



Reducing toxic behaviors in the gaming community: A social-psychological intervention with lasting effects

Ágnes Zsila^{a,*,1}, Gábor Orosz^{b,1}, Zsolt Demetrovics^{a,c,d}

^a Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary

^b Univ. Lille, Univ. Artois, Univ. Littoral Côte D'Opale, ULR 7369 - URePSSS - Unité de Recherche Pluridisciplinaire Sport Santé Société, Lille, France

^c Flinders Institute for Mental Health and Wellbeing, College of Education, Psychology and Social Work, Flinders University, Bedford Park, SA, Australia

^d Centre of Excellence in Responsible Gaming, University of Gibraltar, Gibraltar, Gibraltar

ARTICLE INFO

Handling Editor: Matthieu Guitton

Keywords:

Randomized controlled trial

Toxic behaviors

Video gaming

Wise intervention

ABSTRACT

Toxic behaviors in multiplayer online gaming environments have become a growing concern, as aggression between players can diminish game enjoyment and reduce psychological well-being. This preregistered randomized controlled trial tested the effectiveness of a brief social-psychological intervention designed to reduce both involvement in and the psychological consequences of toxicity by exposing gamers to adaptive strategies against toxicity in the form of peer testimonials. A total of 722 gamers (93.6 % men; $M_{\text{age}} = 31.40$ years, $SD = 8.13$) were randomly assigned to either the treatment group ($n = 372$) or the control group ($n = 350$). The treatment group was exposed to materials focused on toxic behaviors, while the control group read about peer experiences related to disruptive external circumstances in the gaming environment. Maladaptive and adaptive reactions to toxicity, rumination, game enjoyment, and toxic disinhibition were assessed before the intervention and one month later. Compared to the control group, players in the treatment group reported lower levels of rumination and more adaptive responses to toxicity in anticipated situations of toxicity immediately after the intervention. Moreover, participants in the treatment group reported fewer maladaptive reactions to toxicity immediately after the intervention and also one month later compared to the control group. These findings indicate that social-psychological interventions targeting social norms can effectively foster adaptive and reduce maladaptive responses to toxic behavior, which were previously identified as major contributing factors to the social contagion of toxicity in gaming environments.

1. Introduction

1.1. The proliferation of toxicity in the gaming community

Toxic behaviors are forms of interpersonal aggression among players within multiplayer video gaming context that can negatively impact individual or team performance, diminish game enjoyment, and reduce mental well-being (Zsila & Demetrovics, 2025). Kordyaka et al. (2020) proposed that toxic behaviors are frustration-driven, maladaptive reactions to emotionally challenging in-game situations. Some prevalent forms of toxic behaviors include annoying other players by sabotaging their gameplay (i.e., griefing; Kordyaka et al., 2019; Kowert, 2020), using offensive and insulting language (i.e., flaming, harassment, and trash talking; Kou, 2020, pp. 81–92; Kowert, 2020), assisting the enemy

team at the disadvantage of own team members (Blackburn & Kwak, 2014, pp. 877–888; Kwak et al., 2015, pp. 3739–3748), repeating messages or actions to intentionally annoy other players (i.e., spamming; Beres et al., 2021, pp. 1–15; Kowert, 2020), and cheating (Kou, 2020, pp. 81–92; Kowert, 2020). These toxic behaviors can vary in perceived severity: while trash talking or spamming are often considered milder forms of negative acts, verbal abuse and harassment are more serious insults (Laato et al., 2024, pp. 1–18). A recent study involving more than 2000 adolescent and adult video game players has found that 66 % of the players reported being personally affected in toxic behaviors as a target (Zsila et al., 2022). Targets reported higher stress levels and bad mood during the gaming session (Kordyaka et al., 2019), more symptoms of depression and anxiety (Wong & Ratan, 2024), rumination (Fox & Tang, 2017), decreased levels of self-efficacy (Kordyaka et al.,

* Corresponding author. 1064 Budapest, Izabella u. 46. (Hungary).

E-mail addresses: zsilagnes@gmail.com (Á. Zsila), gaborosz@gmail.com (G. Orosz), zsolt.demetrovics@gmail.com (Z. Demetrovics).

¹ Equal contributions. ÁZ and GO are co-first authors.

2023), and social withdrawal (Frommel et al., 2023). Personal experience in toxic encounters has also been associated with a greater risk of gaming disorder (Gandolfi et al., 2023; Zsila et al., 2022). Therefore, it is crucial to provide tools that can increase players' efficacy to cope with toxic behaviors in order to support mental well-being (Laato et al., 2024, pp. 1–18). The present study endeavors to offer such protective strategies to reduce the psychological harms associated with toxicity.

The rapid proliferation of toxic interactions among players in online multiplayer gaming environments has contributed to the normalization of such behaviors in the gaming community. This trend underscores the need for prevention and intervention efforts to enhance the safety of online gaming spheres and support players' well-being (Wijkstra et al., 2024). Confirming the need for addressing toxicity, a recent study has indicated that players generally perceived toxicity as a serious problem in the gaming community (Zsila et al., 2022). Prior studies have suggested that the normalization of toxicity in online gaming spheres can be explained by the online disinhibition effect (Beres et al., 2021, pp. 1–15; Liu & Agur, 2023), which describes that individuals are more likely to engage in hostile or inappropriate behaviors in online settings compared to offline circumstances, due to factors such as anonymity, invisibility, and perceived lack of consequences (Suler, 2004). More recently, the Unified Theory of Toxic Behavior (UTTB) has identified toxic disinhibition as the strongest predictor of exhibiting toxic behaviors (). Disinhibition is further exacerbated by the widespread use of anonymous usernames, which hide players' personal identity and diminish the sense of social responsibility (Laato et al., 2024, pp. 1–18). Based on the UTTB, further predictors of toxic disinhibition include positive attitudes towards disruptive behaviors, frequent experiences in toxicity as a target, and lower perceived control over such controversial situations (Kordyaka, Laato, Jahn, et al., 2023). Previous targets of toxicity often engage in such behaviors later, and the higher propensity for engagement has been associated with perceptions and normative beliefs on how they should respond to stressful in-game situations (Kordyaka, Laato, Jahn, et al., 2023). Therefore, fostering adaptive, immediate behavioral and cognitive responses (e.g., reporting) and reducing maladaptive reactions in players (e.g., fighting back) can possibly contribute to more efficient coping strategies in toxic encounters, which can also decrease the risk of engagement in reactive aggression, a common response to frustrating in-game situations (Kordyaka et al., 2020).

According to Laato et al. (2024, pp. 1–18), game developers apply a range of prevention and intervention strategies to combat toxicity, including proactive controls (e.g., limiting in-game communication channels) and toxicity detection strategies (e.g., utilizing machine learning models to detect toxic language). A recent systematic review on interventions targeting toxic behaviors has found that most interventions (77.8 %) suggested novel approaches (e.g., predicting the occurrence of toxic behaviors based on gaming patterns), while only a minority of studies (22.2 %) tested the efficacy of existing intervention systems (Wijkstra et al., 2024). Several studies used artificial intelligence-based detection systems to identify toxic behavioral patterns (e.g., Canossa et al., 2021; Lee et al., 2018), while only a few studies addressed the mental health of targets (e.g., Fox & Tang, 2017; Reid et al., 2022). Although these studies provided some evidence on the effectiveness of the interventions in terms of improved mood and decreased toxic interactions (Wijkstra et al., 2024), the efforts were either game-specific (e.g., Overwatch), or focused on individual-level outcomes (e.g., mood improvement) rather than addressing a broader context of toxicity (e.g., gaming climate).

Prevention and intervention programs that address peer aggression in online settings mostly apply social skill development, psycho-education, training, climate improvement, and policies (Polanin et al., 2022). Although some of these components could be useful when targeting toxicity, such as social skill development, considering that empathy moderated the association between online disinhibition and cyberbullying perpetration (Wang et al., 2022), these intervention efforts largely assume that group members know each other, the group is

well-established, and norms are settled. However, team members in online multiplayer games often vary greatly, even across matches (Eklund, 2015; Kou & Gui, 2014, pp. 161–169), which can contribute to the online disinhibition effect (Beres et al., 2021, pp. 1–15; Liu & Agur, 2023). Hilvert-Bruce and Neill (2020) proposed that changing the normative belief system within the context of online multiplayer games that supports the acceptability of toxic behaviors can be a key component of prevention and intervention efforts against toxicity. In line with this suggestion, Nexø and Kristiansen (2023) highlighted the importance of targeting two social components in efforts to reduce toxicity: 1) weakening norms that permit toxicity in the gaming community, and 2) encouraging influential players (i.e., social referents) to publicly oppose toxic behaviors, thereby shaping group norms in a broader audience (e.g., followers). Therefore, social interventions that focus on reshaping social norms can possibly be an effective, broader-scope approach.

1.2. Using a brief social-psychological intervention to reduce involvement in and psychological harms associated with toxic behaviors

Responses to other players' toxic behavior can vary widely. Some players may choose to report such behavior, which is an adaptive reaction, while others might engage in reactive verbal aggression, a destructive response (Türkay et al., 2020, pp. 1–13). It was hypothesized that a brief social-psychological intervention, designed to encourage constructive reactions, could catalyze long-term changes in how players respond to toxic behavior, ultimately contributing to a less toxic gaming environment. The present intervention is grounded in the theoretical framework of wise social-psychological interventions (Walton, 2014; Walton & Wilson, 2018). These short-term interventions are specifically designed to shift ways of thinking or feeling in order to produce lasting improvements in social problems (Walton, 2014; Walton & Crum, 2021; Walton et al., 2017). Such interventions employ a wide variety of strategies targeting underlying psychological processes (Walton, 2014) such as the "saying-is-believing" strategy, a form of self-persuasion (Aronson, 1999). Wise social-psychological interventions have been widely applied in educational settings in the US (e.g., to cope with academic challenges; Walton et al., 2023; Yeager et al., 2019). More recently, Orosz et al. (2023, 2024) conducted wise interventions to help recognize fake news and misinformation in the general population, extending the scope of predominantly US wise intervention research tradition to investigations in Europe. Similar to previous research, the aim of applying a wise social-psychological intervention in this study serves prosocial purposes (Orosz et al., 2023). Specifically, the present intervention aims to reduce maladaptive reactions in response to toxicity (i.e., retaliatory behaviors and disengagement) and alleviate associated harms. For the purpose of fostering adaptive strategies in players when encountering toxicity, several strategies were incorporated in the present intervention that can effectively contribute to the adoption of these adaptive reactions in highly stressful in-game situations, including expert positioning of the participants, exposure to descriptive norms and peer testimonials, and the saying-is-believing task in which participants are instructed to write advice to a beginner player based on the peer experiences they are exposed to in the form of testimonials.

To foster participants' willingness to provide detailed advice to beginner players – which was set as the primary objective of the intervention for the participants –, players were initially positioned as experts. This expert positioning, a widely used component of brief social-psychological interventions, help prevent individuals from becoming passive due to the perception of being a "subject" in an intervention in need of education (see Orosz et al., 2023, 2024; Yeager et al., 2016). Therefore, participants were involved as active contributors in the process of creating safer gaming environments.

Subsequently, participants were exposed to testimonials from players who had adopted adaptive responses to toxic behaviors such as muting disruptive players, leaving the match and joining more constructive teams, reporting toxic players, and controlling the urge to

Treatment

3. Report the disruptive player

"I play to get rid of everyday stress, not to encounter more tension while gaming. If I see someone deliberately annoying others or being openly hostile towards them, I think such a player should simply be reported. As a result, several malicious players have already been banned. As a senior high school student, there's a lot of pressure, and we have limited time to relax, so I think it's very important that we truly enjoy the game when we play. If we can take actions against players who clearly just come to annoy others, we take a step towards creating an online gaming environment which is even safer and a more positive experience."

K4T3mage, 18-year-old League of Legends player

Control

1. Ergonomic gaming chair for maximum comfort

"If you hunch over your computer for long hours, every part of your body goes numb at the end, and your grandma's back pain is nothing to yours. A comfortable, ergonomically designed gaming chair can help you focus on your game instead of your stiff limbs or agonizing back pain. There are many types of gaming chairs available, and they cost only a few thousand HUF more than a regular kitchen chair. I personally feel much better after a long gaming session since I bought my own gaming chair. I can focus better on the game, and my posture has improved as well. For me, maintaining my health is essential for the ultimate gaming experience."

RogueMaster09, 35-year-old MMORPG player

Note. Translated from Hungarian to English using ChatGPT (OpenAI, 2023)

Fig. 1. Sample testimonials from the treatment and control conditions.

retaliate (see Fig. 1). The descriptive norms and testimonials guided participants towards a "saying-is-believing" self-persuasion task (Aronson, 1999), in which participants could personalize the intervention's messages and strategies. Originally, Aronson (1999) used this element in a family context in which an older sibling offered advice to a younger family member, and the perceived emotional proximity between them contributed to a greater persuasion effect. Similarly, participants were instructed to offer advice for a beginner player in the present study. By selecting messages they found meaningful from the testimonials (see Yeager et al., 2016 in the field of education), participants were encouraged to contribute their own suggestions for mitigating toxic behaviors (see Fig. 2 for the instruction and detailed psychological mechanisms).

This indirect and subtle persuasive approach, supported by descriptive norms and testimonials, was expected to promote behavioral changes. Specifically, an increase was expected in adaptive behavioral responses (e.g., reporting, muting) and a reduction in maladaptive, predominantly retaliatory responses immediately after the intervention and one month later. Also, positive changes were also anticipated in players' subjective enjoyment of the game, alongside reduced rumination over the feelings and thoughts associated with toxic encounters (Garnefski & Kraaij, 2006), and decreased involvement in toxic interactions.

1.3. Aim of the study

Overall, toxic behaviors in online multiplayer gaming environments have become a frequent and adverse experience among gamers. Intervention efforts focused mainly on the detection and prevention of these behaviors based on observed gaming patterns, while psychological harm

reduction was scarcely addressed (Wijkstra et al., 2024). The present intervention addresses those social norms that sustain toxicity, and offers strategies that can support players' efficacy to cope with highly stressful in-game situations that can increase the likelihood of toxic encounters. Considering the normalization of toxicity in recent years, it was suggested that modifying the social norms that support the general acceptance of toxicity could possibly be a core component of an effective intervention program against toxic behaviors (Hilvert-Bruce & Neill, 2020; Nexø & Kristiansen, 2023).

Indeed, a wise social-psychological intervention addressing norms can be effective as such interventions can change players' perceptions of in-game behaviors (e.g., what is acceptable and what is not tolerated) by actively involving players in the process of shifting common norms (e.g., toxic behaviors are integral part of the online gaming culture), and encourage adaptive, immediate responses to stressful in-game situations without long-term, repeated exposure to these strategies. Therefore, the present study aimed to test the efficacy of a social-psychological intervention developed to reduce the involvement and psychological consequences of toxic behaviors in gamers.

It was expected that the psychological mechanisms integrated into the intervention could lead to immediate and short-term (one-month) changes in players' reactions, which may be more flexible, temporal, and situation-dependent (Kordyaka et al., 2019), thereby sensitive to even a single exposure of adaptive strategies. It can also possibly affect subjective perceptions of game enjoyment, while more pervasive changes in the gaming climate may require more exposure and reinforcement of adopted norms. Therefore, the aim of this intervention was to address the major problems and mechanisms associated with the normalization of toxicity: players' reduced psychological well-being and game enjoyment, maladaptive reactions to stressful in-game situations

Help explain what we can do against toxic behaviors

We are collecting short messages from experienced players like you to provide the best examples for beginner players.¹

We are looking for strategies to quickly and effectively eliminate toxic behaviors so that players can soon get over the unpleasant experiences and enjoy the game.

In an upcoming program, we would like to share some messages (anonymously) with beginner players. **Therefore, we would like to ask you to write some advice following these instructions:**²

Your message:

- If you know a beginner player who could benefit from your advice, you can write your message to them. If you prefer so, you can name them and start your message like this: "Machine2020, I'd like to give you some tips to help you better handle toxic behaviors you encounter while playing."³
- First, summarize the ideas and thoughts on what you just read in a few sentences, in your own words. (You can find their description below the text box.)⁴
- **Then, think a bit about...**
 - what reasons you can use to tell them most effectively **what to do if they experience toxic behaviors from others.**⁵
 - what you suggest they can do to **effectively handle the situation**, even when it's difficult to keep their cool.⁶
- Finally, formulate your letter in a way that presents the most compelling arguments.

There are no right or wrong answers; don't worry about spelling. Just focus on conveying your thoughts.⁷

You can write your message here:

[text box]

Fig. 2. Instruction and psychological mechanisms applied in the treatment condition after reading the testimonials (Orosz et al., 2024).

as a consequence of frustration, and thereby, repeated involvement in toxicity.

1.4. Preregistered hypotheses

Feinstein et al. (2014) found that victims of cyberbullying with higher ruminative tendencies were more likely to experience symptoms of depression. In a video gaming context, Fox and Tang (2017) found that sexual harassment predicted higher rumination, which in turn, was associated with maladaptive reactions, such as women's withdrawal from the game. As the present intervention aimed to reduce maladaptive responses to toxicity (e.g., fighting back), it was anticipated that rumination on such incidents could also diminish, as the cycle of toxicity could be interrupted at an earlier stage (Kordyaka, Laato, Jahn, et al., 2023).

H1. The treatment (in contrast to the control condition) will lead to lower levels of rumination on toxic behaviors prospectively (immediately after the intervention) and in the longer run (one-month follow-up).

Kordyaka et al. (2020) proposed that toxic behaviors should be viewed as frustration-driven maladaptive responses to stressful in-game situations. Liu and Agur (2023) also emphasized that retaliation is a major motivation for toxicity. With regard to peer aggression, Pace et al. (2018, 2021) suggested that negative emotions (e.g., worry and hatred) and cognitive distortions can contribute to the continued use of maladaptive coping strategies, which in turn, are associated with higher

Note. Translated from Hungarian to English using ChatGPT (OpenAI, 2023)

¹ Gamers are positioned as **experts** who can provide the most compelling advice. This strategy was supposed to increase participants' engagement in the intervention as active contributors (Orosz et al., 2023, 2024; Yeager et al., 2016).

² Participants can give advice from an expert position, which can increase participants' perception of a meaningful contribution to the intervention program (Orosz et al., 2023, 2024; Yeager et al., 2016).

³ Participants wrote **personalized messages** to anchor the strategies they read about in the testimonials, and tailor them to their everyday experiences as a strategy of **self-persuasion** by aligning the advice with their behavior (Aronson et al., 1991; Orosz et al., 2023; Stone et al., 1994).

⁴ Prior to the personalization, participants were requested to **summarize the strategies** to ensure that they **understood** the main messages of the intervention. This element was derived from the intervention by Orosz et al. (2023) to further support self-persuasion.

⁵ **Saying-is-believing** self-persuasive strategy was applied based on Aronson (1999).

⁶ The message was framed in the **context** of situations where controlling impulses is particularly challenging, based on the description of toxic behaviors by Kordyaka et al. (2020). Emphasis on these challenging situations could possibly underline the importance of using adaptive strategies against toxicity.

⁷ Elimination of **perfectionist concerns** was also targeted, enabling participants to focus on the content of their message rather than on formal or grammatical details. This element was derived from Orosz et al. (2023).

psychological distress (Chahal et al., 2016). Reid et al. (2022) also suggested that effective coping strategies should be encouraged to cope with stressful in-game situations to combat toxicity in video games. Therefore, as testimonials in the present intervention targeted the enhancement of adaptive responses and the reduction of maladaptive reactions to toxicity, the following hypotheses were proposed.

H2. The treatment (in contrast to the control condition) will lead to fewer maladaptive reactions (e.g., fight back) to toxic behaviors prospectively (immediately after the intervention) and in the longer run (one-month follow-up).

H3. The treatment (in contrast to the control condition) will lead to more adaptive strategies against toxic behaviors (e.g., report, mute) prospectively (immediately after the intervention) and in the longer run (one-month follow-up).

Research found that frequent exposure to toxicity within gaming communities can diminish players' enjoyment (Beres et al., 2021, pp. 1–15), while ineffective or insufficient efforts to address toxicity can undermine the positive atmosphere of the gaming culture (Laato et al., 2024, pp. 1–18). Although some studies pointed out that toxic behaviors (e.g., grieving) can also be rewarding to some perpetrators who derive satisfaction from provoking frustration in other players (Achterbosch et al., 2017; Adinolf & Turkay, 2018, pp. 365–372; Cook et al., 2018), descriptive norms in the present intervention suggested that only a small proportion of players engage in consistently disruptive behaviors for enjoyment. Therefore, it was anticipated that the intervention can enhance overall game enjoyment by promoting more constructive

responses to toxicity, thereby also decreasing the occurrence of retaliation, which can diminish other players' game enjoyment.

H4. The treatment (in contrast to the control condition) will lead to higher subjective enjoyment of multiplayer gaming in longer run (one-month follow-up).

On one hand, an increase in adaptive reactions to toxicity was expected to decrease the occurrence of involvement in toxic interactions. For instance, muting the toxic player can protect individuals from further verbal harassment. On the other hand, a decrease in maladaptive responses was also anticipated, which could mitigate the risk of engagement in toxicity. For instance, fighting back can contribute to the cycle of toxicity in which the toxic player can become a target, and vice versa (Kordyaka, Laato, Jahn, et al., 2023). Therefore, it was proposed that encouraging the use of adaptive rather than maladaptive strategies could contribute to the disruption of the cycle of toxicity (e.g., retaliation), which could decrease toxicity overall.

H5. The treatment (in contrast to the control condition) will lead to less frequent involvement in toxic behaviors in the longer run (one-month follow-up).

According to the UTTB (Kordyaka et al., 2020, 2023), toxic disinhibition was predicted by positive attitudes towards negative online behaviors, frequent involvement in toxic interactions as a target, and lower perceived control. Kordyaka, Laato, Jahn, et al. (2023) emphasized that normative beliefs on how targets of toxicity should respond in emotionally charged in-game situations can influence targets' later engagement in toxicity. The present intervention offered adaptive strategies to respond to such incidents, provided descriptive norms to reinforce social responsibility (Laato et al., 2024, pp. 1–18), and the testimonials illustrated the consequences of toxicity, reducing the perception of invisibility. Therefore, it was expected that the intervention would decrease toxic disinhibition in players.

H6. The treatment (in contrast to the control condition) will lead to lower levels of toxic disinhibition in the longer run (one-month follow-up).

2. Methods

2.1. Participants

The target population of this study included Hungarian gamers above 14 years of age who regularly played video games (at least 1 h a week) and, specifically, multiplayer games. To reach the target population, the study and the link to the online research material were promoted through the website and social media platforms of the leading Hungarian gamer lifestyle magazine, GameStar. The research material was created using Qualtrics software, which enabled the random assignment of participants to either the treatment or the control group. Participants were assigned to either the treatment or control condition with equal probability (50–50 %). Randomization was implemented at the individual level using the built-in randomization tool in Qualtrics. No blocking or stratification procedures were applied, and no randomization seed was set. Shopping vouchers totaling 200,000 HUF (about 578 USD) were randomly awarded to increase participation in the research. Participation was completely voluntary and anonymous, with all participants providing informed consent. Additionally, minors (gamers aged 14 to 17) provided passive parental consent by clicking a box, which indicated that their parents or legal guardians had read the terms and agreed that the underage respondent could participate in the research. This protocol was in line with local ethical standards, which require that both the parent or legal guardian and the underage participant to be informed about the research and provide written consent. Passive parental consent was acceptable for data collections that did not record the participant's name. The study received ethical

approval from the Institutional Review Board of the principal investigator's university (protocol number: 2022_29) and adhered to the principles of the Declaration of Helsinki.

In accordance with the preregistered study plans (see https://osf.io/vwntg/?view_only=bcc1a5f723a54e0d998736e24d134f4c), a total of 506 participants ($n = 253$ for the treatment condition and $n = 253$ for the control condition) was required to adequately test the main hypotheses (H1–H4) based on the a priori power analysis with the following parameters: $d = 0.25$, alpha error probability = 0.05, power = 0.8, and number of predictors = 1, in line with the conventions in brief social-psychological intervention research (Orosz et al., 2023, 2024; Yeager et al., 2016). Due to the lack of previous intervention studies aiming to reduce toxic behaviors and their effect on mental health, power estimation was set to be able to detect even small long-term effects ($d = 0.25$). The power analysis was conducted using R software (pwr package), assuming a two-tailed independent-samples *t*-test with equal sample sizes in the treatment and control groups. This test type was selected because it provided the most straightforward and widely accepted approximation to the preregistered ordinary least-squares (OLS) regression design with one predictor (condition). The power analysis did not incorporate corrections, and applying post-hoc adjustments (e.g., Bonferroni) could have substantially reduced statistical power while increasing the risk of Type II error, particularly with regard to the limited sample sizes in the follow-up. Therefore, the *p*-value of 0.05 was consistently retained as a level of significance.

A total of 722 gamers (93.63 % male; $M_{\text{age}} = 31.40$ years, $SD = 8.13$) were exposed to the randomized intervention or control materials of which 372 ($n = 355$ male, 95.43 %; $M_{\text{age}} = 31.56$ years, $SD = 8.25$) were assigned to the treatment and 350 ($n = 321$ male, 91.71; $M_{\text{age}} = 31.21$ years, $SD = 8.00$) to the control condition. A total of 520 participants (70.08 % of the pre- and post-test participants; 93.27 % male; $M_{\text{age}} = 32.09$ years, $SD = 7.77$) completed the follow-up questionnaire one month later (72.31 %, $n = 269$ of participants from the treatment condition [$n = 256$ male, 95.17 %; $M_{\text{age}} = 32.19$ years, $SD = 7.99$], and 71.71 %, $n = 251$ of participants from the control condition [$n = 229$ male, 91.24 %; $M_{\text{age}} = 32.00$ years, $SD = 7.54$]). Attrition rate was 27.69 % for the treatment condition and 28.29 % for the control condition. The total attrition rate was 27.99 %.

2.2. Procedure

A between-subjects experimental design within a randomized controlled trial was conducted with two conditions (treatment and control). Participants were informed that the research team had been developing a program for beginner gamers, aiming to foster positive experiences in online multiplayer gaming settings. Subsequently, participants were randomly assigned to either the intervention or the control condition.

Participants were initially positioned in an expert role as they had significant experience in video gaming. In the treatment condition, participants were informed that the goal of the program was to offer behavioral and affective strategies for beginner players that can reduce the negative effect of toxic behaviors on the gaming experience. The definition of toxic behaviors was introduced to participants based on the definition proposed by Kou (2020, pp. 81–92). Subsequently, participants were exposed to adaptive strategies against toxicity, which reflected the content of the later testimonials. Then, they read some recent statistics regarding the occurrence and psychological effects of toxic behaviors based on Zsila et al. (2022). In the following step, participants were exposed to four adaptive strategies against toxic behaviors in the framework of peer testimonials (i.e., muting, leaving the match and joining a constructive team, reporting, and controlling the urge to fight back; see an example in Fig. 1). After reading the testimonials, participants were instructed to write a few recommendations for a beginner player. Specifically, they were asked to summarize the four strategies, and consider reasons that can effectively convince the player to use

adaptive strategies in situations of toxic interactions. Participants were also encouraged to write suggestions on how this player could effectively cope with toxic behaviors even in frustrating situations (see Fig. 2 for the complete instruction with psychological mechanisms).

Materials of the control condition followed a similar structure to the intervention materials. However, the program's objective was reframed to focus on offering behavioral and affective strategies for beginner players that can mitigate the impact of disruptive environmental circumstances on the gaming experience. These circumstances were described as factors outside the gaming sphere that can negatively affect players' comfort and attention, thereby deteriorating the gaming experience (e.g., loud noise, uncomfortable chair). Subsequently, examples were provided to illustrate how these factors can deteriorate the gaming experience of many gamers (e.g., loud noise distracts the player). In the following step, participants read four testimonials demonstrating how fellow gamers managed to mitigate these environmental annoyances (i.e., purchasing an ergonomic gaming chair, maintaining physical well-being, creating a peaceful environment, upgrading gaming PC). After reading the testimonials, participants were asked to write recommendations for beginner players. Similar to the intervention condition, participants were instructed to summarize the four strategies, consider the most compelling reasons for employing these strategies to reduce the effect of annoying environmental circumstances, and make suggestions on how the player can effectively eliminate these effects even in challenging situations. The control condition did not include toxic behaviors. The structure of the research material is depicted in Fig. 3. The original Hungarian and the translated English version of the research materials for the treatment and control conditions, and examples for respondents' messages written in the treatment and control conditions can be found in Hungarian and in English in an online repository (https://osf.io/vwntg/?view_only=bcc1a5f723a54e0d998736e24d134f4c). Differences between the preregistered research plan and the actual study are summarized in SM Table 1.

2.3. Intervention rationale

Initially, participants were positioned as experts in both conditions, thereby fostering participants' engagement in the research, reinforcing their active role in shaping the community rather than being "subjects" in need of instruction and education (e.g., Orosz et al., 2023; Orosz, Faragó, Paskuj, & Krekó, 2024; Orosz, Faragó, Paskuj, Rakovics, et al., 2024; Yeager et al., 2016). This strategy was expected to be efficient in the present gaming context where a large proportion of players had extensive experience in online multiplayer gaming (Kordyaka, Laato, Jahn, et al., 2023). Therefore, they were asked to provide feedback on improving the present program.

Before participants could offer advice in the treatment condition, they were exposed to the definition of toxic gaming behavior based on Kou (2020, pp. 81–92), accompanied by statistics on toxic behaviors from a previous study (Zsila et al., 2022). These statistics were used to highlight three descriptive norms (see Goldstein & Cialdini, 2011): (1) the prevalence of players who have experienced or have been exposed to toxic behaviors (82.3 %), (2) the percentage of players who found such behaviors irritating (78 %), and (3) the relatively small proportion of players who engaged in toxic behavior without becoming a target (1.8 %) – which could not potentially be attributed to reactive aggression. These norms were intended to illustrate that although toxic behaviors affect a large proportion of players, the number of those exhibiting toxic behaviors is small, underlining the potential to combat the issue.

Adaptive strategies against toxicity in the treatment condition were selected based on previous studies (e.g., Adinolf & Türkay, 2018, pp. 365–372; Kou, 2020, pp. 81–92; Türkay et al., 2020, pp. 1–13) and semi-structured interviews with three expert psychologists with prior experience in video game research. These experts reviewed the formulated testimonials. Testimonials provided normative information about other gamers' experiences with toxicity and strategies they employed to

successfully navigate these challenging situations. Peer testimonials are commonly applied in social-psychological interventions targeting social norms (see Chen et al., 2024; Orosz et al., 2023, 2024). These peer testimonials can be perceived as credible sources of personal experiences that help people recognize certain behaviors (e.g., toxicity) and provide behavioral guidelines about the appropriate and efficient reactions to these behaviors. Testimonials are also useful to decrease reactance in participants as these texts are framed as personal opinions (Orosz et al., 2023).

In the final step, participants were instructed to write advice to novice players on how they could effectively cope with toxic behaviors (treatment condition) or disruptive circumstances (control condition), in the form of a personalized message. Personalized messages could further enhance participants' engagement by recalling personal experiences, which could indirectly serve as a self-persuasion through the alignment of the advice with their behaviors (Aronson et al., 1991; Orosz et al., 2023; Stone et al., 1994). Drawing on the suggestion by Kordyaka et al. (2020) that toxic behaviors occur in highly frustrating in-game situations, it was emphasized in the treatment condition that the advice should be applied to situations where controlling emotional and behavioral impulses may be challenging. This element could possibly underline the importance of adopting adaptive strategies to cope with these situations.

2.4. Measures

Major demographics (i.e., gender and age) and information regarding video gaming were collected. Specifically, the period of playing video games regularly, time spent video gaming on an average weekday and weekend day, multiplayer experiences, platforms, and preferred genres were reported by participants (see response options in SM Table 2).

Involvement in toxic behaviors was assessed by the question of "Have you experienced behaviors from other players that deteriorated your gaming experience in the past month?" in the pre-test and the follow-up. A few examples were provided for such behaviors, including flaming, hate speech, spamming, cheating, feeding, griefing, and doxxing based on previous studies (i.e., Kou, 2020, pp. 81–92; Kowert, 2020; Monge & O'Brien, 2021). Response options were 1 = yes and 2 = no. Descriptive statistics on involvement in the sample are presented in SM Table 3. All other study-relevant constructs that were assessed and used in the hypothesis testing are presented in Table 1. The instructions and items were administered in Hungarian. Non-validated items and research materials were translated to English using chatGPT (OpenAI, 2023). The accuracy of the translation was carefully checked by the research team. Validated scales (i.e., Rumination About an Interpersonal Offense Scale and Online Disinhibition Scale) were translated to Hungarian by a research team member and back-translated to English by another independent member of the research team. Both had considerable experience in the cross-cultural adaptation of self-report scales.

2.5. Statistical analysis

Data analysis was performed using R software. OLS regressions were used to test the preregistered hypotheses, including immediate post-intervention and long-term effects (one month later). No covariates were used. Regressions were used because the preregistered hypotheses focused on condition effects at each time point (i.e., pre-test, post-test, and follow-up), rather than on condition $\times$ time interactions. Regressions also enabled the standardization of outcomes and the use of non-parametric test statistics when model assumptions were not met. *T* regression test statistics were reported to explore whether the regression coefficient is significantly different from zero, indicating a significant effect of the predictor on the dependent variable. However, non-parametric Welch tests were reported when the assumptions of OLS regression were not met (see the results of the assumption testing in SM 4).

3. Results

There were no significant differences between the control and treatment conditions in any sociodemographic and psychological variables that were assessed before the intervention (i.e., age, gender, level of education, subjective game enjoyment, rumination, and toxic disinhibition, all $ps > 0.770$).

A detailed overview of the hypothesis testing and estimates are presented in Table 2. With regard to rumination post-intervention and one month later, H1 anticipated that the intervention would result in a decrease in expected rumination on involvement in toxic behaviors immediately after the intervention and one month later. The findings supported the immediate effect of the intervention regarding rumination on toxic behaviors, $t(518) = -2.00, p = 0.046, g_{Hedges} = 0.17$. Moreover, the long-term effect did not reach the level of significance based on the subsequent Welch test ($p = 0.680$).

Regarding maladaptive reactions post-intervention and one month later, H2 predicted that the intervention would lead to fewer maladaptive reactions to toxic behaviors immediately after the intervention and in the longer run (one-month follow-up). The results of the Welch test confirmed both the immediate, $t_{Welch}(520) = 2.76, p = 0.006, g_{Hedges} = 0.24$, and the long-term effect, $t_{Welch}(217) = 2.24, p = 0.030, g_{Hedges} = 0.30$, indicating a significant decrease in maladaptive reactions to toxic behaviors.

With regard to adaptive reactions post-intervention and one month later, H3 posited that the intervention, in contrast to the control condition, would result in more adaptive strategies against toxic behaviors immediately and in the follow-up. According to the Welch test, the immediate results revealed a promising trend, with intervention participants showing a greater willingness to use adaptive strategies against toxic behaviors in contrast to the control, $t_{Welch}(520) = -2.96, p = 0.003, g_{Hedges} = -0.26$. Although these effects were not significant in the follow-up ($p = 0.750$), these results suggest the intervention's potential to promote adaptive strategies over time.

Regarding subjective game enjoyment one month later, H4 postulated that the treatment (contrast to the control condition) would lead to higher subjective enjoyment of multiplayer gaming in the longer run (one-month follow-up). However, the Welch test results showed no significant difference between the treatment and the control groups' level of enjoyment ($p = 0.731$).

Regarding involvement in toxic behaviors one month later, H5 suggested that the treatment (contrast to the control) condition would lead to less frequent involvement in toxic behaviors in the longer run (one-month follow-up); however, again, the results did not confirm this hypothesis, as there was no significant difference between the treatment and the control group based on the Welch test ($p = 0.240$).

In relation to toxic disinhibition one month later, H6 postulated that the treatment (contrast to the control) condition would lead to lower levels of toxic disinhibition in the longer run (one-month follow-up). However, no significant difference was found between the treatment and the control groups based on the Welch test ($p = 0.690$). Significant effects of the intervention are presented in Fig. 4.

4. Discussion

Reactive aggression, as a response to frustration, is prevalent in highly competitive online environments, such as multiplayer video gaming (Achterbosch et al., 2017). Existing tools to combat toxic behaviors mostly included the detection and sanctioning of toxicity, while only a few intervention efforts were made to elevate mood after such incidents (see Wijkstra et al., 2024 for a review). The present intervention aimed to reduce the harms associated with toxic behaviors, including lowered psychological well-being and game enjoyment, maladaptive reactions to toxicity, and repeated involvement in toxic encounters. It was found that the intervention reduced maladaptive reactions (e.g., fighting back) to toxic behaviors immediately after the intervention and one month later, and resulted in positive changes with regard to the use of adaptive strategies and rumination after the intervention. Lasting effects in relation to subjective enjoyment, toxic disinhibition, and involvement in toxic encounters were not found.

First, it was postulated that the intervention would lead to decreased ruminative tendencies in players after the intervention and one month later (H1). This hypothesis was partially supported as rumination levels reduced in players only after the intervention. However, this result was close to the level of significance, warranting further confirmation in future research using larger samples. Fox and Tang (2017) found that rumination was unrelated to experiences of general harassment in players; however, general and sexual harassment both predicted withdrawal from gaming. Additionally, another study involving women players indicated that most players did not seek help or social support after being targets of harassment, and could enjoy the gaming session despite negative experiences (Cote, 2017). These findings suggest that rumination may not be a general experience after toxic encounters in the longer run, and players rather become desensitized to toxicity and disengage from the experience or withdraw from participation by leaving the game.

Second, it was hypothesized that the intervention would lead to less maladaptive (H2) and more adaptive (H3) reactions to toxicity. Decreased maladaptive reactions were supported by the results immediately after the intervention and one month later, while the positive

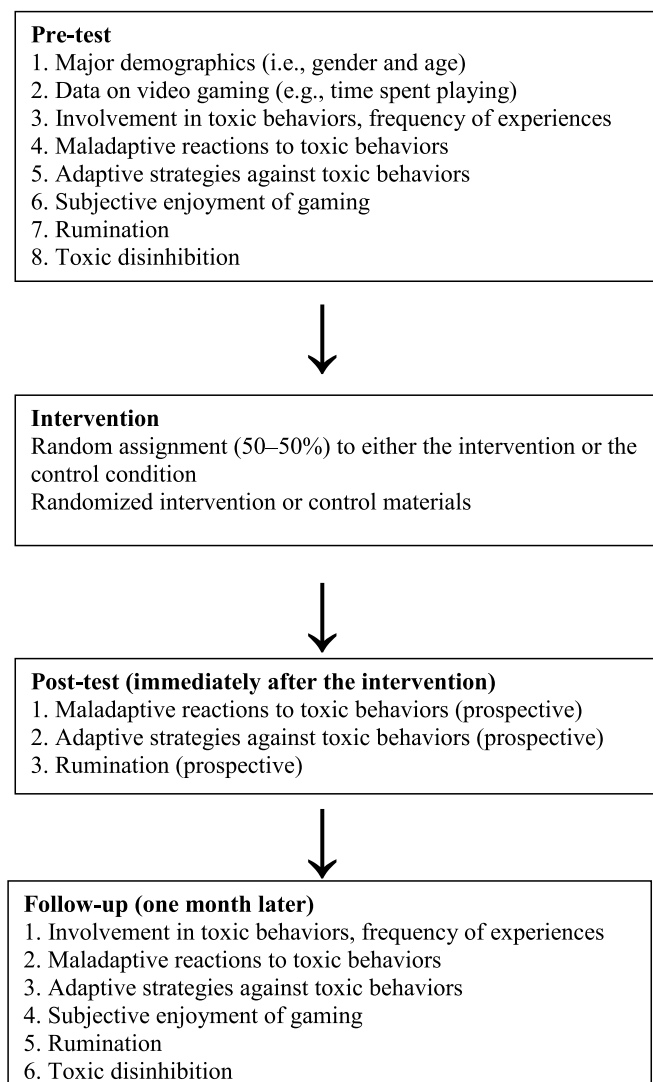


Fig. 3. Structure of the research material.

effect in the use of adaptive strategies was confirmed only after the intervention. Retaliation is a key motivation for engagement in toxic behaviors (Liu & Agur, 2023). Indeed, the substantial overlap between targets and perpetrators (Ruotsalainen & Meriläinen, 2023) indicates that toxicity is contagious within a gameplay (Kim & Tsvetkova, 2021; Shen et al., 2020). Therefore, decreasing maladaptive reactions to other players' toxicity, such as retaliation can be a key component of efforts against the spreading of toxicity. Players with a higher propensity for using maladaptive strategies (e.g., self-blame for poor team performance; Frommel & Mandryk, 2024) can possibly benefit from the integration of adaptive reactions into their coping strategies. Reid et al. (2022) suggested that social support and pursuing positive outcomes, as effective coping strategies in individuals with experiences of discrimination and bullying, should be used in efforts to reduce the harms associated with toxicity. In addition to descriptive norms and the "saying-is-believing" self-persuasion strategy (Aronson, 1999), the testimonials used in this study aimed at presenting such adaptive strategies with later benefits reported by a fellow player. As these testimonials focused mainly on the presentation of these adaptive strategies and benefits, it may not be surprising that the present intervention was most effective at changing these strategies, while reduction in related harms (e.g., reduced subjective game enjoyment) and direct involvement in toxicity was not confirmed. These findings suggest that more targeted components should be incorporated in future intervention materials to reduce these specific outcomes (e.g., dynamic norms suggesting that while toxicity was accepted earlier, such behaviors are not tolerated by the gaming community nowadays).

Third, the expectation that the intervention would lead to higher subjective game enjoyment one month later (H4) was not supported. This result can possibly be explained by previous findings that disruptive players can find enjoyment in toxicity (see Achterbosch et al., 2017; Cook et al., 2018). Similarly, Paul et al. (2015) found no difference in the enjoyment level of grievers and non-grievers, suggesting that disruptive behaviors can also be rewarding (e.g., retaliation), and can increase game enjoyment in some players. In a similar vein, Cote (2017) indicated that women players can maintain game enjoyment in spite of negative experiences. Therefore, game enjoyment may be more strongly associated with basic need fulfillment than toxicity, which can ultimately fulfill social interaction needs (see Cook et al., 2018). Specifically, players differ in their levels of basic need satisfaction and frustration across roles in toxic encounters (e.g., targets and perpetrators; Paul et al., 2015), which may be a more influential component of game enjoyment than the involvement in toxic interactions.

Furthermore, the hypothesis postulating that the intervention would lead to less frequent involvement in toxic behaviors one month later did not receive support. Descriptive norms were presented in the form of statistics preceding the testimonials highlighting that although the occurrence of engagement in toxicity is high and the majority of players find these behaviors irritating, the proportion of disruptive players who had never been targets of such negative acts is small, suggesting that the problem can be addressed effectively. More specifically, these statistics suggest that the number of players who consistently engage in disruptive behaviors throughout their gameplay is quite small. However, the present testimonials, as the main components of the intervention, focused primarily on coping with toxicity, while social norms were not explicitly described in these texts. Future research using social norms directly in such testimonials may be more effective in facilitating long-term changes in personal involvement. Moreover, the present intervention primarily focused on post-incident coping rather than the prevention of being involved in toxicity.

Finally, the assumption that the intervention would lead to lower toxic disinhibition one month later was not confirmed. Although disinhibition has a major role in exhibiting toxicity (Kordyaka et al., 2020), using less maladaptive strategies may not be directly associated with players' fundamental perceptions of the circumstances which can contribute to toxicity (e.g., anonymity). In this sense, toxic disinhibition

can be a form of coping against the harms of toxicity in targets as such a distancing perspective can facilitate the rationalization of toxic players' behaviors, mitigating personal harms (e.g., toxic players may think writing insulting things in the chat is not so hurtful to others).

4.1. Implications

The present findings suggest that social interventions have the potential to efficiently combat toxicity in the gaming community. Therefore, further strategies should be developed. Components that were explicitly included in the present testimonials (e.g., adaptive and maladaptive strategies) proved to be consistently efficient in a short period of time (i.e., immediately after the intervention), resulting in the immediate reduction of retaliatory tendencies. Players exposed to fellow players' adaptive strategies against toxic behaviors and the benefits of using these strategies reported decreased rumination and stronger intention to use more adaptive strategies and less maladaptive strategies when encountering toxicity after the intervention. The present study also provided evidence that this intervention could decrease maladaptive reactions to toxicity in players within a month. Maladaptive reactions to toxic behaviors (e.g., retaliation) were found to be major contributors to the social contagion of toxicity (Kim & Tsvetkova, 2021; Shen et al., 2020). The present study offers a promising intervention strategy to mitigate maladaptive reactions to toxicity, which can prevent the spreading of disruptive behaviors. Therefore, these findings can contribute to a more effective combat against toxicity, and thereby, a safer and more positive gaming climate (see Czako et al., 2023). Another contribution of this study is the preregistered randomized controlled design in a relatively large gamer sample – compared to previous experimental studies conducted in this community (see Wijkstra et al., 2024) –, which enabled the exploration of causal inferences, and contributed to a more nuanced understanding of how exposure to adaptive strategies against toxicity can impact retaliatory tendencies in a short-term period. Furthermore, the openness of the research materials and measures contribute to the replicability of the intervention in a broader, international context to increase the generalizability of the findings.

4.2. Limitations

A major limitation of this study is the lack of significant improvement in game enjoyment, disinhibition, and rumination in the longer run, which may suggest that such distal constructs could not be effectively targeted by this brief, single-session social-psychological intervention in a one-month interval. This finding underlines the importance of considering repeated-exposure interventions to produce and consolidate shifts in players' personal attitudes and perception of the gaming climate. However, another possible explanation for the lack of significant improvement could be that the testimonials in the present intervention mainly included practical suggestions against toxic behaviors (e.g., reporting, muting) instead of emotional support. Reid et al. (2022) highlighted that although players prefer tools that explicitly combat toxic behaviors (e.g., reporting), support tools that combine actions against toxic players and help targets cope with the negative experience would be desirable. Focusing on sanctions against toxic players may lead to inadequate responses (e.g., reporting of beginner players due to their poorer performance), while tools addressing solely the emotional support of targets were found to be more effective (Reid et al., 2022). However, Reid et al. (2022) also emphasized that tools designed for the emotional support of players may not be well-received and generally used due to cynicism and desensitization as consequences of the normalization of toxicity. Future interventions should incorporate targeted strategies aimed at improving the emotional well-being of players impacted by toxicity. Another possible explanation of the lack of improvement in most of the constructs could be that a single session and a one-month follow-up may be insufficient to detect meaningful changes

Table 1
List of measures used in the intervention.

Hypothesis	Variable	Measure	Items	Cronbach's alpha in the present study	Scaling	Source
H1	Rumination	Rumination About an Interpersonal Offense Scale (adapted to toxic behaviors in the past month time interval). In the post-test, participants were asked how likely they would react to other players' toxic behaviors in such ways. Instruction: The following statements refer to reactions players may experience when others insult or intentionally annoy them, thereby ruining their gaming experience. Reflect on your experiences from the past month and indicate how much you agree or disagree with the following statements	6 items 1. ... I couldn't think about anything else for a while other than how badly I was treated in the game. 2. ... I lost my motivation and energy when I thought about what happened to me during the game. 3. ... I had a hard time getting it out of my mind how badly I was treated. 4. ... I tried to figure out why this/these player(s) treated me this way. 5. ... The insult I experienced in the game stayed in my head for a long time. 6. ... I found myself replaying in my mind what happened to me during the game over and over.	Pre-test 0.88 for the treatment and 0.87 for the control group Post-test 0.92 for the treatment and 0.91 for the control group Follow-up 0.88 for the treatment and 0.85 for the control group	Pre-test and follow-up 5-point Likert-scale (from 1 = strongly disagree to 5 = strongly agree) Post-test 6-point scale (1 = very unlikely, 2 = unlikely, 3 = rather unlikely, 4 = rather likely, 5 = likely, 6 = very likely)	Wade et al. (2008)
H2	Maladaptive reactions to toxicity	Self-developed Pre-test and follow-up Indicate how often you responded in the following ways in the past month when you were intentionally insulted, annoyed, or had personal information shared about you without permission during gameplay. Post-test Now imagine that in the coming weeks other players intentionally insult, annoy, or share personal information about you without permission during gameplay. Indicate how likely it is that you would ...	5 items 1. I insulted the toxic player 2. I did the same to the toxic player to make him/her/ them feel its negative impact 3. I responded more harshly to teach him/her/ them a lesson 4. I fantasized about revenge 5. I completely lost interest in the game	–	Pre-test and follow-up 6-point scale (1 = never or almost never [0–10 %], 2 = occasionally [11–35 %], 3 = somewhat less than half of the cases [36–50 %], 4 = somewhat more than half of the cases [51–65 %], 5 = often [66–90 %], 6 = always or almost always [91–100 %]) Post-test 6-point scale (1 = very unlikely, 2 = unlikely, 3 = rather unlikely, 4 = rather likely, 5 = likely, 6 = very likely)	Developed based on Kanetsuna et al., 2006 (i.e., aggressive response and avoidance, including passive behavior and stay away) and Murray-Harvey et al., 2012 on ineffective strategies to cope with peer aggression. Discontinued gameplay (Kordyaka, Laato, Jahn, et al., 2023) and game abandonment as a consequence of toxicity (Zsila et al., 2022) have also been identified as passive, avoidant reactions to exposure to toxic behaviors. These findings justify the inclusion of “completely losing interest in the game” as an ineffective reaction form.
H3	Adaptive reactions to toxicity	Self-developed Pre-test and follow-up Indicate how often you did the following in the past month when you were intentionally insulted, annoyed, or had personal information shared about you without permission during gameplay. Post-test If, in the coming weeks, another player insults, intentionally annoys you, or shares personal information about you without permission during gameplay, how likely is it that you will ...	4 items 1. I reported the player. 2. I muted the player. 3. I did not respond to the insults and ignored the provocation. 4. I looked for another team to play with, where I could play undisturbed.	–	Pre-test and follow-up 6-point scale (1 = never or almost never [0–10 %], 2 = occasionally [11–35 %], 3 = somewhat less than half of the cases [36–50 %], 4 = somewhat more than half of the cases [51–65 %], 5 = often [66–90 %], 6 = always or almost always [91–100 %]) Post-test 6-point scale (1 = very unlikely, 2 = unlikely, 3 = rather unlikely, 4 = rather likely, 5 = likely, 6 = very likely)	Developed based on the suggestion by Türkay et al. (2020), Adinolf and Türkay (2018), and Kou (2020, pp. 81–92). Specifically, reporting, muting, and ignoring were considered. The fourth item was added based on the expert interviews conducted preceding the construction of the present intervention.
H4	Game enjoyment	Self-developed Pre-test and follow-up In the past month, when you played, how often ...	9 items 1. ... were you able to truly enjoy the game? 2. ... were you able to truly immerse yourself in the game? 3. ... were you greatly amused by the game? 4. ... was the gameplay with other players free of	Pre-test 0.88 for the treatment and 0.90 for the control group Follow-up 0.91 for the treatment and 0.91 for the control group	6-point scale 1 = never or almost never (0–10 %), 2 = occasionally (11–35 %), 3 = somewhat less than half of the cases (36–50 %), 4 = somewhat more than half of the cases (51–65 %), 5 = often (66–90 %), 6 = always or almost always (91–100 %).	Developed based on the expert interviews conducted preceding the construction of the present intervention. Item creation and the approval of the final item set were consensually achieved by the expert interviewees and the research team. Items 1, 2 and 3 focused on general enjoyment of gaming,

(continued on next page)

Table 1 (continued)

Hypothesis	Variable	Measure	Items	Cronbach's alpha in the present study	Scaling	Source
			tension? 5. ... were you in harmony with other players? 6. ... was it comfortable to play together with other players? 7. ... did you feel that the atmosphere of other players was supportive? 8. ... did you feel that other players were benevolent? 9. ... did you feel that you could trust in other players?			while other items represented social aspects of game enjoyment.
H5	Frequency of involvement in toxic behaviors	Self-developed Pre-test and follow-up How often did other players do this to you?	5 items 1. Players used rude, offensive language toward you (flaming); mocked you, yelled at you, threatened you, or defamed you. 2. Players insulted you based on your origin, gender, religion, or other personal traits, or engaged in hate speech or incitement against you. 3. Players annoyed you with repetitive chat messages or character movements to distract you (spamming). 4. Players intentionally irritated or upset you by not following the game rules, cheating (cheat codes), switching to the opponent's team (feeding), sabotaging, or hindering your progress (griefing). 5. Players shared personal data or information about you without your consent during gameplay (e.g., your address) (doxxing).	–	6-point scale 1 = never (0 %), 2 = rarely (1–35 %), 3 = somewhat less than half of the cases (36–50 %), 4 = somewhat more than half of the cases (51–65 %), 5 = often (66–99 %), 6 = always (100 %)	Developed based on Beres et al. (2021) – flaming, spamming; Kou (2020, pp. 81–92) – griefing, feeding; and Kowert (2020) – hate speech, cheating, doxxing.
H6	Toxic disinhibition	Online Disinhibition Scale – toxic disinhibition subscale (adapted to online multiplayer context) Instruction: Indicate how much you agree with the following statements.	4 items 1. I don't mind writing or saying offensive things to other players because I remain anonymous. 2. I don't find it difficult to write or say offensive things during the game because there are no consequences. 3. Online games don't have strict rules, so you can do whatever you want. 4. Writing or saying offensive things during the game doesn't count as harassment.	Pre-test 0.83 for the treatment and 0.75 for the control group Follow-up 0.80 for the treatment and 0.79 for the control group	5-point Likert-scale (from 1 = definitely do not believe to 5 = definitely do believe)	Jenkins-Guarnieri et al., 2013; Udris, 2014

Note. Self-developed questions and statements were translated from Hungarian to English using ChatGPT (OpenAI, 2023). The original Hungarian version can be found in the online repository (https://osf.io/vwntg/?view_only=bcc1a5f723a54e0d998736e24d134f4c).

in distal constructs such as game enjoyment and toxic disinhibition. Another limitation is that qualitative analysis on the individual textual responses was not conducted. Future content analysis of these texts could enhance knowledge on the components that can possibly predict more adaptive reactions and effective coping with toxic behaviors. Moreover, the difference between the treatment group and the control group in rumination immediately after the intervention and in maladaptive responses to toxicity in the one-month follow-up was close to the level of significance. Therefore, further research is desirable to confirm these effects using larger samples in which these small effects

can be detected with greater accuracy. Another limitation of this study is the lack of statistical power to analyze possible improvement in subgroups (e.g., targets of toxicity in the pre-test or the one-month follow-up versus players who were not involved in toxicity at these time intervals). Plausible pathways from maladaptive reactions to later involvement with mediating factors could not be meaningfully tested due to this lack of statistical power, either. Also, the present results cannot be generalized to the whole population of players due to the convenience sampling method. The present sample comprised predominantly Hungarian male gamers, which further undermines

Table 2

Summary of the results and estimates related to the hypothesis testing of the intervention.

Hypothesis	Outcome	Estimate (standardized β)	Effect size (Hedges' g)	95 % CI	p -value	q -value (FDR-adjusted)	Test type	Sample size (N)
H1	Rumination (post)	-0.18	$g = 0.17$	[-0.00; -0.34]	$p = 0.046$	$q = 0.104$	t -test	N = 520
	Rumination (follow-up)	-0.06	$g = 0.06$	[-0.21; 0.32]	$p = 0.680$	$q = 0.750$	Welch	N = 217
H2	Maladaptive reactions (post)	-0.26	$g = 0.24$	[-0.07; -0.42]	$p = 0.006$	$q = 0.027$	Welch	N = 520
	Maladaptive reactions (follow-up)	-0.30	$g = 0.30$	[-0.04; -0.57]	$p = 0.030$	$q = 0.090$	Welch	N = 217
H3	Adaptive strategies (post)	0.26	$g = -0.26$	[0.43; 0.09]	$p = 0.003$	$q = 0.027$	Welch	N = 520
	Adaptive strategies (follow-up)	0.04	$g = -0.04$	[-0.31; 0.22]	$p = 0.750$	$q = 0.750$	Welch	N = 217
H4	Game enjoyment (follow-up)	0.03	$g = -0.03$	[-0.20; 0.14]	$p = 0.731$	$q = 0.750$	Welch	N = 217
H5	Involvement in toxic behaviors (follow-up)	-0.16	$g = 0.16$	[-0.11; 0.42]	$p = 0.240$	$q = 0.432$	Welch	N = 217
H6	Toxic disinhibition (follow-up)	-0.04	$g = 0.04$	[-0.14; 0.21]	$p = 0.690$	$q = 0.750$	Welch	N = 217

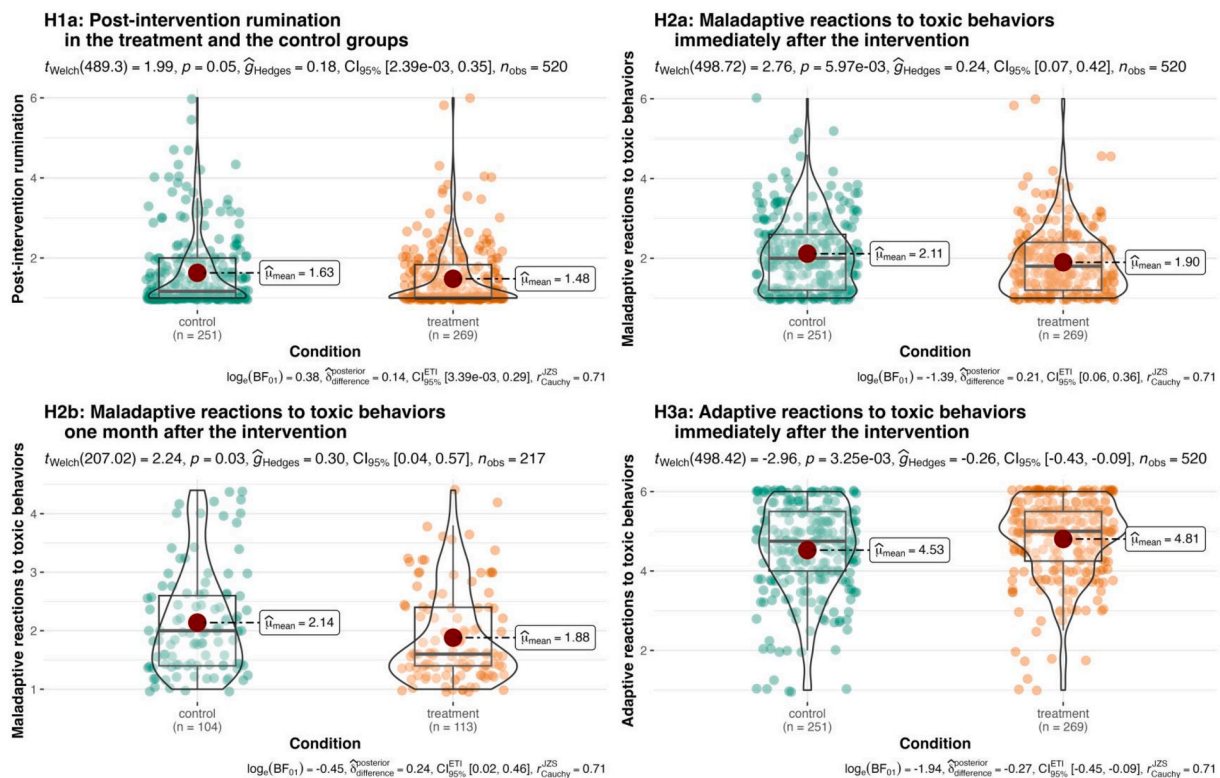


Fig. 4. Significant immediate and long-term effects of the intervention.

generalizability. The validity of self-developed items in the present intervention also needs further empirical support. The measure of exposure to toxicity could not capture severity, modality, or the role of involvement. The present outcomes heavily relied on self-developed measures with limited psychometric evaluation, including checks for measurement invariance across time and condition. Furthermore, this study lacked manipulation checks for perceived norms and coping self-efficacy, and engagement/fidelity data. It should also be acknowledged that the control condition may have inadvertently improved comfort and attention, which could have reduced contrasts on the measure of game enjoyment. Social desirability could influence immediate intentions and reactions to future toxicity. Finally, the present outcomes were self-reported. Immediate post-intervention and

follow-up behavioral outcome measures could make the present results ecologically more robust.

4.3. Future directions

Future interventions should address some important issues to increase efficacy, which could not be addressed in the present investigation due to sample size limitations. First, there are substantial differences among players in the involvement in toxicity. Specifically, beginner players, women, and minority individuals were found to be more frequently targeted (Gandolfi & Ferdig, 2022), suggesting that prevention efforts should consider employing a more differentiated strategy to decrease the risk of involvement and psychological

consequences in specific player groups. For this purpose, prevention and intervention efforts aiming at mitigating social dominance could possibly be effective as social dominance was associated with cyber-aggression (Jagayat & Choma, 2021; Tang et al., 2020).

Based on the present intervention, targeting social norms appears to be an effective strategy for promoting adaptive responses and reducing maladaptive reactions to toxicity. Therefore, game developers may consider establishing prosocial norms and rules within their platforms. Additionally, players should be explicitly informed about the consequences and sanctions associated with engaging in antisocial behaviors. Nevertheless, although responsiveness to toxicity can increase safety in the gaming environment, strict sanctions may undermine the game developer companies' reputation due to players leaving the game, particularly when the act of toxicity is questionable (Reid et al., 2022). Specifically, rigidity of gaming norms was found to be a major contributor to toxic behaviors (Deslauriers et al., 2020), suggesting that tools against toxicity should assess disruptive behaviors in specific contexts, considering multiple factors such as game objectives (e.g., some games instruct players to destroy the enemy team's buildings as part of the game's objective, while this behavior may be perceived as unnecessary and serves only to annoy other players in another game). Therefore, game developers should also consider providing behavioral guidance that is specific to certain game environments. Overall, creating a safe and just gaming environment for players may be challenging which is also financially beneficial for the game developers.

The present intervention utilized descriptive norms suggesting that the proportion of players who consistently engage in toxic behaviors is low, while the majority of players enjoy the benefits of team cooperation and positive social gaming experiences. Incorporating such descriptive norms into the code of conduct provided for players may help shift normative beliefs, challenging the perception that toxicity is an inevitable part of the multiplayer gaming culture (Kordyaka, Laato, Jahn, et al., 2023). Findings that normative beliefs about aggression in cyberspace predict such negative behaviors (Hilvert-Bruce & Neill, 2020) confirm the importance of perceptions in forming social norms in the gaming community. Most players did not believe that gaming environments fundamentally promote aggression between peers in a study by Li and Pustaka (2017), and more than half of these players expressed willingness to play an anti-bullying game, suggesting that clear social norms could contribute to a safer environment that offers rewarding social experiences. Therefore, future interventions should incorporate more types of core social norms to test their efficacy simultaneously (e.g., injunctive norms, which promote behaviors to avoid sanctions, see Byron et al., 2016). Indeed, Irwin et al. (2024) suggested that normative beliefs about undesired behaviors (e.g., toxicity) could discourage players from engaging in toxicity (i.e., descriptive norm) and encourage the avoidance of repercussions (i.e., injunctive norm). Interventions relying on the bandwagon heuristic (i.e., "if others believe it is important, I should believe likewise"; Sundar, 2008) could also be effective (i.e., players internalize the importance of creating and maintaining a safe gaming environment). Incorporating dynamic norms instead of using static norms – which were employed in the present intervention – can also increase the efficacy of future interventions (e.g., toxic behaviors may have been accepted earlier but not tolerated these days), which underline the flexible nature of social norms (see Van Kleef et al., 2019).

The present intervention, in accordance with the suggestion by Reid et al. (2022), primarily focused on the support of targets instead of identifying and sanctioning toxic players. Future research should consider employing a harm-preventing approach to support the emotional well-being of targets of toxic behaviors (Wijkstra et al., 2024). Based on the present intervention, it may be beneficial to provide information for beginner players about how to respond to toxicity besides a general guide related to the gameplay. Specific guidance from experienced, successful players who also encountered toxicity and responded adaptively can possibly encourage beginner players to respond constructively in such situations, which can also foster a more adaptive

emotional coping. Sharing personal experiences and offering practical, adaptive strategies against toxicity by influential players can be a credible source of information to combat toxic behaviors among players. However, the possible short- and long-term effects of this practical guidance on the actual behavior should be carefully monitored and tested in the future.

4.4. Conclusion

Overall, the present findings suggest that social norms-based interventions can be most effective at fostering adaptive reactions and reducing maladaptive responses to other players' toxic behaviors, while their long-term effects on psychological well-being, game enjoyment, and involvement in toxicity were not demonstrated. Observed changes in adaptive and maladaptive reactions to toxicity may suggest that this approach can ultimately discourage the intent of retaliation and promotes more prosocial interactions. Previous studies (Hilvert-Bruce & Neill, 2020; Nexø & Kristiansen, 2023) have suggested that interventions aimed at shifting social norms that tolerate toxic behaviors can effectively reduce toxicity in the gaming culture. The present study provides preliminary empirical evidence that social norms have the potential to diminish the intent to retaliate and other maladaptive reactions to negative acts in emotionally charged in-game situations. This may be a first step towards disrupting the cycle of reactive aggression and hostility – key contributors to the social contagion of toxicity (Kim & Tsvetkova, 2021; Shen et al., 2020).

CRedit authorship contribution statement

Ágnes Zsila: Writing – original draft, Conceptualization. Gábor Orosz: Writing – original draft, Methodology. Zsolt Demetrovics: Writing – original draft, Investigation.

Funding

Ágnes Zsila was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. Gábor Orosz was supported by the ANR grant of the Chaire Professeur Junior of Artois University and by the Strategic Dialogue and Management Scholarship (Phases 1 and 2).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chb.2025.108827>.

Data availability

Data will be made available on request.

References

- Achterbosch, L., Miller, C., & Vamplew, P. (2017). Participant observation of grieving in a journey through the world of warcraft. *Loading*, 10(17), 40–59.
- Adinolf, S., & Turkay, S. (2018). Toxic behaviors in Esports games: Player perceptions and coping strategies. *Proceedings of the 2018 annual symposium on computer-human interaction in play companion extended abstracts*.
- Aronson, E. (1999). The power of self-persuasion. *American Psychologist*, 54(11), 875–888. <https://doi.org/10.1037/h0088188>
- Aronson, E., Fried, C., & Stone, J. (1991). Overcoming denial and increasing the intention to use condoms through the induction of hypocrite. *American Journal of Public Health*, 81(12), 1636–1638. <https://doi.org/10.2105/AJPH.81.12.1636>

- Beres, N. A., Frommel, J., Reid, E., Mandryk, R. L., & Klarkowski, M. (2021). Don't you know that you're toxic: Normalization of toxicity in online gaming. *Proceedings of the 2021 CHI conference on human factors in computing systems*.
- Blackburn, J., & Kwak, H. (2014). Stfu NOOB! predicting crowdsourced decisions on toxic behavior in online games. *Proceedings of the 23rd international conference on world wide web*.
- Byron, M. J., Cohen, J. E., Frattaroli, S., Gittelsohn, J., & Jernigan, D. H. (2016). Using the theory of normative social behavior to understand compliance with a smoke-free law in a middle-income country. *Health Education Research*, 31(6), 738–748. <https://doi.org/10.1093/her/cyw043>
- Canossa, A., Salimov, D., Azadvar, A., Harteveld, C., & Yannakakis, G. (2021). For honor, for toxicity: Detecting toxic behavior through gameplay. *Proceedings of the ACM on Human-Computer Interaction*, 5(CHI PLAY), 1–29. <https://doi.org/10.1145/3474680>
- Chahal, S., Rana, S., & Singh, P. (2016). Impact of coping on mental health of convicted prisoners. *International Journal of Indian Psychology*, 3(2), 66–75.
- Chen, X. Y., Han, J., Brown, C. S., & Usher, E. L. (2024). How does a social belonging intervention affect college students from historically underrepresented groups? A mixed methods investigation. *Journal of Diversity in Higher Education*. <https://doi.org/10.1037/dhe0000613>
- Cook, C., Schaafsma, J., & Antheunis, M. (2018). Under the bridge: An in-depth examination of online trolling in the gaming context. *New Media & Society*, 20(9), 3323–3340. <https://doi.org/10.1177/1461444817748578>
- Cote, A. C. (2017). “I can defend myself” women’s strategies for coping with harassment while gaming online. *Games and Culture*, 12(2), 136–155. <https://doi.org/10.1177/1555412015587603>
- Czakó, A., Király, O., Koncz, P., Yu, S. M., Mangat, H. S., Glynn, J., Romero, P., Griffiths, M. D., Rumpf, H.-J., & Demetrovics, Z. (2023). Safer esports for players, spectators, and bettors: Issues, challenges, and policy recommendations. *Journal of Behavioral Addictions*, 12(1), 1–8. <https://doi.org/10.1556/2006.2023.00012>
- Deslauriers, P., St-Martin, L. I. L., & Bonenfant, M. (2020). Assessing toxic behaviour in dead by daylight: Perceptions and factors of toxicity according to the game’s official subreddit contributors. *Game Studies*, 20(4).
- Eklund, L. (2015). Playing video games together with others: Differences in gaming with family, friends and strangers. *Journal of Gaming & Virtual Worlds*, 7(3), 259–277. <https://doi.org/10.1386/jgvw.7.3.259.1>
- Feinstein, B. A., Bhatia, V., & Davila, J. (2014). Rumination mediates the association between cyber-victimization and depressive symptoms. *Journal of Interpersonal Violence*, 29(9), 1732–1746. <https://doi.org/10.1177/0886260513511534>
- Fox, J., & Tang, W. Y. (2017). Women’s experiences with general and sexual harassment in online video games: Rumination, organizational responsiveness, withdrawal, and coping strategies. *New Media & Society*, 19(8), 1290–1307. <https://doi.org/10.1177/1461444816635778>
- Frommel, J., Johnson, D., & Mandryk, R. L. (2023). How perceived toxicity of gaming communities is associated with social capital, satisfaction of relatedness, and loneliness. *Computers in Human Behavior Reports*, 10, Article 100302. <https://doi.org/10.1016/j.chbr.2023.100302>
- Frommel, J., & Mandryk, R. L. (2024). Toxicity in online games: The prevalence and efficacy of coping strategies. In *Proceedings of the 2024 CHI conference on human factors in computing systems* (pp. 1–12). <https://doi.org/10.1145/3613904.3642523>
- Gandolfi, E., & Ferdig, R. E. (2022). Sharing dark sides on game service platforms: Disruptive behaviors and toxicity in DOTA2 through a platform lens. *Convergence*, 28(2), 468–487. <https://doi.org/10.1177/13548565211028809>
- Gandolfi, E., Ferdig, R. E., Krause, K., Copus, A., Ostrowski-Delahanty, S., & Alemagno, S. (2023). Problematic gaming at a crossroad: Exploring the interplay between internet gaming disorder, toxic attitudes, and empathy in Digital entertainment. *Games and Culture*, Article 15554120231211991. <https://doi.org/10.1177/15554120231211991>
- Garnefski, N., & Kraaij, V. (2006). Cognitive emotion regulation questionnaire—development of a short 18-item version (CERQ-short). *Personality and Individual Differences*, 41(6), 1045–1053. <https://doi.org/10.1016/j.paid.2006.04.010>
- Goldstein, N. J., & Cialdini, R. B. (2011). Using social norms as a lever of social influence. In *The science of social influence* (pp. 167–191). Psychology Press. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780203818565-7/using-social-norms-lever-social-influence-noah-goldstein-robert-cialdini>
- Hilvert-Bruce, Z., & Neill, J. T. (2020). I’m just trolling: The role of normative beliefs in aggressive behaviour in online gaming. *Computers in Human Behavior*, 102, 303–311. <https://doi.org/10.1016/j.chb.2019.09.003>
- Irwin, S. V., Naweed, A., & Lastella, M. (2024). Consuming esports and trash talking: How do social norms and moderating attributes influence behaviour? *Sport in Society*, 27(1), 14–32. <https://doi.org/10.1080/17430437.2023.2200732>
- Jagayat, A., & Choma, B. L. (2021). Cyber-aggression towards women: Measurement and psychological predictors in gaming communities. *Computers in Human Behavior*, 120, Article 106753. <https://doi.org/10.1016/j.chb.2021.106753>
- Jenkins-Guarnieri, M. A., Wright, S. L., & Johnson, B. (2013). Development and validation of a social media use integration scale. *Psychology of Popular Media Culture*, 2(1), 38–50. <https://doi.org/10.1037/a0030277>
- Kanetsuna, T., Smith, P. K., & Morita, Y. (2006). Coping with bullying at school: Children’s recommended strategies and attitudes to school-based interventions in England and Japan. *Aggressive Behavior*, 32(6), 570–580. <https://doi.org/10.1002/ab.20156>
- Kim, J. E., & Tsvetkova, M. (2021). Cheating in online gaming spreads through observation and victimization. *Network Science*, 9(4), 425–442. <https://doi.org/10.1017/nws.2021.19>
- Kordyaka, B., Jahn, K., & Niehaves, B. (2020). Towards a unified theory of toxic behavior in video games. *Internet Research*, 30(4), 1081–1102. <https://doi.org/10.1108/INTR-08-2019-0343>
- Kordyaka, B., Klesel, M., & Jahn, K. (2019). Perpetrators in league of legends: Scale development and validation of toxic behavior. *Proceedings of the 52nd Hawaii international conference on System sciences*. <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1248&context=hicss-52>
- Kordyaka, B., Laato, S., Jahn, K., Hamari, J., & Niehaves, B. (2023). The cycle of toxicity: Exploring relationships between personality and player roles in toxic behavior in multiplayer online battle arena games. In *Proceedings of the ACM on human-computer interaction* (Vol. 7, pp. 611–641). CHI PLAY.
- Kordyaka, B., Laato, S., Weber, S., & Niehaves, B. (2023). What constitutes victims of toxicity-identifying drivers of toxic victimhood in multiplayer online battle arena games. *Frontiers in Psychology*, 14, Article 1193172. <https://doi.org/10.3389/fpsyg.2023.1193172>
- Kou, Y. (2020). Toxic behaviors in team-based competitive gaming: The case of league of legends. *Proceedings of the annual symposium on computer-human interaction in play*.
- Kou, Y., & Gui, X. (2014). Playing with strangers: Understanding temporary teams in league of legends. *Proceedings of the first ACM SIGCHI annual symposium on computer-human interaction in play*. <https://doi.org/10.1145/2658537.2658538>
- Kowert, R. (2020). Dark participation in games. *Frontiers in Psychology*, 11, 2969. <https://doi.org/10.3389/fpsyg.2020.598947>
- Kwak, H., Blackburn, J., & Han, S. (2015). Exploring cyberbullying and other toxic behavior in team competition online games. *Proceedings of the 33rd annual ACM conference on human factors in computing systems*.
- Laato, S., Kordyaka, B., & Hamari, J. (2024). Traumatizing or just annoying? Unveiling the spectrum of gamer toxicity in the starcraft II community. *Proceedings of the 2024 CHI conference on human factors in computing systems*.
- Lee, H.-S., Lee, H.-R., Park, J.-U., & Han, Y.-S. (2018). An abusive text detection system based on enhanced abusive and non-abusive word lists. *Decision Support Systems*, 113, 22–31. <https://doi.org/10.1016/j.dss.2018.06.009>
- Li, Q., & Pustaka, A. (2017). When cyberbullies meet gamers: What do young adults think? *Educational Research*, 59(4), 426–443. <https://doi.org/10.1080/00131881.2017.1369857>
- Liu, Y., & Agur, C. (2023). “After All, They Don’t Know Me” exploring the psychological mechanisms of toxic behavior in online games. *Games and Culture*, 18(5), 598–621. <https://doi.org/10.1177/15554120221115397>
- Monge, C. K., & O’Brien, T. C. (2021). Effects of individual toxic behavior on team performance in League of Legends. *Media Psychology*, 25(1), 82–105. <https://doi.org/10.1080/15213269.2020.1868322>
- Murray-Harvey, R., Skrzypiec, G., & Slee, P. T. (2012). Effective and ineffective coping with bullying strategies as assessed by informed professionals and their use by victimised students. *Journal of Psychologists and Counsellors in Schools*, 22(1), 122–138. <https://doi.org/10.1017/jgc.2012.5>
- Nexo, L. A., & Kristiansen, S. (2023). Players don’t die, they respawn: A situational analysis of toxic encounters arising from death events in league of legends. *European Journal on Criminal Policy and Research*, 29(3), 457–476. <https://doi.org/10.1007/s10610-023-09552-y>
- Orosz, G., Faragó, L., Paskuj, B., & Krekó, P. (2024). Strategies to combat misinformation: Enduring effects of a 15-minute online intervention on critical-thinking adolescents. *Computers in Human Behavior*, Article 108338. <https://doi.org/10.1016/j.chb.2024.108338>
- Orosz, G., Faragó, L., Paskuj, B., Rakovics, Z., Sam-Mine, D., Audemard, G., Modeliar, M. S., & Krekó, P. (2024). Softly empowering a prosocial expert in the family: Lasting effects of a counter-misinformation intervention in an informational autocracy. *Scientific Reports*, 14(1), Article 11763. <https://doi.org/10.1038/s41598-024-61232-x>
- Orosz, G., Paskuj, B., Faragó, L., & Krekó, P. (2023). A prosocial fake news intervention with durable effects. *Scientific Reports*, 13(1), 3958. <https://doi.org/10.1038/s41598-023-30867-7>
- Pace, U., D’Urso, G., & Zappulla, C. (2021). Hating among adolescents: Common contributions of cognitive distortions and maladaptive personality traits. *Current Psychology*, 40(7), 3326–3331. <https://doi.org/10.1007/s12444-019-00278-x>
- Pace, U., Passanisi, A., & D’Urso, G. (2018). Emotional and cognitive correlates of hating among adolescents: An exploratory study. *Journal of Adolescence*, 68, 159–164. <https://doi.org/10.1016/j.adolescence.2018.08.002>
- Paul, H. L., Bowman, N. D., & Banks, J. (2015). The enjoyment of grieving in online games. *Journal of Gaming & Virtual Worlds*, 7(3), 243–258. <https://doi.org/10.1386/jgvw.7.3.243.1>
- Polanin, J. R., Espelage, D. L., Grotper, J. K., Ingram, K., Michaelson, L., Spinney, E., Valido, A., Sheikh, A. E., Torgal, C., & Robinson, L. (2022). A systematic review and meta-analysis of interventions to decrease cyberbullying perpetration and victimization. *Prevention Science*, 23(3), 439–454. <https://doi.org/10.1007/s11212-021-01259-y>
- Reid, E., Mandryk, R. L., Beres, N. A., Klarkowski, M., & Frommel, J. (2022). Feeling good and in control: In-game tools to support targets of toxicity. *Proceedings of the ACM on Human-Computer Interaction*, 6(CHI PLAY), 1–27. <https://doi.org/10.1145/3549498>
- Ruotsalainen, M., & Meriläinen, M. (2023). Young video game players’ self-identified toxic gaming behaviour: An interview study. *Eludamos: Journal for Computer Game Culture*, 14(1), 147–173. <https://doi.org/10.7557/23.7270>
- Shen, C., Sun, Q., Kim, T., Wolff, G., Ratan, R., & Williams, D. (2020). Viral vitriol: Predictors and contagion of online toxicity in World of Tanks. *Computers in Human Behavior*, 108, Article 106343. <https://doi.org/10.1016/j.chb.2020.106343>
- Stone, J., Aronson, E., Crain, A. L., Winslow, M. P., & Fried, C. B. (1994). Inducing hypocrite as a means of encouraging young adults to use condoms. *Personality and*

- Social Psychology Bulletin*, 20(1), 116–128. <https://doi.org/10.1177/0146167294201012>
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology and Behavior*, 7(3), 321–326. <https://doi.org/10.1089/1094931041291295>
- Sundar, S. S. (2008). The main model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger, & A. J. Flanagan (Eds.), *Digital media, youth, and credibility* (pp. 73–100). The MIT Press.
- Tang, W. Y., Reer, F., & Quandt, T. (2020). Investigating sexual harassment in online video games: How personality and context factors are related to toxic sexual behaviors against fellow players. *Aggressive Behavior*, 46(1), 127–135. <https://doi.org/10.1002/ab.21873>
- Türkay, S., Formosa, J., Adinolf, S., Cuthbert, R., & Altizer, R. (2020). See no evil, hear no evil, speak no evil: How collegiate players define, experience and cope with toxicity. *Proceedings of the 2020 CHI conference on human factors in computing systems*.
- Udris, R. (2014). Cyberbullying among high school students in Japan: Development and validation of the Online Disinhibition Scale. *Computers in Human Behavior*, 41, 253–261. <https://doi.org/10.1016/j.chb.2014.09.036>
- Van Kleef, G. A., Gelfand, M. J., & Jetten, J. (2019). The dynamic nature of social norms: New perspectives on norm development, impact, violation, and enforcement. *Journal of Experimental Social Psychology*, 84, Article 103814. <https://doi.org/10.1016/j.jesp.2019.05.002>
- Wade, N. G., Vogel, D. L., Liao, K. Y.-H., & Goldman, D. B. (2008). Measuring state-specific rumination: Development of the rumination about an interpersonal offense scale. *Journal of Counseling Psychology*, 55(3), 419–426. <https://doi.org/10.1037/0022-0167.55.3.419>
- Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science*, 23(1), 73–82. <https://doi.org/10.1177/0963721413512856>
- Walton, G. M., & Crum, A. J. (2021). *Handbook of wise interventions*. Guilford Publications.
- Walton, G. M., Murphy, M. C., Logel, C., Yeager, D. S., & Collaborative, C. T. (2017). *The social-belonging intervention: A guide for use and customization*. New York: Handbook of Wise Interventions. Guilford Press.
- Walton, G. M., Murphy, M. C., Logel, C., Yeager, D. S., Goyer, J. P., Brady, S. T., Emerson, K. T. U., Paunesku, D., Fotuhi, O., Blodorn, A., Boucher, K. L., Carter, E. R., Gopalan, M., Henderson, A., Kroeper, K. M., Murdock-Perriera, L. A., Reeves, S. L., Ablorh, T. T., Ansari, S., ... Krol, N. (2023). Where and with whom does a brief social-belonging intervention promote progress in college? *Science*, 380(6644), 499–505. <https://doi.org/10.1126/science.ade4420>
- Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review*, 125(5), 617–655.
- Wang, X., Qiao, Y., Li, W., & Dong, W. (2022). How is online disinhibition related to adolescents' cyberbullying perpetration? Empathy and gender as moderators. *The Journal of Early Adolescence*, 42(5), 704–732. <https://doi.org/10.1177/02724316211064515>
- Wijkstra, M., Rogers, K., Mandryk, R. L., Veltkamp, R. C., & Frommel, J. (2024). How to tame a toxic player? A systematic literature review on intervention systems for toxic behaviors in online video games. *Proceedings of the ACM on Human-Computer Interaction*, 8(CHI PLAY), 1–32.
- Wong, S., & Ratan, R. (2024). How does exposure to general and sexual harassment relate to female gamers' coping strategies and mental health? *Games and Culture*, 19(5), 670–689. <https://doi.org/10.1177/15554120231177600>
- Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho, C. M., ... Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, 573(7774), 364–369. <https://doi.org/10.1038/s41586-019-1466-y>
- Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., Lee, H. Y., O'Brien, J., Flint, K., & Roberts, A. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology*, 108(3), 374–391.
- Zsila, Á., & Demetrovics, Z. (2025). Taxonomy of toxic behaviors in multiplayer gaming environments: An extension of the context of peer aggression. *Journal of Behavioral Addictions*. <https://doi.org/10.1556/2006.2025.00034>
- Zsila, Á., Shabahang, R., Aruguete, M. S., & Orosz, G. (2022). Toxic behaviors in online multiplayer games: Prevalence, perception, risk factors of victimization, and psychological consequences. *Aggressive Behavior*, 48(3), 356–364. <https://doi.org/10.1002/ab.22023>