Highlights

Playing to Do Research: Exploring the Use of the RAYUELA Video Game to Conduct Prevalence Studies and Risk Factor Analysis of Cyberbullying

Andrea Baños Ramos, María Reneses Botija, Edmond Awad, Mario Castro Ponce, José Ruipérez Valiente, Gregorio López López

- Serious games are presented as an innovative tool for cyberbullying research.
- In-game behaviors are strongly related with validated questionnaire responses.
- In-game measures explain cyberbullying patterns and prevalence.
- In-game behaviors are more sensitive to identifying risk factors instead than validated questionnaires.

Playing to Do Research: Exploring the Use of the RAYUELA Video Game to Conduct Prevalence Studies and Risk Factor Analysis of Cyberbullying

Andrea Baños Ramos, María Reneses Botija, Edmond Awad, Mario Castro Ponce, José Ruipérez Valiente, Gregorio López López

Instituto de Investigación Tecnológica (IIT), Universidad Pontificia Comillas, Madrid, 28015, Spain

Abstract

This research explores the use of a serious game as an alternative or complementary research tool to surveys and validated questionnaires for studying cyberbullying. A scientifically-based video game was designed, incorporating a literature review, analysis of legal sentences, and interviews with perpetrators, victims, and experts. Participants completed questionnaires before and after playing the game, and their in-game behaviors were compared with these responses. The results show that the video game estimates the prevalence of cyberbullying with figures similar to other studies and better approximates the relationship between risk factors, such as time spent on the internet or family communication with cyber aggression. Therefore, the RAYUELA serious game is proposed as a promising research tool for prevalence studies and risk factor analysis in cyberbullying, highlighting its potential and limitations.

Keywords: Cyberbullying, prevalence studies, risk factors, serious games, validated questionnaires.

1. Introduction

Cyberbullying (CB) has emerged as a relevant societal issue, with children and adolescents being particularly vulnerable. Its quantification and characterization are more complex than those of traditional bullying [1], yet these efforts are essential for effective prevention and intervention. Prevalence studies are commonly employed to estimate the proportion of individuals affected, and the sociodemographic characteristics shared by victims are typically examined as risk factors that can guide the development of targeted prevention strategies. Surveys and validated questionnaires (VQs) remain the standard methodological instruments in such research. However, despite their rigorous design and validation processes, these tools may not be fully suited to younger populations [2]. Moreover, although they offer substantial scalability, they are susceptible to observation bias, as respondents, particularly children, may modify their behavior or responses when aware that they are being assessed (Hawthorne effect).

Serious games[1] (SGs), traditionally employed for educational and training purposes, are increasingly recognized as innovative tools for prevention and early detection [4] in the fields of cybersecurity and cybercrime, with growing potential for research-oriented applications. At early stages, these systems aim to enhance awareness and identify common issues that arise in everyday digital contexts, thereby supporting risk prevention and the development of effective interventions [5]. Within this framework, SGs constitute a promising methodological approach for game-based assessment (GBA) [6], as they enable users to make contextually grounded decisions while assuming specific roles. This provides a low-intrusive and user-friendly assessment environment that may help reduce observation bias [7].

Several challenges must be addressed prior to implementing the serious game (SG), including game design, data collection procedures, sample selection, and ensuring that participants possess the necessary digital competencies to engage with a digital game. Following implementation, an additional challenge concerns the validation process, which is essential for establishing the SG as a reliable research instrument and for accurately assessing prevalence in comparison with an already validated questionnaire [8].

The contribution of this article relates to this issue, specifically to evaluating whether SGs can serve as an alternative to traditional methods for collecting data relevant to prevalence studies and risk factor analyses on the cyberbullying phenomenon.

In this study, we examine the potential of using data gathered through a serious game developed within the European project RAYUELA [www.rayuela-

[1] Games that are not designed primarily for entertainment but instead serve educational purposes or function as tools for data collection in empirical research [3].

h2020.eu] to investigate cyberbullying, both in terms of prevalence estimation and risk factor assessment. To evaluate this potential, we formulated three research questions (RQs):

- *RQ1: Are responses to in-game scenarios associated with responses to a validated questionnaire?* To address this question, a set of validation and reliability metrics was applied. Additionally, comparisons were performed between individuals who self-reported experiences of aggression.

- *RQ2: How can the video game be used to conduct prevalence studies?* Indicators of aggression were extracted from the aggression-related scenarios presented throughout the game.

- *RQ3: Can the video game be used to conduct risk factor analyses?* Associations between in-game responses and out-of-game questionnaire items were examined to assess risk factors.

2. State of the Art & Related Work

Serious games (SGs) are designed for specific population groups to help mitigate societal challenges through targeted intervention. The widespread use of digital technologies and social media has increased minors' exposure to online environments, making SGs a potentially valuable tool for early awareness-raising. Originally conceptualized by Clark C. Abt in 1970 [9], SGs are now considered cost-effective, highly scalable, and applicable across diverse domains, while also raising important ethical considerations related to minors' privacy and accessibility.

SGs have demonstrated relevance in several areas: (i) educational contexts, where they support the development of cognitive and social skills; (ii) therapeutic applications, including rehabilitation, emotional regulation, and facilitation of social interaction; (iii) training and simulation, where they enable safe exposure to potentially risky scenarios to foster decision-making and skill acquisition; (iv) gamification and user engagement studies, where they serve to assess accuracy, usability, and user behavior; and (v) accessibility and inclusivity initiatives, by supporting disadvantaged populations in developing cognitive abilities and promoting social collaboration [10].

Several SGs have already been used for research purposes, such as Moral Machine, developed by the Massachusetts Institute of Technology (MIT) in collaboration with Edmond Awad [11]. Implementing such tools, however, requires coordinated work among multidisciplinary experts to ensure methodological soundness and assess the scalability of SGs to other research contexts.

Despite their advantages (see Section 1), SGs also entail notable limitations: (i) high implementation, development, and maintenance costs across institutions; (ii) challenges in designing games that are properly adapted to the target population; and (iii) difficulties in guaranteeing accessibility for users who lack the necessary infrastructure or digital literacy.

If these aspects are not adequately addressed, SGs may themselves introduce vulnerabilities for the populations under study.

To the best of our knowledge, prior research has not attempted to estimate the prevalence of a construct using a serious game as an alternative to traditional assessment instruments, nor for conducting risk factor analyses based on SG-derived measures.

3. Methodology

This section briefly describes the development of the SG and outlines the data collection procedure. It also details the data cleaning process, justifies the final sample for each research question, and describes the variables selected for the analyses. Lastly, the strategies employed to analyze the data using validity and reliability metrics are presented.

3.1. RAYUELA Serious Game Design

The RAYUELA video game was designed based on scientific evidences. Such scientific evidences were gathered by thoroughly reviewing the literature, analyzing 46 European sentences related with cyberbullying and conducting 33 semistructured interviews involving offenders (8), victims (12), and experts (13) (see Figure 1). Two cyberadventures related with cyberbullying were developed including scenarios which aim to measure risk factors or behaviors which may be considered as cyberbullying based on the findings from this research phase (see Figure 2). In addition, before getting the credentials to log in the game, the players must fill a questionnaire providing information such as age, gender, place of birth, immigration history, school type, sexual orientation, or the hours they spend on the Internet. After playing the game, they must fill a validated questionnaire on previous cyberbullying victimization and aggression. No personal information is stored by the game and all participants signed an informed consent especially adapted to minors (depending on the age and the country, their parents also signed such an informed consent).

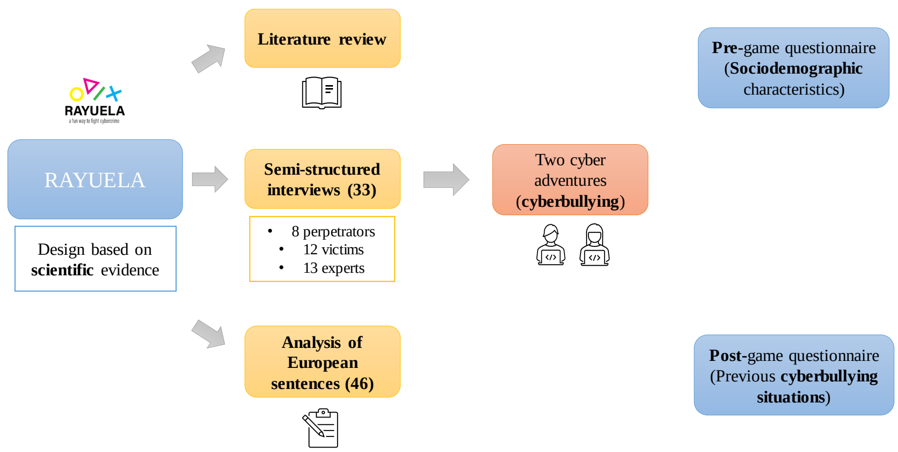

Figure 1: Serious Game Design.

For this reason, we used the validated cyberbullying questionnaire as a baseline to examine its relationship with various in-game scenarios and to validate these associations using statistical metrics for the purpose of conducting prevalence studies and risk factor analyses.

3.2. Data Collection

The collected sample, obtained through various validated questionnaires and SG adventures, included 1859 minors from 10 European countries and required careful processing and cleaning. A thorough data cleaning process was necessary to minimize biases arising from incomplete or inconsistent responses prior to conducting statistical analyses. First, observations from countries with insufficient representation were removed (n = 1851). Additionally, individuals who, despite completing the initial questionnaires and obtaining credentials to configure their avatar, did not to play the video game were excluded (n = 1493). Minors who did not complete the cyberbullying victimization questionnaire were also removed (n = 971). Next, participants

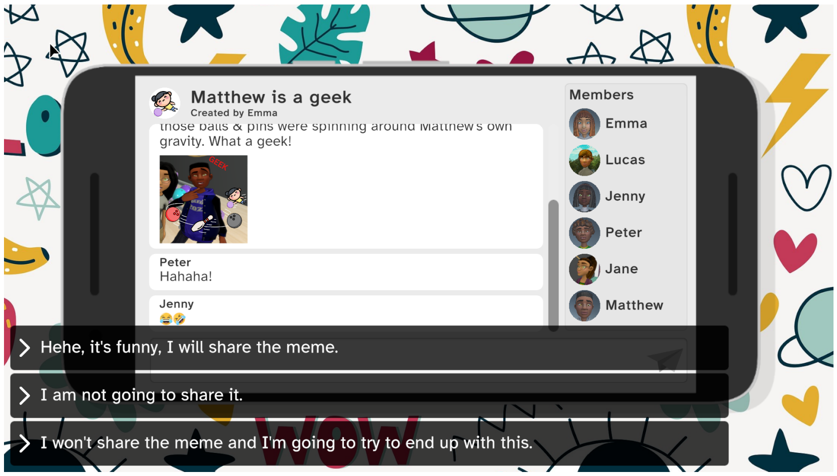

Figure 2: Sharing a Meme Scenario from the RAYUELA Serious Game.

who did not engage with a specific in-game scenario used as a proxy for the cyberbullying victimization questionnaire were excluded (where the players were asked about previous experiences related to cyberbullying, n = 826). Moreover, those who did not play Adventure 3 of the SG (n = 818) and those who did not participate in the cyber-aggression scenario involving sharing a classmate's meme (n = 808) were also excluded (see Figure 3).

Depending on the research question, the final sample varies. For the first research question, the sample was further filtered based on a self-assessment indicator collected after gameplay, which evaluated whether participants behaved in the game as they do in real life. Individuals whose in-game behavior deviated substantially from their real-life behavior were excluded, resulting in a final sample of 684 minors.

The dataset comprised a sample of 684 individuals, including 400 minors who self-identified as males, 259 as females, and 11 as non-binary from 10 European countries (Belgium, Croatia, Czech Republic, Estonia, Hungary, Italy, Netherlands, Portugal, Spain, and United Kingdom) with ages ranging from 12 to 16 years old (Table 1 presents the variables examined in the context of cyberbullying, and Table 2 shows the categories of these indicator variables and their marginal probabilities).

For the second research question, all previous steps were applied, except for the requirement of completing the cyberbullying victimization questionnaire, obtaining a final sample of 687 observations. The sample used for the

Table 1: Indicator variables of cyberbullying identified, illustrating whether each variable was measured in-game or out-of-game.

Type	**Indicator**	**Measured in-game**	**Measured out-of-game**
Environmental	Family communication	✓	✓
	Sociability trait	✓	✓
Personal	Previous CB Offending	✓	✓
	Age	✗	✓
	Gender	✗	✓
Technological	Time spent online	✓	✓

Table 2: Registration, validated questionnaire variables [12], and in-game indicators. The possible response values of each variable and its sample marginal probability (percentage of observation) are shown.

Variable	**Response Values**	**Marginal Probability**
Gender	Male	59.7%
	Female	38.7%
	Non Binary	1.6%
Age	12	16.6%
	13	28.4%
	14	26.4%
	15	16.1%
	16	12.5%
Family Support (out-of-game)	Low	31.1%
	High	68.9%
Family Communication (in-game)	Bad	35.3%
	High	64.7%
Time Spent Online (out-of-game)	Less than 4 h	66.4%
	More than 4 h	33.6%
Time Spent Online (in-game)	Less than 4 h	60.4%
	More than 4 h	39.6%
Sociability Trait (out-of-game)	Sociable	68.3%
	Shy	35.1%
Sociability Trait (in-game)	Sociable	68.3%
	Shy	31.7%
Previous CB Offending	Offender	22.2%
	None	77.8%

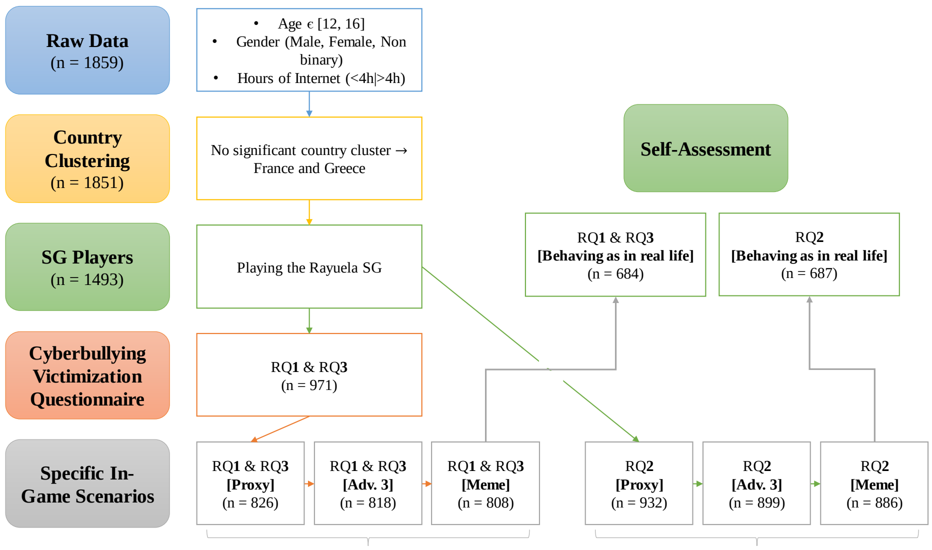

Figure 3: Data Cleaning and Treatment Processes.

first research question was also employed for the third research question.

3.3. Aggression Measures

The aggression-related scenarios considered in the analysis, all of which assessed current aggressive behavior, included the decision to share a meme about a classmate, making fun of a peer's appearance (making a homophobic joke), and witnessing a bullying situation at school. In addition, the SG contained a scenario assessing previous experiences of cyberbullying.

Aggregated measures of aggression-related scenarios were also constructed (sharing a meme, making a homophobic joke, and witnessing of CB). These included a high-specificity indicator (*AND* measure), defined as having acted as an aggressor in all aggression scenarios; a high-sensitivity indicator (*OR* measure), defined as having acted as an aggressor in at least one scenario; and an intermediate criterion (*OR restricted* measure), defined as having acted as an aggressor in two or more scenarios. Measures of current aggression were used to generate these aggregated indicators, to mitigate potential biases associated with retrospective assessments of aggressive behavior.

Table 3: Aggregation-Related Social Interaction Scenarios

Sharing a Meme: After receiving a message from Patty with the meme about Matthew.	**Making a Homophobic Joke:** Playing video games in your room. Your friends joke about Pol's appearance.
1. Offender: Hehe, it's funny, I will share the meme. 2. None: I won't share the meme.	1. Offender: That's funny. 2. None: I don't like this kind of jokes.
Witnessing of CB: I heard that some guys are messing with Paul. I am worried that he may be bullied. What do you think?	**Previously Committed CB:** Your friends start messing with Pol.
1. Offender: They are just having fun; I would not call that bullying. 2. None: I think it's not right...	1. Offender: Yes, it was me who messed with someone else... but it was not such a big deal. 2. None: No, it has never really happened to me, to my knowledge.

3.4. Validation and Reliability

To ensure the consistency of the measured variables, both the aggression-related scenarios and the validated cyberbullying questionnaire were assessed. First, construct validity was examined to identify patterns of correlations among the variables and to determine whether these correlations reflect underlying latent factors, that is, constructs that cannot be directly observed.

An Exploratory Factor Analysis (EFA) was conducted to determine whether common latent factors were being measured by the aggression variables, both in-game and out-of-game [13]. Subsequently, criterion validity was evaluated using a Random Forest model to examine whether the in-game variables were related to an external outcome (out-of-game aggression) and to assess their predictive performance [14]. It is important to note that these measures were administered at different points in time.

After the EFA was conducted, the resulting factorial structure was tested through Confirmatory Factor Analysis (CFA) to establish its reliability [15]. The null hypothesis posits that the potential factors are correlated and that the variables within each factor display high and significant loadings, thereby contributing meaningfully to the corresponding construct.

4. Results

4.1. RQ1: Are responses to in-game scenarios associated with responses to a validated questionnaire?

To address this research question, two variables were analyzed, reflecting previous experiences of cyberbullying perpetration both in-game and out-of-game. The prevalence rates for these measures were 12.1%, and a 22.2%, respectively, indicating the proportion of minors at risk of engaging in cyberbullying. When a contingency table was constructed to examine the joint distribution of these two measures, a prevalence of 6% of offenders emerged.

To refine this assessment, we considered the effect of conditioning the contingency table on prior information. When conditioning on the out-of-game measure, assuming prior knowledge of the responses to the validated questionnaire, the estimated prevalence increased to 27%, suggesting a notable rise in the proportion of potential offenders. Conversely, because the validated questionnaire was administered after the gameplay session, we also conditioned the contingency table on the in-game measure. In this case, the estimated prevalence rose to 49.4%, more than doubling the proportion of potential offenders and indicating a substantial increase when incorporating in-game information.

While examining proportions and conditional probabilities provides meaningful insight into the identification of potential offenders, we further evaluated performance using confusion-matrix indices to ensure robust metrics. The most informative indices for identification were accuracy and specificity, which demonstrated strong performance at 77.6% and 92.1%, respectively.

The association between the two measures was also assessed and found to be substantial. Both frequentist and Bayesian analyses supported this conclusion, yielding a high chi-square statistic of 38.6, a highly significant p-value of 5.2×10^{-10}, and a Bayes factor of 2,723,671 (assuming uniform priors, $\alpha = \beta = 1$). The chi-square test indicates a strong and highly significant association, whereas the Bayes factor provides decisive evidence for a positive directional effect: minors identified as offenders on one measure are more likely to be identified as offenders on the other.

This association is further illustrated through a flexplot descriptive analysis [16], which visualizes the deviation of observed values from their expected

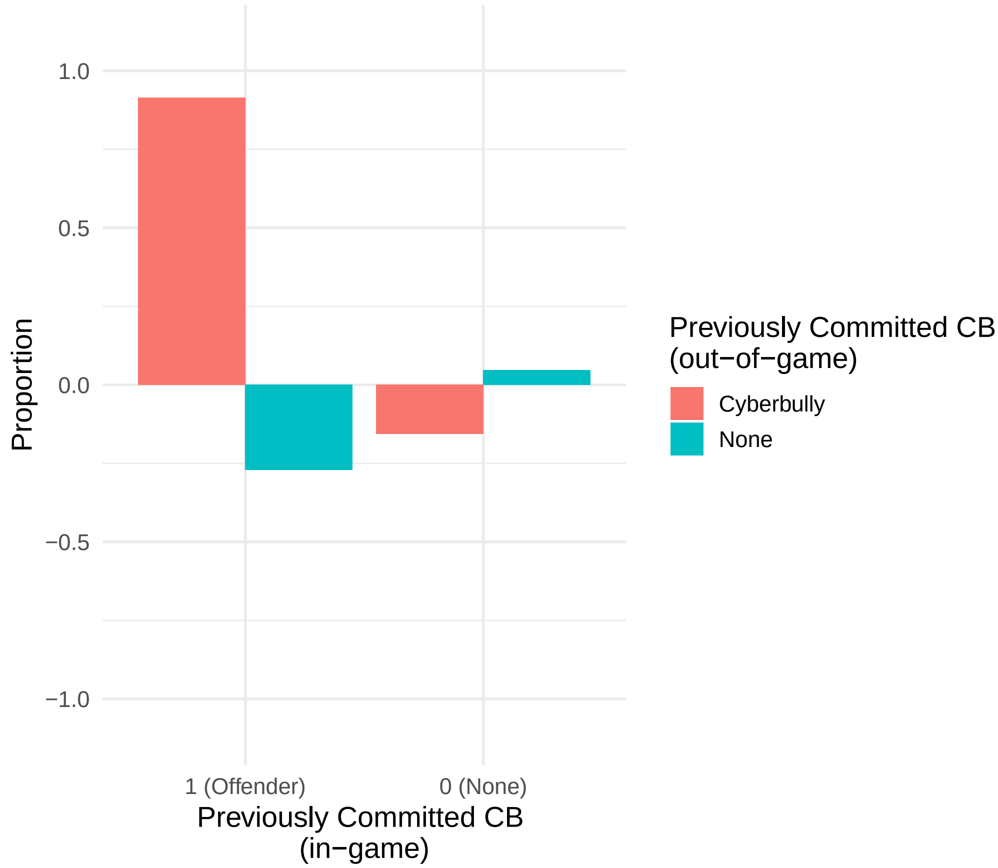

Figure 4: Flexplot between direct aggression measures. The x-axis represents the in-game measure, and the y-axis indicates the out-of-game measure.

frequencies, demonstrating how each variable successfully identifies potential offenders detected by the other (see Figure 4).

4.1.1. Validation

Three aggression-related scenarios of cyberbullying perpetration were included in this analysis alongside previous experiences of cyberbullying measures (in-game and out-of-game). Using these five variables representing aggression, we conducted two type of validity analyses: construct validity and criterion validity.

The construct validity assessment aimed to identify underlying latent factors representing cyber aggression, which are not directly observable. To evaluate whether the five aggression variables were suitable for factor analysis, two preliminary diagnostics were performed. The first was the Kaiser-Meyer-Olkin (KMO) index [17, 18], which compares the magnitude of the observed correlation coefficients with that of the partial correlations. The KMO statistic ranges from 0 to 1 and is typically interpreted as follows:

- Unacceptable $\in$ [0.00, 0.49]
- Miserable $\in$ [0.50, 0.59]
- Mediocre $\in$ [0.60, 0.69]
- Middling $\in$ [0.70, 0.79]
- Meritorious $\in$ [0.80, 0.89]
- Marvelous $\in$ [0.90, 1.00]

Our data yielded a KMO value of 0.66, which falls within the mediocre range, but remains acceptable, especially considering the size of the dataset. Several studies establish 0.60 as a minimum threshold for relevance, indicating that our data are adequate for exploratory factor analysis [19, 20].

The second diagnostic was Bartlett's test of sphericity [21], which evaluates whether the correlation matrix significantly differs from the identity matrix. The null hypothesis was rejected, confirming that the variables are sufficiently correlated to justify factor analysis. Together with the KMO index, this supports the suitability of the data for factor extraction.

An Exploratory Factor Analysis (EFA) was then performed to examine the presence of common latent factors underlying aggression. The number of factors to extract was determined through three complementary methods:

1. Eigenvalues (Kaiser's criterion): factors with eigenvalues greater than 1, explaining a substantial proportion of variation within the data. According to this criterion, a single factor is sufficient to explain cyber aggression.
2. Scree plot: visually identifies the elbow where the curve levels off. In our case, the scree plot suggests retaining two factors (see Figure 5).
3. Parallel analysis: compares observed eigenvalues with those obtained from randomly simulated data. Only two factors exceeded the eigenvalues from the simulated datasets (see Figure 6).

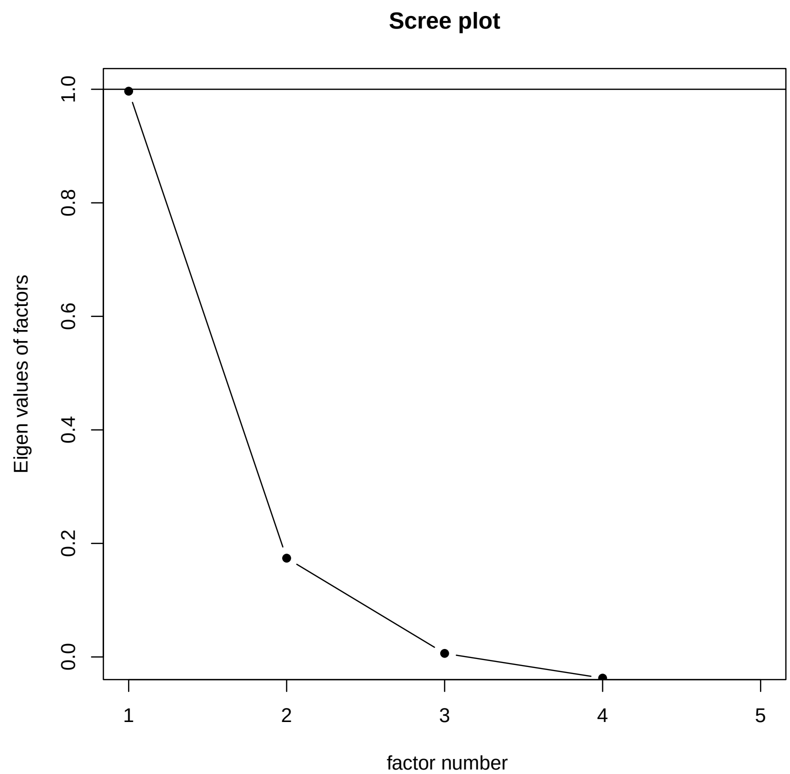

Figure 5: Extraction of factors: Scree plot.

Factor extraction and rotation were guided by the correlations among variables. We set a threshold of 0.20 to examine associations, considering that the distribution of the categories "None" and "Offender" is approximately 80/20%. Because several variables exceeded this threshold, we applied an oblique rotation (promax), which allows factors to correlate.

The results did not reject the hypothesis that two factors are sufficient. These factors were correlated, indicating that they are related but not identical, potentially representing different manifestations of cyber aggression (past vs. current behaviors). Finally, the factor loadings were inspected graphically, illustrating the contribution of each variable (in-game and out-of-game) to each latent factor (see Figure 7).

Regarding criterion validity, our aim was to assess the extent to which the in-game variables were able to predict an external outcome, in this case,

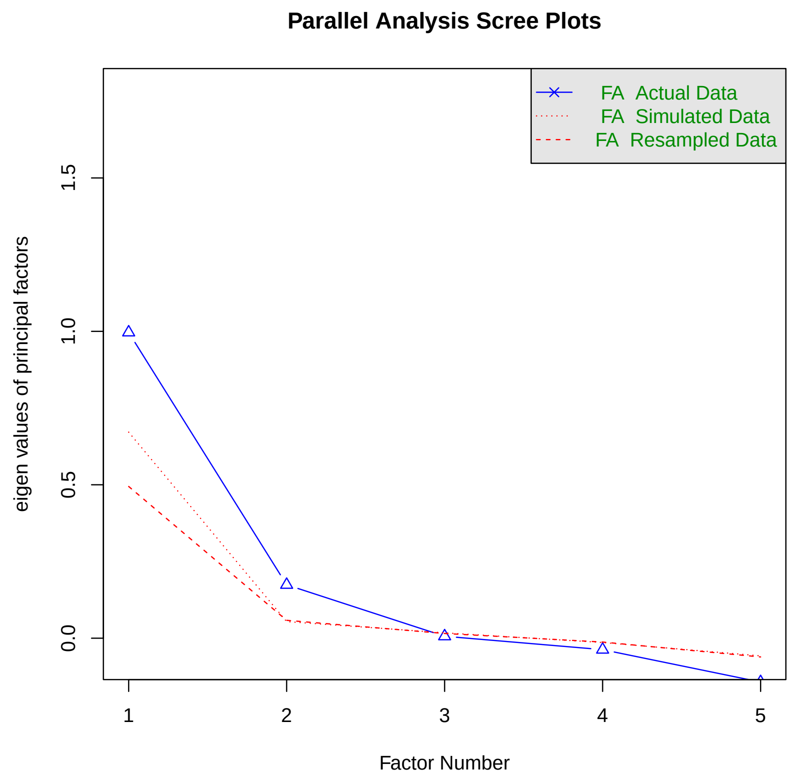

Figure 6: Extraction of factors: Parallel Analysis.

cyberbullying offending outside the game environment. To do so, we examined postdictive validity, evaluating whether the scores derived from the in-game scenarios (current and previous experiences) were associated with an established measure of cyberbullying perpetration collected in the validated questionnaire.

A random forest classifier was implemented to predict out-of-game cyberbullying offending. The dataset was split into a 70% training set (480 observations) and a 30% testing set (204 observations). The training process used 10-fold cross-validation and a hyperparameter search across 10 candidate mtry values, selecting the final model based on the highest ROC performance. The decision threshold for class assignment was set to 0.35, following ROC curve–based optimization (see Figure 8).

Given the unbalanced distribution of the target classes (see Section 3.2),

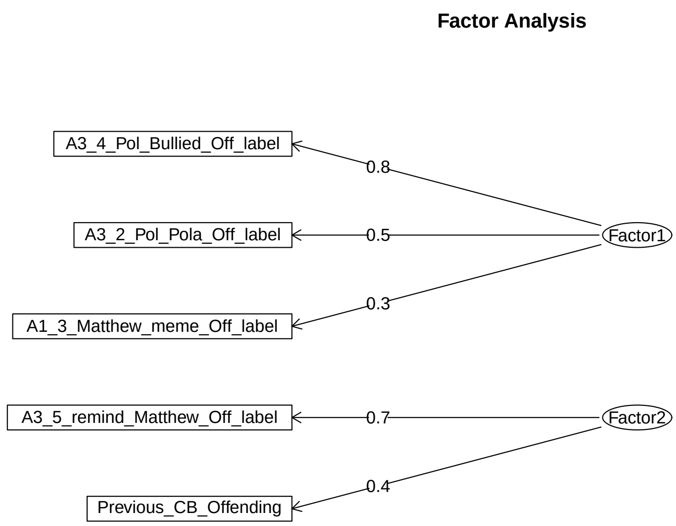

Figure 7: Exploratory Factor Analysis.

the random forest training procedure employed automatic class-balancing within each cross-validation fold, preventing information leakage into the test set. The resulting confusion matrix indicates balanced performance across both classes ("Offender" and "None"), yielding an overall accuracy of 69% and a Cohen's kappa of 0.21 for inter-rater agreement.

The contribution of each in-game predictor to the model was examined through variable importance measures (interpretable machine learning). As shown in Figure 9, the in-game scenario involving previous cyberbullying offending emerged as the strongest predictor, followed by in-game aggression indicators, including the scenario in which participants share a meme of a classmate. These results highlight the substantive role of in-game behaviors in explaining variation in cyberbullying offending outside the game context.

4.1.2. Reliability

To confirm the structure identified in the Exploratory Factor Analysis (EFA), we conducted a Confirmatory Factor Analysis (CFA), assuming as the null hypothesis that the latent factors are correlated. The fit indices demonstrated that the proposed model provided a substantially better fit than the null model (which assumes no latent factors and no covariances

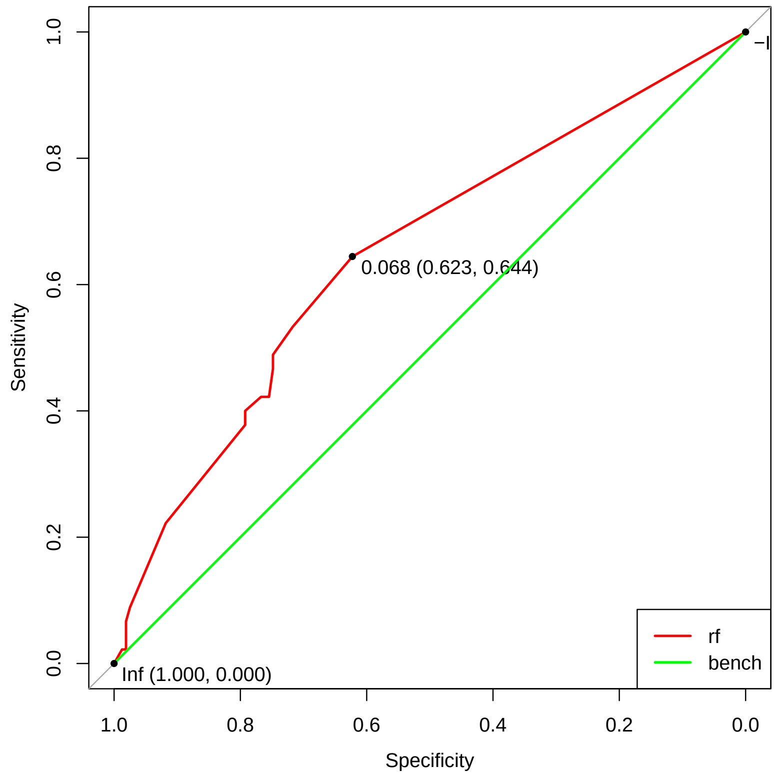

Figure 8: Random Forest: ROC curve.

among observed variables). Both the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI), which range from 0 to 1, exceeded 0.95, indicating excellent model fit relative to the null model.

Within the latent factors, the observed variables exhibited moderate-to-high factor loadings (0.40–0.60), suggesting that each variable contributes meaningfully to its corresponding factor. Additionally, the two correlated factors shared 34% of their variance, supporting the hypothesis derived from the EFA that the factors are related but not identical, and likely represent distinct facets of cyber aggression.

4.2. RQ2: How can the video game be used to conduct prevalence studies?

Once the aggression measures were shown to be reliable and valid, we computed several indices to quantify the prevalence of individuals at risk

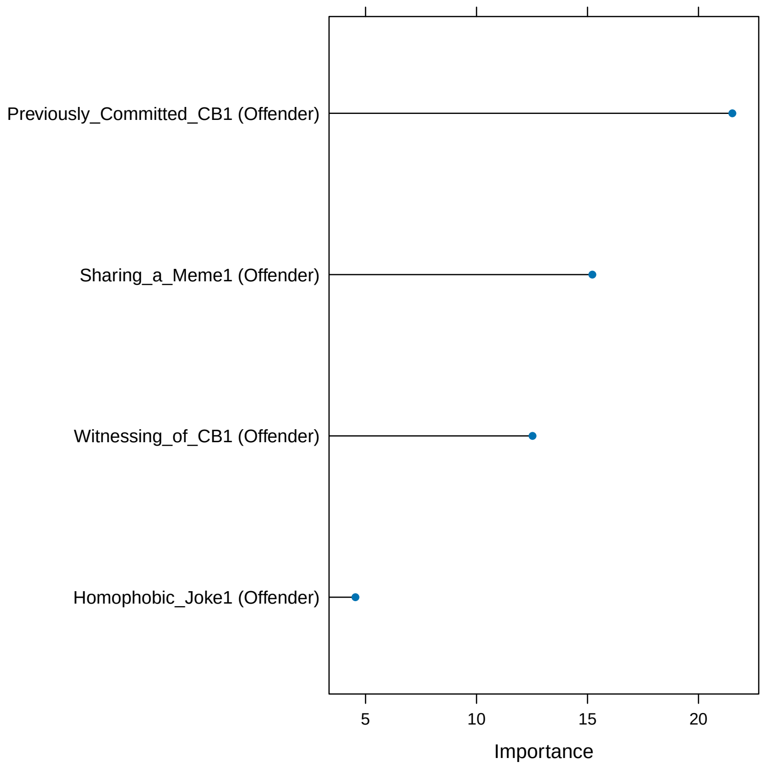

Figure 9: Random Forest: Variable Importance.

of engaging in cyberbullying. The most restrictive metric, *AND*, is highly specific and prioritizes detection accuracy. In contrast, the most inclusive metric, *OR*, maximizes sensitivity and is therefore appropriate for the design of educational and preventive programs. Finally, the *OR restricted* metric provides a balanced, compensated measure that aims to maximize overall accuracy by trading off the strengths of both the *AND* and *OR* criteria.

The estimated prevalence values were 3.3% for the *AND* metric, 34.9% for the *OR* metric, and 11.8% for the *OR restricted* metric. The external, out-of-game validated questionnaire yielded a prevalence of 22.2% of adolescents reporting cyberbullying offending. The aggression-related in-game scenarios capture a prevalence comparable to that obtained with traditional measures. Specifically, the Sharing a Meme scenario showed a prevalence of 20.4% of offenders, and the Making a Homophobic Joke" scenario showed a prevalence of 18.5%, both aligning closely with the validated questionnaire. These values are also within the range reported by international studies, such as the WHO and HBSC survey (2018–2022), which documented an average cyberbullying-offending prevalence of approximately 6% across 44 countries. Additional survey-based studies conducted by our research team using large, representative samples reported a prevalence of 8.7% in Spain and 14.9% in Estonia [22].

The selection of a particular metric depends on the goals of the study—whether focused on prevention, education, awareness, or precision. For this reason, we employed both frequentist and Bayesian statistical frameworks to validate these metrics. Within the frequentist approach, chi-squared tests were conducted, revealing statistically significant associations (p-value < 0.05) and providing effect size estimates through the chi-square coefficient. In the Bayesian framework, Bayes factors were computed to assess the direction and strength of the evidence (see Table 4), using a joint multinomial model with uniform priors ($\alpha = \beta = 1$) and without fixing row or column totals.

Table 4: Jeffreys' scale [23]

Bayes Factor (H_1 vs H_0)	**Grades of evidence**
1 to $10^{1/2}$	Barely worth mentioning
$10^{1/2}$ to 10	Substantial
10 to $10^{3/2}$	Strong
$10^{3/2}$ to 10^2	Very strong
$> 10^2$	Decisive

Other scales for interpreting the Bayes factor, such as the one by Lee and Wagenmakers, were also employed in this type of studies [24]. This scale, which is an adaptation of Jeffreys' scale, is particularly useful for individuals who are new to interpreting Bayes factors (see Table 5) [25].

Overall, the results indicate that any of the proposed metrics can be appropriately used to study cyberbullying prevalence. The *AND* metric is best suited for prevention-focused research, the *OR* metric for designing educational programs, the *OR restricted* metric for precision-oriented analyses, and the scenario-specific measures (Sharing a meme or Making a homophobic joke) for targeted awareness campaigns (see Table 6).

Table 5: Lee and Wagenmakers' scale

Bayes Factor (H_1 vs H_0)	**Grades of evidence**
1 to 3	Anecdotal
3 to 10	Moderate
10 to 30	Strong
30 to 100	Very strong
> 100	Extreme

Table 6: Sensitivity Analysis of Performance Metrics: Frequentist and Bayesian Frameworks.

	Cyberbullying Offending (out-of-game)		
Variable	χ^2	p	BF_{10}
High-specificity	7.56	0.006	2.28
High-sensitivity	32.14	1.44×10^{-8}	1,027,632
Trade-off	17.75	2.52×10^{-5}	373.86
Sharing a Meme	20.09	7.38×10^{-6}	1696.35
Homophobic Joke	11.83	0.0006	35.26

4.3. RQ3: Can the video game be used to conduct risk factor analyses?

The measure selected to evaluate risk factors for in-game cyberbullying offending was the *OR restricted* indicator. Although the "sharing a meme" scenario is the closest conceptual match to the validated out-of-game questionnaire (in terms of prevalence), *OR restricted* was chosen because it represents the most balanced and accurate measure for these analyses.

When examining the relationships between various risk factors and cyberbullying aggression both in-game and out-of-game, the strongest and most significant associations emerged from in-game–to–in-game comparisons (see Table 7). These findings suggest that the serious game is particularly effective for identifying aggressive behaviors among adolescents.

5. Discussion

This paper aims to foster discussion on the use of serious games, specifically in the context of cyberbullying, as an alternative to other research methods such as surveys or questionnaires, and to open new research venues along this line. We believe that serious games presents some characteristics (e.g., immersive experience, context, role play) which may be beneficial

Table 7: Risk Factor Analysis (out-of-game vs. in-game)

Variable		Cyberbullying Offending Out-of-game	In-game
Time Spent Online	Out-of-game	7.57 (0.006) [6.68]	0.83 (0.36) [0.17]
	In-game	36.84 (0.00) [14×10^6]	45.75 (0.00) [1.7×10^{13}]
Family Communication	Out-of-game	6.34 (0.04) [0.21]	3.05 (0.22) [0.04]
	In-game	2.86 (0.09) [0.56]	18.4 (0.00002) [6.26]
Sociable	Out-of-game	0.5 (0.48) [0.16]	0.52 (0.13) [1.87]
	In-game	0.2 (0.65) [0.13]	2.26 (0.47) [727.69]

Note: Chi-squared coefficients, p-values (in parentheses), and Bayes factors (in brackets). The green cell indicates the highest association.

to conduct certain studies or target certain population segments. However, they may also involve issues and introduce errors or biases, as any other measurement or research tool. Thus, more research and discussion is needed.

In general, studies addressing these forms of cybercrime, such as cyberbullying or other cyber-cognitive phenomena, do not typically focus on validating specific instruments or on assessing constructs through games and subsequently comparing them by means of two distinct approaches.

One of them tries to identify different forms of cyberbullying, as a harmful online behavior, including extortion, humiliation or threats, through audio, photos, or videos, among both children and adult populations [26].

Another study considers different aspects of the cyberbullying process, such as the intention to cause harm, its repetitive nature over time, and the use of hate speech, analyzing these through methods such as Term Frequency–Inverse Document Frequency (TF-IDF), profanity, N-grams, and sentiment analysis [27].

However, focusing on validation criteria, studies such as one conducted by Diana R. Sanchez and Markus Langer developed a measure, called Video Game Pursuit (VGPu) and validated it with other game-related scales. Their work proved useful for predicting individuals' in-game performance by employing convergent and divergent validity (through correlations) as well as criterion validity (for prediction) [28]. Other researchers, such as Tianying Feng and Gregory K.W.K. Chung, aimed to validate game-based indicators (GBIs) developed from a theoretical framework, using correlations from post-test gameplay, which proved useful for learning and education [29].

This study is subject to certain limitations. First, the experimental design of the video game was primarily tailored to young populations, namely children and adolescents, which may restrict the generalizability of the findings to other age groups. Second, participation required a minimum level of basic digital skills to adequately interact with the virtual environment. Finally, the estimation of specific constructs through gameplay, their measurement could be improved by examining causal relationships among variables.

As a result of this research, the findings offer several practical implications. One of them is the potential use of serious games in schools to gather information about conflicts that may arise in the classroom. Serious games could also serve to identify the percentage of potential perpetrators or victims over specific periods of time, enabling the implementation of longitudinal studies aimed at mitigating such risks through preventive campaigns in educational settings. This approach may prove more effective than relying solely on questionnaires, where minors are often subject to observational bias and may not respond genuinely. Moreover, obtaining prevalence metrics through serious games may involve less bias than traditional methods with minors, as the immersive nature of gameplay reduces feelings of being evaluated or judged.

6. Conclusion

This study demonstrates that serious games can serve as a new research instrument for conducting prevalence studies and risk factor analyses. However, before implementing this tool, it is essential to validate the metrics derived from this new approach and establish their reliability. In the present research, this was achieved through exploratory and confirmatory factor analyses, as well as by evaluating the predictive capacity of the in-game measures against the traditionally assessed variable.

The exploratory and confirmatory factor analyses revealed the presence of two underlying latent constructs: one capturing current cyber aggression, operationalized through aggression-related scenarios within the serious game, and another reflecting past cyber aggression, assessed through scenarios involving previous cyberbullying experiences in the game and through the corresponding measure from the validated questionnaire (traditional method).

The predictive analysis further showed that the in-game variables are strong predictors of cyberbullying perpetration as measured by the validated questionnaire, and both instruments demonstrated adequate reliability.

Moreover, depending on the specific objectives of a prevalence study (precision, detection capacity, sensitivity, or relevance for educational programs), different measures may be employed, each supported by both frequentist and Bayesian statistical frameworks.

Finally, the measure selected for conducting risk factor analyses was *OR restricted*, chosen for its balanced performance, exhibited the highest precision. The results indicate that the in-game metrics, both for cyberbullying perpetration and for the associated risk factors, yield the strongest and most robust associations.

Overall, based on the analyses presented above, we argue that serious games constitute a promising methodological tool for experimental research. In the context of cyberbullying, they offer an innovative approach for estimating prevalence and investigating risk factors, with the broader goal of informing strategies to mitigate this phenomenon.

Appendix A. Game Decisions Transcript

Note: The RAYUELA video game covers several cybercrimes affecting minors. Only Adventures 1 and 3 deal with cyberbullying, so these were the ones used in this paper.

Adventure 1

Question 1: Photo Sharing

[Talking to Matthew after taking a selfie.]

Now we only have to share and tag the photos. Jane, do you want to share them, or do you prefer me to do it?

- ☐ *I will do it.*
- ☐ *You can do it.*

Question 2: Sociable

[Talking to Robert after sharing the selfie. Dialog depends on the previous answers.]

It seems you like to upload many photos and share stuff on your social network.

- ☐ *I would say I am sociable.*
- ☐ *I consider myself kind of shy.*

Question 3: Matthew Meme

[After receiving a message from Patty with the meme about Matthew.]

- ☐ *Hehe, it's funny, I will share the meme.*
- ☐ *I am not going to share it.*
- ☐ *I won't share the meme and I'm going to try to end up with this.*

Adventure 2

Question 1: Registration Name

[Creating a new profile on a social network. The user has to select a profile name.]

- ☐ *My name and my year of birth.*
- ☐ *My name and surname.*
- ☐ *My favourite music band name.*
- ☐ *Other famous/TV/Book character I like.*

Question 2: Registration Password

[The user has to select a profile password.]

- ☐ *I don't have time for this; I will leave the default password.*
- ☐ *My name and surname.*
- ☐ *I'll use the same password I have on other websites, so it's easier to remember.*
- ☐ *I am going to set a strong password, even if I have to invest some more time.*

Question 3: Registration Profile Type

[The user has to select a profile type.]

- ☐ Public profile.
- ☐ Private profile.

Question 4: Registration Profile Place

[The user has to select a profile place.]

☐ *The name of the city and neighbourhood where I live.*

☐ *The name of my school and country.*

☐ *Something fantastic, as "I am in the clouds", "In the moon" or "Too far away from X".*

☐ *Leave empty.*

Question 5: Registration Profile Photo
[The user has to select a profile photo.]

☐ *A photo of just me.*

☐ *A photo of me and some friends.*

☐ *A photo from the Internet, in which I do not appear.*

Question 6: Comment Patty Post
[Seeing a post from Patty on the social network.]

☐ *Good one!*

☐ *They are awesome.*

☐ *I don't like them, it's so childish.*

☐ *Don't send any comment. I'm sure she won't pay any attention to it.*

Question 7: Use PC
[Using the club's PC after accepting a friend request from a photographer on the social network.]

☐ *View messages.*

☐ *Check photographer's profile.*

Question 8: Friend Request
[Using the club's PC.]

☐ *Accept friend request.*

☐ *Reject friend request.*

☐ *Check photographer's profile.*

Question 9: Send Photos
[Checking messages after accepting the friend request.]

- ☐ *Send photos.*
- ☐ *Not send photos.*

Question 10: More Photos
[Checking messages after sending swimsuit photos.]

- ☐ *Send naked photos.*
- ☐ *Do not send naked photos.*

Question 11: More & More
[Checking messages after sending naked photos.]

- ☐ *Send more naked photos.*
- ☐ *Reject the request and inform Mary.*

Question 12: Ask Help

- ☐ *Ask for help to Mary.*
- ☐ *Say nothing.*

Question 13: Close Case

- ☐ *Ok, I'll check the profile.*
- ☐ *No, I won't check the profile.*

Question 14: Tell Parents
[At the end of the scene. Previous dialog depends on player decisions.]
We should report that profile to the social network. Besides, you should also tell your parents about it and see if we need to talk with the police.

- ☐ *I don't know, Mary, communication with them is not very easy. Lately they get angry about anything and we always end up shouting.*
- ☐ *I'm ashamed! I don't want to tell them something like that. It's better if I try to solve it on my own.*
- ☐ *Yeah, you're probably right. I'm a bit embarrassed but I'll give it a try.*

Question 15: Block Profile

- ☐ *Block the profile.*
- ☐ *Do not block the profile.*

Adventure 3

Question 1: Pirated Content

[Playing video games in your room.]

I know of some sites that pirate the content and then you can download the update for free.

- ☐ *I will download the pirated update through a website.*
- ☐ *I will wait until I have money or until my parents give me the money to buy the new expansion.*

Question 2: Pol or Paula

[Playing video games in your room. Your friends joke about Pol's appearance.]

- ☐ *I don't like this kind of jokes.*
- ☐ *That's funny.*
- ☐ *Say nothing.*

Question 3: Time Overrun

[Playing video games in your room. A warning pops up about the number of hours you have been online]

- ☐ *4 hours are not that much. So, I can keep chatting a bit longer.*
- ☐ *It's time to stop and disconnect for a while, although I might miss some juicy gossiping.*

Question 4: Pol Bullied

I heard that some guys are messing with Paul. I am worried that he may be bullied. What do you think?

- ☐ *They are just having fun; I would not call that bullying.*
- ☐ *I think it's not right... but calling that bullying is a bit of a stretch.*
- ☐ *I think that's unacceptable; we should do something about it.*

Question 5: Remind Matthew

[Your friends start messing with Pol.]

☐ *Yes, I had a similar bad experience... I don't like being picked on.*

☐ *No, it has never really happened to me, to my knowledge.*

☐ *Yes, it was me who messed with someone else... but it was not such a big deal.*

Question 6: Talk to Pol
So, shall we talk to Pol to see how he is?

☐ *It is better to let him be.*

☐ *Of course, we should try to help.*

Question 7: How to Help Pol
We can help you. You are not alone in this. I think...

☐ *We should go to tell the teacher, he should know what to do.*

☐ *We should report the comments to the social network, so that it doesn't happen again.*

☐ *We should not report it, because I don't want to get picked on for being a snitch. . .*

☐ *We should not report it as reporting is usually useless.*

Adventure 4

Question 1: Phishing Email
[You receive an email indicating that your social network account has been compromised.]

☐ *Is my account in danger?! I must act quickly before I lose it. I will follow the instructions in the email.*

☐ *I find it suspicious...I'll better go straight to my social network profile's security settings and change my password there.*

Question 2: New Password
[You proceed to change the password of your account.]

☐ (Password = Name123) *I'm going to leave a password very similar to the one I had before. Otherwise, I'll forget it...*

- ☐ (Password = Football10) *I'm going to make a password with some hobbies or things I like in it. So, I won't forget it!*
- ☐ (Password = Ax/2oP3%nY6) *I'm going to make my password difficult and long. It is more challenging this way, but much safer.*

Question 3: My Account Stolen
[If your account has been phished.]

- ☐ *It does not seem to be that worrying, it's just a social network account. We don't need to tell this to anyone. I can create another account after all.*
- ☐ *It is important to tell someone or report it, since the account contains personal information. It is a crime!*

Question 4: Other Account Stolen
[If John has not been phished.]

- ☐ *It does not seem to be that worrying, it's just a social network account. We don't need to tell this to anyone. He can create another account after all.*
- ☐ *It is important to tell someone or report it, since his account contains personal information. It is a crime!*

Adventure 5

Question 1: Secret Relationship
[You and your friends are commenting that Sheila has a new romantic relationship that is distancing her from her friends and you are worried.]

- ☐ *Love is love, and everyone experiences it in a different way. If she needed help, she would have asked for it, wouldn't she?*
- ☐ *Sounds a bit creepy to me, have you tried looking at her social media?*

Question 2: Biology Paper
[You must meet to do a biology assignment and indicate your preference to meet online or in person.]

- ☐ *Meet at the library this afternoon, so we can go to the cafeteria if we finish earlier.*

- ☐ *Do it by video-conference this afternoon, so we can be more comfortable at home.*

Question 3: Talk Sheila

[John sees Sheila, who is leaning against the wall with her eyes fixed on her mobile phone.]

- ☐ *I will talk to her and let her know she can trust me if she has any problem.*
- ☐ *I should text Mary since she is closer to Sheila.*

Adventure 6

Question 1: Migrant News Check

[When investigating a news item that appears to be false, you must decide which things seem most relevant to verify the information.]

- ☐ *How professional the web page looks like (style, images, design, etc).*
- ☐ *The source itself: is it a known newspaper/website or is it an unknown site?*
- ☐ *If the information looks accurate, for instance with enough numbers and statistics.*
- ☐ *Search on the Internet to contrast the information.*

Question 2: Web Page Looks Like

[Reviewing the website.]

- ☐ *It seems it is true. Definitely not a fake page.*
- ☐ *It looks quite professional, but does it mean it's not fake? We should try other options.*

Question 3: The Source

[Reviewing the source.]

- ☐ *It is a known newspaper, at least I've seen it quite a lot on Social Networks. I would say is not fake.*
- ☐ *Even though it is a kind of famous newspaper, it could contain fake information. We should try other options.*

Question 4: Information Looks Accurate
[Reviewing if the information looks accurate.]

- ☐ *Ok, they are displaying a big amount of data, and look at the graph as it rises. It looks pretty accurate.*
- ☐ *Ok, there are a lot of numbers and graphs, but that does not mean that the data is correct. Data can also be falsified.*

Question 5: Replay post
Ok, so first of all, we should report the content to the social network, and we should probably reply with this information, right?

- ☐ *It is not worth answering. Don't feed the troll!*
- ☐ *Yes, let's add the link to the anti-hoaxes website.*

Question 6: Regarding Charles
What should we do with Charles?

- ☐ *It is a basket case, there is little we can do.*
- ☐ *We should try to talk to him.*

References

[1] Chun, J., Lee, J., Kim, J., and Lee, S.: "An international systematic review of cyberbullying measurements", in *Computers in Human Behavior*, vol. 113, pp. 106485, 2020.

[2] Thomas, H. J., Connor, J. P., and Scott, J. G.: "Integrating traditional bullying and cyberbullying: Challenges of definition and measurement in adolescents–a review", in *Educational Psychology Review*, vol. 27, pp. 135-152, 2015.

[3] Morgan, C. B., Merrill, S. C., Clark, E. M., Wolfson, J. A., and Trubek, A. B.: "A serious games methodology to test solutions for regional food systems inequities", in *Journal of Rural Studies*, vol. 110, pp. 103366, 2024.

[4] Calvo-Morata, A., Alonso-Fernández, C., Freire, M., Martínez-Ortiz, I., and Fernández-Manjón, B.: "Serious games to prevent and detect bullying and cyberbullying: A systematic serious games and literature review", in *Computers & Education*, vol. 157, pp. 103958, 2020.

[5] Martinez, K., Menéndez-Menéndez, M. I., and Bustillo, A.: "Awareness, prevention, detection, and therapy applications for depression and anxiety in serious games for children and adolescents: systematic review", in *JMIR Serious Games*, vol. 9, n. 4, pp. e30482, 2021.

[6] Gomez, M. J., Ruipérez-Valiente, J. A., and Clemente, F. J. G.: "A systematic literature review of game-based assessment studies: Trends and challenges", in *IEEE Transactions on Learning Technologies*, vol. 16, n. 4, pp. 500-515, 2022.

[7] Freese, M., Lukosch, H., Wegener, J., and König, A.: "Serious games as research instruments–Do's and don'ts from a cross-case-analysis in transportation", in *European Journal of Transport and Infrastructure Research*, vol. 20, n. 4, pp. 103-126, 2020.

[8] Kim, Y. J., Knowles, M. A., Scianna, J., Lin, G., and Ruipérez-Valiente, J. A.: "Learning analytics application to examine validity and generalizability of game-based assessment for spatial reasoning", in *British Journal of Educational Technology*, vol. 54, n. 1, pp. 355-372, 2023.

[9] Abt, C. C.: "Serious games" in *University press of America*, 1987.

[10] Damianova, N., and Berrezueta-Guzman, S.: "Serious Games supported by Virtual Reality-Literature Review", in *IEEE Access*, 2025.

[11] Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J-F., and Rahwan, I.: "The moral machine experiment", in *Nature*, vol. 563, n. 7729, pp. 59-64, 2018.

[12] Brighi, A., Ortega, R., Pyzalski, J., Scheithauer, H., Smith, P. K., Tsormpatzoudis, H., and Tsorbatzoudis, H.: "European cyberbullying intervention project questionnaire", in *Computers in Human Behavior*, 2012.

[13] Watkins, M. W.: "Exploratory factor analysis: A guide to best practice", in *Journal of black psychology*, vol. 44, n. 3, pp. 219-246, 2018.

[14] Marengo, D., and Settanni, M.: "Examining the postdictive validity of self-report Big Five personality traits with objective recordings of online behaviors: a ten-year retrospective study using Facebook Page Likes", in *Heliyon*, vol. 10, n. 12, 2024.

[15] Gallagher, M. W., and Brown, T. A.: "Introduction to confirmatory factor analysis and structural equation modeling", in *Handbook of quantitative methods for educational research*, pp. 289-314. Rotterdam: SensePublishers, 2013.

[16] Fife, D. A., Longo, G., Correll, M., and Tremoulet, P. D.: "A graph for every analysis: Mapping visuals onto common analyses using flexplot", in *Behavior Research Methods*, vol. 53, n. 5, pp. 1876-1894, 2021.

[17] Kaiser, M. O.: "Kaiser-Meyer-Olkin measure for identity correlation matrix", in *Journal of the Royal Statistical Society*, vol. 52, n. 1, pp. 296-298, 1974.

[18] Costales, J., Catulay, J. J. J., Costales, J., and Bermudez, N.: "Kaiser-Meyer-Olkin factor analysis: a quantitative approach on mobile gaming addiction using random forest classifier", in *Proceedings of the 6th international conference on information system and data mining*, pp. 18-24, 2022, May.

[19] Binti, P. L., Guci, D. A., Mahmud, S., Sulong, Z., and Abd, M. S.: "THE ACCEPTANCE OF NEW MODEL OF BARAKAH HOUSE FINANCING AMONG PUBLIC IN MALAYSIA", 2018.

[20] Alhadban, R., Awang, Z., and Afthanorhan, A.: "A Measure of Impact of Electronic Human Resource Management (e-HRM) On Organization Performance: The Mediating Effects Of Competitive Advantage: An Exploratory Factor Analysis", in *International Journal of Psychosocial Rehabilitation*, vol. 24, n. 8, pp. 8194-8209, 2020.

[21] Shrestha, N.: "Factor analysis as a tool for survey analysis", in *American journal of Applied Mathematics and statistics*, vol. 9, n. 1, pp. 4-11, 2021.

[22] Reneses, M., Parder, M. L., Riberas-Gutiérrez, M., and Bueno-Guerra, N.: "Cyberbullying and cyberhate as an overlapping phenomenon among

adolescents in Estonia and Spain: Cross-cultural differences and common risk factors", in *Children and Youth Services Review*, pp. 108285, 2025.

[23] Jeffreys, H.: "Theory of Probability", 3rd edition, Oxford University Press, London, pp. 95-103, 1961.

[24] Lee, M. D., and Wagenmakers, E. J.: "Bayesian Cognitive Modeling: A Practical Course", Cambridge University Press, 2014.

[25] Quintana, D. S., and Williams, D. R.: "Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP", in *BMC psychiatry*, vol. 18, n. 1, pp. 178, 2018.

[26] Kopecký, K.: "Cyberbullying and sexting between children and adolescents–comparative study", in *Procedia-Social and Behavioral Sciences*, vol. 149, pp. 467-471, 2014.

[27] Perera, A., and Fernando, P.: "Accurate cyberbullying detection and prevention on social media", in *Procedia Computer Science*, vol. 181, pp. 605-611, 2021.

[28] Sanchez, D. R., and Langer, M.: "Video game pursuit (VGPu) scale development: Designing and validating a scale with implications for game-based learning and assessment", in *Simulation & Gaming*, vol. 51, n. 1, pp. 55-86, 2020.

[29] Chung, G. K., and Feng, T.: "From clicks to constructs: An examination of validity evidence of game-based indicators derived from theory", in *Assessment Analytics in Education: Designs, Methods and Solutions*, Cham: Springer International Publishing, pp. 327-354, 2024.