

doi: <https://doi.org/10.15446/rcp.v32n1.96322>

# The Role of Third Parties in Cyber Conflicts: The SIMA Simulator

**LUISA GARCÍA-VARGAS**

Pontificia Universidad Javeriana, Bogotá, Colombia

**EMILSE DURÁN-APONTE**

Universidad Simón Bolívar, Caracas, Venezuela

**ENRIQUE CHAUX**

Universidad de los Andes, Bogotá, Colombia



The views expressed here are those of the authors. Except where otherwise noted, the contents in this journal is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. Consultation is possible at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

**How to cite this article:** García-Vargas, L., Durán-Aponte, E., & Chaux, E. (2023). The role of third parties in cyber conflicts: The SIMA simulator. *Revista Colombiana de Psicología*, 32(1), 67-81. <https://doi.org/10.15446/rcp.v32n1.96322>

Correspondence concerning this article should be addressed to Enrique Chaux. e-mail: [echaux@uniandes.edu.co](mailto:echaux@uniandes.edu.co)

---

SCIENTIFIC RESEARCH ARTICLE

RECEIVED: JUNE 01, 2021—ACCEPTED: JULY 05, 2022

### **Abstract**

New media creates opportunities to directly measure how adolescents respond to aggressive situations. In this study, we report on SIMA, a simulator of digital chats, created to mimic online aggressive interactions within a conflict situation. Adolescents were invited to be part of a digital chat where all interactions were pre-programmed, except for the participants' interventions, to discuss controversial topics (e.g., animal maltreatment). Two chat members engaged in a conflict programmed to escalate quickly through aggressive insults. Sixth and seventh-grade students (mostly aged 12 to 14; n = 287) from one public school (low SES) and one private school (high SES) participated in the study. Responses to open questions after finishing the chat suggest that the great majority considered that they were taking part in a real on-line interaction. Aggressive offenses towards other members of the chat were very common. SIMA opens many options for further research about bystander behavior in aggressive online interactions.

*Keywords:* cyberaggression, chat, online aggression, conflict, pedagogical intervention.

### **El Papel de los Terceros en los Conflictos Cibernéticos: el Simulador SIMA**

#### **Resumen**

Los nuevos medios de comunicación han creado oportunidades para medir directamente cómo los adolescentes responden a situaciones de agresión. Presentamos a SIMA, un simulador de chats digitales creado para imitar interacciones agresivas en línea en una situación de conflicto. Los adolescentes son invitados a ser parte de un chat digital para discutir temas controversiales (p. ej., maltrato animal). En realidad, las interacciones están todas preprogramadas a excepción de las intervenciones de quienes participan. Dos supuestos miembros del chat tienen un conflicto programado para escalar de forma rápida con insultos agresivos. Participaron estudiantes de sexto y séptimo grado (principalmente de 12 a 14 años; n = 287) de un colegio público (NSE bajo) y uno privado (NSE alto). Las respuestas a preguntas abiertas sugieren que la gran mayoría consideró que estaban participando en una interacción en línea real. Las ofensas agresivas hacia otros miembros del chat fueron muy comunes. SIMA abre muchas opciones para futuras investigaciones sobre el comportamiento de espectadores en interacciones agresivas en línea.

*Palabras clave:* ciberagresión, chat, agresión virtual, conflicto, intervención pedagógica.

## Introduction

CYBER AGGRESSION is unfortunately very common in the everyday lives of many adolescents, and it can be very harmful -even more so than face-to-face aggression- to those subject to it (Sticca & Perren, 2013). Observation and measurement are crucial to understanding cyber aggression and to design better prevention strategies. However, to measure aggression (any action intended to harm others; Parke & Slaby, 1983), or any reaction to it, be it as a victim or a bystander (Macháčková, Dedkova, Sevcikova, & Cerna, 2018), is a great challenge. Many scientific studies of aggression and cyber aggression are based on self-reports (e.g., Burt & Alhabash, 2018; Kim et al., 2020; Kokko et al., 2014; Martínez, Rodríguez-Hidalgo, & Zych, 2020; Menin, Guarini, Mameli, Skrzypiec & Brighi, 2021; Runions, Bak, & Shaw, 2016; Wright, 2017). However, a great limitation of self-reports is the possible social desirability by which participants often do not respond with absolute honesty as they wish to project a positive image of themselves. Another alternative is peer reporting (e.g., Hooijsma, Huitsing, Dijkstra, Flache & Veenstra, 2020), that, although possibly more reliable than self-reporting, can also be biased given that peers may have limited perceptions of the situations, or preferences, and particular relations of friendship with their peers. Reports by teachers (e.g., Dickson et al., 2015) and parents (e.g., Bartels et al., 2018) are common, but they too can be limited, in part because adults often do not find out about situations of aggression that occur between children or adolescents (e.g., Atlas & Pepler, 1998). An important alternative is direct observation of aggressive behavior. Pepler and her group designed a system to observe aggressive situations using wireless cameras and microphones (Pepler & Craig, 1995). By doing so, they were able to measure the frequency of this behaviour very precisely and, as well, identify the social dynamics involved in many of these situations (e.g., Atlas & Pepler, 1998; Craig & Pepler, 1997). However, this methodology is technically complex and requires a great amount

of time, in part, because aggression is usually not so frequent (except during the first years of life; Tremblay, 2002).

Aggression has also been observed in experimental contexts. For example, there are a number of experimental paradigms in which participants can decide on the intensity of a negative stimulus (e.g., noise, electric shock, hot sauce) received by alleged opponents in a competition (e.g., Anderson & Bushman, 1997; Lieberman, Solomon, Greenberg & McGregor, 1999). However, these experiments present limitations too. For example, they differ from common situations where people exercise aggression daily, and this affects the external validity of the experiments, that is, the possibility to generalize the results to other contexts. Anderson and Bushman (1997) argue that this effect is often necessary to guarantee high internal validity; that is, to be sure that the causal relationships between the variables are not due to uncontrolled variables.

To achieve both internal and external validity, it would ideally be necessary to control variables in an experimental context that is very similar to the context where the daily interactions occur. Electronic media present an opportunity to do just this as they provide a specific environment to directly observe aggressive behaviour, and/or reactions to it, in contexts that are similar to how this behaviour occurs daily in current types of aggression, at least, among adolescents. Electronic contexts enable us to simulate very similar interactions to those that occur on social networks and to plan credible aggressive interaction to be able to systematically study participants' reactions to aggressive situations.

In this article, we report the design and early results given by SIMA, a system that simulates online chats and that enables direct observation of adolescents' bystander responses to what they believe is real cyber conflict. We report the reactions of the participants to the SIMA simulator, as well as the evolution over time of their different types of responses and some preliminary findings related to differences by sex, type of school, and topic of the argument.

Given the importance of the role of bystanders in the dynamics of aggression, we focus on their bystander responses to these situations. Observational research as that conducted by Pepler and her group has shown that what peer bystanders do has a great impact on whether the aggression increases or reduces (Atlas & Pepler, 1998; Craig & Pepler, 1998). In fact, several prevention programs have been aimed at promoting assertive (i.e., emphatically defending their own and others' rights but without harming others; Lange & Jakubowski, 1980) and empathetic (i.e., feeling something similar or compatible to what others are experiencing; Hoffman, 2000) bystander intervention (Chaux et al., 2017; Kärnä et al., 2011; Wölfer et al., 2013). We also analyzed responses to situations of cyber conflict given that almost all the studies conducted on the topic have focused on cyberbullying; that is, repeated and systematic cyber aggression against someone, usually in a situation of vulnerability and power imbalance (David-Ferdon & Feldman, 2007; Del Rey, Mora-Merchán, Casas, Ortega-Ruiz & Jaén, 2018; Grigg, 2010; Redondo, Luzardo & Rangel, 2016; Wright, 2015; Wright, et al., 2015). We found no studies on cyber aggression where there is no clear power imbalance among those involved and where aggression could go in both directions, despite that this type of cyber aggression is also very common in the lives of today's adolescents, as pointed out by Corcoran, Mc Guckin & Prentice (2015).

Some studies have shown that it is possible to take advantage of digital contexts to study the way that different variables can affect behaviour. Cheng, Bernstein, Danescu-Niculescu-Mizil & Lezkovec (2017), for example, conducted an experiment where participants interacted on a video chat on controversial political topics (e.g., the 2016 US presidential camping). During the experiment, the researchers sought to manipulate: 1) the participants' mood (before interacting on the chat, half of the participants had to take a very easy quiz, and the other half, a very difficult one); and 2) the context prior to the discussion (half had been subjected

to prior insulting comments, and the other half to neutral comments). Cheng et al. (2017) found that those that began the chat with a negative affect or in a context of having received prior offenses resorted to offenses more frequently than those who received prior positive affect in a neutral context. Therefore they showed that it is possible to influence the frequency of cyber aggression by manipulating the affective and contextual variables.

Similarly, Burt, Kim & Alhabash (2020) developed a simulated chat to measure digital aggression. In a reaction-time game adapted from the Taylor Aggression Paradigm (TAP; Chester & Lasko, 2019), participants received offensive chats from fictitious co-players who they thought were competitors, and their responses in the chat were recorded and analysed. Aggressive responses in their chats were found to correlate significantly to other measures of self-reported aggression and aggressive content in their Twitter posts (Burt et al., 2020; Kim, Clark, Donnellan & Burt 2020).

With this study, we sought to contribute to our understanding of adolescents' behavior in online contexts by piloting a simulator that allows the direct manipulation of the content of the interactions. Specifically, we intended to conduct a pilot implementation of the SIMA simulator and answer the following research questions:

1. How do participants respond to cyber conflict in SIMA?
2. How do participants' different types of intervention evolve during the SIMA dynamic?
3. How much do participants believe that the chat is real?
4. Can SIMA be used to find differences in responses to cyber conflicts according to sex, type of school, or topic of discussion?

## Method

### Participants

The study population was made up of 287 6<sup>th</sup> and 7<sup>th</sup> grade students from one public and one

private school in Bogotá (Colombia) in October 2016. The public school children totalled 172 (82 male and 90 female) and were aged between 10–15 years, whereas the private school students were 115 (61 were male and 54 female) and were aged between 12–14 years. Most of the students from the private school came from high socio-economic families (i.e., 92 % reported coming from strata 4 to 6 out of six SES-strata in Colombia), whereas those from the public school came from low socio-economic levels (i.e., 91% reported coming from strata 1 or 2) and contexts of vulnerability with high levels of community violence.

### The sima Instrument

The SIMA (Aggressive Interactions Simulator–Simulador de Interacciones Agresivas) instrument is a simulator designed to lead every participant to believe that he or she is interacting in a real online chat where, in fact, all interventions are previously planned except the participants' interventions. SIMA was constructed using a web application that can be accessed from any computer, tablet or cell phone (the following were used to develop the instrument: nodejs, npm, express, html 5, socket io, and MongoDB for the database; the simulator was uploaded to the OpenShift open source container application platform by Red Hat for the initial pilot tests to generate multiple access for data gathering).

Each participant was randomly assigned to one of four possible interaction topics: animal maltreatment, the environment, tattoos and piercings, and traffic. In the digital chat each participant interacts with three other people whose interventions are predefined by the application.

Each interaction lasts seven minutes and includes 46 interventions by fictitious participants (with intervals of 2 to 15 seconds between interventions). The first interventions are greetings (e.g., "hello everyone") or related to the topic (e.g., "in my neighbourhood, we look after the gardens"). As from interaction 12, two of the fictitious participants make offensive interventions (e.g., "and they

are ignorant, bull-fighters, for example") and they become involved in a dispute that escalates in its level of aggression (e.g., "stop pretending, are you saying that you don't shower and don't throw away rubbish. Where do you keep it, in your pigsty?" in interaction 26; "Stupiiid woman! I would paint your face so you can see how it feels, hahaha" in interaction 43). The four interactions are identical in the 46 interventions in terms of the type of intervention and aggressive escalation. Given that the interactions are previously programmed, the participants' interventions do not modify those of the other three alleged members of the chat group. All the participants' interventions were recorded to be analysed subsequently.

To finish, each student was asked to answer seven open questions on their experience in the chat: "What did you think of the interaction? What did you think of the other people in the chat group? What happened during the interaction? What did you think you would do? What did you do? Did you think of doing something different to what you did? Why didn't you do it?" The main purpose of these questions was to indirectly check whether students thought that the chat sessions were real.

The first version of the interactions was developed based on informal interviews conducted by the first author with students aged 12 to 15. We compiled as many interaction topics to motivate arguments among them, as we did comments, responses, and reactions frequently used in their discussions, based on these we selected the four topics and built the conversations. When the first version of the simulator was ready, a pilot was conducted with nine adolescents who commented on the time between the interventions in the interactions, the terms used, and the formality of some of the responses. All the participants of the pilot test explained that they thought that they were interacting with other adolescents. Their comments were considered when constructing the version of the SIMA used in this study.

In addition to the SIMA instrument we applied a demographic questionnaire (i.e., alias or fictitious

name, age, sex, and socio-economic level of the participants' households), an instrument dealing with the creative generation of options whose results are not presented here, and a 22-item questionnaire on attitudes related to animal maltreatment, the environment, tattoos and piercings, and traffic that sought to reinforce the participants' belief that this was a study of opinions on the topics. These other instruments were applied prior to the SIMA.

### Procedure

The instrument was applied to all grades on the same day, in computer rooms at both schools. The different groups participated sequentially so that they would not have time to speak with other students. Each student had his or her own computer and they were asked to sit a determined distance from each other to not be able to see the information entered by other students.

A month before applying the instrument, the students' parents or guardians were sent an informed consent form. Only those students whose parents had signed and returned a consent form were allowed to take part in the study. Only 3 of the 118 students (2.5%) in the private school and 8 of the 180 students (4.4%) in the public school did not receive parental consent. Furthermore, before starting the students were told that their participation was voluntary and that they could withdraw from the study at any given point and that only if they agreed with this, they should sign an assent form. None withdrew.

They were given a piece of paper with the Internet addresses they should use to access the instrument. To guarantee randomness, and to avoid them sitting next to other students working on the same topic, the sheets of paper had different addresses for each one of the different topics. The students did not choose the topics they were going to discuss. At the end of the application, they were given a verbal explanation about their interactions not being real.

The study involved deception since participants were made believe that the chat discussion was real. According to APA (2010), deception in

research is only justified if: 1) the study has "prospective scientific, educational, or applied value" and "nondeceptive alternative procedures are not feasible"; 2) it is reasonably expected that the study will not "cause physical pain or severe emotional distress"; and 3) debriefing occurs "as early as is feasible, preferably at the conclusion of their participation". The method complied with these conditions as it represents a substantial improvement in comparison to existing self-report methods, there are no reasons to consider that it can cause any harm, and debriefing occurred immediately after their participation in the simulation.

### Data Analysis

Initially, the data gathered were imported to Excel, and then each of the participants' interventions were classified by two coders based on the following eleven categories. Interventions by fictitious participants are presented in braces.

1. Thematic intervention ( $\kappa = .929$ ): comments on the topic of interaction (e.g., "in my residential complex throwing rubbish is prohibited"; "even if we need it, we mustn't harm the environment").
2. Social interaction ( $\kappa = .933$ ): greetings, goodbyes, and conventional interactions (e.g., "hello"; "how are you?"; "good thanks"; "goodbye"; "CU").
3. Offends other members of the chat group ( $\kappa = .897$ ): direct offenses of those who (the participant believes) take part in the interaction (e.g., "you don't even know what you're talking about so, shhhhh"; {Camio4: comon, they're walking like idiots, you are like horses with blinkers on} "that's very true").
4. Offends others who are not members of the chat group ( $\kappa = .914$ ): direct offenses at others who are not part of the chat group (e.g., "people who don't accept [tattoos] are monsters, I mean it's 2017, accept them already"; "Yes it's true, why did [pedestrians] have to be such animals").
5. Aggressively defends her/himself ( $\kappa = .851$ ): uses offenses to defend him or herself

- and/or tries to deescalate the conflict using offenses (e.g., “it’s not about making yourself feel important, stupid. We all have different tastes and if you don’t want to get a tattoo then just don’t, idiot”; “get lost”).
6. Aggressively defends other members of the chat group ( $\kappa = .881$ ): uses offenses to defend those who (the participant believes) are part of the interaction (e.g., “and this is why there are campaigns aimed at correcting this, you should read a bit more. You are IGNORANT!!”; “NO, YOU ARE! Killing is not an art”).
  7. Aggressively defends others who are not members of the chat group ( $\kappa = .893$ ): uses offenses to defend someone who is not part of the interaction (e.g., “{AnakinSW: really? The builders are idiots?} “No they are not, they only follow orders... Their bosses are idiots”; “Don’t be so rude, STUPIDDDDD”).
  8. Assertively defends him/herself ( $\kappa = .859$ ): defends him or herself in a non-offensive manner and/or seeks to deescalate the conflict (e.g., “am I less of a human being than you for having a tattoo or piercing?”; “I’m not stupid either and I love tattoos; they are brilliant”).
  9. Assertively defends other members of the chat group ( $\kappa = .829$ ): defends those who (the participant believes) are part of the interaction in a non-offensive manner (e.g., “WOW calm down, you don’t have to go that far...”; “Stop, don’t treat him like that”; “You’re arguing too much, calm down”; “Give her some respect she can express her opinions”).
  10. Assertively defends others who are not members of the chat group ( $\kappa = .920$ ): defends someone who is not part of the interaction in a non-offensive manner (e.g., “We are all free to do what we want as long as we don’t harm others”; “they’re not idiots”; “they’re not animals, it’s just that they were never taught to value nature”).
  11. Other interventions ( $\kappa = .784$ ): not intelligible, not clear what they are referring

to, different possible interpretations, or does not correspond clearly to any other category (e.g., “?????”; “oohhhhh”; “we’ll talk about this tomorrow”).

To verify the consistency of the classifications, an assessment of the inter-observer reliability between two researchers (the first and second authors) was conducted. To do so, the  $\kappa$  coefficient for consistency between observers was calculated for each of the categories. The disagreements were analyzed by all authors, to adjust the definitions of the categories. The procedure was repeated four times until  $\kappa$  coefficients were above .78 for all categories, that means high consistency according to Fleiss (1981).

Once the interventions were well classified, quantitative analyses were conducted based on the cumulative number of interventions per participant for each category. Finally, using one-way ANOVAS, means were compared between types of school (public vs. private), sex (male vs. female) and topics of interaction (tattoos & piercing, the environment, animal maltreatment, traffic).

The participants’ answers to two of the open questions (“What did you think of the interaction?” and “What did you think of the other people in the chat group?”) were assessed independently by the first and second authors depending on whether the answer indicated that they believed that the chat group was real or not: 1) he/she definitely did not believe it to be real; 2) it seems that he/she did not believe it was real (e.g., “a little strange”, “very strange”); 3) he/she is not sure whether they believed it was real (e.g., “I don’t know”; no answer); 4) it seems that he/she did believe it was real (e.g., “they started to fight”, “there were ignorant”); 5) he/she definitely did believe it was real (e.g., “I think that there are some well-educated people and some ignorant people”, “Athena has to change, it’s because of people like her that we are in the mess we are”). The answers of the two evaluators were significantly correlated ( $r = .618, p = <.001$ ). Following this, the two evaluators together reviewed

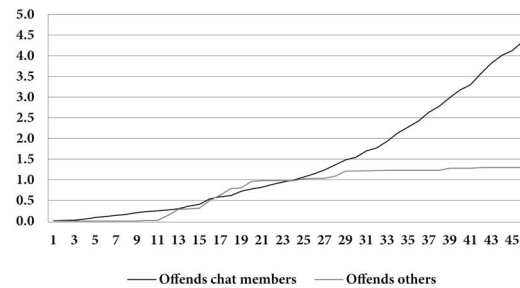
the classification related to what the difference was 2 points or above (14.6 % of the answers) and they reached a consensus on each one.

### Results

Participants' offenses against the alleged members of the chat group were very frequent. On average, each participant used 4.3 offenses during the simulation. The frequency of aggressive interventions increased continually during the interaction. In fact, in the last 10 interactions, there were approximately the same number of offenses as there were in the first 35 interactions (see Figure 1). There were no significant differences between public and private schools (see Table 1), or between male and female (see Table 2). But

differences were observed for topics: those who were assigned to discussion about tattoos made more offensive comments than those who had to discuss environmental topics (see Table 3).

**Figure 1.** Average cumulative counts for offenses directed at those supposedly in the chat group and others not present in the chat group.



**Table 1.** ANOVAS for Differences among Students from Private and Public Schools

	Private	Public	F	Sig.
Aggression	6.34	6.52	.06	.802
Offends members of the chat group	3.97	4.59	1.17	.281
Offends others	1.33	1.27	.07	.786
Aggressively defends him/herself	.69**	.37	7.63	.006
Aggressively defends chat members	.16	.16	0.00	.993
Aggressively defends others	.20	.13	1.77	.184
Assertiveness	2.43***	1.27	17.43	.000
Assertively defends him/herself	.37	.41	.22	.640
Assertively defends chat members	1.17***	.51	12.39	.001
Assertively defends others	.90***	.35	20.84	.000
Social interaction	3.76	4.47**	8.79	.003
Thematic intervention	11.61	15.37***	21.06	.000
Other interventions	4.03	4.37	.47	.493

Note: \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

**Table 2.** ANOVAS for Differences between Male and Female Adolescents

	Male	Female	F	Sig.
Aggression	6.44	6.45	.00	.980
Offends members of the chat group	4.49	4.19	.28	.594
Offends others	1.28	1.31	.01	.911
Aggressively defends him/herself	.35	.64**	6.76	.010
Aggressively defends chat members	.14	.17	.51	.476
Aggressively defends others	.17	.14	.40	.526



	Male	Female	F	Sig.
Assertiveness	1.44	2.03*	4.34	.038
Assertively defends him/herself	.33	.45	2.17	.142
Assertively defends chat members	.70	.85	.65	.421
Assertively defends others	.42	.72*	6.37	.012
Social interaction	4.20	4.16	.03	.865
Thematic intervention	13.16	14.57	2.89	.090
Other interventions	4.17	4.30	.07	.786

Note: \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

**Table 3.** ANOVAS for Different Topics of Interaction Assigned

	Environment	Animal maltreatment	Tattoos & piercing	Traffic	F	Sig.
Aggression	6.00	6.77	6.74	6.10	.34	.800
Offends members of the chat group	3.26 <sup>b</sup>	4.14	5.55 <sup>a</sup>	3.86	3.11	.027
Offends others	2.15 <sup>a</sup>	2.08 <sup>a</sup>	.01 <sup>b</sup>	1.47	34.39	.000
Aggressively defends him/herself	.54	.14 <sup>b</sup>	.75 <sup>a</sup>	.49	5.48	.001
Aggressively defends chat members	.02 <sup>b</sup>	.30 <sup>a</sup>	.16	.12	4.74	.003
Aggressively defends others	.04 <sup>b</sup>	.11	.27 <sup>a</sup>	.15	3.46	.017
Assertiveness	1.22 <sup>b</sup>	1.11 <sup>b</sup>	2.58 <sup>a</sup>	1.68	6.57	.000
Assertively defends him/herself	.39	.37	.44	.36	.20	.986
Assertively defends chat members	.43 <sup>b</sup>	.31 <sup>b</sup>	1.35 <sup>a</sup>	.79	7.16	.000
Assertively defends others	.41	.44	.80	.53	2.37	.071
Social interaction	5.07 <sup>a</sup>	5.44 <sup>a</sup>	3.18 <sup>b</sup>	3.52 <sup>b</sup>	29.58	.000
Thematic intervention	13.02	13.24	13.92	15.01	1.10	.349
Other interventions	3.91	3.86	4.34	4.71	0.63	.593

Note: Cells with different letters (superscript) represent significant differences ( $p < .05$ ) in Bonferonni post hoc test.

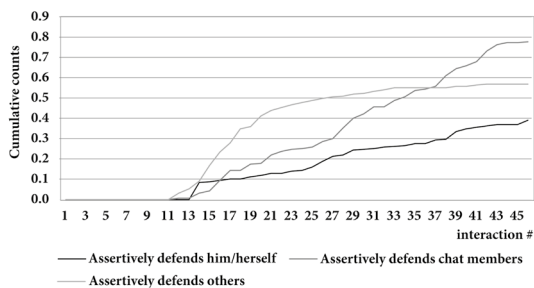
Offenses against those who were not part of the chat were less frequent (1.3 on average) and they were concentrated in the second quarter of the time for the interaction (see Figure 1); that is, when the planned offensive comments began (e.g., 12: “and, they are ignorant, for example, the bullfighters...”; 13: “and what about those who go to see bullfighting? Are they ignorant too?”). There were no differences between types of school or between sexes (see Tables 1 and 2), but differences were observed for topics: those who discussed environmental topics or animal maltreatment

offended others to a significantly greater extent than those who argued about tattoos (see Table 3).

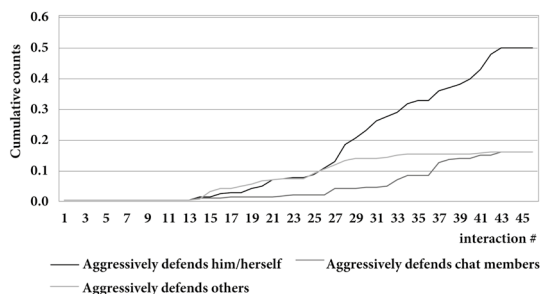
As can be seen in figures 2 and 3, defending assertively him/herself and the other members of the chat increased substantially in the second half of the simulation. Assertively defending chat members was two times more frequent than assertively defending themselves. In contrast, aggressively defending themselves was more than three times more frequent than aggressively defending chat members. Significant and substantial differences were found between the types of school in three categories

related to the defence. In particular, interventions that assertively defended members of the chat session, assertively defending others who were not in the chat groups, and the participants aggressively defending themselves were much more common among participants from public schools (see Table 1). On the other hand, assertively defending others who were not in the chat group and aggressively defending themselves were much more common among female than male participants (see Table 2).

**Figure 2.** Average cumulative counts for assertively defending themselves, defending those supposedly in the chat group, and defending others not present in the chat group.



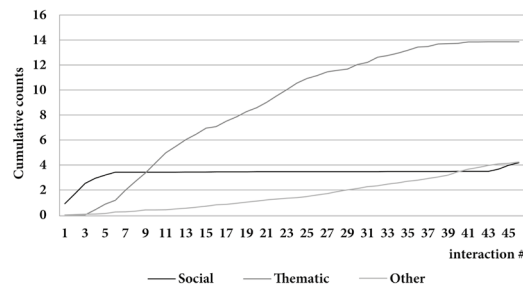
**Figure 3.** Average cumulative counts for aggressively defending her/himself, aggressively defending other members of the chat group, and aggressively defending others who are not members of the chat group.



Both conventional social interactions, and thematic interventions were more frequent among public school students than among private school students (see Table 1). We found no significant differences by sex (see Table 2). Social interactions occurred almost exclusively in the first 6 and last 3 iterations, whereas the thematic interventions were

more frequent in the first interactions than in the last. Less than 10 % of the thematic interventions took place in the last quarter of the time allocated to the interaction (see Figure 4).

**Figure 4.** Average cumulative counts for social, thematic, and other unclassified interventions.



After finishing the chat, the participants were asked to answer open questions. An analysis of their answers to the questions “What did you think of the interaction?” and “What did you think of the other people in the chat group?” indicated that the majority seemed to have believed that the chat session was real. None of the answers indicated that they had not believed that the interaction was real, and only 8.3 % of the participants responded in a way that suggests that they may have suspected that the interaction was not real (e.g., “it was a bit strange”; “very strange”). In contrast, 69 % responded in ways that suggest that they did think that the interaction was real (e.g., “they are very disrespectful”; “some of them are respectful but they become aggressive when defending their point of view”).

**Discussion**

Electronic media provide a great opportunity to directly measure behaviour in digital contexts like those where a large part of today’s interactions among adolescents take place. These measurements can be much more reliable than some methods currently used to measure behaviour in contexts of aggression, as self-reporting or reports by adults. They allow this analysis to be conducted

in a context where variables can be manipulated to directly observe the causal effect of certain variables on behaviour. For example, specific variables could be manipulated to see whether this influences the participants' behaviour (e.g., whether those who received an intervention to promote empathy intervene more when it comes to deescalating conflict). Anderson and Bushman (1997) suggest that when seeking to increase external validity (i.e., results that can be generalized to real contexts) conducting studies in real contexts often ends up affecting internal validity (i.e., being confident that the variables are, in fact, casually related) given that in real contexts it is impossible to control the variables that can affect the results (e.g., confound variables). Computer simulations like the ones presented here simultaneously allow high internal and external validity, given that they can be conducted in experimental contexts where the variables are well-controlled and, at the same time, in context similar to real ones where online interactions take place on a daily basis. The fact that most of the participants seemed to have believed that they were taking part in a real online conversation indicates that simulators like SIMA have great potential to simultaneously achieve both types of validity.

In this study, we found that cyber aggression was very common. Following the thematic interventions, the offenses against the other (alleged) participants in the chat group were the most common interventions. On average, each participant wrote more than 6 offenses, 4 of which were directed at other members of the chat group. Also, 85% of the participants wrote at least one offence. This highlights a worrisomely high frequency of aggression, being consistent with studies that suggest that the distance to the victim and anonymity on social networks favour the use of aggression (Arab & Díaz, 2015), in part, because online contexts seem to be associated with lower levels of empathy (Carrier, Spradlin, Bunce & Rosen, 2015).

No differences by sex were found in terms of the total amount of cyber aggression. This is

consistent with studies carried out in different cultures that show that there are few differences between men and women in the frequency of indirect aggression (i.e., when the victim does not know who began the aggression), in contrast to direct physical and verbal aggression that is more common among men than it is among women (Card, Stucky, Sawalani & Little, 2008; Österman et al., 1998; Salmivalli & Kaukiainen, 2004). In this case, the anonymity of chats allows people to offend others assuming that victims will never know who offended them. More studies are needed to verify whether this lack of difference between men and women also occurs in other cultures.

Assertive bystander intervention is fundamental to stopping situations of aggression in face-to-face contexts (Atlas & Pepler, 1998; Craig & Pepler, 1997; Salmivalli, 2010, Salmivalli, 2014) and probably also in online contexts (Constantino & Álvarez, 2010). In this study, however, only 38% of the participants intervened assertively to defend those who they thought were other adolescents being victims of cyber aggression. Also, these assertive interventions were significantly more frequent between students from the private school than those from the public school. It is not clear whether this difference is associated with more empathy or assertiveness among students from the private school, with a different perception of the seriousness of the offenses, or something else. In fact, many comments from the students from the public school suggest that some of them did not consider that the attacks were particularly alarming. For example, when asked what happened during the interaction, one of the participants answered, "nothing important", and another, "a few fights but it was very entertaining". It would be interesting to conduct a qualitative exploration of whether the students from different contexts could be interpreting the level of seriousness of the offences differently and, therefore, the need to intervene to defend those being attacked. The results also showed a greater frequency of assertive interventions among female adolescents than male

adolescents, when assertively defending other people that were not part of the chat group (e.g., youngsters with tattoos). Another important element to explore in future studies is whether this difference is associated to greater sensitivity or empathy, a greater capacity to intervene assertively or some other difference.

It is important to recognize that the study had several limitations that could be corrected in future versions. On the one hand, answers to the open questions suggest that the great majority of participants believed that the chat session was real, but we cannot rule out that some may have realized that it was a simulation. For example, some students may have seen their classmates' chats and discovered that they had very similar interactions in the public school where there was very little space from one computer to another. We tried to avoid this by staggering the topics between the computers and changing most of the aliases of the fictitious characters, but we cannot rule out that they may have seen each other's interactions anyway. Future technological adjustments will help to minimize the possibility of these risks occurring.

Another limitation of this study is that the classification of the participants' interventions depended on our interpretations of them. We may have misunderstood what the participants might have wanted to say. This is particularly possible given the ambiguity in short interactions typical of online chat sessions, as well as the generational difference between participants (adolescents) and the assessors (adults). Even though the KAPPAS obtained were high, we cannot ensure that what coders interpreted in a coherent manner correspond to what the participants intended to say. Additionally, coders were conscious of the hypotheses of the study, which could have biased their interpretations. Furthermore, it was necessary that coders reached agreements related to some of the entries since students used local expressions that were not known by both researchers. For this reason, coding was not totally independent and KAPPAS may be overestimated. Future studies might

need independent coders, closer in age to the participants, from their local contexts, and unaware of the hypotheses of the study. Interviewing the participants associated to their interventions may help to confirm the validity of the interpretations. Also, future studies could include emoticons to highlight the participants' emotional experiences during the chat session.

We also found significant differences in some of the categories among the groups assigned to different topics of interaction. For example, there were more assertive interventions amongst those assigned the topic of tattoos than among those assigned the topic of animal maltreatment and the environment. This is surprising given that the interventions of the alleged peers were planned to be equivalent in terms of the type of intervention (e.g., topic), their level of aggression, and when these took place. It is possible that the different topics may have aroused different levels of empathy and, therefore, different dispositions to intervening assertively to defend the alleged participants of the chat group. Future research should determine whether it is possible to obtain greater coherence among the discussion topics, or whether the simulation should be limited to single topics to guarantee comparability between the interventions by the different participants.

Another important aspect to explore in future research is a possible adaptation of the system to simulate other types of situations. In fact, we have already constructed a variation to the system that corresponds more closely to a situation of cyberbullying than one of cyber-conflict. In the situation reported here, the aggression is bilateral and there is no clear power imbalance between the two mutually aggressive alleged peers. This corresponds more to a situation of conflict than it does one of bullying (Chaux, 2012). In contrast, in cyberbullying, all the aggression of the alleged peers is directed at one specific peer who, in turn, does not seem to be able to defend him/herself in a clear show of power imbalance. We do not know yet whether cyberbullying or cyber-conflict

generates more assertive interventions by the participant-bystanders.

Despite the limitations, SIMA provides a great opportunity to directly explore adolescents' behaviour in digital spaces very similar to those where situations of aggression occur on a daily basis. These simulators could lead to a new drive towards research in social psychology that could not previously be conducted for ethical reasons (e.g., Burger, 2009). In addition to its usefulness as a methodological tool through which to observe behaviour, the simulator has many practical implications. On the one hand, it could be used to assess the impact of programs aimed at promoting assertive bystander intervention in situations of online aggression. On the other, it can serve as a pedagogical strategy; for example, for students to practice how they could assertively intervene in situations of online aggression, as the system would allow them to analyse their own reactions and identify what they could improve on and how. Finally, identifying variables that can explain the bystander behaviour in situations of cyber aggression can inform future programs where variables should be taken into account to further promote assertive bystander intervention. All this can contribute not only to a better understanding of aggression but also to promoting more peaceful and constructive online environments.

### Acknowledgments

We are very grateful to all the adolescents who very willingly participated in the study, as well as with their parents, teachers, and schools. We thank all the colleagues and reviewers who carefully read the study and offered insightful suggestions. We are also very grateful for the financial support provided by Pontificia Universidad Javeriana–Bogotá, Universidad Simón Bolívar–Caracas, and Universidad de los Andes–Bogotá.

### References

- American Psychological Association (2010). *Ethical Principles of Psychologists and Code of Conduct*.
- Anderson, C. A., & Bushman, B. J. (1997). External validity of “trivial” experiments: The case of laboratory aggression. *Review of General Psychology*, 1, 19-41. <https://doi.org/10.1037/1089-2680.1.1.19>
- Arab, E., & Díaz, A. (2015). Impacto de las redes sociales e internet en la adolescencia: aspectos positivos y negativos. *Revista Médica Clínica Las Condes*, 26, 7-13. <https://doi.org/10.1016/j.rmclc.2014.12.001>
- Atlas, R. S., & Pepler, D. J. (1998). Observations of bullying in the classroom. *Journal of Educational Research*, 92, 86-99. <https://doi.org/10.1080/00220679809597580>
- Bartels, M., Hendriks, A., Fung I., H., Hagenbeek, F., Nivard, M., van Dongen ... Rose, R. (2018). Childhood aggression and the co-occurrence of behavioural and emotional problems: Results across ages 3-16 years from multiple raters in six cohorts in the EU-ACTION project. *European Child & Adolescent Psychiatry*, 27, 1105-1121. <https://doi.org/10.1007/s00787-018-1169-1>
- Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist*, 64, 1-11. <https://doi.org/10.1037/a0010932>
- Burt, A., & Alhabash, S. (2018). Illuminating the nomological network of digital aggression: Results from two studies. *Aggressive Behavior*, 44, 125-135. <https://doi.org/10.1002/ab.21736>
- Burt, S. A., Kim, M., & Alhabash, S. (2020). A novel in vivo measure of cyberaggression. *Aggressive Behavior*, 46, 449-460. <https://doi.org/10.1002/ab.21911>
- Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development*, 79, 1185-1229. <https://doi.org/10.1111/j.1467-8624.2008.01184.x>
- Carrier, L. M., Spradlin, A., Bunce, J. P., & Rosen, L. D. (2015). Virtual empathy: Positive and negative impacts of going online upon empathy in young adults. *Computers in Human Behavior*, 52, 39-48. <https://doi.org/10.1016/j.chb.2015.05.026>
- Chaux, E. (2012). *Educación, convivencia y agresión escolar*. Bogotá, Colombia: Ediciones Uniandes. Taurus, Santillana. <https://doi.org/10.7440/2012.62>
- Chaux, E., Barrera, M., Molano, A., Velásquez, A. M., Castellanos, M., Chaparro, M. P., & Bustamante, A.

- (2017). Classrooms in Peace within violent contexts: Field evaluation of *Aulas en Paz* in Colombia. *Prevention Science*, 18, 828-838. <https://doi.org/10.1007/s11121-017-0754-8>
- Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Anyone can become a troll: Causes of trolling behavior in online discussions. *cscw Conf Comput Support Coop Work*, 1217-1230. <https://doi.org/10.1145/2998181.2998213>
- Chester, D. S., & Lasko, E. N. (2019). Validating a standardized approach to the Taylor aggression paradigm. *Social Psychological and Personality Science*, 10, 620-631. <https://doi.org/10.1177/1948550618775408>
- Constantino, G. D., & Álvarez, G. (2010). Conflictos virtuales, problemas reales: caracterización de situaciones conflictivas en espacios formativos online. *Revista Mexicana de Investigación Educativa*, 15(44), 65-84. Recuperado de <https://www.redalyc.org/pdf/140/14012513005.pdf>
- Corcoran, L., Mc Guckin, C., & Prentice, G. (2015). Cyberbullying or Cyber aggression?: A review of existing definitions of cyber-based peer-to-peer aggression. *Societies*, 5, 245-255. <https://doi.org/10.3390/soc5020245>
- Craig, W., & Pepler, D. (1997). Observations of bullying and victimization in the schoolyard. *Canadian Journal of School Psychology*, 2, 41-60. <https://doi.org/10.1177/082957359801300205>
- David-Ferdon, C. M., & Feldman, M. (2007). Electronic media, violence, and adolescents: An emerging state health problem. *Journal of Adolescent Health*, 41, 51-55. <https://doi.org/10.1016/j.jadohealth.2007.08.020>
- Del Rey, R., Mora-Merchán, J. A., Casas, J. A., Ortega-Ruiz, R., & Jaén, P. E. (2018). Programa "Asegúrate": Efectos en ciberagresión y sus factores de riesgo. *Comunicar*, 56, 39-48. <https://doi.org/10.3916/C56-2018-04>
- Dickson, D. J., Richmond, A. D., Brendgen, M., Vitaro, F., Laursen, B., Dionne, G., & Boivin, M. (2015). Aggression can be contagious: Longitudinal associations between proactive aggression and reactive aggression among young twins. *Aggressive Behavior*, 41, 455-466. <https://doi.org/10.1002/ab.21582>
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. Wiley.
- Grigg, D. W. (2010). Cyber-aggression: Definition and concept of cyberbullying. *Journal of Psychologists and Counsellors in Schools*, 20, 143-156. <https://doi.org/10.1375/ajgc.20.2.143>
- Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511805851>
- Hooijsma, M., Huitsing, G., Dijkstra, J. K., Flache, A., & Veenstra, R. (2020). Being friends with or rejected by classmates: Aggression toward same- and cross-ethnic peers. *Journal of Youth and Adolescence*, 49(3), 678-692. <https://doi.org/10.1007/s10964-019-01173-1>
- Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa anti-bullying program: Grades 4-6. *Child Development*, 82, 311-330. <https://doi.org/10.1111/j.1467-8624.2010.01557.x>
- Kim, M., Clark, S. L., Donnellan, M. B., & Burt, S. A. (2020). A multi-method investigation of the personality correlates of digital aggression. *Journal of Research in Personality*, 85, 103923. <https://doi.org/10.1016/j.jrp.2020.103923>
- Kokko, K., Simonton, S., Dubow, E., Lansford, J. E., Olson, S. L., Huesmann, L. R., Pettit, G. S. (2014). Country, sex, and parent occupational status: Moderators of the continuity of aggression from childhood to adulthood. *Aggressive Behavior*, 40, 552-567. <https://doi.org/10.1002/ab.21546>
- Lange, A., & Jakubowski, P. (1980). *Responsible assertive behavior: Cognitive, behavioural procedures for trainers*. Champaign: Research Press. <https://doi.org/10.2307/583463>
- Lieberman, J. D., Solomon, S., Greenberg, J., & McGregor, H. A. (1999). A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior*, 25, 331-348. [https://doi.org/10.1002/\(SICI\)1098-2337\(1999\)25:5<331::AID-AB2>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1098-2337(1999)25:5<331::AID-AB2>3.0.CO;2-1)
- Macháčková, H., Dedkova, L., Sevcikova, A., & Cerna, A. (2018). Bystanders' supportive and passive responses to cyberaggression. *Journal of School Violence*, 17, 99-110. <https://doi.org/10.1080/15388220.2016.1222499>
- Martínez, J., Rodríguez-Hidalgo, A. J., & Zych, I. (2020). Bullying and cyberbullying in adolescents from disadvantaged areas: Validation of questionnaires;

- prevalence rates; and relationship to self-esteem, empathy and social skills. *International Journal of Environmental Research and Public Health*, 17, 6199. <https://doi.org/10.3390/ijerph17176199>
- Menin, D., Guarini, A., Mameli, C., Skrzypiec, G., & Brighi, A. (2021). Was that (cyber) bullying? Investigating the operational definitions of bullying and cyberbullying from adolescents' perspective. *International Journal of Clinical and Health Psychology*, 21. <https://doi.org/10.1016/j.ijchp.2021.100221>
- Österman, K., Björkqvist, K., Lagerspetz, K. M. J., Kaukiainen, A., Landau, S. F., Fraczek, A., & Caprara, G. V. (1998). Crosscultural evidence of female indirect aggression. *Aggressive Behavior*, 24, 1-8. [https://doi.org/10.1002/\(sici\)1098-2337\(1998\)24:1<::aid-abi>3.0.co;2-r](https://doi.org/10.1002/(sici)1098-2337(1998)24:1<::aid-abi>3.0.co;2-r)
- Parke, R. D., & Slaby, R. G. (1983). The development of aggression. En P. H. Mussen & E. M. Hetherington (Eds.), *Handbook of Child Psychology. Socialization, Personality, and Social Development*. (pp. 567-641). New York, USA: Wiley.
- Pepler, D. J., & Craig, W. M. (1995). A peek behind the fence: Naturalistic observations of aggressive children with remote audiovisual recording. *Developmental Psychology*, 31, 548-553. <https://doi.org/10.1037/0012-1649.31.4.548>
- Redondo, J., Luzardo, M., & Rangel, K. (2016). Ciberagresión: un estudio sobre la prevalencia en estudiantes universitarios Colombianos. *Informes Psicológicos*, 16, 85-99. <http://dx.doi.org/10.18566/infpsicv16n1a05>
- Runions, K. C., Bak, M. & Shaw, T. (2016). Disentangling functions of online aggression: The Cyber-Aggression Typology Questionnaire (CATQ). *Aggressive Behavior*, 43(1). 74-84. <https://doi.org/10.1002/ab.21663>
- Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior*, 15, 112-120. <https://doi.org/10.1016/j.avb.2009.08.007>
- Salmivalli, C., & Kaukiainen, A. (2004). 'Female aggression' revisited: Variable- and person-centered approaches to studying gender differences in different types of aggression. *Aggressive Behavior*, 30, 158-163. <https://doi.org/10.1002/ab.20012>
- Salmivalli, C. (2014) Participant roles in bullying: How can peer bystanders be utilized in interventions? *Theory Into Practice*, 53, 286-292. <https://doi.org/10.1080/00405841.2014.947222>
- Sticca, F. & Perren, S. (2013). Is cyberbullying worse than traditional bullying? Examining the differential roles of medium, publicity, and anonymity for the perceived severity of bullying. *Journal of Youth and Adolescence*, 42, 739-750. <https://doi.org/10.1007/s10964-012-9867-3>
- Tremblay, R. E. (2002). Prevention of injury by early socialization of aggressive behavior. *Injury Prevention*, 8, 17-21. [https://doi.org/10.1136/ip.8.suppl\\_4.iv17](https://doi.org/10.1136/ip.8.suppl_4.iv17)
- Wölfer, R., Schultze-Krumbholz, A., Zagorscak, P., Jäkel, A., Göbel, K., & Scheithauer, H. (2013). Prevention 2.0: Targeting cyberbullying @ school. *Prevention Science*, 15, 879-887. <https://doi.org/10.1007/s11121-013-0438-y>
- Wright, M. F. (2015). Adolescents cyber aggression perpetration and cyber victimization: the longitudinal associations with school functioning. *Social Psychology of Education*, 18(4), 653-666. <https://doi.org/10.1007/s11218-015-9318-6>
- Wright, M. F. (2017). The role of technologies, behaviors, gender, and gender stereotype traits in adolescents' cyber aggression. *Journal of Interpersonal Violence*, 35 (7-8), 1719-1738. <https://doi.org/10.1177/0886260517696858>
- Wright, M. F., Aoyama, I., Kamble, S. V., Li, Z., Soudi, S., Lei, L., & Shu, C. (2015). Peer attachment and cyber aggression involvement among Chinese, Indian, and Japanese adolescents. *Societies*, 5, 339-353. <https://doi.org/10.3390/soc5020339>