ORIGINAL ARTICLE

Bullying-Related Tweets: a Qualitative Examination of Perpetrators, Targets, and Helpers

Karla Dhungana Sainju[1] · Akosua Kuffour[1] · Lisa Young[1] · Niti Mishra[2]

Abstract
Bullying literature notes that aside from the dyadic relationship of target and perpetrator, there are other participant roles in the bullying process including those that reinforce the perpetrator and those that stand up for the target. Most examinations of bullying roles have relied on self-reported data, which suffer from key limitations such as response and recall bias. Twitter data provides a way to overcome these limitations and extend our current understanding of bullying roles. The current study provides one of the first qualitative examinations of tweets to analyze the disclosure and sharing of bullying-related online and offline episodes. Through a qualitative content analysis, the study examines 780 tweets to analyze the descriptions and characteristics of three participant roles: the perpetrator, target, and helper. The results provide multidimensional insights into the context and relationships between bullying roles. The results reveal that each of the bullying role players tweet to share varying perspectives and the discussions transcend beyond just online exchanges. The results also confirm that Twitter is used not only as a channel for bullying but also as a tool for connection between the different role players. Implications of how Twitter can be leveraged to promote anti-bullying initiatives to educate and inform users about bullying, while also helping build resilience and emotional regulation, are discussed. Additionally, the study also has implications for artificial intelligence and can help to build improved classifiers to detect bullying-related discourse and content online.

Keywords Bullying roles · Twitter · Social media · Qualitative analysis

Introduction

Bullying is defined as behavior that involves unwanted behavior, a power imbalance, and repetition (Olweus, 1993). Bullying can occur in many forms, including verbal, physical, social, emotional, and cyber bullying. The literature states that aside from the dyadic relationship of target and perpetrator, there are other participant roles in the bullying process. These include those that actively or passively reinforce the bullying behavior, those that directly or indirectly help the target, and those who may change their behavior according to the social context (Bellmore et al., 2015; Levy & Gumpel, 2018; Olweus & Limber, 2010; Salmivalli, Karhunen, et al., 1996; Salmivalli, Lagerspetz, et al., 1996; Salmivalli, 1999; Salmivalli, 2010; Wójcik & Flak, 2019; Xu et al., 2012). However, previous examinations of bullying participant roles have largely been limited to self-report surveys or interviews (Lambe et al., 2019; Salmivalli, 1999; Salmivalli, Karhunen, et al., 1996; Salmivalli, Lagerspetz, et al., 1996). Relying solely on self-report data to understand bullying roles has some disadvantages including several sources of bias. Response bias may lead to biased estimates of self-assessed behavior including social desirability bias, where respondents want to look good by providing a response that they think is socially or morally correct (Rosenman et al., 2011). Furthermore, this approach expects participants to recall behaviors or emotions retrospectively which can result in recall bias. It has been suggested that

✉ Karla Dhungana Sainju
 Karla.dhungana-sainju@ontariotechu.ca

 Akosua Kuffour
 akosua.kuffour@ontariotechu.net

 Lisa Young
 lisanichole.young@ontariotechu.net

 Niti Mishra
 nitimkc@gmail.com

[1] Faculty of Social Science and Humanities, Ontario Tech University, 2000 Simcoe St N, Oshawa, ON L1G 0C5, Canada

[2] Rotman School of Management, University of Toronto, 105 St George St, Toronto, ON M5S 3E6, Canada

the longer the recall period, the less accurate participants' responses may be (Stull et al., 2009). Self-report data can also be limited to small sample sizes and have a narrow focus based on the scope of the survey.

In the last decade, merging the fields of social science and computer science, researchers have begun to analyze bullying roles in online spaces such as Twitter through the use of machine learning algorithms (Bellmore et al., 2015; Chelmis et al., 2017; Chatzakou et al., 2017b; Dhungana Sainju et al., 2021; Xu et al., 2012). Twitter, a microblogging and social networking site launched in 2006, allows users to write out 280 character posts also known as "tweets" that appear on their timeline. Tweets can include links to websites and hashtags, words, or phrases preceded by a hash (#) sign, which helps to categorize the tweet and allow it to pop up on a Twitter search. Users can also follow other Twitter users to see their tweets and can "retweet" or share other users' tweets. Twitter data is primarily public-facing, and its design allows users to share their own experiences or follow a specific incident or topic in real-time. It is also one of the most popular social media platforms, available in 33 languages with over 330 million monthly active users worldwide (Clement, 2019). Anyone who is 13 years or older can open an account. Recent statistics indicate that about 80.2% of Twitter users worldwide are between 18 and 49 years old; 18–24-year-olds represent 25.2%, 25–34-year-olds make up 26.6%, and 35–49-year-olds were the largest age group of Twitter users at 28.4%. Those 50 and over account for 12% and 13–17-year-olds were the smallest age group representing 7.8% of all Twitter users (Tankovska, 2021).

While both traditional examinations and machine learning approaches expand our understanding of the different bullying roles, we still have limited contextual information about how and why these roles transpire in online spaces. To our knowledge, no qualitative research exists on descriptions of bullying episodes and roles using social media data from Twitter. Given the growth and impact of social media platforms, it is important to understand how bullying episodes and its' role players are described in online spaces. Additionally, since most prior research on bullying tends to focus on online spaces as a conduit for cyberbullying perpetration, it is equally important to understand how online platforms such as Twitter can also serve as therapeutic, cathartic, and sympathetic spaces where individuals can interact, support, and share information with others. Combining Twitter data and qualitative analysis will allow us to advance our contextual understanding of both online and offline bullying episodes and roles and could point to important considerations for leveraging Twitter to address the issue of bullying. Finally, this work also has implications for how natural language processing (NLP) and machine learning (ML) techniques can be used to better understand bullying and cyberbullying roles as well as build richer classifiers. The approach used in this study to identify bullying traces and bullying roles within tweets can help provide valuable context for machine learning and artificial intelligence engineers, in particular by helping to provide clarity in bullying-related terminology, recognizing abusive or bullying content, and accounting for context within the tweets to optimize bullying detection NLP models and ML classifiers (Vidgen et al., 2019).

Literature Review

Who are the Perpetrators?

According to a systematic analysis drawing on 30 years of bullying research, the typical bullying perpetrator has externalizing behaviors, negative attitudes and beliefs about others, trouble resolving problems, and is negatively influenced by peers (Cook et al., 2010). Kowalski et al. (2014) meta-analysis of cyberbullying research among youth indicates that those who cyberbully tend to also bully in face-to-face settings. Risky online behavior (Kowalski et al., 2014), limited self-control (Vazsonyi et al., 2012), aggressive offline behavior (Ang et al., 2011), and low self-esteem (Dhungana Sainju, 2020) are also linked to increased cyberbullying perpetration. Moreover, bullying perpetrators are not limited to youth; adults can also engage in bullying behavior in work, social, and familial settings (Bartlett & Bartlett, 2011; Piotrowski, 2016). Characteristics such as narcissism, Machiavellianism, verbal aggression, and domineering and self-centered behaviors are linked to adult bullies (Piotrowski, 2015, 2016).

What are the Risk Factors for Becoming a Target of Bullying?

Anyone can become a target of bullying; however, certain characteristics may increase the risk of victimization with bullying often targeted at minority groups. Studies indicate that individuals who identify, or are perceived, as lesbian, gay, bisexual, transgender, or questioning (LGBTQ) are twice as likely to get bullied on school property (Kann et al., 2018) and significantly more likely to get bullied or harassed online compared to non-LGBTQ youth (GLSEN, CiPHR, & CCRC, 2013). Additionally, perpetrators may also target overweight and obese children (Janssen et al., 2004), children with attention-deficit-hyperactivity disorder (Holmberg & Hjern, 2008), learning disabilities (Mishna, 2003), autism spectrum disorder (Zeedyk et al., 2014), intellectual disabilities (Christensen et al., 2012), and those who have lower levels of peer acceptance and perceived popularity (de Bruyn et al., 2010).

Who are the "Helpers" in a Bullying Incident?

According to Salmivalli (1999), individuals who support the target and stand up to the perpetrator are "defenders." Xu et al. (2012) introduced two additional helper roles within the social media context. A "reporter" is someone who shares information about a bullying episode but is not involved in the incident, and an "accuser" is someone who directly accuses a perpetrator in a social media post (Bellmore et al., 2015; Xu et al., 2012). The helper roles have also been described along a continuum and prior research point to fluidity in the roles based on the social and peer context (Levy & Gumpel, 2018; Olweus & Limber, 2010; Wójcik & Flak, 2019). Wójcik and Flak (2019) identified the role of "frenemy" as individuals who befriend targets when not in the presence of perpetrators or peers but ignore and disengage in social contexts such as the classroom. Additionally, Levy and Gumpel's (2018) study referred to the "help-seeker" as individuals who assist indirectly; these helpers may ask others to assist or may report the incident but do not directly challenge or approach a bullying perpetrator for fear of retaliation. Though the help-seekers' indirect behavior may not stop the aggression, their actions provide important support for the victim and their wellbeing (Levy & Gumpel, 2018). The literature examining offline helper roles is extensive compared to examinations of online defending. A recent systematic review by Lambe et al. (2019) of defending included a final selection of 130 articles on offline defending versus 25 articles on online defending. According to the review, assessments of online defending have almost exclusively relied on self-reports from youth, with 47% of online research relying on early adolescents, 42% focusing on high school students, and a few studies conducted with university student samples. The review results indicate that in both online and offline scenarios, the key characteristics of a defender included being female, having high empathy and low moral disengagement, being popular and well-liked by peers, and having perceived supportive relationships with their parents, teachers, and schools (Lambe et al., 2019). While these studies provide insight into the characteristics of online helpers, they are limited to self-report data and run the risk of both response and recall bias. Moreover, although Xu et al. (2012) identified two new roles related to online defending, little is known about the contextual nature of these roles as prior studies have focused on creating machine learning algorithms to automatically categorize these roles (Bellmore et al., 2015; Chelmis et al., 2017; Dhungana Sainju et al., 2021; Xu et al., 2012) but none have attempted to qualitatively examine the social media post itself to discern the characteristics of the helper's posts.

Where Does Bullying Behavior Occur?

Bullying behavior can occur in both in-person and online settings. Previous studies indicate that bullying among youth tends to happen most frequently in schools, in the classroom, playground, hallways, and cafeteria (Dake et al., 2003; Kartal & Bilgin, 2009; Olweus, 1991). Bullying can also take place at home. Up to 40% of children experience sibling bullying, a form of intrafamilial aggression, which may also increase their risk of being involved in peer bullying (Wolke et al., 2015). Cyberbullying through digital technology has also seen a rise in recent years. A 2018 report from the Pew Research Center found that 59% of US teens experienced some form of cyberbullying (Anderson, 2018). A global survey conducted across 28 countries found that 33% of parents reported knowing a child in their community who had been cyberbullied and 17% said their child had experienced cyberbullying (Newall, 2018). Social network platforms such as Twitter, Snapchat, Instagram, and Facebook have been reported as common venues for cyberbullying (UNICEF, 2019; Whittaker & Kowalski, 2015). Whittaker and Kowalski's (2015) study found that cyber targets were most commonly victimized via texting (56.8%), followed by Twitter (45.5%), Facebook (38.6%), Instagram (13.7%), and YouTube (11.4%). Among adults, bullying occurs most frequently at work. Statistics from the Workplace Bullying Institute indicate that 19% of adult Americans, or 60.3 million US workers, are affected by workplace bullying (Workplace Bullying Institute, 2020), and a 2018 survey of over 1800 Canadians found that more than half had been bullied or knew a co-worker who had experienced workplace bullying (Forum Research Inc., 2018).

Do Perpetrators, Targets, and Helpers Know Each Other?

Bullying research that focuses on face-to-face aggression often emphasizes bullying occurring in schools and operationalizes the perpetrator as someone known within the peer group. Similarly, research also indicates that the majority of cyberbullying takes place between classmates or known individuals (Felmlee & Faris, 2016; Newall, 2018), and cyberbullies also tend to engage in face-to-face bullying (Kowalski et al., 2014). However, with the growth of cyberbullying and the anonymity that online spaces offer (Hinduja & Patchin, 2014), it is also reasonable to expect that target-cyber bully relationships may extend beyond the peer group (Pyzalski, 2011). Findings from Pyzalski (2012) suggest that in addition to bullying targets known offline, online spaces also allow for the cyberbullying of strangers, vulnerable individuals (for example, the homeless or alcoholics), celebrities, and specific groups (for example, fans

of a certain band or football team). A qualitative analysis of cyberbullying-related tweets found that more than half (56.3%) included individuals known to the users, 29% had no clear subject, and 15% was referring to a celebrity or someone not personally known (McHugh et al., 2019).

Why Do Individuals Participate, Share, or Self-disclose About Bullying Experiences Online?

It is suggested that computer-mediated communication eliminates people's social presence cues, and as such, sharing on social networking sites may decrease apprehension about communication (Hunt et al., 2012). Cyberbullies may engage in the behavior due to the anonymity online platforms afford (Hinduja & Patchin, 2014). They may also be less empathetic (Steffgen et al., 2011) since they cannot see their targets' reactions and may feel less remorse (Slonje, Smith & Frisen, 2012). However, online platforms may also serve as a venue to discuss both offline and online bullying interactions. Findings from Dhungana Sainju et al. (2021) reveal that Twitter may serve as a space to validate a target's bullying experience and allow helpers, in particular reporters and defenders, to spread awareness and stand up against bullying behavior. Similarly, McHugh et al. (2019) found that information sharing and activism were themes found within cyberbullying related tweets.

Previous Twitter Studies on Bullying

The majority of prior studies examining bullying discourse on Twitter utilize machine learning algorithms to classify and detect bullying within tweets (Blanco et al., 2014; Chatzakou et al., 2017a, 2017b; Devaneyan, 2016). These studies highlight how machine learning methodologies can analyze big data to effectively detect bullying behavior on Twitter; however, they do not provide any contextual information on the bullying behavior. A handful of studies have extended this approach by employing machine learning to identify specific bullying roles within tweets including bullies, victims, defenders, reporters, and accusers (Bellmore et al., 2015; Chelmis et al., 2017; Chatzakou et al., 2017b; Dhungana Sainju et al., 2021; Xu et al., 2012). Other studies such as Calvin et al. (2015) analyzed hashtags associated with bullying-related tweets and found that it was used to discuss high profile suicides, current television programs, and anti-bullying campaigns. Based on our search of the literature, McHugh et al.'s (2019) study is one of the only qualitative examinations of bullying-related tweets; however, the focus of the study was on cyberbullying and the study did not specifically investigate characteristics of the different bullying roles. Rather, their qualitative analysis focused on a set of 300 cyberbullying-related tweets to identify themes within the tweets including the purpose, focus, tone, intent,

time frame, and the nature of URLs included with the tweets. These studies all serve to confirm the utility of Twitter data to expand our understanding of bullying and notably underscore the need to employ qualitative analysis to advance the literature on bullying roles and its characteristics.

The Current Study

The current study expands on previous examinations of bullying roles through the lens of social media data. Most of the extant literature on bullying roles and their characteristics rely on solicited and prompted self-report data and have key limitations. A systematic review conducted by Vessey and colleagues (2014) on published self-report measures of youth bullying found that the psychometric soundness was limited among most self-report instruments. As well, others have suggested that the reliability and validity of self-reported bullying roles may be limited as not all bullying participants may admit their bullying behavior or realize that their behavior constitutes bullying (Salmivalli, Karhunen, et al., 1996; Salmivalli, Lagerspetz, et al., 1996; Solberg & Olweus, 2003). This social desirability bias can confound study results by presenting inaccurate relationships between variables or altering prevalence rates (Van de Mortel, 2008). The stability of recall has also been questioned; Rivers' (2001) study on retrospective reports of school bullying found that while participants' memories were stable during a 1-year span, he cautions that over time there are reductions in the amount of detail that can be recalled. The characteristics of bullying roles may also be limited due to sample size, variation in bullying terminology, or if a participant's actions do not fit within a priori definitions and descriptions of bullying roles and behaviors (Jeffrey & Stuart, 2020; Modecki et al., 2014; Vivolo-Kantor et al., 2014).

By utilizing public Twitter data, the current study overcomes most of these self-report limitations by observing bullying discourse in real-time, on a larger scale, from a wide range of individuals, and allows us to gather information on bullying episodes that is unsolicited and unprompted (McCormick et al., 2017). Through a qualitative content analysis, the key goal of the study is to analyze descriptions of discrete offline and online bullying episodes provided in tweets and discern the description and characteristics related to three bullying participant roles: the perpetrator, target, and helper. To our knowledge, this study is one of the first of its kind to qualitatively examine bullying discourse and participant roles using Twitter data, reflecting characteristics of observational and ethnographic research. The current study is primarily exploratory. Based on our prior knowledge of bullying roles and characteristics and to validate and extend the literature, the current study is situated within the following broad research questions: (1) What is the context of

Table 1 List of keywords used to identify bullying-related tweets

Primary keywords	Secondary keywords
Bullied, bully, bullying, cyberbullies, cyberbully, and cyberbullying	Aggressive, anonymous, assault, bash, bashing, beat, bystander, coercion, defending, degrade, degrading, destroy, destroyed, embarrass, embarrassing, embarrassed, exclude, excluded, exclusion, Facebook, force, forced, gossip, gossiped, group text, group chat, harass, hit, hitting, humiliating, humiliated, ignore, ignored, ignoring, Instagram, insult, insulted, intimidate, intimidated, isolate, isolated, jealous, laugh, laughing, manipulate, manipulating, manipulated, mean, mislead, misleading, mock, mocked, mocking, online, perpetrated, perpetrator, pretend, pretended, rejection, reject, rejected, repeatedly, repeat, revenge, rumor, rumour, scare, scared, shove, shoved, shun, shunned, Snapchat, social media, spread, stressful, suicide, target, targeted, tease, teasing, teased, text, texted, threat, threatened, WhatsApp

the bullying participant's tweet? (2) What is the relationship between the bullying participant roles? (3) What characteristics of bullying episodes are mentioned within the tweets?

Method

Data

The data for the current study was retrieved via Twitter's streaming Application Program Interface (API), a free and automated retrieval service that allows access to up to 1% of the population tweets. The data for the current study was collected as part of a larger data collection (see Dhungana Sainju et al., 2021). This section describes how the larger dataset was collected and the qualitative analysis section explains how a subset of randomly selected tweets was qualitatively analyzed for the current study. Tweets between August 7, 2019, and March 31, 2020, were collected using a list of primary keywords "bullied, bully, bullying, cyberbullies, cyberbully, and cyberbullying." Next, a list of secondary keywords was applied to the tweets to further refine the dataset. See Table 1 for a full list of keywords. Only tweets that matched both a primary and secondary keyword(s) were retained. The dataset was cleaned to remove the data of noise and spam accounts by removing re-tweets, non-English tweets, tweets with six or more hashtags, and tweets that only contained a web address. A total of 847,548 tweets were collected after the keyword filtering and data processing.

Next, logistic regression and support vector machines (SVM) machine learning algorithms and TF-IDF based natural language processing methods (Pedregosa et al., 2011; Xu et al., 2012; Zhu & Goldberg, 2009) were applied to the tweets to classify them into a bullying trace, which represented any tweet where the author participated in or mentioned a discrete bullying episode (Bellmore et al., 2015; Dhungana Sainju et al., 2021; Xu et al., 2012). Tweets were taken at face value and did not adhere to the traditional definition of bullying which tends to include an imbalance of power and repetition (Olweus, 1993). Rather, tweets were classified as bullying traces based on first-hand experiences and interpretation of the tweet authors themselves. Tweets were not categorized as bullying traces if they included a news headline that was copy and pasted without any additional original content or commentary, if it referred to an episode that sounded like bullying but was not explicitly defined as bullying by the author, if it referred to a future event, or if it was referring to an opinion about bullying rather than a discrete bullying episode. This resulted in 240,018 or 28.58% of the sample being identified as a bullying trace. See Dhungana Sainju et al. (2021) for full data processing and machine learning procedures.

Tweet Author Roles

Tweets identified as bullying traces were further categorized to identify the role of the tweet authors. Guided by Salmivalli (1999), Xu et al. (2012), and Bellmore et al. (2015), each tweet was classified as being posted by one of the following. A "target" referred to an episode where they were currently being bullied or had been bullied in the past. A "perpetrator" engaged in past or ongoing bullying. A "defender" stood up against a perpetrator, a "reporter" shared information about a bullying episode but was not involved, and finally, an "accuser" accused someone of bullying but did not identify himself or herself as a target, defender or other role. For the current study, the roles of defender, reporter, and accuser were included in one category identified as a "helper."

Qualitative Analysis

For the current study, a qualitative content analysis was utilized to examine the characteristics of tweets from three bullying roles: the perpetrator, target, and helper. Content analysis tends to focus on the content or contextual meaning

of texts including oral, print, or electronic texts (Kondracki & Wellman, 2002). The use of public tweets represents a purposeful random sampling strategy where those who are knowledgeable or have experienced a specific phenomenon of interest are targeted (Palinkas et al., 2015). The keyword selection and tweet classification were used to categorize the tweet author roles and a random sample of tweets for each role was selected to be qualitatively analyzed. For the first step, using a directed content analysis approach (Hsieh & Shannon, 2005) and guided by the broad research questions noted above, the study authors read through tweets within each role to identify key categories and develop an initial coding scheme. During this stage, related conceptual categories were found within the tweets in all bullying roles including the timing and location of the reported bullying episodes and the relationship between bullying roles.

Next, two of the study authors independently coded a set of 25 randomly selected tweets for each category according to the initial coding scheme. After coding, the two sets of labels were compared to examine the level of agreement between the coders. During the recontextualization phase, the labels were reviewed to ensure that all tweets fit into the identified categories and codes and the aims of the study (Bengtsson, 2016). Coding categories were discussed at length to amend and add in new codes after the initial round of coding. A codebook, which cataloged and defined the categories and codes, was created to ensure reliability among the coders during the categorization phase (Bengtsson, 2016). To further increase the stability and reliability of the coding process, two additional rounds of coding were conducted with new randomly selected tweets until a final coding scheme was established and all coding categories had a high level of agreement, exceeding 80% agreement or more. The categorization process revealed several themes about the key characteristics related to perpetrators, targets, and helpers. See Table 2 for the list of categories and codes derived for each of the roles.

Once the final coding scheme was created, we randomly selected a new set of 260 target tweets, 260 helper tweets, and 260 perpetrator tweets from tweets that were identified as bullying traces for a total of 780 tweets to be included in the final content analysis sample. The tweets were divided between the two coders who utilized the qualitative software program Dedoose (Dedoose, 2018) to conduct the qualitative analysis. Once all the tweets were labeled, the coded tweets were reviewed to ensure agreement with the coding. Qualitative content analysis allows for the quantification of categories and codes which helps to further highlight the phenomena being examined (Bengtsson, 2016). Thus, the results section below highlights the categories, codes, and frequencies within each set of bullying role tweets.

Results

Perpetrator Characteristics

Analyzing the context of the tweets posted by a perpetrator revealed that almost one-third (31.92%) were engaging in or discussing bullying someone in the tweet, for example, "We need to keep bullying her until she logs off," "That's how I'm having fun, you are meant to be bullied," and "I bully all the first years in my dorm. I'm so mean but they are stupid."[1] Almost a quarter (23.46%) were admitting that they had bullied someone in the past: "me and these kids bullied this homophobe in gym class," "I used to bully people bc I knew I can't get beat up," and "I bullied him to the point where it started to spread it is what he deserved." A total of 12.30% were apologizing or expressing remorse for past bullying behavior: "I personally was very homophobic when I was in the closet. I used the f word and bullied other gays because I couldn't accept that about myself" and "I know I bully you sometimes and I wanna say I don't mean it so sorry." Exactly 8.46% was denying involvement in a bullying episode, writing tweets such as "@user why are you accusing me of bullying when I'm defending someone being hounded on social media" and "I would never actually bully anyone, I was just playing but she thought I was mean. Another 8.46% were expressing intent to harm seen in tweets such as "you think you can bully me? Just for this threat I will bully you twice as hard." Finally, 15.38% of tweets did not have enough information to ascertain the context of the post.

Almost 7 out of 10 (65.76%) perpetrators knew their targets, directly mentioning their user name or referring to the target. In 25% of the tweets, there was no mention of a target so the relationship was unclear and about 1 in 10 (9.61%) tweets were referring to a celebrity or public figure. Analysis of where the perpetrator engaged in the behavior revealed that a majority of the tweets (61.92%) did not mention a specific location. Almost one-third (31.92%) mentioned bullying online, and small percentages noted bullying someone at home (2.3%) and school (3.84%). Almost 4 in 10 (37.69%) perpetrators were referring to an ongoing bullying incident, while 37.30% referenced a past incident. A quarter of perpetrator tweets did not have enough information to determine if it was a past or ongoing incident. See Fig. 1 for the perpetrator tweet characteristics.

---

[1] To avoid traceability of an example tweet, all user names were removed and tweets have been shortened instead of including it verbatim.

Table 2 Categories and codes for each bullying role

Perpetrator characteristics	
What is the context of the perpetrator's post?	Apologizing or expressing remorse
	Expressing intent to harm
	Engaging in bullying: direct engagement of bullying in the tweet
	Past bullying: admitting to having bullied in the past
	Denial: denying involvement in a bullying episode
	Not clear or known
Does the perpetrator know the target?	Knows the target: direct mention or reference to a target
	Not known: no specific target mentioned
	Not personally known: referring to celebrities or public figures not known in real life
Is there any mention of where the bullying is occurring?	School
	Home
	Work
	Online
	No specific location mentioned
Is it an ongoing or past event?	Past: tweet refers to a bullying incident that has already occurred
	Ongoing: tweet referring to bullying that is currently happening
	Not known: tweet does not include a time frame
Target characteristics	
What is the context of the target's tweet?	Sharing own bullying story: tweeting about the bullying experience(s) they encountered
	Sharing own story to console: telling their own story to make someone feel better, sometimes in response to a specific incident
	Call out a perpetrator
	Not know or not clear
Does the target know their perpetrator?	Knows the perpetrator: direct mention or reference to perpetrator
	Not mentioned but known: no specific mention of a perpetrator but it is clear that the target knows the perpetrator
	Not known
	Not personally known: referring to celebrities or public figures not known in real life
Does the target identify why they are being bullied/were bullied?	Demographic characteristics: e.g., sexual orientation, age, race, gender
	Disability: e.g., autism, mental illness, learning disability
	Physical looks: e.g., body size, specific physical feature
	Engagement in a particular activity: e.g. gamer, vegan, nerd, geek
	Stan/fandom: e.g. K-pop, Ariana Grande fans, Beyonce fans
	Not known: no mention of why they are/were being bullied
Is there any mention of where the bullying is occurring?	School
	Home
	Work
	Online
	No specific location mentioned
Is it an ongoing or past event?	Past: tweet refers to a bullying incident that has already occurred
	Ongoing: tweet referring to bullying that is currently happening
	Not known: tweet does not include a time frame
Helper characteristics	
What is the context of the helper's post?	Reporter: sharing information about a bullying episode not involved in
	Defender: standing up against a perpetrator
	Accusing: accusing someone of bullying
	Helper turned bully: helper engaging in bullying-like behavior through aggressive language
Does the helper know the target?	Knows the target: direct mention or reference to a target
	Not known: no specific target mentioned
	Not personally known: referring to celebrities or public figures not known in real life

Table 2 (continued)

Perpetrator characteristics	
What is the context of the perpetrator's post?	Apologizing or expressing remorse Expressing intent to harm Engaging in bullying: direct engagement of bullying in the tweet Past bullying: admitting to having bullied in the past Denial: denying involvement in a bullying episode Not clear or known
Does the helper identify why the target was/is being bullied?	Demographic characteristics: e.g., sexual orientation, age, race, gender Disability: e.g., autism, mental illness, learning disability Physical looks: e.g., body size, specific physical feature Engagement in a particular activity: e.g., gamer, vegan, nerd, geek Stan /fandom: e.g., K-pop, Ariana Grande fans, Beyonce fans Not known: no mention of why they are/were being bullied
Does the helper know the perpetrator?	Knows the perpetrator: direct mention or reference to perpetrator Not known: no specific perpetrator mentioned Not personally known: referring to celebrities or public figures not known in real life
Is it an ongoing or past event?	Past: tweet refers to a bullying incident that has already occurred Ongoing: tweet referring to bullying that is currently happening Not known: tweet does not include a time frame

Target Characteristics

The analysis of target tweets found that more than one-third (36.22%) were tweeting to share their bullying story. For example, "During high school I got bullied and beat up for being queer," and "I have a bow cut, glasses that are too large and overweight and I am bullied for being weird so I have to pretend to like things I don't to try and fit in." Another 27.95% were calling out a perpetrator, "This has been an ongoing issue for some time now and I have sat by while it continued. I refuse to be treated like that and watch others be treated the same," and "@user I hope you are punished for making all the lies about me. This is targeted harassment and has been 6 months of online abuse and bullying." About 7.08% were sharing their own story to console someone going through similar struggles. Tweets such as "Look people bullied n hit me, n I didn't wanna go to school, but I stayed positive and kept trying. Trust me it will get better," and "I saw your video @user and I cried, I'm a mama of 2 and I got bullied a lot. Don't let haters win. You're amazing." Three in 10 target tweets (28.74%) did not contain enough information for us to discern the context of the tweet.

A majority of target tweets (64.56%) did not identify the reason for being bullied. However, 11.81% of tweets suggested being bullied due to engagement in a specific activity such as gaming, for being a vegan, "nerd," or a "geek." Another 7.4% of tweets pointed to their physical looks as the reason they were targeted. Demographic characteristics such as someone's sexual orientation, race, and age were mentioned in 7.08% of tweets. A disability such as autism or mental illness was noted in 6.29% of tweets, and finally,

a small percentage (2.75%) pointed to bullying within fandoms.

Half (52.73%) of the targets directly mentioned or referenced their perpetrator in their tweets. Three in 10 target tweets made it clear that they knew their perpetrator but did not specifically mention them. In 15.62% of targets' tweets, it was not clear if the perpetrator was known and 1.56% of tweets suggested that the perpetrator was not someone they knew but rather a celebrity or a public figure. A total of 40.31% of targets did not specify a bullying location. However, more than a quarter (28.29%) mentioned being bullied online, and another 27.13% mentioned being bullied in school. A small proportion of targets pointed to bullying at work (2.32%) and at home (1.93%). Lastly, 67% of target tweets were referring to a past bullying experience, while more than a quarter (27.19%) mentioned an ongoing bullying episode. Exactly 5.43% of target tweets made no time-frame reference. See Fig. 2 for target tweet characteristics.

Helper Characteristics

An examination of the helper tweets found that a little more than one-third (36.23%) were tweeting to defend a target of bullying, for example, "I saw a video trending online where a couple was being bullied. That's wrong in many ways. However much we practice hate, let's not forget we're all human" and "@user sorry you're having to deal with this bully. Just be thankful that you are an amazing person and will never be a person like her who says mean things to make themselves feel better" About one-third (31.52%) of helpers were reporting about a bullying incident that they were not involved with, including both high-profile incidents reported
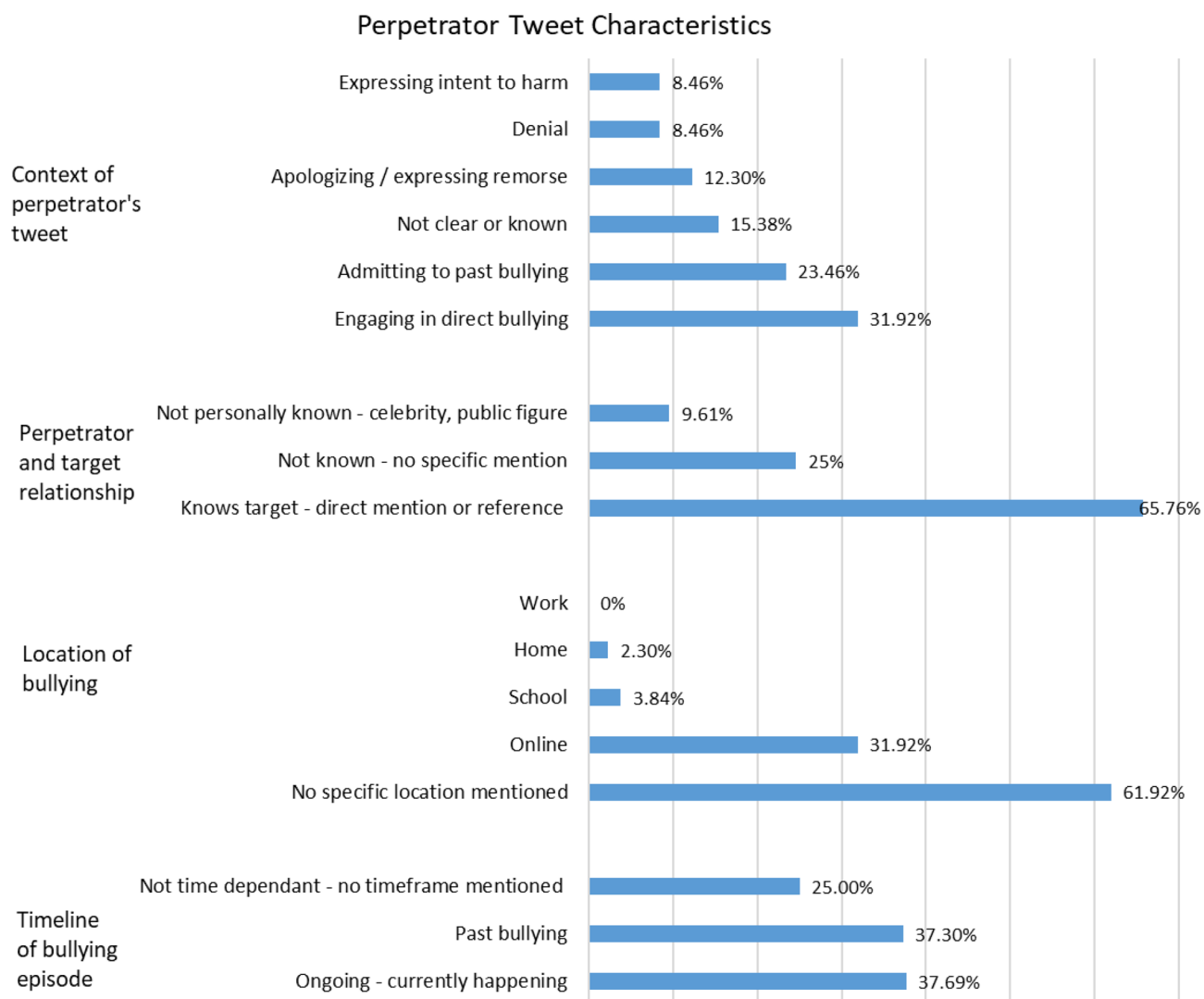
Perpetrator Tweet Characteristics

Fig. 1 Perpetrator tweet characteristics

in the news and incidents they were personally familiar with. For instance, "This is heartbreaking to watch a 9 year old to the point he wants to commit suicide" and "we had kids at my school commit suicide over bullying. It was never talked about." A little over a quarter (26.81%) were accusing someone of being a perpetrator; "you sound like an ugly bully @user and I doubt he's your only target." The analysis also showed that a small percentage (5.43%) of helpers used aggressive language while standing up for a target, potentially engaging in bullying-like behavior themselves through tweets such as "@user I really want to beat her up for you I hate her. She's a big ass bully" and "@user stop cyberbullying I'll call my attorney to sue your ass you come after my daughter again and I will destroy you."

Helpers made more direct references to perpetrators compared to targets. A little more than half (55.83%) directly mentioned a perpetrator by their user name or referred to them in the tweet. One-fifth (20.49%) of helpers referred to a perpetrator who was a celebrity or public figure. Meanwhile, 23.67% of helper tweets did not mention a specific perpetrator so it was unclear if they knew the perpetrator. With regard to the helper-target relationship, 37.23% specifically mentioned a target in their tweet, while 37.94% did not note any specific target. About a quarter (24.82%) referred to targets that were a celebrity or a public figure. About 7 in 10 helper tweets did not mention a specific reason why the target was being bullied. The remaining tweets were divided between being targeted for fandom involvement (6.04%), physical looks (7.11%), engagement in a particular activity (4.98%), demographic characteristics (4.98%), and disabilities (4.27%). Analyzing the time reference of the helper tweets indicates that 48.76% was referring to a past bullying incident, while 34.27% was referring to an ongoing incident. 16.96% of
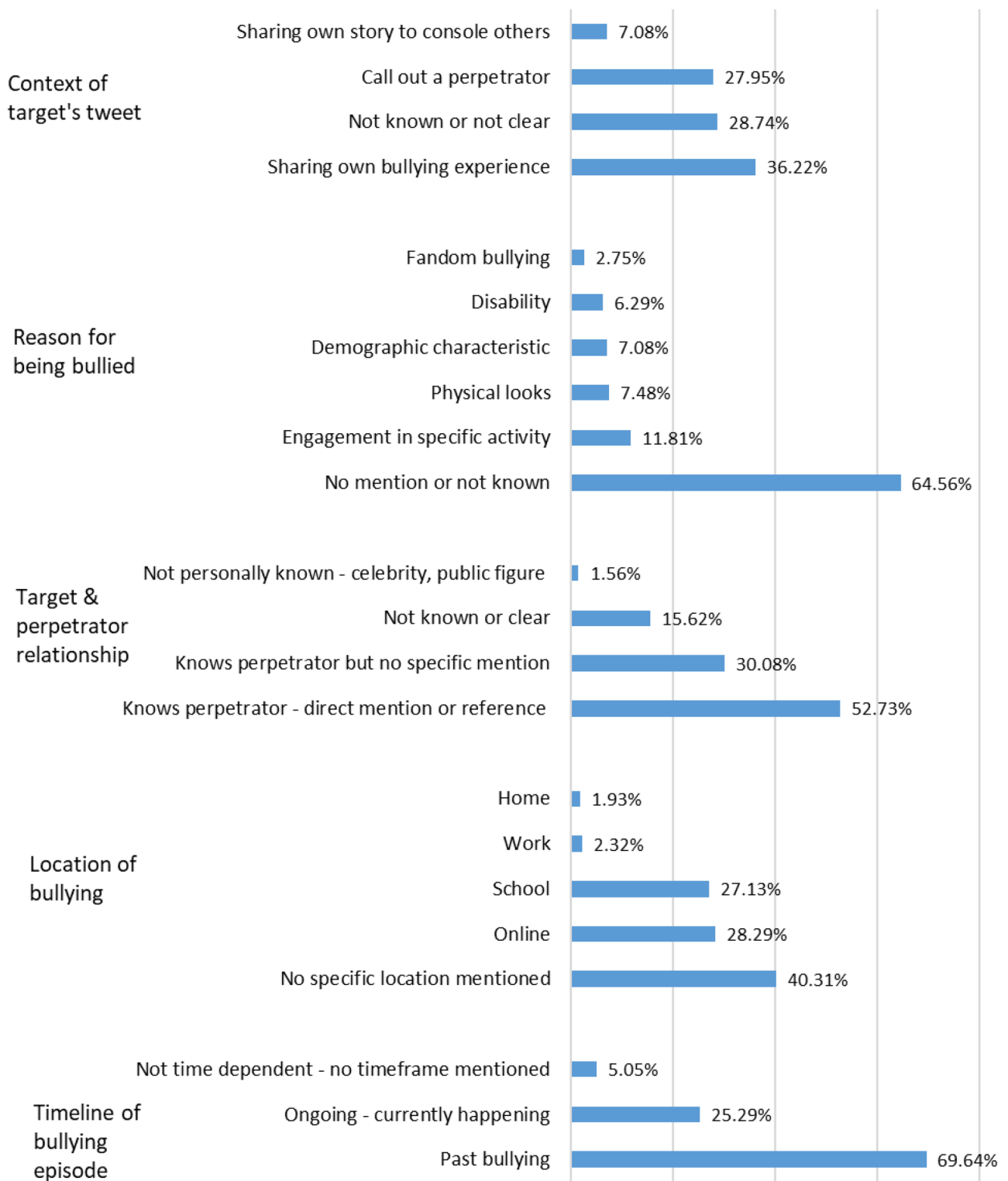
Fig. 2 Target tweet characteristics

Helper Tweet Characteristics

Helper tweet context
- Helper becoming perpetrator through aggressive language — 5.43%
- Accuser — 26.81%
- Reporter — 31.52%
- Defender — 36.23%

Helper & perpetrator relationship
- Not personally known - celebrity, public figure — 20.49%
- Not known - no specific mention — 23.67%
- Knows perpetrator - direct mention or reference — 55.83%

Helper & target relationship
- Not personally known - celebrity, public figure — 24.82%
- Knows target - direct mention or reference — 37.23%
- Not known - no specific mention — 37.94%

Reason target is being bullied
- Disability — 4.27%
- Engagement in specific activity — 4.98%
- Demographic characteristics — 4.98%
- Fandom bullying — 6.04%
- Physical looks — 7.11%
- No mention or not known — 72.59%

Timeline of bullying episode
- Not time dependent - no timeframe mentioned — 16.96%
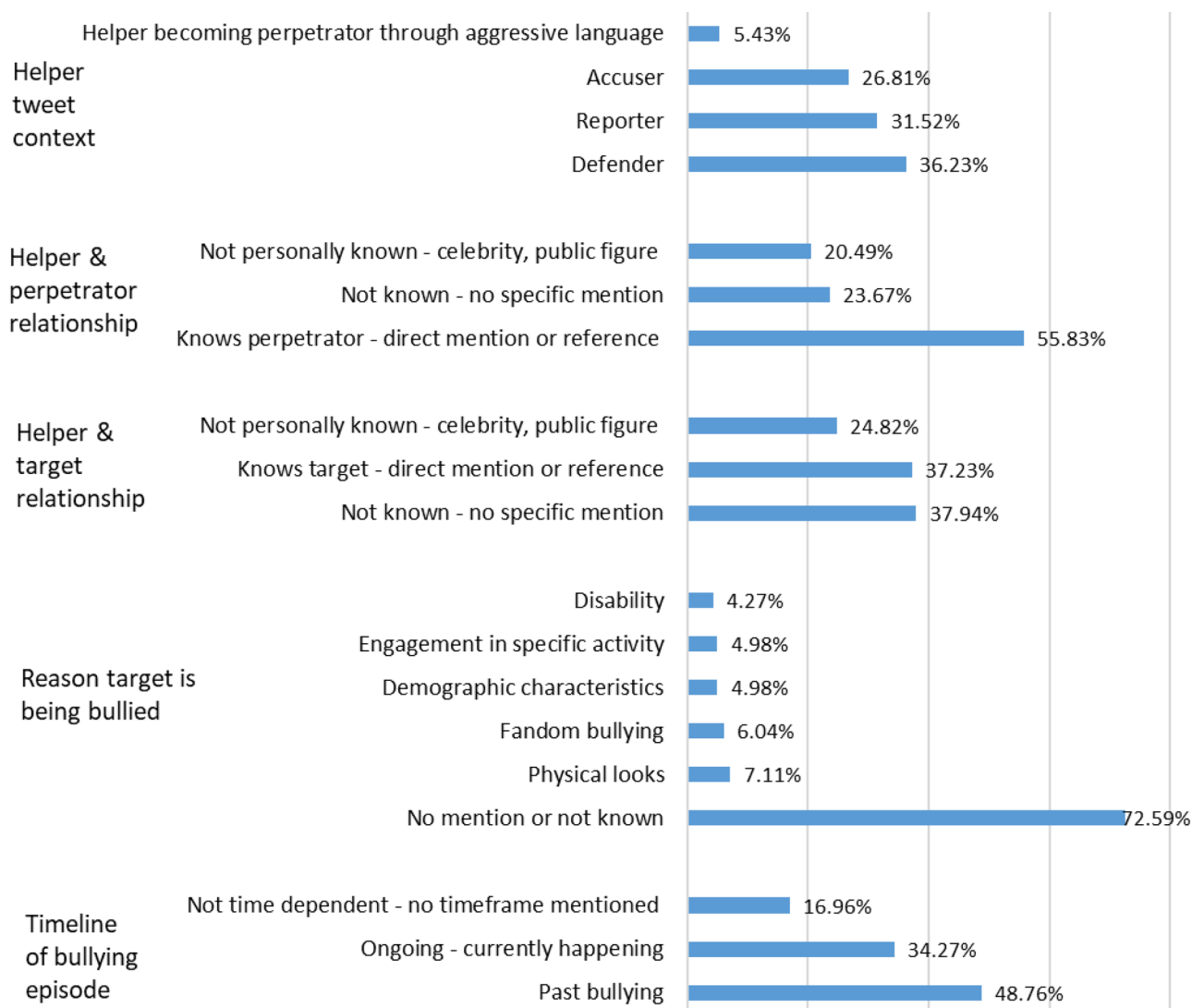- Ongoing - currently happening — 34.27%
- Past bullying — 48.76%

Fig. 3 Helper tweet characteristics

helper tweets made no timeframe reference. See Fig. 3 for helper tweet characteristics.

Discussion

This study provides one of the first qualitative examinations of bullying discourse and characteristics of bullying roles using Twitter data and advances the current literature in three primary ways. First, the study utilized unprompted and unsolicited social media data which addressed key issues related to self-report data, most notably social desirability bias and recall bias. Secondly, the results provide multidimensional insights into the context and relationships between bullying roles. By conducting a qualitative content analysis of bullying-related tweets, the results reveal that each of the bullying role players tweets to share varying perspectives and the discussions transcend beyond just online exchanges. The tweets suggest that offline experiences are intertwined with digital interactions; most bullying role players know each other, they are tweeting about current and past episodes of bullying, and the tweets highlight varied reasons for being bullied. Finally, the results confirm that Twitter is used not only as a channel for bullying but also as a tool for connection between the different role players. This suggests that Twitter can be leveraged to promote anti-bullying initiatives to

educate and inform users about bullying, while also helping build resilience and emotional regulation.

Twitter provides bullying participants with a computer-mediated communication platform to navigate different contexts, and this was reflected in the content of the tweets. Our findings support the notion that perpetrators are aware and intentional about their behaviors (Olweus, 1993); a third of perpetrators were tweeting to bully others and a little more than a quarter were admitting to past bullying. Prior studies on bullies and emotions also suggest that perpetrators tend to express less empathy (Hymel et al., 2010) and cyberbullies have been found to express less remorse than traditional bullies (Slonje, Smith & Frisen, 2012). Yet, our findings demonstrate self-reflection and awareness among some perpetrators evident through tweets where they took responsibility for their actions and expressed remorse. Studies examining victims' willingness to self-disclose experiences of bullying indicate that most feel reluctant in face-to-face settings to tell anyone due to shame and embarrassment (Menesini & Camodeca, 2008), fear of retaliation (Slee, 1993), or a belief that authorities, such as their teachers, tolerate the behavior (Unnever & Cornell, 2004). Our analysis, however, reveals that rather than suppressing victim voices and discouraging self-disclosure, Twitter is amplifying their voices. Targets are speaking out about their experiences by sharing painful past experiences and using their tweets to comfort others. This may speak to the influence of computer-mediated communication on self-disclosure. Prior research reveals that online communication promotes higher levels of self-disclosure as compared to face-to-face interactions (Jiang et al., 2011) and may increase one's sense of belonging (Davis, 2012). Additionally, Šléglová1 and Cerna's (2011) interviews with adolescents revealed that victims of cyberbullying develop coping strategies similar to those reflected in our findings. The public nature of Twitter and the ability to directly mention a user signals a direct marker of addressivity (Honey & Herring, 2009; Werry, 1996). These include activities directed at the aggressor, akin to the tweets in our study where targets called out a perpetrator, and seeking social support, analogous to the tweets where targets shared previous bullying experiences to validate and console others.

Relatedly, prior research on bystanders often points to the concept of diffusion of responsibility, which suggests that the more bystanders there are, the less likely someone will help (Gini et al., 2015; Lambe et al., 2019). In an online context, bystanders are unable to anticipate how many other viewers are witnessing the bullying, and thus, one could assume a reduced likelihood of intervening. Additional studies also indicate that bystanders are more likely to intervene privately than publicly online (Bastiaensens et al., 2015; Patterson et al., 2017). Yet, similar to the target tweets, the analysis revealed that helpers are present and active on Twitter, publicly defending a target, reporting an episode of bullying, or accusing a perpetrator. While we are unable to construe intent beyond what is presented at face-value in the 280 character tweet, findings from Machácková et al. (2013) suggests that a bystander's emotional response, such as feeling upset after witnessing cyberbullying and targets' direct request for help, are connected to a higher likelihood of bystander support online. The same study also infers that fear of intervening had no impact on bystander support (Machácková et al., 2013). This stands in contrast to research on face-to-face bystander support, which notes that the fear of becoming a target may reduce the likelihood of helping (Kanetsuna & Smith, 2002) or result in indirect forms of assistance such as the "help-seeker" (Levy & Gumpel, 2018). This could mean that helpers are more inclined to support targets on Twitter due to the physical distance afforded between the role players. The use of "re-tweets" may also provide a less confrontational and indirect form of calling out a bully but it could still show support for a target. It could also be reflecting the scope and reach of the platform. Global and local bullying incidents can instantly start trending on Twitter, prompting re-tweets, hashtags, and support worldwide. Yoo et al.'s (2014) exploration of Twitter usage indicates that social conformity, the response to an external force, influences a user's sharing activities. Accordingly, if a story about bullying is trending, triggers an emotional response from bystanders, and serves to endorse a desired social image in line with their group or community, it may promote increased support, activism, and information sharing from helpers.

The findings also highlight the relationship between the bullying roles and the overlap between online and offline spaces. Most targets, perpetrators, and helpers knew each other and directly mentioned user name(s) or referenced them in their tweets. It could be that some users only know each other virtually; however, less than a third of perpetrator and target tweets were referring to online bullying, and more than a quarter of target tweets referred to school-based bullying. A small number of targets also mentioned bullying at home and work. This supports prior research that a majority of cyberbullying likely takes place between individuals known in real life (Felmlee & Faris, 2016; McHugh et al., 2019; Newall, 2018). Additionally, those who experience (Waasdorp & Bradshaw, 2015) or perpetrate cyberbullying (Kowalski et al., 2014) may also be participants in offline bullying. Thus, the belief that someone can simply turn off their computer or cell phone to make bullying stop is misplaced (Sabella et al., 2013). Stepping away from technology does not solve the issue and does not consider the intertwined nature of offline and online behaviors.

Another dimension to this overlap is the time frame within the tweets. Not all tweets were referring to ongoing bullying episodes. While perpetrator tweets were more likely to be related to an ongoing bullying episode, a

significant majority of targets and helpers were referring to past incidents. This ties back to the implication that in the context of bullying, Twitter is a multidimensional space. While it may provide a space for perpetrators to enact bullying behavior, it also serves as a venue for targets and helpers to speak out. Dore et al. (2017) study infers that those who engage in helping others online may improve their own emotional well-being in stressful situations. Likewise, Levy-Gigi and Shamay-Tsoory's (2017) findings surmise that rather than trying to deal with an issue individually, help from an outside perspective can help reduce stress and improve emotional regulation. Consequently, Twitter may serve as a safe space and community for the social regulation of emotions; in standing up and speaking out, targets and helpers are supporting each other as they navigate both offline and online bullying situations. The use of hashtags can help to categorize Twitter posts and can also be used by targets and helpers to direct attention to specific situations, to show support, and to raise awareness around bullying.

Finally, our analysis revealed numerous reasons for being bullied. The findings that characteristics such as demographics, physical looks, disabilities, and engagement in certain activities are related to both offline and online bullying have been well established in the extant literature (de Bruyn et al., 2010; Janssen et al., 2004; Kann et al., 2018; Mishna, 2003) and was similarly echoed in our findings. However, one aspect of bullying that was reflected within the tweets but has received limited attention is the topic of fandom bullying. Fandoms are communities built around shared interests such as specific musical artists, sports teams, or unique interests such as anime, video games, comic books, and manga. These communities are strongly represented on social media (Guo, 2016). An example of participatory culture (Jenkins et al., 2009), these communities of super fans or "stans" obsessively support, promote, and defend their respective fandom. Toxic fandom behavior can result in pitting rival stars against each other, sparing with rival fans, and defending their favorite artist from any form of criticism, and often materializes on social media platforms such as Twitter (Hunt, 2019). Moreover, the references made to celebrities and public figures within the tweets suggest that fans are not the only ones at the receiving end of aggressive and unwanted behavior. Given the rise in fandom communities and the increasing prevalence of "cancel culture" where celebrities and public figures face backlash and are "canceled" or culturally blocked from having a prominent platform or career due to their actions or even foot-in-mouth moments (Romano, 2019), this aspect of our findings necessitates that academic consideration is given to understanding the impact that fandom bullying has on bullying roles and how it may further exacerbate the issue of bullying.

Limitations and Future Research

While analyzing Twitter data provided a novel way to expand on the bullying roles literature, it also posed some limitations. Analyzing 280 character tweets required us to take it at face-value without any contextual or background knowledge; thus, we were unable to determine if any specific characteristics are related to each role. We were also unable to garner the intent behind the tweet, thus potentially missing out on sarcasm, jokes, or harmless teasing done in jest. Demographic characteristics of the authors were also missing. While most prior research on bullying roles focuses on children, we have no way of knowing the age of the Twitter users captured in the current study. However, it should be noted there were references made to school-based bullying and bullying that occurred in the past so some of the data reflect youth bullying. Future studies should consider expanding on our approach by connecting any relevant contextual and background knowledge about Twitter users. Additionally, while recall bias is noted as a limitation of self-report measures, it could potentially also impact shared experiences of bullying online. Rivers (2001) found that retrospective reports of bullying were generally stable and consistent across 1 year. For tweets that refer to current or recent bullying episodes, the recall bias may not be as applicable. However, since we do not know the age of the Twitter users and the exact timeline of the bullying episode they are referring to, we acknowledge that this could be a limitation as well.

The study was also limited to tweets captured through the keyword selection. Bullying is verbally expressed in many ways, and our selection of keywords only captured a snippet of bullying-related tweets. Additionally, perpetrators are not likely to use the word "bully" when engaging in direct acts of bullying. While we captured some instances of direct bullying engagement in the perpetrator tweets, we found that most were referring to past episodes, apologizing or expressing remorse, expressing intent to harm, or did not have enough information. Future studies should look into incorporating an expanded keyword selection, which would be better suited to identifying actual bullying episodes rather than just disclosure of bullying episodes. The qualitative nature of the analysis also assumes that the coders of the tweets may bring their own experiences and background into the analysis; thus, we made certain to review the coded tweets for agreement. Finally, using a TF-IDF-based NLP method, a machine-learning model was built to classify the tweets in this study as a bullying trace and to classify the bullying traces into the different tweet author

roles (Dhungana Sainju et al., 2021). Future studies can utilize the approach and insights generated from Dhungana Sainju et al. (2021) and the current study to build improved classifiers to detect bullying-related discourse and content online. The performance of the machine learning models can be used as a benchmark to improve on and semantic complexity can be added to the NLP model through deep learning techniques to further strengthen its performance. Moreover, oversampling and increasing the number of labeled tweets can also help improve the classification task. Increased sample size and an evenly distributed or roughly similar number of annotated data within classification categories can improve the supervised machine-learning model and optimize future bullying and cyberbullying detections models and classifiers. This type of qualitative study that examines how different bullying role players compose tweets that can qualify as bullying traces provides opportunities for the systematic incorporation of domain knowledge within the NLP models which can further improve their performance.

Conclusion

The current study implies that despite our acceptance of online spaces as venues for bullying, it can also provide a platform to humanize the behavior and experience of being bullied. By being able to discover, connect, and share with a geographically diverse set of users about a common shared experience or event, Twitter functions as a platform for offering social support, encouraging emotional regulation, and building a sense of community. At a time when social media networks are becoming a significant part of mobilizing and catalyzing social action and activism, understanding how bullying roles emerge on Twitter can be used as part of the fight against bullying.

Our results suggest that bystanders may be more willing to become helpers in online spaces and we know from prior research that when bystanders intervene it can greatly reduce instances of bullying (Hawkins et al., 2001) and create a safer environment (Nickerson et al., 2008). Furthermore, hearing from others who have experienced similar situations can also make others feel better. As Wagner et al. (2015) found in their study, sharing emotional experiences with others, whether positive or negative in context, helps to buffer the impact of the negative stimuli and enhance the positive stimuli. Thus, in addition to offline components, anti-bullying initiatives should consider leveraging Twitter to encourage positive bystander behavior and promote the sharing of bullying experiences. Anti-bullying campaigns, coupled with hashtags, especially during high-profile bullying incidents can serve to increase awareness and conversations around bullying (Dhungana Sainju et al., 2021; McHugh et al., 2019). Several recent examples, including

the Black Lives Matter movement (#BlackLivesMatter), the Me Too movement (#MeToo), and the UN's global Women campaign (#HeForShe) showcase how Twitter hashtags were used to help raise awareness across varied social justice issues. In addition, anti-bullying campaigns can also collaborate with social media influencers (SMIs) or celebrities to help promote and increase dialogue around bullying issues. SMIs and celebrities often have a large social media following and have built credibility in a specific industry (Freberg et al., 2011). They have the power to persuade and influence consumer behavior (Booth & Matic, 2011) and play a key role in digital activism (Hutchinson, 2021). For example, the COVID-19 pandemic led several countries including China, Australia, Koran, and Japan to partner with country-specific influencers to promote responsible pandemic etiquette including social distancing, hand washing, and wearing masks (Abidin et al., 2021). As one of the most popular social media networks, the use of hashtags along with the exposure from key influencers on Twitter could offer a culturally relevant and low-cost option to educate and inform users about bullying and encourage more dialogue around bullying issues while also helping build resilience and emotional regulation among its users.

To conclude, the study provides an original contribution to the literature on bullying roles by providing one of the first qualitative analyses using Twitter data. A significant innovation of the study was being able to examine the disclosure of actual behavior and intentions, rather than self-reported conduct. By examining bullying roles through a different data source, the study was able to augment our prior understanding of bullying roles and presents an opportunity to leverage Twitter to encourage behavior that may help reduce the likelihood of bullying behaviors.

Declarations

Ethical Approval The current study was reviewed and approved by the University of Ontario Institute of Technology's Research Ethics Board (REB) as a secondary use of data study (REB# 15,917). All procedures performed in the study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

Abidin, C., Lee, J., Barbetta, T., & Miao, W. S. (2021). Influencers and COVID-19: Reviewing key issues in press coverage across Australia China Japan and South Korea. *Media International Australia, 178*(1), 114–135.

Anderson, M. (2018). A majority of teens have experienced some form of cyberbullying. *Pew Research Center*. Retrieved from https://www.pewresearch.org/internet/2018/09/27/a-majority-of-teens-have-experienced-some-form-of-cyberbullying/

Ang, R. P., Tan, K. A., & Talib Mansor, A. (2011). Normative beliefs about aggression as a mediator of narcissistic exploitativeness and cyberbullying. *Journal of Interpersonal Violence, 26*(13), 2619–2634.

Bartlett, J. E., & Bartlett, M. E. (2011). Workplace bullying: An integrative literature review. *Advances in Developing Human Resources, 13*(1), 69–84.

Bastiaensens, S., Vandebosch, H., Poels, K., Van Cleemput, K., DeSmet, A., & De Bourdeaudhuij, I. (2015). Can I afford to help How affordances of communication modalities guide bystanders' helping intentions towards harassment on social network sites. *Behaviour & Information Technology, 34*(4), 425–435.

Bellmore, A., Calvin, A. J., Xu, J. M., & Zhu, X. (2015). The five W's of bullying on Twitter: who, what, why, where and when. *Computers in Human Behavior, 44*, 305–314.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open, 2*, 8–14.

Blanco, K., Briceno, A., Steele, A., Tapia, J., McKay, J., Towers, S., & Yong, K. E. (2014). The Dynamics of Offensive Messages in the World of Social Media: the Control of Cyberbullying on Twitter. arXiv preprint. arXiv:1408.0694

Booth, N., & Matic, J. A. (2011). Mapping and leveraging influencers in social media to shape corporate brand perceptions. *Corporate Communications: an International Journal., 16*(3), 184–191.

Calvin, A. J., Bellmore, A., Xu, J. M., & Zhu, X. (2015). # bully: Uses of hashtags in posts about bullying on Twitter. *Journal of School Violence, 14*(1), 133–153.

Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., & Vakali, A. (2017). Hate is not binary: Studying abusive behavior of #GamerGate on Twitter. arXiv:1705.03345 [cs.SI]

Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., & Vakali, A. (2017). Mean birds: Detecting aggression and bullying on Twitter. 13–22. https://doi.org/10.1145/3091478.3091487

Chelmis, C. Zois, D. & Yao. (2017). Mining patterns of cyberbullying on Twitter. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* New Orleans LA 2017. 126–133. https://doi.org/10.1109/ICDMW.2017.22

Christensen, L. L., Fraynt, R., Neece, C. L., & Baker, B. L. (2012). Bullying adolescents with intellectual disability. *Journal of Mental Health Research in Intellectual Disabilities, 5*(1), 49–65. https://doi.org/10.1080/19315864.2011.637660

Clement, J. (2019). Number of monthly active Twitter users worldwide from 1st quarter 2010 to first quarter 2019. *Statistica*. Retrieved from: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Cook, C. R., Williams, K. R., Guerra, N. G., Kim, T. E., & Sadek, S. (2010). Predictors of bullying and victimization in childhood and adolescence: A meta-analytic investigation. *School Psychology Quarterly, 25*(2), 65–83. https://doi.org/10.1037/a0020149

Dake, J. A., Price, J. H., Telljohann, S. K., & Funk, J. B. (2003). Teachers' perceptions and practices regarding school bullying prevention activities. *Journal of School Health, 73*, 347–354.

Davis, K. (2012). Friendship 2.0: Adolescents' experiences of belonging and self-disclosure online. *Journal of Adolescence*, *35*(6), 1527–1536.

de Bruyn, E. H., Cillessen, A. H. N., & Wissink, I. B. (2010). Associations of peer acceptance and perceived popularity with bullying and victimization in early adolescence. *The Journal of Early Adolescence, 30*(4), 543–566. https://doi.org/10.1177/0272431609340517

Dedoose Version 8.0.35, web application for managing, analyzing, and presenting qualitative and mixed method research data (2018)**.** Los Angeles, CA: SocioCultural Research Consultants, LLC. www.dedoose.com

Devaneyan, P. (2016). Automatic detection of cyberbullying from Twitter. *International Journal of Computer Science and Information Security*, *6*(6).

Dhungana Sainju, K. (2020). Beyond the schoolyard: A multilevel examination of individual, school and school district variables associated with traditional and cyber peer aggression. *Child Youth and Care Forum, 49*, 769–797. https://doi.org/10.1007/s10566-020-09555-4

Dhungana Sainju, K., Mishra, N., Kuffour, A., & Young, L. (2021). Bullying discourse on Twitter: An examination of bully-related tweets using supervised machine learning. *Computers in Human Behavior, 120*, 106735. https://doi.org/10.1016/j.chb.2021.106735

Doré, B. P., Morris, R. R., Burr, D. A., Picard, R. W., & Ochsner, K. N. (2017). Helping others regulate emotion predicts increased regulation of one's own emotions and decreased symptoms of depression. *Personality & Social Psychology Bulletin, 43*(5), 729–739. https://doi.org/10.1177/0146167217695558

Felmlee, D., & Faris, R. (2016). Toxicities: Networks of friendship, dating, and cyber victimization. *Social Psychology Quarterly, 79*(3), 243–262. https://doi.org/10.1177/0190272516656585

Forum Research Inc. (2018). 1 in 2 Canadians experience bullying in the workplace. Retrieved from https://poll.forumresearch.com/data/2b79a59f-a126-4c5d-a90a-4c700e8e610eBullying%20Nov%2012.pdf

Freberg, K., Graham, K., McGaughey, K., & Freberg, L. (2011). Who are the social media influencers?: A study of public perceptions of personality. *Public Relations Review*, *37*, 90–92.

GLSEN, CiPHR, & CCRC, . (2013). *Out online: The experiences of lesbian, gay, bisexual and transgender youth on the Internet*. GLSEN.

Gini, G., Pozzoli, T., & Bussey, K. (2015). The role of individual and collective moral disengagement in peer aggression and bystanding: A multilevel analysis. *Journal of Abnormal Child Psychology, 43*(3), 441–452.

Guo, T. (2016). Fandom Bullying—'Why can't we just enjoy [insert favourite TV show/film/book/etc.] without being attacked?' *Medium.* Retrieved from https://medium.com/@tgforuws/fandom-bullying-why-cant-we-just-enjoy-insert-favourite-tv-show-film-book-etc-a9c0cb826411

Hawkins, D. L., Pepler, D. J., & Craig, W. M. (2001). Naturalistic observations of peer interventions in bullying. *Social Development, 10*(4), 512–527. https://doi.org/10.1111/1467-9507.00178

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288.

Hinduja, S., & Patchin, J. W. (2014). *Bullying beyond the schoolyard: preventing and responding to cyberbullying* (2nd ed.). Sage Publications.

Holmberg, K., & Hjern, A. (2008). Bullying and attention-deficit–hyperactivity disorder in 10-year-olds in a Swedish community. *Developmental Medicine & Child Neurology, 50*(2), 134–138.

Honey, C. & Herring, S.C. (2009). Beyond microblogging: Conversation and collaboration via Twitter. 2009 42nd Hawaii International Conference on System Sciences, Waikoloa, HI, USA,1–10. https://doi.org/10.1109/HICSS.2009.89

Hunt, D., Atkin, D., & Krishan, A. (2012). The influence of computer-mediated communication apprehension on motives for Facebook use. *Journal of Broadcasting & Electronic Media, 56*(2), 187–202.

Hunt, E. (2019). Stan by me: how super-fan culture turned sour. *NME*. Retrieved from https://www.nme.com/features/stan-super-fan-culture-turned-toxic-2457815

Hutchinson, J. (2021). Micro-platformization for digital activism on social media. *Information, Communication & Society, 24*(1), 35–51.

Hymel, S., Schonert-Reichl, K. A., Bonanno, R. A., Vaillancourt, T., & Henderson, N. R. (2010). Bullying and morality: Understanding how good kids can behave badly. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 101–118). Routledge.

Lambe, L. J., Della Cioppa, V., Hong, I. K., & Craig, W. M. (2019). Standing up to bullying: A social ecological review of peer defending in offline and online contexts. *Aggression and Violent Behavior*. https://doi.org/10.1016/j.avb.2018.05.007

Levy, M., & Gumpel, T. P. (2018). The interplay between bystanders' intervention styles: An examination of the "bullying circle" approach. *Journal of School Violence, 17*(3), 339–353.

Levy-Gigi, E., & Shamay-Tsoory, S. G. (2017). Help me if you can: Evaluating the effectiveness of interpersonal compared to intrapersonal emotion regulation in reducing distress. *Journal of Behavior Therapy and Experimental Psychiatry, 55*, 33–40. https://doi.org/10.1016/j.jbtep.2016.11.008

Janssen, I., Craig, W. M., Boyce, W. F., & Pickett, W. (2004). Associations between overweight and obesity with bullying behaviors in school-aged children. *Pediatrics, 113*(5), 1187–1194.

Jeffrey, J., & Stuart, J. (2020). Do research definitions of bullying capture the experiences and understandings of young people? A qualitative investigation into the characteristics of bullying behaviour. *International Journal of Bullying Prevention, 2*, 180–189.

Jenkins, H., Purushotma, R., Weigel, M., Clinton, K. & Robison, A. (2009). *Confronting the challenges of participatory culture*. The John D. and Catherine T. MacArthur Foundation Reports on Digital Media and Learning.

Jiang, L. C., Bazarova, N. N., & Hancock, J. T. (2011). The disclosure-intimacy link in computer -mediated communication: An attributional extension of the hyperpersonal model. *Human Communication Research, 37*(1), 58–77.

Kann, L., McManus, M. S., Harris, W. A., Shanklin, S. L., Flint, K. H., Queen, B., Lowry, R., Chyen, D., Whittle, L., Thornton, J., Lim, C., Bradford, D., Yamakawa, Y., Leon, M., Brener, N., & Ethier, K. (2018). Youth risk behavior surveillance—United States 2017. *MWR Surveillance Summary, 67*(8), 1.

Kartal, H., & Bilgin, A. (2009). Bullying and school climate from the aspects of the students and teachers. *Egitim Arastirmalari-Eurasian Journal of Educational Research, 36*, 209–226.

Kanetsuna, T., & Smith, P. K. (2002). Pupil insights into bullying, and coping with bullying: A bi-national study in Japan and England. *Journal of School Violence, 1*, 5–29.

Kondracki, N. L., & Wellman, N. S. (2002). Content analysis: Review of methods and their applications in nutrition education. *Journal of Nutrition Education and Behavior, 34*, 224–230.

Kowalski, R., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin, 140*(4), 1073–1137.

Macháčková, H., Dedkova, L., Sevcikova, A., & Cerna, A. (2013). Bystanders' support of cyberbullied schoolmates. *Journal of Community & Applied Social Psychology, 23*, 25–36.

McCormick, T. H., Lee, H., Cesare, N., Shojaie, A., & Spiro, E. S. (2017). Using Twitter for demographic and social science research: tools for data collection and processing. *Sociological Methods & Research, 46*(3), 390–421. https://doi.org/10.1177/0049124115605339

McHugh, M. C., Saperstein, S. L., & Gold, R. S. (2019). OMG U #Cyberbully! An exploration of public discourse about cyberbullying on Twitter. *Health Education & Behavior, 46*(1), 97–105.

Menesini, E., & Camodeca, M. (2008). Shame and guilt as behaviour regulators: Relationships with bullying, victimization and prosocial behaviour. *British Journal of Developmental Psychology, 26*, 183–196.

Mishna, F. (2003). Learning disability and bullying: Double jeopardy. *Journal of Learning Disabilities, 36*(4), 336–347.

Modecki, K. L., Minchina, J., Harbaugh, A. G., Guerra, A. G., & Runions, K. C. (2014). Bullying prevalence across contexts: A meta-analysis measuring cyber and traditional bullying. *Journal of Adolescent Health, 55*(5), 602–611.

Newall, M. (2018). Global views on cyberbullying. *Ipsos*. Retrieved from https://www.ipsos.com/en/global-views-cyberbullying

Nickerson, A. B., Mele, D., & Princiotta, D. (2008). Attachment and empathy as predictors of roles as defenders or outsiders in bullying interactions. *Journal of School Psychology, 46*, 687–703. https://doi.org/10.1016/j.jsp.2008.06.002

Olweus, D. (1991). Bully/victim problems among school children: Basic facts and effects of a school-based intervention program. In K. Rubin & D. Pepler (Eds.), The development and treatment of childhood aggression. Hillsdale, NJ: Lawrence Erlbaum.

Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Blackwell Publishers.

Olweus, D., & Limber, S. P. (2010). Bullying in school: Evaluation and dissemination of the Olweus Bullying Prevention Program. *American Journal of Orthopsychiatry, 80*(1), 124–134.

Patterson, L. J., Allan, A., & Cross, D. (2017). Adolescent bystander behavior in the school and online environments and the implications for interventions targeting cyberbullying. *Journal of School Violence, 16*(4), 361–375. https://doi.org/10.1080/15388220.2016.1143835

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health, 42*(5), 533–544.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Piotrowski, C. (2015). Adult bully syndrome: A bibliometric analysis on concordance with personality disorder traits. *Journal of Instructional Psychology, 42*(1), 1–3.

Piotrowski, C. (2016). Adult bully syndrome: An integrative conceptualization based on a personality disorders framework. *Psychology and Education: an Interdisciplinary Journal, 53*(1–2), 91–98.

Pyzalski, J. (2011). Electronic aggression among adolescents: An old house with a new façade (or even a number of houses). In Youth culture and net culture: Online social practices, ed. C. Hällgren, E. Dunkels and G.-M. Frånberg, 278–95. Hershey, PA: IGI Global.

Pyzalski, J. (2012). From cyberbullying to electronic aggression: typology of the phenomenon. *Emotional and Behavioural Difficulties, 17*(3–4), 305–317.

Rivers, I. (2001). Retrospective reports of school bullying: Stability of recall and its implications for research. *British Journal of Developmental Psychology, 19*, 129–142.

Romano, A. (2019). Why we can't stop fighting about cancel culture. *Vox*. Retrieved from https://www.vox.com/culture/2019/12/30/20879720/what-is-cancel-culture-explained-history-debate

Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural & Healthcare Research, 2*(4), 320–332. https://doi.org/10.1504/IJBHR.2011.043414

Sabella, R., Patchin, J., & Hinduja, S. (2013). Review: Cyberbullying myths and realities. *Computers in Human Behavior, 29*, 2703–2711. https://doi.org/10.1016/j.chb.2013.06.040

Salmivalli, C., Karhunen, J., & Lagerspetz, K. M. (1996). How do the victims respond to bullying? *Aggressive Behavior, 22*, 99–10.

Salmivalli, C., Lagerspetz, K. M. J., BjoÈrkqvist, K., & OÈ sterman, K. and Kaukiainen, A. . (1996). Bullying as a group process: Participant roles and their relations to social status within the class. *Aggressive Behavior, 22*, 1–15.

Salmivalli, C. (1999). Participant role approach to school bullying: Implications for interventions. *Journal of Adolescence, 22*, 453–459.

Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior, 15*(2), 112–120. https://doi.org/10.1016/j.avb.2009.08.007

Slee, P. T. (1993). Bullying: A preliminary investigation of its nature and the effects of social cognition. *Early Child Development and Care, 87*(1), 47–57. https://doi.org/10.1080/0300443930870105

Šléglová, V., & Cerna, A. (2011). Cyberbullying in adolescent victims: Perception and coping. Cyberpsychology: *Journal of Psychosocial Research on Cyberspace*, 5(2), Article 4. Retrieved from https://cyberpsychology.eu/article/view/4248/3294

Slonje, R., Smith, P. K., & Frisén, A. (2012). Processes of cyberbullying, and feelings of remorse by bullies: A pilot study. *European Journal of Developmental Psychology, 9*, 244–259.

Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior, 29*, 239–268.

Steffgen, G., König, A., Pfetsch, J., & Melzer, A. (2011) Are cyberbullies less empathic? Adolescents' cyberbullying behavior and empathic responsiveness. *Cyberpsychology, Behavior, and Social Networking*, *14*(11), 643–648. https://doi.org/10.1089/cyber.2010.0445

Stull, D. E., Leidy, N. K., Parasuraman, B., & Chassany, O. (2009). Optimal recall periods for patient-reported outcomes: Challenges and potential solutions. *Current Medical Research and Opinion, 25*, 929–942.

Tankovska, H. (2021). Twitter: Distribution of global audiences, 2021, by age group. Statistica. Retreived from: https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/

UNICEF. (2019). UNICEF poll: More than a third of young people in 30 countries report being a victim of online bullying. Retrieved from https://www.unicef.org/press-releases/unicef-poll-more-third-young-people-30-countries-report-being-victim-online-bullying

Unnever, J. D., & Cornell, D. G. (2004). Middle school victims of bullying: Who reports being bullied? *Aggressive Behavior, 30*, 373–388.

Van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, *25*(4).

Vazsonyi, A. T., Macháčková, H., Sevcikova, A., Smahel, D., & Cerna, A. (2012). Cyberbullying in context: Direct and indirect effects by low self-control across 25 European countries. *European Journal of Developmental Psychology, 9*(2), 210–227. https://doi.org/10.1080/17405629.2011.644919

Vessey, J., Strout, T. D., DiFazio, R. L., & Walker, A. (2014). Measuring the youth bullying experience: A systematic review of the psychometric properties of available instruments. *Journal of School Health, 84*, 819–843.

Vidgen, B., Harris, A., Nguyen, D., Tromble, R., Hale, S., & Margetts, H. (2019). Challenges and frontiers in abusive content detection. *Proceedings of the Third Workshop on Abusive Language Online*, 80–93.

Vivolo-Kantor, A. M., Martell, B. N., Holland, K. M., & Westby, R. (2014). A systematic review and content analysis of bullying and cyber-bullying measurement strategies. *Aggression and Violent Behavior, 19*(4), 423–434.

Waasdorp, T. E., & Bradshaw, C. P. (2015). The overlap between cyberbullying and traditional bullying. *Journal of Adolescent Health, 56*(5), 483–488. https://doi.org/10.1016/j.jadohealth.2014.12.002

Wagner, U., Galli, L., Schott, B. H., Wold, A., van der Schalk, J., Manstead, A. S. R., Scherer, K., & Walter, H. (2015). Beautiful friendship: Social sharing of emotions improves subjective feelings and activates the neural reward circuitry. *Social Cognitive and Affective Neuroscience*, *10*(6), 801–808.

Werry, C. C. (1996). Linguistic and interactional features of Internet Relay Chat. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (pp. 47–63). John Benjamins.

Whittaker, E., & Kowalski, R. M. (2015). Cyberbullying via social media. *Journal of School Violence, 14*, 11–29.

Wójcik, M., & Flak, W. (2019). Frenemy: A new addition to the bullying circle. *Journal of Interpersonal Violence*. https://doi.org/10.1177/0886260519880168

Wolke, D., Tippett, N., & Dantchev, S. (2015). Bullying in the family: Sibling bullying. *The Lancet Psychiatry, 2*(10), 917–929. https://doi.org/10.1016/S2215-0366(15)00262-X

Workplace Bullying Institute. (2020). Retrieved from https://workplacebullying.org/

Xu, J. M., Jun, K. S., Zhu, X., & Bellmore, A. (2012). Learning from bullying traces in social media. In Proceedings of the conference of the North American chapter of the association for computational linguistics: Human language technologies (pp. 656–666). Montreal, Quebec, Canada: Association for Computational Linguistics. Retrieved from http://www.proceedings.com/15547.html.

Yoo, J., Choi, S., Choi, M., & Rho, J. (2014). Why people use Twitter: Social conformity and social value perspectives. *Online Information Review, 38*(2), 265–283. https://doi.org/10.1108/OIR-11-2012-0210

Zeedyk, S. M., Rodriguez, G., Tipton, L. A., Baker, B. L., & Blacher, J. (2014). Bullying of youth with autism spectrum disorder, intellectual disability, or typical development: Victim and parent perspectives. *Research in Autism Spectrum Disorders, 8*(9), 1173–1183.

Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 3*, 1–130. https://doi.org/10.2200/S00196ED1V01Y200906AIM006