Computers & Education 189 (2022) 104594

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Improving adolescent moral reasoning *versus* cyberbullying: An online big group experiment by means of collective intelligence

Pablo Bautista [a], Jacobo Cano-Escoriaza [b], Eva Vicente Sánchez [a], Ana Cebollero-Salinas [b], Santos Orejudo [a,*]

[a] Department of Psychology and Sociology. Faculty of Education, University of Zaragoza, Spain
[b] Department of Educational Science. Faculty of Education, University of Zaragoza, Spain

ARTICLE INFO

Keywords:
Collective intelligence
Adolescents
Moral reasoning
Online interaction
Cyberbullying

ABSTRACT

Collective Intelligence (CI) is identified as the group capacity to come up with responses to complex tasks that are not accomplished or are of worse quality if performed cooperatively or individually. Based on this premise, we considered knowing if adolescents' moral reasoning would increase when a big group faces a topical moral dilemma: a sexting-centered cyberbullying case. To do so, the CI platform Thinkhub was used with an online group of 793 simultaneously connected year-1 Higher Secondary Education students. This platform contemplates an initial individual work phase, and another dynamic work group phase in which responses gradually appear. The system finally presents the most valued responses to the participants. The obtained results revealed a significant rise in the individual moral development levels that emerged during the interaction process when the participants engaged in one of the three posed questions, for which the response rate was also higher. The promising potential of the CI generated during online interaction processes followed to solve complex tasks, their weak points and future research in this field, are discussed.

1. Introduction

1.1. Collective intelligence

Collective Intelligence (CI) emerges as a construct to explain how a group of people performs with different types of collaborative group tasks based on the notion that the mean achieved performance with this format is superior to each member's individual performance (Woolley et al., 2010). This occurs with tasks involving a certain level of complexity that are designed to aggregately obtain responses. In empirical terms, the work of Woolley et al. (2010) is considered a starting point to obtain evidence for this conceptual proposal. By analyzing the results obtained by 192 work groups made up of between two and five people who did these tasks, those authors found that a single "C"-factor emerged to explain performance, which they called CI, and how it exceeded an individual's mean and was barely related to the group members' overall intelligence (Woolley et al., 2010). The groups solved different types of tasks, such as mathematical problems, spatial reasoning or moral-type tasks. A single factor also emerged from the factor analysis performed

* Corresponding author.
E-mail addresses: pbautista@posta.unizar.es (P. Bautista), jcano@unizar.es (J. Cano-Escoriaza), evavs@unizar.es (E.V. Sánchez), anacebollero@posta.unizar.es (A. Cebollero-Salinas), sorejudo@unizar.es (S. Orejudo).

https://doi.org/10.1016/j.compedu.2022.104594
Received 29 November 2021; Received in revised form 5 July 2022; Accepted 8 July 2022
Available online 12 July 2022

with the results, which also backed the CI construct.

Replicas of the study by Woolley et al. (2010) have been done with new more complex tasks, such as playing chess against an artificial intelligence (AI) program, conflict negotiation or solving moral dilemmas (Engel et al., 2014; Hjertø & Paulsen, 2016; Meslec et al., 2016). They obtained similar results, and some were obtained by research groups, but they included a main novelty: a shift from the face-to-face interaction applied in initial studies to online interaction which allows the number of participants in groups to substantially increase (Engel et al., 2014; Woolley & y Aggarwal, 2020). In any case, it is worth pointing out that one study did not replicate these results, that by Bates and Gupta (2017), who found that group performance is closely related to the intelligence of the participants in the group. Nonetheless, the possibility of presenting poorly defined problems to large groups to find solutions comes over in another closely related construct to CI, the so-called "crowd intelligence" (Bernstein et al., 2018). Such big group contexts offer the possibility of finding solutions to problems being conditioned by certain factors that emerge during interactions, such as some participants not collaborating, or lack of originality due to over-reproduction of responses or them spreading. Toyokawa et al. (2019) experimentally describe these limitations in online settings.

Nevertheless, a relevant idea comes over when studying these contexts: that of original and quality responses appearing in accordance with some kind of moderator or leader who manages the interaction to obtain quality results (Salganik et al., 2006). Bigham and Bernstein (2018) analyzed the idea-spreading process by means of "crowd intelligence" studies and pointed out a similar role to help to spread ideas known as the facilitator, which allows the group to display high productivity known as "crowd fertilization". The presence of facilitators in CI studies reveals the importance of social interaction in groups (Bigham and Bernstein, 2018). Many dimensions need to be dealt with, which range from modeling and copying ideas, to studying facilitation roles or the role that leadership may play in spreading ideas.

Apart from the facilitator role, researchers have shown an interest in knowing which variables favor CI appearing. For this purpose, two analysis levels have been considered (Woolley & y Aggarwal, 2020): one based on the interaction processes that emerge in the group (top-down); another that takes group members' characteristics (bottom-up). Research works related to top-down factors focus mainly on interaction variables among group members (Bernstein et al., 2018), such as task duration, turns to speak, the diversity of the provided responses, the time spent on searching for a consensus or groups' heterogeneity (Dai et al., 2020; De Vincenzo et al., 2017). At the bottom-up level, some of the analyzed variables are gender (Curşeu et al., 2015), cognitive diversity (Aggarwal et al., 2019), emotional intelligence (Hjertø & Paulsen, 2016), social sensitivity (Woolley & y Aggarwal, 2020) or individual intelligence (Bates & Gupta, 2017).

Nonetheless, research into the conditions under which quality or more limited responses may appear remains an open line of work, as is the possibility of checking the feasibility of big groups of people being able to interact to solve problems and to avoid some of the problems that arise in these models (Toyokawa et al., 2019).

1.2. Moral reasoning in adolescence

As previously mentioned, Woolley et al. (2010) evidenced that a CI factor emerged when facing different kinds of collaborative tasks, including moral reasoning tasks. Based on a theorical cognitive-evolutionary framework, means to improve morale have been proposed, such as exposing people to a higher moral reasoning level than that they are normally at (Kohlberg, 1989; Turiel, 1966) by putting into practice moral conflict experiments and peer discussion to encourage moral reasoning or to impose their own perspectives on other people (Etxebarría, 1999; Kohlberg, 1976, 1989). From this perspective, and by means of CI, a feasible positive evolution of moral reasoning has been produced by creating an experiment that causes moral conflict and encourages peer discussion (Etxebarría, 1999; Kohlberg, 1976, 1989). However, to be able to state that better quality moral reasoning emerges through collaborative group tasks (i.e., CI), it is necessary to know how moral reasoning evolves, particularly in adolescence, which is the evolutionary stage that the present work focuses on. Kohlberg's moral development theory (Kohlberg, 1971) offers a suitable framework to know how moral development and its different levels evolve. Kohlberg (1976) refers to several maturity stages on a 6-phase scale divided into three levels:

- Preconventional: actions are guided to avoid punishment and oriented in defending your personal interest. Acting in according with society is oriented to get your goals, not to benefit others.
- Conventional: actions are guided by society which is guided by law and order to define what is bad or good.
- Postconventional: actions are guided by your reflection of the norms and law stipulated by the society. You are capable to modify them if those are injustice with other groups and generating new universal-ethical principles.

These levels are closely linked with personal goals (values acquired during development), norms and the social context in which one lives (Kohlberg, 1981). These levels and phases are measured by different moral dilemmas. Each moral dilemma represents a complex situation where the participant should answer some questions trying to resolve it. The answers to the questions determined in which of the three levels and in which of the six phases of moral development the respondent was (Kohlberg, 1981). The robust classification of moral development and the use of moral dilemmas developed by Kohlberg (1976, 1981), have continued to be used to date in research in different fields such business ethics, professional ethics in pharmacy (Martí-Vilar et al., 2021) and moral education (Zhang & Zhao, 2017) and it is still considered as a relevant tool for measuring moral development, although both, its theory and its moral dilemmas have been evolving (Martí-Vilar et al., 2021).

Although some authors defend that this theory does not adapt to children's and adolescents' complexity (Walker, 2006), nowadays a more up-to-date line exists that was developed by Beauchamp et al. (2013), Doolet et al. (2010) and Vera-Estay et al. (2014): the

so-called So-Moral. These authors developed updated versions of the original moral dilemmas, but in new formats like video, and kept the moral reasoning levels set out by Kohlberg (1981) as a means to analyze them. The moral dilemmas that center on the reality of those who face it are a relevant pedagogic tool to improve moral reasoning levels (Balakrishnan, 2011; Cortes, 2002).

The moral development of adolescence should lie at a preconventional or conventional level (Kohlberg, 1976; Rest, 1979), although a very important qualitative leap takes place between the first adolescence stage (12–14 years) and mid-adolescence (15–17 years) (Chiasson et al., 2017). Indeed, it is during this period when adolescents can develop concepts like justice, rights and social well-being (Jambon & y Smetana, 2015), especially when they become aware of the consequences of their actions and situations in which their help has a positive effect (Paciello et al., 2013). This means that moral reasoning is linked with personal values, direct influences (Schwartz, 2010; Paciello et al., 2013) and the social context (Myyry et al., 2013). For instance, those adolescents concerned about others' well-being are led to act and offer help directly or indirectly without minding the cost of their action (Paciello et al., 2013).

Conversely, those adolescents whose moral reasoning is linked with becoming powerful and dominating others condition their help to their own benefit without considering others' benefits (Schwartz, 2010; Paciello et al., 2013). Indeed, one of the conducts involved in acquiring power or domineering action, which is unfortunately experienced too often in adolescence, is cyberbullying, whose prevalence varies principally in accordance to the context, significantly influenced by cultural contexts and the perspective of measure (prevalence of victims, aggressors or both) obtaining a medium prevalence of 23.8% in Canada, 23% in China or 6.3% in Germany and 5.2% in Sweden (Brochado et al., 2017). Bearing this premise in mind, this experience can condition adolescents' moral development. Currently, several research works have managed to establish a direct and significant relation between moral reasoning and cyberbullying in adolescence (Lo Cricchio et al., 2021).

1.3. Cyberbullying and adolescent moral reasoning

Cyberbullying is understood as aggressive intentional action that an individual take against someone who cannot defend him/herself and is repeated in time using a device connected to the Internet (Patchin & y Hinduja, 2006; 2015; Smith et al., 2008), which usually emerges into the victim's immediate environment (Chan, Ahrumugam, Scheithauer, Schultze-Krumbholz, & Ooi, 2019; Ortega & Zych, 2016) such us schoolmates or friends. Cyberbullying effects on the psychological health of adolescent victims and its directly related with anxiety, depression, stress or frustration (Evangelio et al., 2022). There is a large number of possible representations of cyberbullying (Lucas et al., 2016), converting this phenomenon in one which is very difficult to identify. For example, when a single message becomes viral, its impact on the victim is considered cyberbullying, just as it would happen when the victim is repeatedly insulted or attacked (Patchin & y Hinduja, 2015).

One of the most frequent ones is sexting, which centers on the voluntary creation and delivery of text messages, photos, or videos with personal sexual content via the Internet or mobile phones without the permission to do so (Lounsbury et al., 2011). Sexting behavior prevalence among adolescents aged between 12 and 17 years is 15%, increasing to 36% at 17 years of age (Gámez-Guadix et al., 2017), however, other studies exposed that the 4.8% of adolescents engaged in abusive sexting behaviors, principally boys, and 4.3% are victims of non-consensual sexting behaviors (Barroso et al., 2021). This behavior is associated to traits of extraversion and conscientiousness and linked to impulsivity and lack of self-control and predicted by low agreeableness and higher neuroticism (Gámez-Guadix et al., 2017) and its related with behavioral and emotional problems callousness, experiences of neglect and abuse in childhood, and various forms of aggression (Barroso et al., 2021). However, other studies suggest that sexting behavior in girls were not related with psychological well-being changes, concluding that is part of sexual development during middle to late adolescence (Burić et al., 2021).

To prevent cyberbullying and its multiple representations, exist a large number of prevention strategies. The technological ones such us blocking the agresor or modifying passwords or choosing to don't react or avoid the cyberaggressor do not help to solve the case (De la Caba & López, 2013). Other strategies to overcome cyberbullying and awareness-raising against it, like training the cyber-coexistence (Ortega & Zych, 2016), and e-competencies (Cebollero-Salinas, Cano-Escoriaza, & Orejudo, 2022) talking with teachers or families, as well as seeking a positive social support (De la Caba & López, 2013; Ozansoy et al., 2018), are extremely effective to solve cases but, the education community must be coordinated (Bautista y Vicente, 2020).

The relation between cyberbullying and moral reasoning is linked with moral disconnection, which focuses on deactivating moral conduct to perform immoral actions without feeling any remorse (Bandura, 1990). This means that performing cyberbullying actions is related to high moral disconnection levels (Lo Cricchio et al., 2021; Paciello et al., 2020; Perren et al., 2012; Perren & Gutzwiller, 2012) because the cyberbully inhibits remorse about his/her victim, mainly because (s)he does not see how his/her action can make the victim suffer (Leduc et al., 2018; Wang et al., 2016). Conversely, low moral disconnection levels are directly linked with not performing cyberbullying actions because the harm they can cause to someone is well understood (Paciello et al., 2020; Perren & Gutzwiller, 2012; Wang et al., 2016).

If we bear in mind today's literature that links cyberbullying with an individual's moral reasoning, it is feasible to state that those people whose moral reasoning lies between conventional or postconventional levels (Kohlberg, 1976) will continue to be far-removed from carrying out such conduct (Paciello et al., 2020; Perren & Gutzwiller, 2012; Wang et al., 2016), while those whose moral reasoning lies at the preconventional level (Kohlberg, 1976) will be more prone to perform this conduct (Lo Cricchio et al., 2021; Paciello et al., 2020; Perren et al., 2012; Perren & Gutzwiller, 2012).

1.4. Objectives and hypothesis

In order to generate an online interaction area where CI emerges and intervenes, a large group of adolescents faced a moral dilemma set out in a presented cyberbullying case. To analyze if CI helps in the evolution of adolescents' moral development, and in line with the research work by Wang et al. (2016), the responses they gave with a higher moral reasoning level were expected to encourage solving the case. The two specific objectives that allow this goal to be met were:

1. From the moral perspective, check whether the answers that emerge in the successive phases of interaction within the CI platform are less complex in the first phase of individual resolution of the moral dilemma and increase in complexity and popularity once the participants interact with the rest and the IC acts by eliminating those responses that are not valued and accepted by all the participants until the end of the experiment. As a hypothesis, we propose that the responses of the individual phase will be less complex in moral reasoning than the responses of the group interaction phase and that as the interaction increases or the system eliminates less popular responses, the resulting ones will be better moral quality.
2. Analyze the responses given by the participants during the experiment by comparing the first and last responses given by each participant to find that the complexity and moral level of their responses positively evolves towards responses with a higher level of moral development that favor the resolution of the dilemma exposed. As a hypothesis, we expect to find that complexity and answer moral level evolves positively.

2. Material and method

2.1. Participants

In order to carry out this experiment, a public call was organized with all the education centers that wished to participate from the Spanish Autonomous community of Aragón where year 1 of Higher Secondary Education was taught. No direct data on age or gender were taken to guarantee maximum anonymity but, in the Spanish educational system, students access this educational level mostly at the age of 17. By gender, national statistics show very similar percentages, specifically, in 2018, 53.1% were girls and 46.9% boy (Ministerio de Educación y Formación Profesional, 2020). Nineteen different education centers showed an interest in participating in this study. The final sample included 793 students whose gender and exact age were not known at any time point. The sample was grouped into 33 classrooms from the 19 education centers in Aragón. The number of participants per class ranged from 14 in the smallest class, to 33 in the biggest class. The Thinkhub platform provides an answer for each participant in each question (3) and each phase (7), so, for each participant we obtained 21 responses during the experiment, becoming a total of 16.653 responses (between 2 and 174 words each one). As it was impossible to manage such a large sample and the 16.653 open responses given by the participants during the experiment (an extremely large number to perform the necessary qualitative classification analysis), a decision was made to randomly select 100 (2.100 responses) students of the 793 participants to analyze and classify the presented data. This number guarantees a sufficient representation of responses, providing sufficient statistical power for the viability of the study.

2.2. Instruments

2.2.1. Thinkhub, the CI tool

Based on the CI idea according to Woolley et al. (2010), researchers from both the University Institute of Research into the Biocomputing and Physics of Complex Systems (BIFI) of the University of Zaragoza (Spain) and the company Kampal Data Solutions created the Thinkhub tool to create high-quality solutions for problems using a model of successive social interactions (Orejudo et al., 2021, Unpublished results) and, by doing so, to avoid the most usual problems in these contexts (Toyokawa et al., 2019). To this end, the system employs different means (copy, user permutation, visualizing others' responses and eliminating responses) and seven phases that modulate work by starting with a first individual phase, followed by six interaction phases, which are briefly set out as follows:

- ***Phase 1***: students' individual participation in solving the task without the rest intervening
- ***Phases 2, 3, 4 and 5***: in these phases, students view the responses from the previous phase with four anonymous neighbors, and have the chance to see their responses, copy them, amend them or keep their own response. In this way, the ideas that have been, and will be, produced move through a network of users, are classified and obtain an internal popularity indicator based on the number of times they have been copied
- ***Phase 6***: in this phase, the system eliminates the less frequent responses to show only the top 10 responses. This allows the students to copy and amend a response or keep their own
- ***Phase 7***: in this last phase, the top 10 are updated and shown to all the students. On this occasion, the students can only copy responses or keep their own

This interaction model attempts to solve key questions: for instance, the fact that only the responses of the four neighbors (other four students participating in the experiment which are selected randomly by the AI) are shown in Phases 2, 3, 4 and 5 avoids information overload and, at the same time, allows information to spread all along the network (Orejudo et al., 2021, Unpublished results). Online interaction eliminates monopolized turns to speak (Mann & Helbing, 2017). The system also offers an indicator of

responses' popularity, which aims to come close to the prestige effect. The system can be executed in the background among Phases 2, 3, 4 and 5, and has an influence on creating a top 10 of responses in Phases 6 and 7.

The prestige effect, or presence of leaders, can lead to positive effects in interaction contexts by reducing and selecting information to favor a consensus by acting as a facilitator element. It can also lead to poor-quality responses, low diversity and a high confidence level when backed by leaders (Bernstein et al., 2018; Lorenz et al., 2011). In the Thinkhub case, prestige is generated from responses' popularity, and prestige is created according to how frequently responses appear in the network due to shared preparation, which is feasible in closed responses, or by copying other network members' responses, when prestige equals popularity.

Finally, Thinkhub modulates responses' heterogeneity with a popularity system. As from Phase 6, a gradual process that eliminates responses is generated according to their popularity. Previous studies (Orejudo et al., 2021 Unpublished results) have found a relation between popularity and quality. This means that its responses elimination process acts as the group's virtual moderator by transferring the previously described moderator role to AI (Bigham and Bernstein, 2018).

2.2.2. The moral dilemma on the thinkhub platform

The work team behind this study created a complex moral dilemma for adolescents by extrapolating what Woolley et al. (2010) did in their study to provide a wide range of responses and interactions to generate CI on the platform. Following recent recommendations about moral development research by means of real moral dilemmas that youths may be familiar with (Balakrishnan, 2011; Beauchamp et al., 2013; Cortes, 2002; Doolet et al., 2010; Vera-Estay et al., 2014), the students had to face an adolescent sexting situation which, as previously mentioned, is one of the most frequent types of cyberbullying in this life stage.

Apart from the interest that the cyberbullying construct arouses in adolescents and how close it is to them, the robust theoretical sustentation established by the link between higher moral development levels with not performing and preventing cyberbullying actions (Paciello et al., 2020; Perren & Gutzwiller, 2012; Wang et al., 2016) made this moral dilemma a suitable one to be assessed. The applied moral dilemma was as follows: a girl (Pilar) takes an intimate photo of herself for her boyfriend (Alex), who sends it to a male friend (Quino) for a challenge. Quino ends up posting it. Three questions are asked: a) What do you think about Pilar taking a photo and giving it to her boyfriend? b) Should Alex send the photo to Quino? c) Should Quino have posted Pilar's photo? These questions encourage the participants to answer with open texts and without word limits.

2.3. Procedure

In order to conduct the study, information about it was sent to all the education centers in the city of Zaragoza. The participating centers had to propose a teacher as the activity coordinator. During two sessions, the teachers were trained in the conditions of the experiment to be run, how the platform works and how to coordinate all the participating centers to perform the task in a synchronized manner. The families of the participants were informed by letter that the experiment was to take place, its interest as an academic activity, and the anonymity and voluntary conditions. If the families or youths did not wish to participate, they were not included. This was the only factor to determine if a participating student was included or excluded.

Finally, to guarantee the anonymity of participation and that the participants were actually the selected people, the 19 centers received a personalized password for each participant. This identifying password, which had to be used to be identified on the Thinkhub platform where the experiment took place, comprised six numbers to ensure user's anonymity. This number was formed by their school number, their classroom's number and each student's number. As a task design, students were invited to solve problems together and collaboratively and there was also a prize for the classes that provided the best answers, the most original and had the most popular answer. That is, there was a double combination of internal and external incentives to perform the task with the highest performance.

After determining which centers would participate, all the students were given the same time to connect to the Thinkhub platform by a computer. The experiment lasted 2 h, during which time two types of test had to be solved (1 h for each test). The first was related to mathematics, and the second to the aforementioned moral dilemma, which is the object of this work. The specific moral dilemma phase is set out in Table 1, along with time specifications and the characteristics of the Thinkhub interaction model.

2.3.1. Classifying the moral dilemma

To analyze the obtained responses, a classification system was created following the usual recommendations (Anguera Argilaga, 1986; Rodríguez, 2017). One of the authors of the study created a first classification system by analyzing the responses of 10 randomly taken participants (210 total answers) from 100 to perform the analysis with and specified general criteria:

Table 1
Summary of the phases and actions to perform in each phase.

Phase	Edit	Responses View	Copy	User permutation	Change responses	Responses ruled out	Time per phase
1	Yes	No	No	No	No	No	10:15–10:23
2	Yes	Captured neighbor	Yes	No	No	No	10:23–10:27
3	Yes	Real-time neighbor	Yes	Yes	No	No	10:27–10:34
4	Yes	Real-time neighbor	Yes	Yes	Yes	No	10:34–10:42
5	Yes	Real-time neighbor	Yes	Yes	Yes (with neighbors)	Yes	10:42–10:49
6	Yes	Top 10	Yes	Yes	Yes (with the Top10)	Yes	10:49–10:55
7	No	Top 10	Yes	No	No	No	10:55–11:00

- If in the same sentence the are two categories, considering the context and the circumstances, the one with the highest moral level will be the scored.
- If several units are found in the same sentence, considering the context, the dominant unit is chosen, which conditions or influences the behavior.
- When in any question, the answer refers to another question, it is categorized omitting what refers to the other question.

This analysis provided the following seven categories, which were defined following the theory by Kohlberg (1976):

- Level 1: preconventional phase, orientation toward disobedience and punishment: the immediate consequence of action is avoiding experiences linked with punishment and seeking to meet one's own needs. Brief examples: "she deserves it", "that's what you get for sending nudes".
- Level 2: preconventional phase, orientation toward one's own interest: defending personal interest and act in good conscience to achieve this. Brief examples: "As much as you trust, you don't know how it will end, it can be dangerous", "She doesn't know what her boyfriend can do".
- Level 3: conventional phase, orientation toward approval: own effort so that actions fit in with what society believes is good. Brief example: "It has to be consistent, but if there is trust, there is nothing to worry about", "I'm fine with that if you're confident enough, but I wouldn't do it".
- Level 4: conventional phase, orientation toward law and order: good and bad come from adhering norms or not. Brief examples: "It is a personal decision; she is aware enough of the consequences", "He has no authority over her"
- Level 5: postconventional phase, orientation toward social contract: reflecting on whether laws and norms are right for society or not. Brief example: "It is your right; you lose control of your privacy (loss of principle)".
- Level 6: postconventional phase, orientation toward universal-ethical principles: very abstract reasoning based on creating universal-ethical principles that differ from laws themselves. This moral level was not found and categorized in the whole experiment.
- Level 0, cannot be coded: a response that cannot be calculated: this category includes all the unclassifiable responses, like answers associated with topics not related with the sexting case (jokes about politics, videogames or other social media content) or the absence or answer. Brief examples: "no"; "I want to sell a car"; "I want to trade some Pokémons in Pokémon Go".

To test the classification system's reliability, categories were classified by two researchers within a 2-week time interval. Categorize each answer on each moral level defined by the categorization system was complex due to the variety of possible answers generated by the collective intelligence process. To categorize them it took in count short phrases and words which specified moral actions or reflections about the sexting behavior and its consequences on the fictional characters of the case. Every moral actions or reflection were linked to the different Kohlberg's moral levels. After the 2-week time interval, both researches compared the first 10 random participants answers an it's punctuation into the categorization system. Even with the complexity of the system and the answers, the obtained agreement was 0.849 according to the Kappa Index and it was not necessary to readjust the categorization system.

2.4. Data analysis

In order to verify if the responses that appeared in the successive interaction phases were more or less complex from the moral perspective, an analysis with three different phases was carried out. First a descriptive analysis was done of the responses that appeared in all the experiment's phases. As the Introduction indicates, every time that these responses emerged in the platform interaction context could be a previous condition for CI to appear.

Then to test our first hypothesis, we analyzed how responses evolved according to the contents in the different phases. To do so, we compared the values of the seven phases following an independent samples procedure by one factor Anova with the Brown-Forsythe test because groups' variances cannot be assumed. Finally, to check our second hypothesis, we compared the first response of the 100 participants obtained in the experiment (i.e., the individual participation phase), which corresponded to Phase 1, to the last given response (i.e., after finalizing all the interaction phases). To make this comparison, we used the Wilcoxon test for related samples. The employed software was SPSS, which offers these options in a block of non-parametric tests adapted with an ordinal measurement level,

Table 2
Descriptive results. Number of responses per phase.

Phase	Question 1		Question 2		Question 3	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
1	121	36.7	98	34.8	96	35.2
2	24	7.3	26	9.2	24	8.8
3	45	13.6	41	14.5	43	15.8
4	51	15.5	35	12.4	32	11.7
5	46	13.9	59	20.9	57	20.9
6–7	43	13.0	23	8.2	21	7.7
Total	330	100.0	282	100.0	273	100.0

like that set out with the moral dilemma analysis. In all the hypothesis tests, a critical value for rejecting the null hypothesis and committing a type 1 error of 0.05 is established. To know the size of the effect, eta squared has been calculated (Ellis, 2010). For the Wilcoxon test, the calculation procedure proposed by Lenhard and Lenhard (2016) has been used.

3. Results

Table 2 sets out the number of responses for all three coded questions. In the first question, 330 responses were obtained, with 282 in the second question and 273 in the third. The distribution of these questions in all the experiment's phases is noteworthy. In the first phase (the only phase involving individual work), the 100 participants produced more than one third of all the responses in each phase; that is, 121 responses in the first question, 98 in the second and 96 in the third. Yet in the other phases, the number of responses considerably lowered, and the fewest different responses appeared in Phase 2 (24, 26 and 24 responses, respectively, for each question), but a gradual increase was noted in the other phases, especially in Question 1. Finally, in the last phase, Phase 7, once again a different response pattern appeared for all three questions, with 43 in Question 1, and only 23 and 21 in Questions 2 and 3, respectively.

According to contents (Table 3), the most frequent responses were found in the intermediate/high moral reasoning values. Specifically, in Question 1, 56.4% of the responses appeared between levels 3 and 4, but this percentage in Questions 2 and 3 rose to 84.7% and 79.94%, respectively. It is highlighted that a percentage of responses between 11.2% and 8.9% could not be coded; in other words, responses not related to the moral reasoning task.

It is worth noting that the distribution of the responses in Question 1 was wider for both number of responses (Table 2) and the moral reasoning level (Table 3). It was also in this question where the responses with more moral complexity were found; i.e., level 5 moral reasoning (i.e., postconventional phase) with 5.5%. It was interesting to note that these 18 answers appeared in the last experiment phase. That is to say, these responses were not given by the 100 participants in the experiment's previous phases, but they were created using the generated dynamics and spread through the system to reach the participants, who then selected them in the last phase.

Table 4 shows the evolution of the type of responses per phase and moral reasoning level. Statistically significant differences were only found in Question 1, and were associated with being further spread and with the quality of the responses to this question. This meant that the mean values obtained in Phase 7 of this question increased compared to the other phases ($F_{5,324} = 6.805$, $p < .001$). For the other questions, no statistically significant differences were observed among phases.

The last analysis of the evolution of the moral reasoning levels that occurred throughout the experiment derived from the comparison of each participant's first response (Hypothesis 2). This comparison is shown in Table 5, which also presents the Wilcoxon rank test results. We can see that as a participant's reasoning level improved, we obtained a positive rank, a negative rank if it worsened, and a draw if it remained the same. In Table 5, statistically significant differences appear in Question 1 given that the number of positive ranks (i.e., participants whose moral reasoning level improved) is much higher (39) than that of the participants whose moral reasoning level becomes worse (7). Although the number of positive ranks for Questions 2 and 3 (23 and 24, respectively) was larger than the negative ranks (11 and 19, respectively), we cannot conclude that any improvement took place during this process. Graphically, this comparison is shown in Fig. 1, in which it can be seen how the greatest changes occur in the question 1.

4. Discussion

The aim of this work was to verify if the moral reasoning of adolescents who faced a moral dilemma could improve by the interactions generated by a platform designed according to CI principles. The results partially supported the considered hypotheses because, in at least one of the three questions that included the dilemma, we found differences while the experiment continued with participants' higher moral reasoning levels. This rise in their reasoning levels took place within the reference framework of Kohlberg's theory (1976). As all the participants were taken into account with the same question, their moral reasoning levels improved.

In order to explain these results, which favored the hypotheses set out in Question 1, but did not favor the results in Questions 2 and 3, we put forward two hypotheses. First, analyzing activity on the platform offers interesting information because in Question 1, the

Table 3
Distribution of responses per reasoning level.

Moral reasoning	Question 1		Question 2		Question 3	
Level	N	%	N	%	N	%
Cannot be coded	37	11.2	25	8.9	28	10.3
Level 1	7	2.1	0	0.0	2	0.7
Level 2	50	15.2	11	3.9	7	2.6
Levels 2-3	28	8.5	0	0.0	3	1.1
Level 3	52	15.8	96	34.0	78	28.6
Levels 3-4	44	13.3	57	20.2	59	21.6
Level 4	90	27.3	86	30.5	81	29.7
Levels 4-5	4	1.2	7	2.5	15	5.5
Level 5	18	5.5	0	0.0	0	0.0
Total	330	100.0	282	100.0	273	100.0

Table 4
Anova: Test Phase x Level of moral reasoning.

Phase	Question 1			Question 2			Question 3		
	N	M	SD	N	M	SD	N	M	SD
1	121	3.27	1.18	98	3.22	0.88	96	3.18	1.16
2	24	3.17	1.04	26	3.54	0.47	24	3.31	0.83
3	45	3.49	1.38	41	3.22	1.07	43	3.36	1.06
4	51	4.03	1.67	35	3.21	1.13	32	2.81	1.57
5	46	3.82	1.93	59	2.81	1.36	57	2.96	1.35
6–7	43	4.58	1.17	23	2.91	1.61	21	3.21	0.85
F Brown-Forsythe		6.805			1.972			1.211	
Sig		.000			.087			.338	
η^2		.099			.038			.021	

Table 5
Comparison of the First vs. Last response. The Wilcoxon signed rank test.

	N	Average rank	Sum of ranks	Z	Sig.	η^2
Question 1						
Negative ranks	7	17.64	123.50	−4.578	.000	.235
Positive ranks	39	24.55	957.50			
Draws	43					
Total	89					
Question 2						
Negative ranks	11	17.05	187.50	−1.930	.054	.019
Positive ranks	23	17.72	407.50			
Draws	53					
Total	87					
Question 3						
Negative ranks	19	23.50	446.50	−.328	.743	.001
Positive ranks	24	20.81	499.50			
Draws	42					
Total	85					

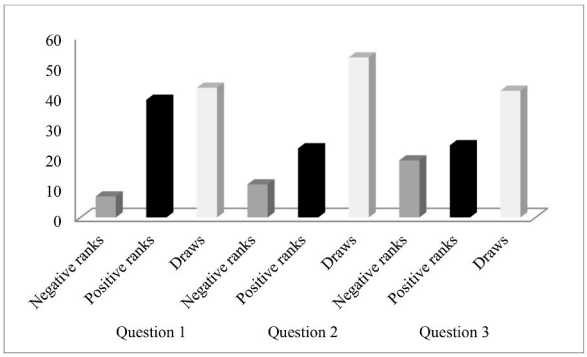

Fig. 1. Pre-post rank comparison.

participants performed more activity (330 different responses vs. 283 and 270 in Questions 2 and 3, respectively), as they did in all the experiment's phases, especially in the first individual phase and in the last two phases when the participants got to know the most popular 10 responses in the experiment. Their activity in the last phase should be the most important because, as previously indicated, this phase is when several participants are presented with more morally complex responses, which they could take as their own.

Along with this explanation about performing more activity in Question 1, which could be due to a contextual factor, i.e., it being the first question and given the task's design, the participants paid more attention. Thus, we could add a second explanation about the nature of the questions. Question 1 was also the question that resulted in more diverse responses at Kohlberg's levels, with 17.3% of responses at preconventional levels (levels 1 and 2) and 6.7% at postconventional levels (levels 4 and 4–5). In Questions 2 and 3, the responses concentrated much more at the conventional level, with only 3.9% and 3.3% of them at preconventional levels and 2.5% and 5.5% at postconventional levels, respectively. Indeed Question 1 was the only one with a response at a postconventional moral reasoning level (Level 5), where an adolescent would not normally be found according to the reference theory (Kohlberg, 1976; Rest, 1979). More specifically, the analysis of change showed that a change took place in the experiment in Question 1 with more

participants moving from one phase to another, but this situation did not occur in Questions 2 and 3. This means that the statement and the moral reasoning that youths worked on could have had a ceiling effect in Questions 2 and 3, which could have limited the results obtained with the experiment, whose possible responses were more closed and less complex than in Question 1, and would imply less engagement by the participants (Toyokawa et al., 2019).

This model also seems to limit the impact of certain phenomena that occur in interaction models with lots of participants, such as lack of responses, lack of original responses or obtaining extreme responses (Toyokawa et al., 2019). However, it ensures a good number of different responses and the rates at which the agents involved in the model participate tend to be high. Nonetheless, it can have perverse effects. For example, around 10% of the obtained responses could not be coded either because they lacked minimum information or were far-removed from the task to be performed. Such responses can be explained by the adolescents' anonymity context, who actually sought to achieve popularity by "injecting" into the network "funny" or "striking" responses. However, the task's demands and the generated dynamics meant that these responses were not majority responses.

An important aspect of the system is that in order to achieve a final consensus, it works with an answer selection system based on popularity. This mechanism can have a positive and a negative effect. On the one hand, it manages to reach said consensus in the most democratic way possible, above all by guaranteeing the distribution of information through the system, although it may also imply a certain imposition of responses on participants who do not make elections or who are very infrequent, generating certain tyranny of the elder. On the other hand, the system may also have a negative effect by eliminating high-quality responses. They are removed because they appear in the final interaction phases or are not popular enough. We have no data to confirm this statement, but we can back the quality of the responses that continue in the system and appear to be the most popular ones. Question 1 was the only question that obtained one response at the postconventional moral reasoning level (level 5) during the experiment, where adolescents do not tend to be found according to the reference theory (Kohlberg, 1976; Rest, 1979). This response ended the experiment being the top 1 of Question 1, and two factors backed this result.

The first one, which could have influenced the final responses of Questions 2 and 3, was adolescents facing a case that they might be familiar with (Balakrishnan, 2011; Cortes, 2002), and one that centered on a socially close context (Myyry et al., 2013), such as social networks and the harmful impact of sexting (Lo Cricchio et al., 2021). The second one focused on lack of moral disconnection with the cybervictim in the presented case (Lo Cricchio et al., 2021; Perren et al., 2012); that is, an empathic relation with the victim could be established by understanding his/her basic rights and the harm that (s)he suffered (Paciello et al., 2020; Wang et al., 2016). The other more frequent responses in all three questions were at a conventional level (Kohlberg, 1976). This was very positive for the adolescents, whose collective moral reasoning intended to reject any sexting action because it was illegal, antisocial and harmful (Wang et al., 2016) for the victim by empathizing with him/her (Paciello et al., 2020; Perren & Gutzwiller, 2012; Wang et al., 2016).

Although Kohlberg's theory (1981) has certain limitations because it centers on what is cognitive and not on what is emotional (Walker, 2006), the fact that a group of adolescents reached a postconventional moral development level when, according to Kohlberg (1981), adolescents should arrive at a conventional level at the most, implies that the CI which emerged during the experiment offered a very high potential for the evolution of moral development (Woolley et al., 2010) when such questions were answered.

This study is not free of limitations, which are mostly marked by experimental phase development. The sample initially included 1000 participants for the study at the macrolevel, but the problems with connecting to the platform that arose during the experiment meant having to lower this number to the cited 793 participants. Some education centers' poor Internet connections and mass access of the participants to the platform initially saturated the network, which prevented everyone not only connecting when the experiment began, but also being able to spend the time set to finish the first phase. It was also necessary to coordinate many different education centers to perform the task in a synchronized manner. However, loss of some participants should not affect the results because this loss did not affect the sample's characteristics, but could narrow down the network's size. No former studies with a large sample size were found to compare our work to.

We can also find some threats to the internal and external validity. On the one hand, the interactions are carried out in an online environment and anonymously, but the participants are in classrooms with other peers with whom they can make face-to-face interactions, for which it is difficult to know the impact on the results of the experiment. On the other hand, there is another important variable, the motivation of the participants to perform the task. In this case, the students were invited to solve problems together and there was also a prize for the classrooms that provided the best answers, the most original and had the most popular answer. This type of structure depends on the number of participants and their personal motivations, which should be controlled in future experiments.

Another limitation lies in the implications of learning with the experiment. That is, we assume that a learning process takes place when an adolescent prepares a response and writes it down. This is valid for the original responses that appear in the system, although copying processes may occur, which are particularly important in the final interaction phases. So, we can assume that when the interaction ends, the participants do not randomly select these responses, but a previous reasoning process comes into play that facilitates discovering better responses. However, this assumption should be investigated in the future considering if long-term changes remain by following qualitative methodologies to help to know which process youths follow during the experiment.

A last limitation is associated with the theoretical approach used. Although Kohlberg's theory can be considered a classic theory of moral development (Killen & Smetana, 2015), with certain limitations, it highlights the interactive and social role of reasoning. In this sense, it is assumed that people in a group can evolve their moral reasoning when faced with the point of view of other people, as has been shown in this study (Mercier, 2011). Moreover, the approach carried out in this study does not assume that the levels of moral reasoning reached must be present in all the situations that we could analyze, they have simply served us to be able to establish an analysis of the evolution of reasoning during the exposed interaction. Even when these approaches can support the use of Kohlberg's (1981) model, we cannot recognize any limitations of it. For example, the scarce emphasis placed on the emotional responses that could emerge in the process, and that other authors, such as Haidt (2001) or Walker (2006), highlight and that could well be a

necessary approach in future works on this subject. Proposed interaction model. A similar approach could be represented by Ferreira et al. (2021) or Tanner (2021). This last approach, likewise, highlights the role attributed to moral reasoning and moral judgment in behavior, also questioning the role attributed by cognitive-social theories to it, being recognized the criticism regarding whether reasoning is prior to the behavior or if, on the contrary, it is a product of it. The analysis of this approach, as well as other philosophical approaches to moral reasoning, such as those of Appiah (2008), offer interesting alternatives on this point, but which already exceed the limits of this work.

5. Conclusions

Thus, we conclude that the Thinkhub platform designed to generate CI may improve adolescents' reasoning levels. Several of the platform's design elements might explain this potential. On the one hand, the platform is based on a neighbor's interaction model that measures the quantity of the information presented to the participants. It also helps to diffuse information all along the network, as shown in the experiment's last phases when the highest quality responses reach the participants and they can select them. In the interaction context, this means that our experiment demonstrates that new responses are given by the participants at a rate of almost one response to every two phases. This interaction model seems to ensure a sufficient level of activity in the network, which is an implicit condition in CI models (Toyokawa et al., 2019) and other online collaborative learning contexts (Castellanos et al., 2017; Järvelä et al., 2021; Scheffel et al., 2017), which is an education context that shares certain premises with the herein presented CI model.

On the other hand, it also includes a responses selection module based on AI in Phases 6 and 7. It selects the most popular responses in the system and presents them to the participants, who are in charge of selecting which ones can better respond to the questions set out. This last mechanism of the system is based on two key principles of the "crowd intelligence" model: one is the moderator's role in guaranteeing the task being performed (Bernstein et al., 2018), while the other is the importance of social influence as a mechanism involved in the participants' decision making. It is true that this social influence is different in nature because, given the participants' anonymity, reaching a consensus by the participants' choices is the factor that determines responses' popularity.

Precisely thank to the anonymity, the link between the participants to seek the best possible answer to the problem and the action of the AI as a mediator of the group, they can avoid situations that lead to a tyranny of the majority or negative responses that do not help to overcome the conflict can become the most popular. The use of collective intelligence as a new educational methodology for solving highly complex problems, such as the moral dilemma presented in this study, can help adolescents to understand the risks involved of the abusive use of Internet, as well as its consequences, that can generate negative behaviors such as cyberbullying or to be exposed to practices derived from it, as is the case of sexting. Through collective intelligence and platforms like Thinkhub, it is possible to educate and train adolescents by confronting a large group of them with different real situations to which they may be exposed. As well as the adolescents can be educated, active and future teachers who must train them, can improve their ability to take the right decisions in the digital world against cyberbullying cases presented their classrooms thanks to collective intelligence learning.

Credit author statement

Pablo Bautista Alcaine, Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing. Jacobo Cano Escoriaza, Conceptualization, Validation, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Project administration. Eva Vicente Sánchez, Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing. Ana Cebollero Salinas. Conceptualization. Methodology, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing. Santos Orejudo Hernández, Conceptualization, Methodology, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision, Funding acquisition.

Funding and acknowledgements

Kampal Data Solutions; Research Project: Collective intelligence. Application in collaborative environments for problem solving. Advance Intelligence Chair United States Foreign Trade Institute. Gobernement of Aragon. Educaviva research group: S57_20R

References

Aggarwal, I., Woolley, A. W., Chabris, C. F., & Malone, T. W. (2019). The impact of cognitive style diversity on implicit learning in teams. *Frontiers in Psychology, 10*, 112. https://doi.org/10.3389/fpsyg.2019.00112

Anguera Argilaga, M. T. (1986). La investigación cualitativa. *Educar*, (10), 23–50.

Appiah, K. A. (2008). *Experiments in ethics*. Harvard University Press.

Balakrishnan, V. (2011). *Real-life dilemmas in moral education*. The University of Malaya Press. https://doi.org/10.3109/02699052.2013.775486

Bandura, A. (1990). Mechanisms of moral disengagement in terrorism. *Origins of terrorism: Psychologies, ideologies, states of mind, 28*(7), 161–191. https://doi.org/10.5860/choice.28-4136

Barroso, R., Ramião, E., Figueiredo, P., & Araújo, A. M. (2021). Abusive sexting in adolescence: Prevalence and characteristics of abusers and victims. *Frontiers in Psychology, 12*, 380. https://doi.org/10.3389/fpsyg.2021.610474

Bates, T. C., & Gupta, S. (2017). Smart groups of smart people: Evidence for IQ as the origin of collective intelligence in the performance of human groups. *Intelligence, 60*, 46–56. https://doi.org/10.1016/j.intell.2016.11.004

Bautista Alcaine, P., & Vicente Sánchez, E. (2020). How the education community perceives cyberbullying: A comparison of students, teachers and families. *Journal of New Approaches in Educational Research, 9*(2), 230–244. https://doi.org/10.7821/naer.2020.7.554

Beauchamp, M. H., Dooley, J. J., & y Anderson, V. (2013). A preliminary investigation of moral reasoning and empathy after traumatic brain injury in adolescents. *Brain Injury, 27*, 1–7. https://doi.org/10.3109/02699052.2013.775486

Bernstein, E., Shore, J., & y Lazer, D. (2018). How intermittent breaks in interaction improve collective intelligence. *Proceedings of the National Academy of Sciences of the United States, 35*, 8734–8739. https://doi.org/10.1073/pnas.1802407115

Bigham, J. P., Bernstein, M. S., & y Adar, E. (2018). Human-computer interaction and collective intelligence. In E. T. W. Malone, & M. S. Bernstein (Eds.), *Handbook of collective intelligence* (pp. 57–83). Cambridge, Massachusetts, London: The MIT Press.

Brochado, S., Soares, S., & Fraga, S. (2017). A scoping review on studies of cyberbullying prevalence among adolescents. *Trauma, Violence, & Abuse, 18*(5), 523–531. https://doi.org/10.1177/1524838016641668

Burić, J., Garcia, J. R., & Štulhofer, A. (2021). Is sexting bad for adolescent girls' psychological well-being? A longitudinal assessment in middle to late adolescence. *New Media & Society, 23*(7), 2052–2071. https://doi.org/10.1177/1461444820931091

Castellanos, J., Haya, P. A., & Urquiza-Fuentes, J. (2017). A novel group engagement score for virtual learning environments. *IEEE Transactions on Learning Technologies, 10*(3), 306–317. https://doi.org/10.1109/TLT.2016.2582164

Cebollero-Salinas, A., Cano-Escoriaza, J., & Orejudo, S. (2022). Are emotional e-competencies a protective factor against habitual behavior (media multitasking, cybergossip, phubbing) in Spanish students of secondary education? *Computers & Education, 181*, Article 104464. https://doi.org/10.1016/j.compedu.2022.104464

Chan, N. N., Ahrumugam, P., Scheithauer, H., Schultze-Krumbholz, A., & Ooi, P. B. (2019). A hermeneutic phenomenological study of students' and school counsellors' "lived experiences" of cyberbullying and bullying. *Computers & Education, 146*, Article 103755. https://doi.org/10.1016/j.compedu.2019.103755

Chiasson, V., Vera-Estay, E., Lalonde, G., Dooley, J. J., & y Beauchamp, M. H. (2017). Assessing social cognition: Age-related changes in moral reasoning in childhood and adolescence. *The Clinical Neuropsychologist, 31*(3), 515–530. https://doi.org/10.1080/13854046.2016.1268650

Cortes, A. (2002). La contribución de la psicología ecológica al desarrollo moral. Un estudio con adolescentes. *Anales de Psicología, 18*, 111–134. https://doi.org/10.6018/analesps

Curşeu, P., L., Pluut, H., Boroş, H., & Meslec, N. (2015). The magic of collective emotional intelligence in learning groups: No guys needed for the spell! *British Journal of psychology, 106*, 218–234. https://doi.org/10.1111/bjop.12075

Dai, A., Zhao, Z., Li, R., Zhang, H., & Zhou, Y. (2020). Evaluation mechanism of collective intelligence for heterogeneous agents' group. *IEEE Access, 8*, 28385–28394. https://doi.org/10.1109/ACCESS.2020.2971278

De Vincenzo, I., Giannoccaro, I., Carbone, G., & Grigolini, P. (2017). Criticality triggers the emergence of collective intelligence in groups. *Physical Review, 96*(2). https://doi.org/10.1103/PhysRevE.96.022309

De la Caba, M. A., & López, R. (2013). La agresión entre iguales en la era digital: Estrategias de afrontamiento de los estudian-tes de último ciclo de Primaria y del primero de Secundaria. *Revista de Educación, 362*, 247–272. https://doi.org/10-4438/1988-592X-RE-2011-362-160.

Ellis, P. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511761676

Engel, D., Woolley, A. W., Jing, L. X., Chabris, C. F., & Malone, T. W. (2014). Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face. *PLoS One, 9*(12), e115212. https://doi.org/10.1371/journal.pone.0115212

Evangelio, C., Rodríguez-González, P., Fernández-Río, J., & Gonzalez-Villora, S. (2022). Cyberbullying in elementary and middle school students: A systematic review. *Computers & Education, 176*, Article 04356. https://doi.org/10.1016/j.compedu.2021.104356

Ferreira, P. C., Veiga Simão, A. M., Paiva, A., Martinho, C., Prada, R., Ferreira, A., & Santos, F. (2021). Exploring empathy in cyberbullying with serious games. *Computers & Education, 166*, Article 104155. https://doi.org/10.1016/j.compedu.2021.104155

Gámez-Guadix, M., De Santisteban, P., & y Resett, S. (2017). Sexting entre adolescentes españoles: Prevalencia y asociación con variables de personalidad. *Psicothema, 29*(1), 29–34. https://reunido.uniovi.es/index.php/PST/article/view/11585.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Hjertø, K. B., & Paulsen, J. M. (2016). Beyond collective beliefs: Predicting team academic performance from collective emotional intelligence. *Small Group Research, 47*(5), 510–541. https://doi.org/10.1177/1046496416661236

Jambon, M., & y Smetana, J. G. (2015). *Moral development, theories of* (pp. 788–795). International Encyclopedia of the Social & Behavioral Sciences. https://doi.org/10.1016/b978-0-08-097086-8.23201-7

Järvelä, S., Malmberg, J., Sobocinski, M., & Kirschner, P. A. (2021). Metacognition in collaborative learning. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *Vol. 19. International handbook of computer-supported collaborative learning. Computer-supported collaborative learning series*. Cham: Springer. https://doi.org/10.1007/978-3-030-65291-3_15.

Killen, M., & Smetana, J. G. (2015). Origins and development of morality. In M. E. Lamb (Ed.) (7th ed.), *Vol. 3. Handbook of child psychology and developmental science* (pp. 701–749). Wiley-Blackwell.

Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel (Ed.), *Psychology and genetic epistemology* (pp. 151–235). New York: Academic Press. https://doi.org/10.1016/b978-0-12-498640-4.50011-1.

Kohlberg, L. (1976). Moral stages and moralization. In E. T. Lickona (Ed.), *Moral development and behavior. Theory, research and social issues. Nueva Cork*. Holt & Winston.

Kohlberg, L. (1981). *The meaning and measurement of moral development.* Worcester, MA: Cark University Press.

Kohlberg, L. (1989). In E. E. Turiel, I. Enesco, & I. y J Linza (Eds.), *Estadios morales y moralización. El enfoque cognitivo-evolutivo*. Madrid: Alianza Psicología (1989). El mundo social en la mente infantil (71-100).

Leduc, K., Conway, L., Gomez-Garibello, C., & y Talwar, V. (2018). The influence of participant role, gender, and age in elementary and high-school children's moral justifications of cyberbullying behaviors. *Computers in Human Behavior, 83*, 215–220. https://doi.org/10.1016/j.chb.2018.01.044

Lenhard, W., & Lenhard, A. (2016). *Computation of effect sizes*. Psychometrica. https://doi.org/10.13140/RG.2.2.17823.92329

Lo Cricchio, M. G., García-Poole, C., te Brinke, L. W., Bianchi, D., & y Menesini, E. (2021). Moral disengagement and cyberbullying involvement: A systematic review. *European Journal of Developmental Psychology, 18*(2), 271–311. https://doi.org/10.1080/17405629.2020.1782186

Lorenz, J., Rauhut, H., Schweitzer, F., & y Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences, 108*(22), 9020–9025. https://doi.org/10.1073/pnas.1008636108

Lounsbury, K., Mitchell, K. J., & y Finkelhor, D. (2011). *The true prevalence of Sexting.* Children Crimes Against Research Center, University of Hampshire. April 2011.

Mann, R. P., & Helbing, D. (2017). Optimal incentives for collective intelligence. *Proceedings of the National Academy of Sciences, 114*(20), 5077–5082. https://doi.org/10.1073/pnas.1618722114

Martí-Vilar, M., Escrig-Espuig, J. M., & Merino-Soto, C. (2021). A systematic review of moral reasoning measures. *Current Psychology*, 1–15. https://doi.org/10.1007/s12144-021-01519-8

Mercier, H. (2011). What good is moral reasoning? *Mind & Society, 10*(2), 131–148. https://doi.org/10.1007/s11299-011-0085-6

Meslec, N., Aggarwal, I., & Curşeu, P. L. (2016). The insensitive ruins it all: Compositional and compilational influences of social sensitivity on collective intelligence in groups. *Frontiers in Psychology, 7*, 676. https://doi.org/10.3389/fpsyg.2016.00676

Ministerio de Educación y Formación Profesional. (2020). *Igualdad en cifras MEFP 2020. Aulas por la igualdad.* Secretaría General Técnica.

Myyry, L., Juujärvi, S., & y Pesso, K. (2013). Change in values and moral reasoning during higher education. *European Journal of Developmental Psychology, 10*(2), 269–284. https://doi.org/10.1080/17405629.2012.757217

Orejudo, S., Cano, J., Salinas, A. B., Bautista, P., Clemente, J., Rivero, P., & Rivero, A.y T. A. (2021). *Evolutionary generation of Collective Intelligence in very large groups of students.* Unpublished results.

Ortega, R., & Zych, I. (2016). La ciberconducta y la psicología: Retos y riesgos. *Psicologia Educativa, 22*, 1–4. https://doi.org/10.1016/j.pse.2016.04.001

Ozansoy, K., Altınay, Z., & Altınay, F. (2018). Developing strategies to prevent "cyber-bullying". *Eurasia Journal of Mathematics, Science and Technology Education, 14* (5), 1925–1929. https://doi.org/10.29333/ejmste/8549910.29333/ejmste/85499

Paciello, M., Fida, R., Tramontano, C., & Cole, E.y C. L. (2013). Moral dilemma in adolescence: The role of values, prosocial moral reasoning and moral disengagement in helping decision making. *European Journal of Developmental Psychology, 10*(2), 190–205. https://doi.org/10.1080/17405629.2012.759099

Paciello, M., Tramontano, C., Nocentini, A., Fida, R., & y Menesini, E. (2020). The role of traditional and online moral disengagement on cyberbullying: Do externalising problems make any difference? *Computers in Human Behavior, 103*, 190–198. https://doi.org/10.1016/j.chb.2019.09.024

Patchin, J. W., & y Hinduja, S. (2006). Bullies move beyond the schoolyard. *Youth Violence and Juvenile Justice, 4*, 148–169. https://doi.org/10.1177/1541204006286288

Patchin, J. W., & y Hinduja, S. (2015). Measuring cyberbullying: Implications for research. *Aggression and Violent Behavior, 23*, 69–74. https://doi.org/10.1016/j.avb.2015.05.013

Perren, S., Gutzwiller-Helfenfinger, E., Malti, T., & y Hymel, S. (2012). Moral reasoning and emotion attributions of adolescent bullies, victims, and bully-victims. *British Journal of Developmental Psychology, 30*(4), 511–530. https://doi.org/10.1111/j.2044-835X.2011.02059.x

Perren, S., & y Gutzwiller-Helfenfinger, E. (2012). Cyberbullying and traditional bullying in adolescence: Differential roles of moral disengagement, moral emotions, and moral values. *European Journal of Developmental Psychology, 9*(2), 195–209. https://doi.org/10.1080/17405629.2011.643168

Rodríguez, R. M. (2017). Directrices para la construcción de sistemas de categorías válidos. *Apuntes de Psicología, 35*(1), 25–34. http://www.apuntesdepsicologia.es/index.php/revista/article/view/645.

Salganik, M. J., Dodds, P. S., & y Watts, D. J. (2006). Experimental study of inequalityand unpredictability in an artificial cultural market. *Science, 311*, 854–856. https://doi.org/10.1126/science.1121066

Scheffel, M., Drachsler, H., & de Kraker, J. (2017). Widget, widget on the wall, am I performing well at all? *IEEE Transactions on Learning Technologies, 10*(1), 42–52. https://doi.org/10.1109/TLT.2016.2622268

Schwartz, S. H. (2010). Basic values: How they motivate and inhibit prosocial behavior. In M. Mikulincer, & P. Shaver (Eds.), *Prosocial motives, emotions, and behavior: The better angels of our nature* (pp. 221–241). Washington, DC: American Psychological Association Press. https://doi.org/10.1037/12061-012.

Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & y Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry, 49*(4), 376–385. https://doi.org/10.1111/j.1469-7610.2007.01846.x

Toyokawa, W., Whalen, A., & y Laland, K. N. (2019). Social learning strategies regulate the wisdom and madness of interactive crowds. *Nature Human Behaviour, 3*(2), 183–193. https://doi.org/10.1038/s41562-018-0518-x

Turiel, E. (1966). An experimental test of the sequentiality of developmental stages in the child's moral judgments. *Journal of Personality and Social Psychology, 3*(6), 611–618. https://doi.org/10.1037/h0023280

Vera-Estay, E., Beauchamp, M., & y Dooley, J. (2014). Cognitive underpinnings of moral reasoning in adolescence: The contribution of executive functions. *Journal of Moral Education, 44*, 17–33. https://doi.org/10.1080/03057240.2014.986077

Walker, L. J. (2006). Gender and morality. In M. Killen, & J. Smetana (Eds.), *Handbook of moral development* (pp. 93–115). Mahwah, NJ: Erlbaum.

Wang, X., Lei, L., Liu, D., & y Hu, H. (2016). Moderating effects of moral reasoning and gender on the relation between moral disengagement and cyberbullying in adolescents. *Personality and Individual Differences, 98*, 244–249. https://doi.org/10.1016/j.paid.2016.04.056

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & y Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330*, 686–688. https://doi.org/10.1126/science.1193147

Woolley, A. W., & y Aggarwal, I. (2020). *Collective intelligence and group learning*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263362.013.46

Zhang, Q., & Zhao, H. (2017). An analytical overview of Kohlberg's theory of moral development in college moral education in mainland China. *Open Journal of Social Sciences, 5*, 151–160. https://doi.org/10.4236/jss.2017.58012