84	1.36
EF	5.12	0.98
CF	4.15	1.35
OF	5.36	0.62

Table 9. Mixed Design ANOVA for the Gender Scores of Usability and Satisfaction Questionnaire

Variable	Sum sqr	d.f.	F-value	p-value	η^2
Global Score	250.365	12	3.022	0.001	0.745
SF	98.362	12	0.256	0.125	0.698
EF	71.325	12	1.598	0.174	0.264
CF	652.069	12	1.319	0.359	0.305
OF	127.653	12	1.402	0.003	0.259

of the mobile APP and the content presentation have been very well rated by the participants, in terms of usability.

The ANOVA tests indicated significant differences in gender and age. Table 9 shows the participants satisfaction depending on the gender including both the general satisfaction with the app and the satisfaction grouped by types of questions. This table shows that there exists significant differences in gender only in the global score. Women scored the app higher than men. Similarly to the results shown above in terms of AS-score, the motivation and will of the female participants for acquiring new knowledge about ambivalent sexism has achieved to dissociate the average results obtained in the game matches. The players were allowed to improve their scores by repeating the matches of the game, and it was observed in the female participants a higher proactivity, attention, and skills in the playing of the mini-games of the Li@ados APP. These differences led to the bias shown in Table 9 for the gender variable.

Table 10 shows the participants satisfaction depending on the age, including both the general satisfaction with the app and the satisfaction grouped by types of questions. This table shows that there are significant differences not only in the global scores, but also in other factors. Surprisingly, the use of Tukey post-hoc tests revealed that the age range of 13–14 years provided completely different answers to the rest of age ranges at p-value <0.05. The reason for these differentiated results in this age range could be that the age of 13–14 years correspond to the stage of personal growth when they start to explore the unknown, without direct supervision, as it happened in their childhood. The first couple relationships appear, mainly for the female population, which would justify the results shown above. It is an age where parents start to concede certain autonomy, trusting that it will make them responsible. That is, the behavior at this age may not be completely aligned to other immediately earlier or later age ranges.

Finally, Figure 6 shows (as a directed graph) the correlation analysis for the responses given by the participants. The results of this analysis include the correlation factor and the significance level p. The correlation coefficient is calculated as a number between −1 and 1, with 1 being the strongest possible positive correlation and −1 being the strongest possible negative correlation. A positive

Table 10. Mixed Design ANOVA for the Age Scores of Usability and Satisfaction Questionnaire

Variable	Sum sqr	d.f.	F-value	p-value	η^2
Global Score	47.136	13	1.251	0.021	0.366
SF	851.219	13	2.364	0.015	0.001
EF	61.201	13	0.297	0.584	0.025
CF	64.257	13	1.364	0.364	0.135
OF	222.213	13	2.205	0.027	0.843

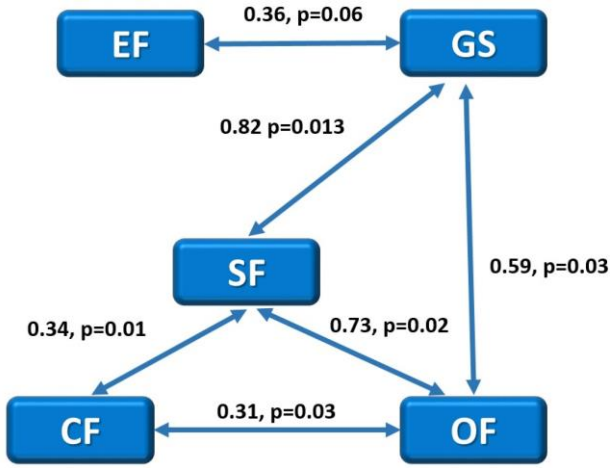

Fig. 6. Significant correlations among the responses given.

correlation means that as the values obtained in the answers for these factors increase, the values obtained for the other factor will also increase. The p-value evaluates how well the results data rejects the null hypothesis, which states that there is no relationship between two compared groups. Successfully rejecting this hypothesis tells you that the results may be statistically significant. In academic research, p-value is defined as how likely it is that we would obtain the results (or more dramatic results) we did assuming that there is no correlation or relationship among the subjects. A significance cutoff, known as the alpha value, is specified to check whether the p-value is significant or not. Alpha is usually set to 0.05, meaning that the probability of achieving the same or more extreme results assuming the null hypothesis is 5%. Therefore, Figure 6 shows that the obtained score (denoted as GS) is strongly correlated to all the considered factors except the **ergonomics factor** (**EF**). This correlation (especially noticeable between the obtained score and the sensory factors, $r=0.82, p=0.013$) is also supported by many personal opinions of the participants in the experiments, collected in the PostTest2 questionnaires. In this sense, numerous participants emphasized that, unlike other online games for educational purposes, Liad@s app was not cumbersome to use (some participants made even reference to the term "light application" in their comments) and did not require much mental effort to have a fun time. Moreover, other users told us that information provided by the application was easy to read and understand, as well as the playability was high because there were not elements in play that hinder or distract the users action.

6 CONCLUSIONS

In this article, we have carried out a comparative study (based on real users) of the Liad@s app versus a traditional workshop about sexism for teenagers, to measure which intervention would produce the best impact on users. The results do not show statistically significant differences between the knowledge acquired with each intervention, since both groups significantly improved their knowledge. However, significant differences in learning appeared between gender and between the two different procedures when separately evaluating the data collected from both HS and BS questionnaires. These results validate our primary hypothesis. The theoretical and practical implications of the findings of this work are that the game would help boys because there are more male players, and they are more sexists and aggressive than girls in partner relationships [22], and also that that any app with interactive, 3D elements, and so on, designed for sexism prevention will improve the learning about this topic at least in a similar way than traditional interventions. Additionally, the authors think that the app could help to protect women, because they are more prone to download entertainment apps like Li@dos [37]. The authors recommendations for practitioners is that it is used as a useful pedagogical tool for group tutoring sessions in high schools, since this app has been very well received by teenagers.

Also, we evaluated the usability of the app and the user satisfaction with this application. From the 36 users, 31 (86.11%) preferred the mobile app in front of the traditional workshop. These results validate our secondary hypothesis. A more detailed study about the participants satisfaction depending on the gender shows that women scored the app higher than men, and the age range of 13–14 years provided completely different answers to the rest of age ranges.

As a future work to be done, we plan to add new tests including new variables of sexism like ambivalence or BS, which make gender inequalities become chronic.

REFERENCES

[1] Plan International Ltd. (Version 1.1.2). 2019. Sheboard. Retrieved 15 November, 2017 from https://play.google.com/store/apps/details?id=fi.intodigital.sheboard.

[2] Jeanne Alhusen, Tina Bloom, Amber Clough, and Nancy Glass. 2015. Development of the MyPlan safety decision app with friends of college women in abusive dating relationships. *Journal of Technology in Human Services* 33, 3 (2015), 263–282. DOI: https://doi.org/10.1080/15228835.2015.1037414

[3] Gordon W. Allport. 1954. *The Nature of Prejudice*. Addison-Wesley Publishing Company.

[4] Theodore W. Anderson. 2011. *Anderson–Darling Tests of Goodness-of-Fit*. Springer Berlin Heidelberg, 52–54. DOI: https://doi.org/10.1007/978-3-642-04898-2_118

[5] Laurent Begue, Elisa Sarda, Douglas A. Gentile, Clementine Bry, and Sebastian Roché. 2017. Video games exposure and sexism in a representative sample of adolescents. *Frontiers in Psychology* 8 (2017), 466. DOI: https://doi.org/10.3389/fpsyg.2017.00466

[6] Mons Bendixen and Leif Edward Ottesen Kennair. 2017. When less is more: Psychometric properties of Norwegian short-forms of the ambivalent sexism scales (ASI and AMI) and the Illinois Rape Myth Acceptance (IRMA) Scale. *Scandinavian Journal of Psychology* 58, 6 (2017), 541–550. DOI: https://doi.org/10.1111/sjop.12392

[7] Gerd Bohner, Katrin Ahlborn, and Regine Steiner. 2010. How sexy are sexist men? Women's perception of male response profiles in the ambivalent sexism inventory. *Sex Roles* 62, 7 (01 Apr 2010), 568–582. DOI: https://doi.org/10.1007/s11199-009-9665-x

[8] Jason T. Bowey, Ansgar E. Depping, and Regan L. Mandryk. 2017. Don'T talk dirty to me: How sexist beliefs affect experience in sexist games. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 1530–1543. DOI: https://doi.org/10.1145/3025453.3025563

[9] Lynae M. Brayboy, Alexandra Sepolen, Taylor Mezoian, Lucy Schultz, Benedict S. Landgren-Mills, Noelle Spencer, Carol Wheeler, and Melissa A. Clark. 2017. Girl talk: A smartphone application to teach sexual health education to adolescent girls. *Journal of Pediatric and Adolescent Gynecology* 30, 1 (2017), 23–28.

[10] J. W. Creswell. 2012. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Pearson. Retrieved from https://books.google.es/books?id=4PywcQAACAAJ.

[11] Usue de la Barrera, Estefanía Mónaco, Silvia Postigo-Zegarra, José-Antonio Gil-Gómez, and Inmaculada Montoya-Castilla. 2021. EmoTIC: Impact of a game-based social-emotional programme on adolescents. *PLOS ONE* 16, 4 (April 2021), 1–17. DOI: https://doi.org/10.1371/journal.pone.0250384

[12] Robert A. Emmons and Michael E. McCullough. 2003. Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life. *Journal of Personality and Social Psychology* 84, 2 (Dec. 2003), 377–389. DOI: https://doi.org/10.1037//0022-3514.84.2.377
[13] E. Galarza Fernández, R. Cobo Bedía, and M. Esquembre Cerdá. 2016. The media and the symbolic violence against women. *Revista Latina de Comunicación Social* 71 (2016), 818–832.
[14] Alessandro Gabbiadini and Tobias Greitemeyer. 2019. Fitness mobile apps positively affect attitudes, perceived behavioral control and physical activities. *Journal of Sports Medicine and Physical Fitness* 59 (2019), 407–414. DOI: https://doi.org/10.23736/S0022-4707.18.08260-9
[15] P. Glick and S. T. Fiske. 1996. The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology* 70, 3 (1996), 491–513. DOI: https://doi.org/10.1037/0022-3514.70.3.491
[16] Kristyn Gorton. 2007. Theorizing emotion and affect: Feminist engagements. *Feminist Theory* 8, 3 (2007), 333–348. DOI: https://doi.org/10.1177/1464700107082369
[17] Catherine G. Greeno. 2002. Major alternatives to the classic experimental design. *Family Process* 41, 4 (2002), 733–736. DOI: https://doi.org/10.1111/j.1545-5300.2002.00733.x
[18] Simon P. Hammond, Neil Cooper, and Peter Jordan. 2018. Social media, social capital and adolescents living in state care: A multi-perspective and multi-method qualitative study. *The British Journal of Social Work* 48, 7 (Feb. 2018), 2058–2076. DOI: https://doi.org/10.1093/bjsw/bcx144
[19] R. Harrison, D. Flood, and D. Duce. 2013. Usability of mobile applications: Literature review and rationale for a new usability model. *Journal of Interaction Science* 1, 1 (2013). DOI: https://doi.org/10.1186/2194-0827-1-1
[20] Kasper Hornbaek. 2013. Some whys and hows of experiments in human–computer interaction. *Foundations and Trends® in Human–Computer Interaction* 5, 4 (2013), 299–373. DOI: https://doi.org/10.1561/1100000043
[21] Sarah L. Hutson-Comeaux and Janice R. Kelly. 2002. Gender stereotypes of emotional reactions: How we judge an emotion as valid. *Sex Roles* 47, 1 (01 Jul 2002), 1–10. DOI: https://doi.org/10.1023/A:1020657301981
[22] Ernest N. Jouriles, Hye Jeong Choi, Caitlin Rancher, and Jeff R. Temple. 2017. Teen dating violence victimization, trauma symptoms, and revictimization in early adulthood. *The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine* 61, 1 (July 2017), 115–119. DOI: https://doi.org/10.1016/j.jadohealth.2017.01.020
[23] Joseph E. Kahne and Susan E. Sporte. 2008. Developing citizens: The impact of civic learning opportunities on students' commitment to civic participation. *American Educational Research Journal* 45, 3 (2008), 738–766. DOI: https://doi.org/10.3102/0002831208316951
[24] Korina Katsaliaki and Navonil Mustafee. 2015. Edutainment for sustainable development: A survey of games in the field. *Simulation & Gaming* 46, 6 (2015), 647–672. DOI: https://doi.org/10.1177/1046878114552166
[25] Roger E. Kirk. 2011. *Simple Random Sample*. Springer Berlin Heidelberg, Berlin, 1328–1330. DOI: https://doi.org/10.1007/978-3-642-04898-2_518
[26] Celia Kitzinger and Hannah Frith. 1999. Just say no? The use of conversation analysis in developing a feminist perspective on sexual refusal. *Discourse & Society* 10, 3 (1999), 293–316. DOI: https://doi.org/10.1177/0957926599010003002
[27] Derek A. Kreager, Richard B. Felson, Cody Warner, and Marin R. Wenger. 2013. Women's education, marital violence, and divorce: A social exchange perspective. *Journal of Marriage and Family* 75, 3 (2013), 565–581. DOI: https://doi.org/10.1111/jomf.12018
[28] Jennifer E. Lansford, Ann T. Skinner, Emma Sorbring, Laura Di Giunta, Kirby Deater-Deckard, Kenneth A. Dodge, Patrick S. Malone, Paul Oburu, Concetta Pastorelli, Sombat Tapanya, Liliana Maria Uribe Tirado, Arnaldo Zelli, Suha M. Al-Hassan, Liane Peña Alampay, Dario Bacchini, Anna Silvia Bombi, Marc H. Bornstein, and Lei Chang. 2012. Boys' and Girls' relational and physical aggression in nine countries. *Aggressive Behavior* 38, 4 (2012), 298–308. DOI: https://doi.org/10.1002/ab.21433
[29] Campbell Leaper and Christia Spears Brown. 2014. Chapter six - sexism in schools. In *Proceedings of the Role of Gender in Educational Contexts and Outcomes*, Lynn S. Liben and Rebecca S. Bigler (Eds.), Advances in Child Development and Behavior, Vol. 47. JAI, 189–223. DOI: https://doi.org/10.1016/bs.acdb.2014.04.001
[30] Soledad de Lemus, Miguel Castillo, Miguel Moya, José Luis Padilla, and Estrella Ryan. 2008. Elaboración y validación del inventario de sexismo ambivalente para adolescentes. *International Journal of Clinical and Health Psychology* 8, 2 (2008), 537–562.
[31] Yoan-Daniel Malinov, David Millard, and Tom Blount. 2021. StuckInSpace: Exploring the difference between two mediums of play in a multi-modal virtual reality game. In *Proceedings of the 2021 IEEE Virtual Reality and 3D User Interfaces*. 5–18.
[32] F. J. Massey. 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association* 46, 253 (1951), 68–78.
[33] Kerry McKellar, Elizabeth Sillence, and Michael A. Smith. 2017. Exploring the preferences of female teenagers when seeking sexual health information using websites and apps. In *Proceedings of the 2017 International Conference on Digital Health* ACM, New York, NY, 43–47. DOI: https://doi.org/10.1145/3079452.3079497

[34] Magdalena Mendez-Lopez, M Carmen Juan, Ramon Molla, and Camino Fidalgo. 2021. Evaluation of an augmented reality application for learning neuroanatomy in psychology. *Anatomical Sciences Education* (April 2021). DOI: https://doi.org/10.1002/ase.2089

[35] J. J. Navarro and E. Pastor. 2018. De los riesgos en la socialización global: Adolescentes en conflicto con la ley con perfil de ajuste social. *Convergencia: Revista de Ciencias Sociales* 76 (2018), 119–145. DOI: https://doi.org/10.29101/crcs.v25i76.4442

[36] José Navarro-Pérez, Amparo Oliver, Ãngela Carbonell, and Barry H Schneider. 2020. Effectiveness of a mobile app intervention to prevent dating violence in residential child care. *Psychosocial Intervention* 29, 2 (2020), 59–66. Retrieved from http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1132-05592020000200001&nrm=iso.

[37] José J. Navarro-Pérez, Angela Carbonell, and Amparo Oliver. 2019. The effectiveness of a psycho-educational app to reduce sexist attitudes in adolescents. *Revista de Psicodidáctica (English ed.)* 24, 1 (2019), 9–16. DOI: https://doi.org/10.1016/j.psicoe.2018.07.002

[38] Rachael O'Byrne, Susan Hansen, and Mark Rapley. 2008. If a girl doesn't say no: young men, rape and claims of insufficient knowledge. *Journal of Community & Applied Social Psychology* 18, 3 (2008), 168–193. DOI: https://doi.org/10.1002/casp.922

[39] A. Pollatsek and A. D. Well. 1995. On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21, 3 (1995), 785–794. DOI: https://doi.org/10.1037/0278-7393.21.3.785

[40] Tamara Ramiro-Sánchez, María Teresa Ramiro, María Paz Bermúdez, and Gualberto Buela-Casal. 2018. Sexism and sexual risk behavior in adolescents: Gender differences. *International Journal of Clinical and Health Psychology* 18, 3 (2018), 245–253. DOI: https://doi.org/10.1016/j.ijchp.2018.04.002

[41] A. C. Rekdal, A. Karahasanovic, V. By Kampenes, J. E. Hannay, N. Liborg, D. K. Sjoberg, and O. Hansen. 2005. A survey of controlled experiments in software engineering. *IEEE Transactions on Software Engineering* 31, 09 (Sep. 2005), 733–753. DOI: https://doi.org/10.1109/TSE.2005.97

[42] Judy Robertson. 2012. Likert-type scales, statistical methods, and effect sizes. *Communications of the ACM* 55, 5 (May 2012), 6–7. DOI: https://doi.org/10.1145/2160718.2160721

[43] Silvia Semenzin and Lucia Bainotti. 2020. The use of telegram for non-consensual dissemination of intimate images: Gendered affordances and the construction of masculinities. *Social Media + Society* 6, 4 (2020). DOI: https://doi.org/10.1177/2056305120984453

[44] Samuel Sanford Shapiro and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3/4 (1965), 591–611.

[45] Patti M. Valkenburg and Jochen Peter. 2007. Preadolescents' and adolescents' online communication and their closeness to friends. *Developmental Psychology* 43, 2 (2007), 267–277. DOI: https://doi.org/10.1037/0012-1649.43.2.267

[46] Geert P. Verheijen, William J. Burk, Sabine E. M. J. Stoltz, Yvonne H. M. van den Berg, and Antonius H. N. Cillessen. 2018. Friendly fire: Longitudinal effects of exposure to violent video games on aggressive behavior in adolescent friendship dyads. *Aggressive Behavior* 44, 3 (2018), 257–267. DOI: https://doi.org/10.1002/ab.21748

[47] Joshua H. West, Lindsay M. Belvedere, Rebecca Andreasen, Christine Frandsen, P. Cougar Hall, and Benjamin T. Crookston. 2017. Controlling your "app" etite: How diet and nutrition-related mobile apps lead to behavior change. *JMIR Mhealth Uhealth* 5, 7 (10 Jul 2017), e95. DOI: https://doi.org/10.2196/mhealth.7410

[48] Bob G. Witmer, Christian J. Jerome, and Michael J. Singer. 2005. The factor structure of the presence questionnaire. *Presence: Teleoperators and Virtual Environments* 14, 3 (2005), 298–312. DOI: https://doi.org/10.1162/105474605323384654

[49] Bob G. Witmer and Michael J. Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (1998), 225–240.

Received February 2021; revised June 2021; accepted June 2021