JOURNAL OF
Research on Adolescence
Society for Research on Adolescence
www.s-r-a.org

Does defending affect adolescents' peer status, or vice versa? Testing the moderating effects of empathy, gender, and anti-bullying norms

Lydia Laninga-Wijnen[1] | Sarah T. Malamut[1,2] | Claire F. Garandeau[1] | Christina Salmivalli[1]

[1]Developmental Psychology, INVEST Flagship, University of Turku, Turku, Finland

[2]Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, The Netherlands

Correspondence
Lydia Laninga-Wijnen, Developmental Psychology, INVEST Flagship, University of Turku, Assistentinkatu 7, 20500 Turku, Finland.
Email: lalawi@utu.fi

Abstract

This study examined bidirectional associations between students' bully-directed defending behavior and their peer status (being liked or popular) and tested for the moderating role of empathy, gender, and classroom anti-bullying norms. Three waves of data were collected at 4–5-month time intervals among 3680 Finnish adolescents ($M_{age} = 13.94$, 53.0% girls). Cross-lagged panel analyses showed that defending positively predicted popularity and, to a larger degree, being liked over time. No moderating effect of empathy was found. Popularity was more strongly predictive of defending, and defending was more strongly predictive of status among girls than among boys. Moreover, the positive effects of both types of status on defending were—albeit to a limited extent—stronger in classrooms with higher anti-bullying norms.

KEYWORDS
defending, norms, peer status

INTRODUCTION

Although bystanders are present in about 80% of bullying situations in school (Jones et al., 2015), only a minority (about 25%) attempts to help victimized peers (e.g., Craig & Pepler, 1998; Laninga-Wijnen et al., 2022). The silence of passive bystanders may be more hurtful for victims than the bullying itself (Jones et al., 2015). Passivity from bystanders could signal tacit approval or condoning of the bullying, which gives perpetrators a green light to continue or to pursue their behavior. Therefore, anti-bullying programs increasingly encourage *peer defending*, which refers to prosocial actions undertaken in response to a bullying situation, such as being kind and supportive to the victim (victim-directed defending) or publicly confronting bullies and actively trying to stop bullying by asking others to intervene (bully-directed defending; Lambe & Craig, 2020). Though defending may help victims (Laninga-Wijnen et al., 2022; Sainio et al., 2011), it could be risky for defenders,

especially when defending occurs in bully-directed ways, as such defending attempts challenge and contest the behaviors of generally powerful bullies and their allies (Lambe & Craig, 2020). Therefore, bully-directed defending is assumed to require high peer status—such as being liked or popular—but also to be potentially costly in terms of status (Salmivalli et al., 2021).

Given the growing call to encourage youth to defend their victimized peers, more insight is needed into the bidirectional links between defending and peer status. Most studies on the associations between defending and peer status have been cross-sectional. A recent meta-analysis detected a small, positive link between the two, but there was also substantial heterogeneity in effect sizes across studies (Ma et al., 2019). Only three studies have examined prospective bidirectional links between peer status and defending (Meter & Card, 2015; Pozzoli & Gini, 2021; van der Ploeg et al., 2017) and only one of them considered both being liked and popularity as indicator of status (Pozzoli

& Gini, 2021). Their findings vary considerably from each other both in the magnitude and in the direction of effects. This may be an indication that the bidirectional associations between adolescents' status and defending behavior vary as a function of individual and contextual factors. We propose that these bidirectional associations depend on students' empathy, gender, and classroom anti-bullying norms, as these factors may determine the anticipated and actual social costs and benefits of defending (Pouwels et al., 2019). Thus, the current study aimed to clarify whether defending requires high status and has consequences for adolescents' status, and whether these associations are moderated by students' gender and empathy, as well as classroom anti-bullying norms. We focus on bully-directed ways of defending, given their potential relevance for adolescents' status.

Is high status a prerequisite for defending?

Adolescence is marked by a heightened concern about one's status (LaFontana & Cillessen, 2010). Youth may strive both for belonging to the peer group (being liked; also referred to as peer acceptance) and for being noticed, admired, or powerful among their peers (i.e., being popular; Dawes & Xie, 2017). Qualitative studies have indicated that adolescents may consider a certain level of peer status a prerequisite for daring to stand up against a powerful bully (Spadafora et al., 2020). There are at least three possible reasons for this. First, adolescents with high status—both in terms of being liked and being popular—may feel more confident to defend victimized peers with no fear of social repercussions from bullies, as high-status youth generally have more friends (Laninga-Wijnen & Veenstra, 2021), which can protect them against victimization (Hodges & Perry, 1999). Second, adolescents who are well-liked or popular tend to be more assertive (Poorthuis et al., 2021) and have better social–emotional skills (Laninga-Wijnen et al., 2021), allowing them to better understand the effect of their behaviors on their peers (Hawley, 2014), making them more likely to defend. Third, costly signaling theory (Grafen, 1990)—which applies mainly to popularity—proposes that people tend to behave more prosocially when they are being observed by others. In addition, being prosocial can be a way to demonstrate one's social power, especially if the prosocial acts are perceived as "risky." The visibility of popular youth may therefore explain why they defend victims—it demonstrates their power to others (Pronk et al., 2019).

Only three studies have tested whether high peer status predicts more defending over time, while also testing for opposing directions (i.e., whether defending predicts higher status). One study found that Finnish 4th–6th graders with higher popularity were more likely to defend victims about 5 months later (van der Ploeg et al., 2017), whereas another study on American 6th and 7th graders did not detect any significant effect of being liked on students' defending behavior 1 year later (Meter & Card, 2015). The third study found that neither being popular *nor* being liked predicted defending 1 year later among Italian 6th–8th graders (Pozzoli & Gini, 2021). Thus, previous work is inconclusive regarding whether high status is a prerequisite for defending. Based on the theoretical arguments mentioned above, we expected that both being popular and being liked would positively predict bully-directed defending over time in the current study.

Is status a consequence of defending?

The act of defending may have consequences for defenders' peer status. However, there are theoretical arguments for both positive and negative consequences. On the one hand, it is conceivable that defending increases adolescents' status. Costly signaling theory (Grafen, 1990) posits that engaging in public prosocial behaviors, such as generosity, is not only a way to display one's status, but also to enhance it. To qualify as a costly signal, these prosocial behaviors should be observable by others (i.e., public; Smith & Bliege Bird, 2000) and costly in terms of time, energy, or risk, so as to display the prosocial actors' qualities, which in turn should heighten their social power (Boone, 1998). Bully-directed defending can qualify as 'costly signal' because it is observable by peers and it demonstrates that defenders dare to take a risk: they are powerful and brave enough to take a stance against bullying. This may foster respect and admiration among peers and thereby increase defenders' popularity (Reijntjes et al., 2016). Defending may also affect how well-liked defenders are by peers, because defenders undertake prosocial actions that are helpful for others (Pronk et al., 2020).

On the other hand, defending may have negative repercussion for adolescents' peer status. Preventing aggressors from achieving their goals (Pozzoli & Gini, 2013), undermining their control over resources (Olthof et al., 2011), and challenging their power could result in retaliation from the bullies and their allies (*retaliation hypothesis*; Huitsing et al., 2014; Spadafora et al., 2020). This retaliation might entail social exclusion, physical attacks, and damage to the reputation of defenders. Such treatment should logically lower defenders' status in the peer group. Indeed, in qualitative studies, fear of losing their popularity or fear of becoming the next target are the main reasons that adolescents cite for their reluctance to defend victimized peers (Strindberg et al., 2020).

Findings of longitudinal studies that examine whether and how defending contributes to status (while controlling for the predictive role of status in defending) vary considerably. Defenders *decreased* in being liked in one study (Meter & Card, 2015) and *increased* in being liked in another (Pozzoli & Gini, 2021). Defending predicted higher popularity in one study (van der Ploeg et al., 2017) but was unrelated to future popularity in another (Pozzoli & Gini, 2021). As there are theoretical reasons to expect both that defending could lead to higher status and to lower

DEFENDING AND PEER STATUS

JOURNAL OF
Research on Adolescence SRA Society for Research on Adolescence
www.s-r-a.org

3

status, and because prior work is inconclusive, we explored in the current study whether defending predicted being liked and popularity over time among adolescents, with no directional hypothesis.

Moderators of the bidirectional links between status and defending

A key goal of our study was to examine whether the prospective links between defending and peer status depend on individual and contextual factors. As outlined above, defending victimized peers can entail either social costs or benefits (Do et al., 2017). Therefore, before intervening in a bullying situation, students may weigh the potential costs and benefits of doing so (cost–benefit analysis; Armstrong-Carter et al., 2023; Nickerson et al., 2014). This may be particularly true for high-status youth, as they may have more to lose (i.e., more costs, as the higher one's status is, the more one can lose of it), but may also be particularly motivated to engage in defending as a means to further increase their status (more driven to gain benefits—costly signaling theory; Grafen, 1990). Moreover, the extent to which defending actually predicts a higher or lower status may depend on the extent to which these behaviors align with social (gender) or normative expectations (Nielson et al., 2017; Peets et al., 2015). Based on previous work (Lucas-Molina et al., 2018; Pouwels et al., 2019), we considered empathy, gender, and peer norms as important determinants for the anticipated and actual social costs and benefits of defending.

Empathy

Following the bystander intervention model, it can be argued that empathy shapes the anticipated costs and benefits of defending (Batson, 1988; Nickerson et al., 2014) because an empathetic individual should place more emphasis on the benefits of helping others and thereby should see the costs of non-intervention as high. Previous longitudinal work has identified empathy as an important predictor of defending behavior (Deng et al., 2021). We argue that empathy may also interact with youths' peer status in predicting their defending behavior, as a strong capacity to sympathize with victims may override high-status youths' hesitance to defend; i.e., the potential benefit of ending victims' suffering may outweigh the potential anticipated cost of losing status. In line with this reasoning, some studies have found that the concurrent link between status and defending was stronger for empathic youth (e.g., Caravita et al., 2009; Pöyhönen et al., 2010), yet other studies did not find this effect (e.g., Kim et al., 2013; Lucas-Molina et al., 2018). Despite inconsistencies in cross-sectional studies, we expected that the prospective effect of students' status (being popular and being liked) on defending would be stronger for adolescents with higher levels of empathy.

Gender

There are reasons to expect that status may be both a stronger predictor and a stronger outcome of bully-directed defending among boys than among girls. First of all, boys may have a stronger preference to show their status to others by means of using public and risky prosocial behaviors. Indeed, the costly signaling principle (being prosocial to show how powerful one is) has more often been detected among men (McAndrew, 2019). For example, an experimental study showed that boys were more likely than girls to demonstrate publicly generous behaviors as a means to demonstrate their status (Dreher et al., 2016). As costly signaling theory is about status in terms of social power or popularity, we hypothesize that popularity may be a stronger predictor of defending among boys than among girls, and we will explore whether the same pattern is found for being liked.

Furthermore, confrontational defending behaviors may fit better to what is socially expected from boys than from girls; therefore, boys' status may be more strongly affected by bully-directed defending than girls' status. Social role theory of gender and helping (Eagly & Crowley, 1986) argues that traditionally, boys and girls are socialized in gendered ways to help others. Masculine prosocial behavior tends to be characterized by strength and assertiveness, whereas feminine prosocial behavior is more focused on kindness and being attuned to others' needs (Hine, 2017; Nielson et al., 2017). Consistent with this reasoning, studies about defending in bullying situations have shown that boys are more likely to publicly stand up against the bully, whereas girls are more likely to comfort victims or provide advice in more indirect, private ways (Lambe & Craig, 2020; Reijntjes et al., 2016). Thus, defending others by standing up to the bully may be viewed as more socially acceptable for boys. Consequently, we hypothesized high peer status to be a stronger predictor *and* a stronger outcome of bully-directed defending among boys than among girls.

Classroom anti-bullying norms

Classroom anti-bullying norms may also moderate the extent to which status predicts defending and vice versa. Social norms determine which behaviors are expected or appropriate in a particular context (Shaw, 1981), and previous work has shown that high-status youth are more likely to conform to these social norms than low status youth (Lucas-Molina et al., 2018; Peets et al., 2015). This effect has often been explained by the idea that high status youth are concerned about losing their status (i.e., an anticipated cost) if they do not conform to these norms (Peets et al., 2015; Yun & Graham, 2018). If bullying is considered inappropriate in a classroom (i.e., a strong injunctive anti-bullying norm) high-status youth may consider it relatively safe to defend victimized peers, because such defending behaviors are compatible

4

JOURNAL OF
Research on Adolescence | SRA Society for Research on Adolescence
www.s-r-a.org

LANINGA-WIJNEN ET AL.

with the norm. Thus, high status youth may consider the potential social costs of defending to be lower in classrooms with high anti-bullying norms, and possibly even anticipate social benefits for aligning with group norms.

Classroom norms may also determine the actual social costs and benefits of defending. Following social misfit theory, students who display behaviors that align with group norms are more likely to be accepted by their peers than students who do not conform to norms (Wright et al., 1986). Therefore, defending will likely result in higher status (being liked and popular) in classrooms with higher anti-bullying norms.

A few cross-sectional studies have tested whether anti-bullying norms moderated the link between status and defending in primary schools. These studies found that popularity was more strongly related to defending in classrooms with stronger anti-bullying norms (Garandeau et al., 2022; Peets et al., 2015). Less consistent findings emerged for the role of norms in the link between being liked and defending. One study found that the positive link between being liked and defending was stronger in classrooms with higher anti-bullying norms (Lucas-Molina et al., 2018) whereas another study did not detect any significant moderating effect of norms in the link between being liked and defending (Garandeau et al., 2022). The cross-sectional design of these studies could not provide insights into whether anti-bullying norms affected the extent to which either (1) high-status youth were more likely to defend victims over time, or (2) defenders were more likely to increase in status over time. Moreover, the role of norms may be more pronounced in adolescence, when adult supervision wanes, paving the way for stronger influences of peers (Laursen & Veenstra, 2021). Based on both the theoretical arguments and prior findings mentioned above, we expected that higher status (popularity and being liked) would predict more defending, and that more defending would predict increased status in classrooms where anti-bullying norms are higher.

Current study

Using a cross-lagged panel model, we examined bidirectional associations between bully-directed defending and two types of status simultaneously to identify the unique and relative role of popularity and being liked over time. We expected higher status to predict higher levels of defending over time. We also examined if and how defending behavior is associated with future peer status. We did not formulate directional hypotheses about these effects as prior empirical studies are scarce and inconclusive in this regard.

Our second aim is to test whether the bidirectional associations between defending and status vary as a function of individual (empathy, gender) and contextual (anti-bullying norm) factors. We expect that empathy moderates the effect of status on future defending. Specifically, we hypothesize that the positive prospective links between both types of status and defending are stronger at higher levels of empathy. We hypothesize popularity to be a stronger predictor of

defending among boys and we explore whether gender moderates the prospective effect of being liked on defending. We also expect both status types to be stronger outcomes of defending among boys. Moreover, we expect that high-status youth are more likely to defend victims and that defenders are more likely to increase in status over time in classrooms with stronger anti-bullying norms.

METHOD

Participants and procedure

Data were collected from a sample of seventh and eighth grade students who took part in the randomized controlled trial conducted for the evaluation of the KiVa program in Finnish secondary schools (see Kärnä et al., 2013). Three waves of data were collected in May 2008, December 2008 and May 2009. The academic year in Finland is from mid-August until the end of May, thus data was collected across two academic years. For the analyses, we selected only control schools and classrooms that participated from the first wave onwards ($n = 5738$). In total, 88.0% of students ($n = 5048$) had active parental consent and were included in our analyses. Not all classroom compositions remained stable. Classrooms with more than 20% change in composition were excluded, resulting in a sample of $n = 4833$ students. To ensure reliability and validity of peer-nomination scores, we selected classrooms with at least 10 participating students at all waves ($n = 4376$; cf. Laninga-Wijnen et al., 2022) and with a participation rate of at least 40%, resulting in a sample of 3680 students in 211 classrooms (including 112 seventh-grade classrooms). Participants were on average 13.94 years of age (SD = 0.74), 53.0% identified themselves as girls, and the majority of participants were born in Finland (83.9%).

The online questionnaires were completed during regular teaching hours and was supervised by teachers who had been thoroughly instructed 2 weeks prior to data collection. The participants were ensured of the confidentiality of their answers and it was made clear to them that participation to the study was voluntary and they could withdraw at any time. The order of items and scales was randomized. At the beginning of each questionnaire session, students learned about the definition of bullying (Olweus, 1996), emphasizing the repetitive nature of bullying as well as the power imbalance between the bully and the victim. The data collection was done in accordance with the Declaration of Helsinki and the recommendations of the Ethics Board of the University of Turku with written informed consent from all subjects and their parents.

Measures

Defending

Defending was assessed by means of the Participant Role Questionnaire (PRQ; Salmivalli & Voeten, 2004), including

two items that described bully-directed forms of defending (i.e., "tries to make others stop bullying" and "tells the others to stop bullying"). All students could nominate an unlimited number of classmates for each item. For each student, the number of received nominations was divided by the number of possible nominations within each class and multiplied by 100, in order to retrieve scores that represent the percentage of classmates that nominated a student for that item. These percentages could theoretically range from 0 to 100. The two items correlated strongly and significantly with each other across time-points, varying from $r = .768$ to $r = .821$ (all $p < .001$). Consequently, these two items were averaged to create one scale for defending.

Popularity

Popularity was assessed across the three waves using a peer-nominated question: "who are the most popular?" (cf. Peeters et al., 2021). Students could nominate an unlimited number of classmates. Again, the number of received nominations was divided by the number of possible nominate within each class, and multiplied by 100, resulting in percentage scores.

Being liked

The degree to which each student was liked by peers was assessed across all three waves using the peer-nominated item: "who are the ones you like the most?" (cf. Laninga-Wijnen et al., 2019). Students were allowed to nominate an unlimited number of classmates. Percentage scores of being liked were computed by dividing the number of received nominations by the number of potential nominations within each classroom, and multiplied by 100.

Empathy for victims at T1 and T2

We assessed empathy for victims of bullying with the following four items: When the bullied student starts to cry, I also feel bad; When someone is bullied, I start to get angry on his/her behalf; When the bullied student feels sad, I want to comfort him/her; When the bullied student is sad, I also feel sad (see Kärnä et al., 2013). These items cover affective empathy, which refers to the capacity to sympathize with victims. We assessed affective, rather than cognitive empathy (the capacity to understand the emotions of victims), as affective empathy has been found to be a more important predictor of defending (Peets et al., 2015; van der Ploeg et al., 2017). Answers were provided on a 4-point Likert scale from 0 (*never true*) to 3 (*always true*). The scale was internally consistent ($\alpha_{t1} = .88$, $\alpha_{t2} = .90$).

Anti-bullying injunctive norms at T1

Participants indicated the extent to which they agreed or disagreed with three anti-bullying propositions that were based on Rigby's and Slee's pro-victim scale (1991). Students answered on a 5-point Likert scale, ranging from 0 = *totally disagree* to 4 = *totally agree*; thus higher scores indicate stronger anti-bullying attitudes. An example item is: "It is wrong to join in bullying." The reliability of this subscale was Cronbach's $\alpha = .70$, indicating adequate internal consistency. For each participant, these items were averaged into a mean score. In turn, these mean scores were aggregated at the classroom level, to create a measure of anti-bullying injunctive norms. There was high between-classroom variability in anti-bullying norms, with means ranging from 1.81 to 3.65 ($M = 2.88$, SD = 0.31). Anti-bullying norms were lower in eighth grade than in seventh grade ($M_{grade7} = 2.93$, SD = 0.29; $M_{grade8} = 2.83$, SD = 0.33), as indicated by an independent t-test, with $t(209) = 2.25$, $p = .026$. Anti-bullying norms were also lower in classrooms with a higher percentage of boys ($r = -.291$, $p < .001$).

Covariates

Students were asked about their gender, which was coded as 0 = girl, 1 = boy. Age in years was also entered as covariate, as were anti-bullying attitudes at the individual level (Rigby & Slee, 1991).

Analyses

First, because our selection criteria (e.g., having stable classroom composition and sufficient participants per classroom, cf. Garandeau et al., 2021; Kaufman et al., 2022) led to a relatively high sample reduction, we examined whether included students differed from the excluded students in our variables of interest. Multivariate analyses indicate that included students did not significantly differ from excluded students in empathy, anti-bullying attitudes, popularity, being liked, and defending at T1 (all p's > .088). At T2, excluded students scored significantly higher on being liked than included students ($F(4698) = 4.82$, $p = .028$), however, the effect size was very small, with partial $\eta^2 = .001$. A similar pattern for being liked was present at T3 ($F(4422) = 23.73$, $p < .001$), but again, the effect size was very small (partial $\eta^2 = .005$). The excluded group was also slightly more popular ($F(4422) = 6.95$, $p = .008$, partial $\eta^2 = .002$). In the total sample included in the analyses, on average, 5.3% of the data was missing on the variables of interest across waves.

Second, to test our hypotheses, we conducted cross-lagged panel analyses on the three data waves using path analyses in M*plus* version 8.3 (Muthén & Muthén, 2017). This allowed us to examine the predictive role of popularity and being liked in defending—as well as the predictive role of defending in the two types of status—with time intervals of about 4–5 months, while controlling for the prospective links in the other temporal direction. In the analyses without classroom norms, we controlled for the interdependence of students being nested within classrooms using

the CLUSTER option. In the analyses including classroom norms, we conducted multilevel analyses. We centered all individual-level variables at the group mean, because we were interested in students' status positions and behaviors relative to others in their classroom. Also, applying group-mean centering is appropriate when testing cross-level interactions (Enders & Tofighi, 2007). Classroom-level variables (e.g., anti-bullying norms) were centered at the grand mean.

To test the moderating role of empathy in the extent to which status predicts defending, we calculated four interaction terms (e.g., empathy T1*popularity T1; empathy T2*being liked T2). These interaction terms were simultaneously included in the model.

Multi-group cross-lagged panel analyses were performed to test whether bidirectional associations differed for boys and girls. Using the Satorra Bentler scaled chi-square difference test (Satorra & Bentler, 1994), we compared a model in which all parameters were constrained to be equal for boys and girls (constrained model) with a model in which all parameters were freely estimated for boys and girls (unconstrained model). In case the unconstrained model fits the data significantly better than the unconstrained model, the beta estimates of this unconstrained model and their 95% CIs for the cross-lagged effects were compared for boys and girls. Differences between beta estimates were considered significant only when the 95% CIs of the beta estimates for boys did not include the beta estimates for girls and vice versa (Pfister & Janczyk, 2013). Lastly, multi-level cross-lagged panel analyses with cross-level interactions were conducted to examine the moderating role of anti-bullying norms in the longitudinal bidirectional links between status and defending.

We used maximum likelihood estimations with robust standard errors (MLR; Mueller & Hancock, 2008). Model fit precision was examined using the chi-square statistic (χ^2), comparative fit index (CFI), Tucker-Lewin Index (TLI), root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR; Kline, 2005). We used full information maximum likelihood estimation to handle missing data.

RESULTS

Descriptive statistics

Means, standard deviations, and individual-level correlations are presented in Table 1. Correlations are provided for the whole sample, as well as for boys and girls separately. Correlations between defending and being liked were moderately positive, and correlations between defending and being popular were positive yet small across time. Correlation coefficients of defending and status were positive and moderate for girls, and positive and small for boys.

Bidirectional associations between status and defending

First, a cross-lagged panel model including popularity, being liked and defending at the three time points was run. The model fit was excellent, with $\chi^2(24) = 235.717$, $p < .001$, RMSEA = 0.051, CFI = 0.983, TLI = 0.950, SRMR = 0.037. Results of this baseline model (with AIC = 219,722.67 and $BIC_{adjusted}$ = 219,917.80) are displayed in the first column of Table 2.

TABLE 1 Means, standard deviations, and correlations among main variables of interest for the total sample and for boys and girls separately.

	M (SD) all students	M (SD) boys	M (SD) girls	1.	2.	3.
1. Age	13.94 (0.74)	13.97 (0.75)[a]	13.93 (0.72)[a]		.002	.014
2. Def T1	9.8 (11.2)	6.5 (8.1)[a]	13.1 (12.8)[b]	.020/.000		.693*
3. Def T2	8.6 (10.5)	5.6 (7.5)[a]	11.6 (12.2)[b]	.057*/−.032	.692*/.586*	
4. Def T3	6.6 (8.9)	4.7 (7.1)[a]	8.7 (10.1)[b]	.077*/.032	.578*/.437*	.674*/.543*
5. Lik T1	25.5 (14.8)	25.7 (15.4)[a]	26.5 (14.1)[a]	.039/−.001	.377*/.320*	.336*/.225*
6. Lik T2	25.8 (14.2)	24.9 (14.2)[a]	27.6 (14.1)[b]	−.015/.007	.271*/.239*	.343*/.254*
7. Lik T3	22.3 (14.4)	21.4 (14.3)[a]	24.1 (14.3)[b]	.030/.051*	.278*/.263*	.320*/.234*
8. PopT1	12.4 (18.7)	13.4 (19.4)[a]	11.7 (18.5)[b]	.078*/.060*	.275*/.147*	.270*/.142*
9. Pop T2	11.1 (17.2)	12.3 (18.0)[a]	10.3 (16.6)[b]	.064*/.060*	.271*/.133*	.304*/.141*
10. Pop T3	9.7 (15.3)	10.2 (15.9)[a]	9.2 (15.0)[a]	.071*/.041	.286*/.132*	.308*/.131*
11. Emp T1	1.16 (0.73)	0.93 (0.69)[a]	1.39 (0.69)[b]	−.011/.002	.275*/.264*	.246*/.200*
12. Emp T2	1.17 (0.75)	0.91 (0.72)[a]	1.39 (0.70)[b]	.031/.013	.239*/.224*	.234*/.206*

Note: T1, T2, and T3 refer to Timepoint 1, 2, and 3 respectively. Lik, being liked; Def, defending; Pop, popularity. $N \approx 3167$. Mean values are the non-centered values. For status and defending variables, means reflect the average percentage of classmates that nominated someone for these variables. Pearson correlations in the upper diagonal are for the entire sample. In the lower diagonal, correlations are for girls/boys. Different superscripts across rows 2 and 3 indicate significant differences between boys and girls.

*$p < .05$.

Status as predictor of future defending

Peer status was not predictive of future defending, with only one exception: being popular at T1 positively predicted defending at T2. The other findings are in contrast with our hypotheses that both being popular and liked would positively predict defending.

Defending as a predictor of status

Defending behavior at T1 had a positive effect on both popularity and being liked at T2, and defending at T2 had a positive effect on being liked at T3. Although the effect of defending at T2 on being popular at T3 was also in a positive direction, it did not reach statistical significance ($p = .094$). Thus, defending predicted a relative increase in popularity and—especially—in being liked.

The moderating role of empathy

In the next model, we tested whether empathy at T1 moderated the extent to which status at T1 predicted defending behavior at T2, and whether empathy at T2 moderated the extent to which status at T2 predicted defending behavior at T3. The model included four interaction terms (empathy*popularity and empathy*being liked at T1 and T2). Only one out of the four interaction terms was significant (empathy T1*popularity T1; $B = 0.022$, $SE = 0.010$, $p = .027$) indicating that, consistent with our expectations, popular youth were more likely to defend if they scored higher on affective empathy. However, this interaction barely explained any additional variance in defending

($<0.01\%$). Therefore, we decided not to interpret the interactions with empathy further and, for parsimony, not include them in subsequent analyses. Thus, findings did not support the hypothesis that the prospective link between status and defending would be moderated by empathy for victims.

The moderating role of gender

To test whether the bidirectional links between status and defending differed for boys and girls, we compared a fully constrained model (AIC = 219,146.22, $BIC_{adjusted} = 219,367.95$, $\chi^2(87) = 381.95$, $p < .001$, Correction Factor = 1.36) with an unconstrained model (AIC = 219,003.63; $BIC_{adjusted} = 219,340.66$, $\chi^2(48) = 255.17$, $p < .001$, Correction Factor = 1.17). The Satora-Bentler Scaled Chi-Square Difference was $\Delta\chi^2_{SB}(\Delta 39) = 138.60$, $p < .001$, indicating that the unconstrained model was preferred over the constrained model. The results of the unconstrained model with a model fit of RMSEA = 0.050, CFI = 0.984, TLI = 0.959, and SRMR = 0.039 are displayed in the second and third column of Table 2 and a summary of findings is presented in Figure 1.

Status as a predictor of defending

The unconstrained model indicated that the effect of popularity on future defending differed between girls and boys. Among girls, higher popularity consistently predicted higher defending across time-points, whereas for boys, only popularity at T1 positively predicted defending at T2. The confidence intervals of betas of popularity T1 on defending at T2

4.	5.	6.	7.	8.	9.	10.	11.	12
.049*	.018	−.008	.036*	.070*	.063*	.057*	−.013	.012
.568*	.339*	.270*	.283*	.200*	.191*	.204*	.332*	.300*
.659*	.284*	.317*	.298*	.197*	.215*	.218*	.292*	.288*
	.265*	.311*	.413*	.169*	.183*	.231*	.248*	.245*
.292*/.237*		.617*	.512*	.405*	.392*	.356*	.080*	.075*
.329*/.259*	.610*/.623*		.618*	.320*	.356*	.340*	.078*	.077*
.418*/.393*	.516*/.508*	.634*/.591*		.251*	.277*	.338*	.100*	.099*
.235*/.109*	.360*/.454*	.284*/.371*	.210*/.307*		.803*	.741*	.017	.015
.268*/.102*	.357*/.431*	.315*/.413*	.248*/.321*	.806*/.800*		.816*	.015	.022
.327*/.130*	.341*/.374*	.310*/.382*	.319*/.368*	.751*/.729*	.820*/.811*		.028	.029
.234*/.134*	.094*/.056*	.067*.037	.088*/.060*	.038./.027	.031/.036	.040/.044		.580*
.225*/.136*	.087*/.045	.055*/.039	.085*/.052*	.032/.026	.044/.031	.044/.032	.585*/.477*	

TABLE 2 Cross-lagged panel analyses testing bidirectional associations between defending and peer status for the entire sample and for boys and girls separately.

Outcome regressed on predictor	Total sample (n=3427)			Girls (n=1828)			Boys (n=1599)		
	B	SE	Beta (CI)	B	SE	Beta (CI)	B	SE	Beta (CI)
Defending T1 on boy	−5.928***	0.446	−0.631 (−0.701, −0.562)	–	–	–	–	–	–
Defending T1 on age	0.447**	0.160	0.035 (0.011, 0.060)	0.538	0.336	0.038 (−0.007, 0.084)	0.341	0.265	0.038 (−0.020, 0.097)
Defending T1 on anti-bullying attitudes	0.799***	0.198	0.080 (0.041, 0.120)	0.551	0.335	0.044 (−0.011, 0.099)	0.729***	0.159	0.101 (0.061, 0.142)
Defending T2 on defending T1	0.613***	0.028	0.682 (0.641, 0.723)[a]	0.618***	0.031	0.688 (0.642, 0.733)[a]	0.586***	0.047	0.645 (0.574, 0.717)[a]
Defending T2 on being liked T1	0.001	0.011	0.001 (−0.030, 0.032)	0.016	0.016	0.022 (−0.020, 0.063)[a]	−0.014	0.010	−0.031 (−0.074, 0.012)[b]
Defending T2 on popularity T1	0.028***	0.008	0.062 (0.030, 0.094)	0.040**	0.012	0.077 (0.033, 0.121)[a]	0.019**	0.007	0.059 (0.016, 0.102)[a]
Defending T2 on empathy T1	0.798***	0.183	0.065 (0.036, 0.093)	1.057**	0.279	0.069 (0.033, 0.105)[a]	0.309	0.241	0.032 (−0.016, 0.080)[a]
Defending T3 on defending T2	0.451***	0.028	0.571 (0.513, 0.629)	0.472***	0.031	0.591 (0.528, 0.654)[a]	0.398***	0.063	0.490 (0.405, 0.575)[a]
Defending T3 on defending T1	0.150***	0.020	0.211 (0.157, 0.266)	0.130***	0.021	0.181 (0.122, 0.240)[a]	0.184***	0.043	0.249 (0.176, 0.322)[a]
Defending T3 on being liked T2	0.010	0.008	0.019 (−0.011, 0.049)	0.007	0.011	0.012 (−0.024, 0.047)[a]	0.015	0.010	0.038 (−0.011, 0.086)[a]
Defending T3 on popularity T2	0.007	0.005	0.018 (−0.009, 0.045)	0.025**	0.009	0.053 (0.014, 0.092)[a]	−0.009	0.006	−0.032 (−0.072, 0.008)[b]
Being liked T1 on boy	−0.880	0.642	−0.069 (−0.167, 0.030)	–	–	–	–	–	–
Being liked T1 on age	0.120	0.204	0.007 (−0.016, 0.030)	0.422	0.415	0.025 (−0.022, 0.071)[a]	−0.203	0.415	−0.012 (−0.058, 0.035)[a]
Being liked T2 on being liked T1	0.579***	0.018	0.605 (0.577, 0.633)	0.595***	0.024	0.610 (0.571, 0.649)[a]	0.564**	0.020	0.607 (0.572, 0.642)[a]
Being liked T2 on defending T1	0.091***	0.024	0.070 (0.035, 0.105)	0.083**	0.027	0.071 (0.027, 0.115)[a]	0.106**	0.040	0.058 (0.015, 0.100)[a]
Being liked T3 on being liked T2	0.470***	0.023	0.482 (0.440, 0.523)	0.502***	0.028	0.507 (0.459, 0.555)[a]	0.434***	0.029	0.454 (0.396, 0.511)[a]
Being liked T3 on being liked T1	0.193***	0.021	0.207 (0.163, 0.251)	0.182***	0.025	0.189 (0.137, 0.240)[a]	0.207***	0.027	0.232 (0.174, 0.291)[a]
Being liked T3 on defending T2	0.095***	0.027	0.067 (0.028, 0.107)	0.080**	0.030	0.062 (0.016, 0.108)[a]	0.125**	0.049	0.065 (0.011, 0.119)[a]
Popularity T1 on boy	1.615*	0.625	0.088 (0.021, 0.155)	–	–	–	–	–	–
Popularity T1 on age	1.589***	0.314	0.064 (0.040, 0.088)	1.700***	0.006	0.069 (0.031, 0.107)[a]	1.470*	0.537	0.059 (0.017, 0.101)[a]
Popularity T2 on popularity T1	0.728***	0.019	0.796 (0.774, 0.819)	0.709***	0.026	0.793 (0.759, 0.827)[a]	0.746***	0.023	0.799 (0.774, 0.824)[a]

TABLE 2 (Continued)

Outcome regressed on predictor	Total sample (n = 3427)			Girls (n = 1828)			Boys (n = 1599)		
	B	SE	Beta (CI)	B	SE	Beta (CI)	B	SE	Beta (CI)
Popularity T2 on defending T1	0.072*	0.030	0.040 (0.007, 0.073)	0.095**	0.036	0.062 (0.017, 0.107)[a]	0.027	0.047	0.010 (−0.025, 0.045)[b]
Popularity T3 on popularity T2	0.544***	0.026	0.615 (0.561, 0.669)	0.541***	0.034	0.609 (0.539, 0.679)[a]	0.546***	0.033	0.619 (0.549, 0.690)[a]
Popularity T3 on popularity T1	0.200***	0.022	0.248 (0.193, 0.319)	0.199***	0.028	0.250 (0.179, 0.321)[a]	0.201***	0.031	0.244 (0.171, 0.317)[a]
Popularity T3 on defending T2	0.043†	0.026	0.025 (−0.005, 0.063)	0.058*	0.029	0.038 (0.000, 0.077)[a]	0.005	0.044	0.002 (−0.032, 0.036)[a]
Covariances									
Defending T1 with being liked T1	34.63***	3.27	0.306 (0.263, 0.349)	43.24***	4.57	0.334 (0.281, 0.386)	24.82***	2.99	0.283 (0.233, 0.334)
Defending T1 with popularity T1	42.24***	4.47	0.263 (0.218, 0.307)	57.31***	6.96	0.310 (0.251, 0.368)	24.69***	4.14	0.199 (0.142, 0.257)
Being liked T1 with popularity T1	99.68***	6.45	0.426 (0.388, 0.464)	86.31***	7.55	0.387 (0.337, 0.437)	114.94***	8.37	0.466 (0.421, 0.512)
Defending T2 with being liked T2	5.94***	1.13	0.111 (0.071, 0.151)	9.382***	1.67	0.157 (0.105, 0.208)	2.13	1.22	0.049 (−0.006, 0.103)
Defending T2 with popularity T2	5.33***	1.24	0.096 (0.054, 0.138)	7.24***	1.85	0.122 (0.064, 0.181)	3.24*	1.39	0.068 (0.013, 0.114)
Being liked T2 with popularity T2	12.40***	1.94	0.133 (0.094, 0.172)	7.93***	2.30	0.091 (0.041, 0.142)	17.47***	3.01	0.175 (0.119, 0.231)
Defending T3 with being liked T3	5.42***	1.06	0.139 (0.092, 0.185)	7.48***	1.48	0.171 (0.113, 0.229)	3.17**	1.10	0.098 (0.035, 0.150)
Defending T3 with popularity T3	4.42***	0.88	0.125 (0.079, 0.171)	6.38***	1.32	0.171 (0.106, 0.235)	2.22**	0.88	0.071 (0.017, 0.126)
Being liked T3 with popularity T3	13.09***	1.60	0.178 (0.137, 0.219)	11.96***	2.07	0.173 (0.119, 0.226)	14.48**	2.49	0.186 (0.123, 0.249)
Variance explained									
Defending T1	0.118	0.012		0.003	0.003		0.012	0.005	
Defending T2	0.555	0.023		0.533	0.029		0.429	0.043	
Defending T3	0.577	0.027		0.575	0.031		0.463	0.053	
Being liked T1	0.001	0.002		0.001	0.001		0.000	0.001	
Being liked T2	0.405	0.018		0.406	0.025		0.391	0.020	
Being liked T3	0.435	0.021		0.444	0.023		0.411	0.026	
Popularity T1	0.006	0.002		0.005	0.003		0.003	0.003	

TABLE 2 (Continued)

Outcome regressed on predictor	Total sample (n = 3427)			Girls (n = 1828)			Boys (n = 1599)		
	B	SE	Beta (CI)	B	SE	Beta (CI)	B	SE	Beta (CI)
Popularity T2	0.654	0.017		0.664	0.023		0.643	0.020	
Popularity T3	0.696	0.016		0.706	0.020		0.685	0.023	

Note: In the rows of girls and boys, different superscripts imply significant differences in parameters between boys and girls. In this Table, we left out the effects of control parameters on defending and status at T2 and T3 for parsimony, but these can be retrieved upon request.

$^{†}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

for boys and girls overlapped. However, the beta of popularity at T2 on defending behavior at T3 among girls did not fall in the confidence interval of the beta of popularity at T2 on defending T3 among boys and vice versa. Thus, across time points, popularity was a consistent, positive predictor of defending among girls, whereas it was less consistently predictive of defending among boys, which was in contrast to our hypothesis.

Being liked was mostly unrelated to future defending for both genders, with the exception of being liked at T1 being predictive of defending at T2 for boys. However, the beta estimate of being liked at T1 on defending for boys did fall in the confidence interval of the beta estimate of this effect for girls. Thus, being liked barely predicted the extent to which students engaged in defending, and this did not differ significantly between boys and girls.

Defending as a predictor of status

In contrast to our hypothesis that defending would be more strongly predictive of status among boys, our results indicate that defending predicted significantly higher levels of popularity over time only for girls. Moreover, the beta for the effect of defending at T1 on popularity at T2 for girls did not fall into the confidence interval of the beta for the effect of defending at T1 on popularity at T2 for boys, and vice versa. The beta for the effect of defending at T2 on popularity at T3 for girls also did not fall into the confidence interval of the beta for the effect of defending at T2 on popularity at T3 for boys, yet the beta of boys did fall into the confidence interval of girls. Together, these findings indicated that defending was predictive of popularity among girls but not among boys, which is in contrast to our hypothesis that defending would be a stronger predictor of popularity among boys.

Next, defending was consistently predictive of increases in being liked among both girls and boys across time points, and the beta estimates of these effects each fell into the confidence intervals for the other gender, indicating that the effects did not significantly differ for boys and girls. Thus, findings do not support our hypothesis that defending would be a stronger positive predictor of being liked among boys than among girls.

The moderating role of anti-bullying norms

To test whether anti-bullying norms moderate the extent to which popularity and being liked predict or result from defending, we first estimated the random slopes for eight paths (e.g., T2 popularity on T1 defending, T3 defending on T2 being liked, etc.). Almost all of these random slopes significantly varied across classrooms, except the random slope of T3 defending on T2 popularity ($p = .095$). Next, we included cross-level interactions to predict classroom-level variance in the eight random slopes.
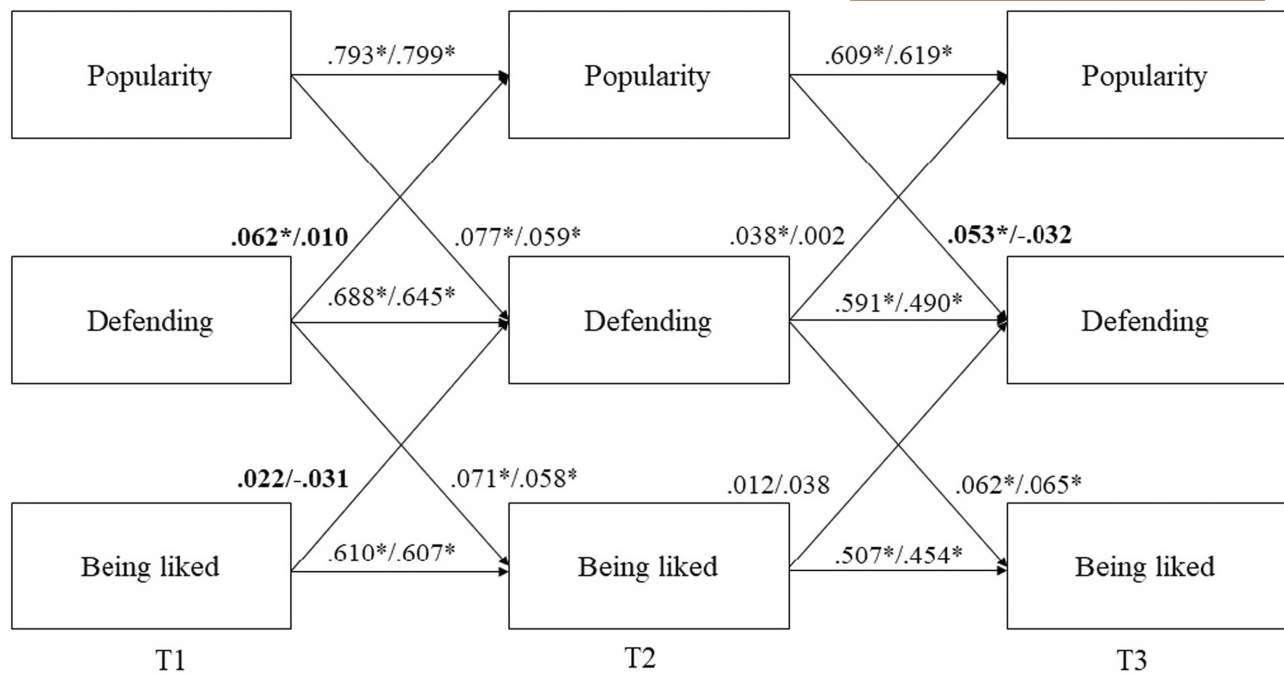
FIGURE 1 Bi-directional association between status and defending for girls/boys. *Note*: Standardized cross-lagged paths between status defending, separated for boys and girls. Bolded estimates vary significantly between boys and girls.

Status as predictor of defending across classrooms

Anti-bullying norms significantly moderated the link between being liked at T2 and relative increases in defending at T3. The moderating role of anti-bullying norms in the link between being popular at T2 and relative increases in defending at T3 touched upon significance levels ($p = .05$; Table 3). These two cross-level interactions are depicted in Figures 2 and 3, respectively. For the slope of defending T3 regressed on being liked at T2, the region of significance for the moderator was −.90 to .26. The simple slopes were significant for classrooms that had values on grand-mean centered anti-bullying norms *outside* of this region. The value of $M = −0.90$ was rare in our data ($n = 1$) but there were multiple classrooms ($n = 42$) scoring $M = 0.26$ or higher on anti-bullying norms. For these classrooms, the link between being liked at T2 and defending at T3 was significantly positive. For the slope of defending at T3 regressed on popularity at T2, simple slopes were significant for classrooms that had values on grand-mean centered anti-bullying norms *lower* than $M = −0.09$. Thus, in classrooms with low anti-bullying norms, popular youth were particularly likely to refrain from defending. The effects can be considered as small, as the variance in the random slopes explained by defending norms was only 7.7% for being liked and 17.2% for popularity. The role of anti-bullying norms in the other random slopes of T1 status to T2 defending was non-significant.

Defending as predictor of status across classrooms

There were no significant cross-level interactions for the effect of anti-bullying norms on the extent to which defending is *predictive* of status (Table 3). Thus, we did not find support for our hypothesis that anti-bullying norms would strengthen the rewarding effects of defending on students' status.

DISCUSSION

Even though defending may be helpful for victims of bullying (Laninga-Wijnen et al., 2022; Sainio et al., 2011), it is thus far unclear whether bully-directed forms of defending place defenders themselves at risk. These types of defending go against the goals and behaviors of powerful bullies, and therefore may require a high peer status *and* be potentially costly in terms of status. As anti-bullying interventions increasingly encourage youth to defend victims, the current study sought to clarify bidirectional associations between adolescents' peer status and bully-directed defending, and tested the moderating role of empathy, gender, and anti-bullying norms in these prospective links. Our findings show that it is generally safe for youths' peer status to defend victims—defending contributed to increases in being liked, as well as higher popularity among girls. Moreover, popularity was more strongly predictive of defending among girls than among boys; and classroom anti-bullying norms partly

TABLE 3 Multi-level cross-lagged panel analyses testing the role of anti-bullying norms in bidirectional associations between defending and peer status.

Outcome regressed on predictor	Random intercept model			Cross-level interaction model		
	B	SE	CI	*B*	SE	CI
Individual-level						
Defending T1 on boy	−5.933***	0.445	−6.806, −5.061	−5.933***	0.445	−6.806, −5.061
Defending T1 on age	0.820**	0.291	0.251, 1.390	0.820**	0.291	0.251, 1.390
Defending T1 on anti-bullying attitudes	0.798***	0.198	0.410, 1.186	0.798***	0.198	0.410, 1.186
Defending T2 on defending T1	0.609***	0.028	0.555, 0.663	0.609***	0.028	0.555, 0.663
Defending T2 on being liked T1	0.000	0.010	−0.020, 0.020	−0.001	0.010	−0.021, 0.020
Defending T2 on popularity T1	0.029***	0.008	0.013, 0.044	0.029**	0.008	0.012, 0.045
Defending T2 on empathy T1	0.814***	0.178	0.466, 1.162	0.823***	0.178	0.475, 1.171
Defending T3 on defending T2	0.448***	0.029	0.392, 0.504	0.449***	0.029	0.392, 0.507
Defending T3 on defending T1	0.148***	0.019	0.110, 0.186	0.148***	0.020	0.109, 0.187
Defending T3 on being liked T2	0.008	0.009	−0.009, 0.025	0.008	0.008	−0.009, 0.025
Defending T3 on popularity T2	0.007	0.005	−0.004, 0.018	0.006	0.005	−0.005, 0.016
Being liked T1 on boy	−0.877	0.639	−2.131, 0.376	−0.877	0.639	−2.131, 0.376
Being liked T1 on age	0.058	0.403	−0.732, 0.848	0.058	0.403	−0.732, 0.848
Being liked T2 on being liked T1	0.575***	0.018	0.539, 0.610	0.575***	0.018	0.539, 0.610
Being liked T2 on defending T1	0.113***	0.025	0.065, 0.161	0.110***	0.025	0.061, 0.158
Being liked T3 on being liked T2	0.468***	0.023	0.423, 0.512	0.463***	0.023	0.418, 0.508
Being liked T3 on being liked T1	0.192***	0.021	0.152, 0.232	0.192***	0.021	0.152, 0.232
Being liked T3 on defending T2	0.086**	0.008	0.030, 0.143	0.083***	0.032	0.022, 0.145
Popularity T1 on boy	1.587***	0.620	0.370, 2.803	1.587***	0.620	0.370, 2.803
Popularity T1 on age	3.404***	0.617	2.195, 4.616	3.404***	0.617	2.195, 4.616
Popularity T2 on popularity T1	0.724***	0.019	0.686, 0.761	0.724***	0.019	0.687, 0.761
Popularity T2 on defending T1	0.073*	0.030	0.015, 0.132	0.069*	0.031	0.008, 0.131
Popularity T3 on popularity T2	0.541***	0.025	0.491, 0.583	0.538***	0.026	0.487, 0.589
Popularity T3 on popularity T1	0.201***	0.022	0.157, 0.244	0.202***	0.022	0.158, 0.246
Popularity T3 on defending T2	0.033	0.026	−0.017, 0.083	0.017	0.027	−0.036, 0.070
Classroom-level						
Classroom-level defending T2 on anti-bullying norms	3.574*	1.378	0.873, 6.274	3.573*	1.378	0.873, 6.247
Classroom-level defending T3 on anti-bullying norms	3.134*	1.566	0.065, 6.202	3.134*	1.566	0.065, 6.202
Classroom-levels of being liked T2 on anti-bullying norms	1.049	1.281	−1.463, 3.560	1.049	1.281	−1.463, 3.560
Classroom-levels of being liked T3 on anti-bullying norms	2.170	1.748	−1.256, 5.596	2.170	1.748	−1.256, 5.596
Classroom-levels of being popular T3 on anti-bullying norms	1.476†	0.895	−0.278, 3.229	1.476†	0.895	−0.278, 3.229
Classroom-levels of being popular T2 on anti-bullying norms	1.688†	0.883	−0.043, 3.419	1.688†	0.883	−0.043, 3.419
Cross-level interactions						
Defending T2 on being liked T1*anti-bullying norm				−0.007	0.039	−0.084, 0.070
Defending T3 on being liked T2*anti-bullying norms				0.074*	0.035	0.005, 0.143
Defending T2 on being popular T1*anti-bullying norms				0.013	0.023	−0.033, 0.059

TABLE 3 (Continued)

Outcome regressed on predictor	Random intercept model			Cross-level interaction model		
	B	SE	CI	B	SE	CI
Defending T3 on being popular T2*anti-bullying norms				0.037[†]	0.019	0.000, 0.073
Being liked T2 on defending T1*anti-bullying norms				−0.014	0.083	−0.178, 0.149
Being liked T3 on defending T2*anti-bullying norms				0.087	0.084	−0.077, 0.252
Popularity T2 on defending T1*anti-bullying norms				−0.132	0.093	−0.314, 0.050
Popularity T3 on defending T2*anti-bullying norms				0.074	0.072	−0.067, 0.215

Note: Intraclass correlations for defending varied from .265 to .412 across waves, for being liked .194 to .247 across waves, and for being popular from .002 to .010. In this Table, we left out the effects of control parameters on defending and status at T2 and T3 for parsimony, but these can be retrieved upon request.

[†]$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$.
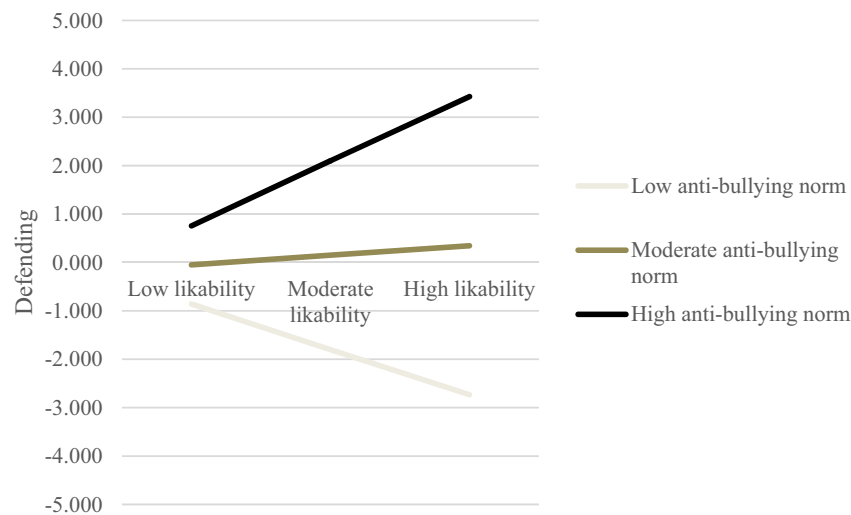
FIGURE 2 The moderating role of anti-bullying norms in the link between defending and acceptance, with low values being 2 SD under the mean, and high values being 2 SD above the mean.

enhanced the positive role of being well-liked and popular in defending victimized peers.

Is high status a prerequisite for defending? The role of gender, empathy, and norms

In our study, we found limited evidence that high peer status is a prerequisite of defending, and the extent to which this was the case mostly depended on gender and classroom norms. In the general analyses (without gender and norms as moderators), only the effect of being popular at T1 on defending at T2 was significant. This finding could indicate that having a larger social network may be perceived as a more solid position from which youth dare to stand up against bullies (Hodges & Perry, 1999). Yet, the extent to which status contributes to defending can be considered small, and no effect of being liked on defending

was detected after controlling for all other effects in the model.

The extent to which status predicts defending partly depended on adolescents' gender as well as on classroom norms. Specifically, and in contrast to our hypothesis, popularity was positively predictive of defending behavior among girls, but unrelated to defending among boys. One reason for this unexpected finding could be that, compared to boys, girls may feel more strongly that they need an influential social position before they dare to stand up against powerful bullies and their allies (Strindberg et al., 2020). It might also be that engaging in prosocial risky behavior as a way to demonstrate one's status is more common among girls than in boys. This contrasts with our expectations that bully-directed defending behaviors might be a way to demonstrate status, especially for boys. We based this expectation on the findings of prior work showing that boys were more likely to engage in risky
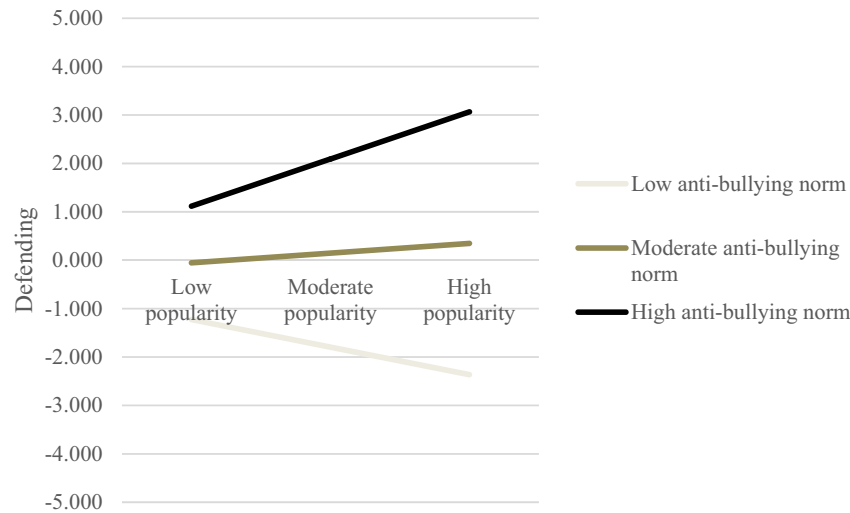
FIGURE 3 The moderating role of anti-bullying norms in the link between defending and popularity, with low values being 2 SD under the mean, and high values being 2 SD above the mean.

prosocial behaviors to "show off" with their status (Dreher et al., 2016; McAndrew, 2019). There may be reasons why our finding diverges from previous work. For instance, even though the bully-directed defending behaviors we assessed may be riskier than victim-directed defending behaviors (such as comforting victims), it could still be that they are not considered risky enough among boys, and therefore would not be perceived as appropriate for demonstrating their power. For instance, having aggressive confrontations with bullies is more likely to qualify as a risky, costly signal (indeed, these behaviors are more common among boys; Lambe & Craig, 2020) and we did not specifically assess whether attempts to stop the bully involved such aggressive or confrontational ways of defending. Thus, it is possible that we assessed bully-directed defending behaviors that are considered as more appropriate for girls, which is an interesting avenue for future studies.

Furthermore, we did not find evidence supporting the hypothesis that high status youth would be more likely to defend if they are high on empathy. This finding is surprising, as empathy has consistently been found to be positively associated with defending, and is considered to be a prerequisite for defending (e.g., Ma et al., 2019). However, our results suggest that empathy and status may both contribute to defending behavior, independently of each other. Thus, even when youth have low status, high levels of empathy may still predict increases in defending, as suggested by studies showing that low status youth (such as victims) still defend others (Malamut et al., 2021)—possibly due to their empathy.

Lastly, the extent to which high status predicted increases in defending behavior was—albeit to a limited extent—dependent on classroom norms: in classrooms with high anti-bullying norms, well-liked youth were more likely to show a relative increase in defending than in classrooms with low anti-bullying norms. In classrooms with

low anti-bullying norms, popular youth were more likely to refrain from defending than in classrooms with high anti-bullying norms. Even though the effects were not very strong, these findings suggest that both popular *and* well-liked youth may engage in cost–benefit analyses (Nickerson et al., 2014) before they defend victimized peers. In classrooms with high anti-bullying norms, higher status youth may perceive defending as safer and therefore may engage in defending more than in classrooms with low anti-bullying norms (Pouwels et al., 2019).

Is status a consequence of defending? The role of gender and norms

Our results quite consistently indicate that youth who defend victimized peers are more likely to show a relative increase in peer status over time. Thus, in contrast to what many youth may fear (Strindberg et al., 2020), defending can be *beneficial* rather than disadvantageous for one's peer status. The effect of defending on being liked was more consistently present than the effect of defending on being popular. Possibly, defending signals youths' genuine care about others' safety and wellbeing, which may elicit an affectionate response among those who profit from these defending behaviors or from those who are witnesses of these behaviors. Moreover, defending behaviors may be helpful in stopping bullying (Saarento et al., 2015) and fostering a positive classroom climate (Laninga-Wijnen et al., 2021). Therefore, defending may evoke feelings of appreciation and gratitude—not only among victims, but also among other classmates – and enhance the likability of defenders over time.

The finding that defending can also boost youths' popularity aligns with costly signaling theory (Grafen, 1990) which posits that engaging in public, risky prosocial behaviors can be a means to improve one's status over time.

DEFENDING AND PEER STATUS

JOURNAL OF
Research on Adolescence SRA Society for Research on Adolescence
www.s-r-a.org

15

Standing up against a powerful bully may foster respect and admiration among other peers, which in turn, would boost one's peer status. Importantly, and in contrast to our hypothesis, the beneficial effects of defending on popularity were only detected among girls and not among boys. Future studies should investigate why defending may be more strongly prospectively related with popularity among girls than among boys.

Inconsistent with our hypothesis, we did not detect a moderating effect of classroom anti-bullying norms on the extent to which defending contributes to status over time. These findings are not in line with the results from cross-sectional studies showing that the link between defending and status varied depending on classroom norms (Garandeau et al., 2022; Lucas-Molina et al., 2018; Peets et al., 2015). It is possible that the link detected in these studies was due to status being a stronger *predictor* of defending in classrooms with high anti-bullying norms, which we also found in the current study (even though effects were limited). It is possible that classroom anti-bullying norms only predict youth's anticipated reward from defending rather than the actual rewards of defending—future research should consider this possibility.

Yet, it should also be acknowledged that some measurement issues may have prevented us from detecting significant moderating effects of classroom norms. We assessed classroom bullying injunctive norms by aggregating individual scores in anti-bullying attitudes, hence we did not capture whether youth were accurately aware of their classmates' anti-bullying attitudes. However, we still considered aggregating individual attitudes to be a more valid way to capture norms than asking about classmates' anti-bullying attitudes. Indeed, many youth may not be accurate when reporting their classmates' anti-bullying attitudes, a phenomenon referred to as norm misperception or pluralistic ignorance (Perkins et al., 2011). Indeed, research has shown that children and adolescents do not sufficiently express their private anti-bullying attitudes leading their peers to underestimate how much their classmates disapprove of bullying and misperceive what is normative in their classroom (Dillon & Lochman, 2022; Perkins et al., 2011). Another limitation of our measurement of classroom anti-bullying norms was that the classroom average of individual anti-bullying attitudes does not capture the extent to which a classroom is heterogenous or homogenous with respect to these attitudes. For instance, norms may be different among girls versus boys in the same classroom (see Isaacs et al., 2013). Furthermore, in some classrooms, the average level of anti-bullying attitudes may be high because all classmates endorse anti-bullying attitudes (i.e., homogeneity), whereas a similarly high level of anti-bullying norms in other classrooms may result from a few adolescents reporting extremely strong anti-bullying attitudes (i.e., heterogeneity). In heterogeneous classrooms, it may be more difficult for youth to identify what the norm is, and therefore, behaving in ways that deviate from the average may have fewer consequences in terms of status.

Thus, the detection of the effects of norms on the extent to which defending affects future status may have been prevented by possible misperception of anti-bullying norms (Dillon & Lochman, 2022; Perkins et al., 2011), as well as high heterogeneity of anti-bullying attitudes within classrooms.

Strengths and limitations

The current study has several strengths. First, prospective links between status and defending in both temporal directions were tested. This does not only provide insights into the temporal precedence of effects but also allows a better understanding of how relative increases in status may relate to relative increases in defending and vice versa. The current study used a traditional cross-lagged panel model, rather than random-intercept cross-lagged panel models, because we were interested in predicting *relative*, between-person differences. Peer status is a *relative* construct in and of itself, and only has meaning when it is examined in relation to the status of others within a particular classroom, which warrants examining relative between-person changes. Second, the few previous studies that examined bidirectional links between defending and status often assessed only one type of status and focused on defending in general. Our study extends upon this work by capturing two distinct types of status (popularity and being liked) and by focusing on bully-directed ways of defending as these may be most risky for youths' status. Third, we aimed at explaining inconsistencies in prior work by examining both individual-level and contextual-level moderators in the prospective links between defending and status, and our findings provide key insights into how bidirectional associations between defending and popularity vary depending on students' gender and classroom norms.

The current study also has some limitations. First, our operationalization of classroom norms by aggregating individual-level attitudes may not be ideal. Even though it is a commonly used measure of norms (see Laninga-Wijnen & Veenstra, 2021), using peer nominations or youths' perceptions of what their classmates generally think about bullying may have allowed us to detect a stronger moderating effect of norms. Second, we did not examine by *whom* defenders were viewed as popular or well-liked. For instance, it is plausible that defenders are better-liked by the victims they defend and by bystanders who disapprove of the bullying; however, this may not be true for bullies and their allies. At the same time, the goal of the current study was to examine whether defending predicts youths' *general* reputation in the classroom and whether their general reputation predicts their defending behavior; and the finding that it is relatively safe for youths' general reputation to defend victims is important from an intervention perspective. Future studies are encouraged to investigate among which classmates defenders gain (or lose)

in popularity or acceptance over time. Third, the theoretical arguments that we presented to explain why high status may predict future defending, and vice versa, were not directly tested in the current study. Future work should investigate the mechanisms underlying the positive association between peer status and defending. For example, a desire to demonstrate one's social power to others or a feeling of being protected by a network of peers may play a role in students' cost–benefit analysis. The reasons why defending benefits youths' status also remain unclear and should be examined in future studies.

CONCLUSION AND PRACTICAL IMPLICATIONS

The current study indicates that, in contrast to what many adolescents may fear, defending victimized peers does not jeopardize one's status among peers. In fact, defending may even increase one's acceptance and—for girls—one's popularity. These findings are good news for anti-bullying interventions: Defending victimized peers does not put youth at risk of losing status and may even be rewarding. Thus, these findings suggest that the practice of anti-bullying programs to encourage youth to defend victims is not harmful, at least, not in terms of their status. Importantly, our findings require further validation by also examining whether defending might affect other aspects of defenders' social adjustment (see also Malamut et al., 2022). Moreover, whether victims indeed profit from being defended should be critically evaluated. Even though some work suggests that defended victims have better psychological adjustment (Sainio et al., 2011) and higher feelings of belonging to the classroom (Laninga-Wijnen et al., 2022) than non-defended victims, there is increasing recognition that defending may not be as helpful for victims as desired (Gaffney et al., 2021). Thus, more research should be devoted to whether, how, and when defending helps victims by stopping the bullying and improving their adjustment. Standing up for others in general (i.e., when somebody is being discriminated or threatened) is an important citizen skill which should be taught to youth, not only to help others, but also to help oneself: a previous review has shown that adolescents reap emotional benefits from helping others (Fuligni, 2019). The current study has expanded this evidence by showing that defenders also reap social benefits from helping victims.

ORCID

Lydia Laninga-Wijnen https://orcid.org/0000-0001-6158-8950
Sarah T. Malamut https://orcid.org/0000-0001-5907-2752
Claire F. Garandeau https://orcid.org/0000-0002-0395-020X

REFERENCES

Armstrong-Carter, E., Do, K. T., Duell, N., Kwon, S. J., Lindquist, K. A., Prinstein, M. J., & Telzer, E. H. (2023). Adolescents' perceptions of social risk and prosocial tendencies: Developmental change and individual differences. *Social Development*, *32*, 188–203. https://doi.org/10.1111/sode.12630

Batson, C. D., Dyck, J. L., Brandt, R., Batson, J. G., Powell, A. L., McMaster, M. R., & Griffitt, C. (1988). Five studies testing two new egoistic alternatives to the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, *55*, 52–77. https://doi.org/10.1037/0022-3514.55.1.52

Boone, J. L. (1998). The evolution of magnanimity: When is it better to give than to receive? *Human Nature*, *9*, 1–21.

Caravita, S., DiBlasio, P., & Salmivalli, C. (2009). Unique and interactive effects of empathy and social status on inovlvement in bullying. *Social Development*, *18*, 140–163.

Craig, W. M., & Pepler, D. J. (1998). Observations of bullying and victimization in the school yard. *Canadian Journal of School Psychology*, *13*(2), 41–59. https://doi.org/10.1177/082957359801300205

Dawes, M., & Xie, H. (2017). The trajectory of popularity goal during the transition to middle school. *Journal of Early Adolescence*, *37*(6), 852–883. https://doi.org/10.1177/0272431615626301

Deng, X., Yang, J., & Wu, Y. (2021). Adolescent empathy influences bystander defending in school bullying: A three-level meta-analysis. *Frontiers in Psychology*, *12*, 690898. https://doi.org/10.3389/fpsyg.2021.690898

Dillon, C. E., & Lochman, J. E. (2022). Correcting for norm misperception of anti-bullying attitudes. *International Journal of Behavioral Development*, *46*, 443–452. https://doi.org/10.1177/0165025419860598

Do, K. T., Guassi Moreira, J. F., & Telzer, E. H. (2017). But is helping you worth the risk? Defining prosocial risk taking in adolescence. *Developmental Cognitive Neuroscience*, *25*, 260–271. https://doi.org/10.1016/j.dcn.2016.11.008

Dreher, J.-C., Dunne, S., Pazderska, A., Frodl, T., Nolan, J. J., & O'Doherty, J. P. (2016). Testosterone causes both prosocial and antisocial status-enhancing behaviors in human males. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 11633–11638. https://doi.org/10.1073/pnas.1608085113

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, *100*(3), 283–308. https://doi.org/10.1037/0033-2909.100.3.283

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Fuligni, A. J. (2019). The need to contribute in adolescence. *Perspectives on Psychological Science*, *14*(3), 331–343. https://doi.org/10.1177/1745691618805437

Gaffney, H., Ttofi, M. M., & Farrington, D. P. (2021). What works in anti bullying programs? Analysis of effective intervention components. *Journal of School Psychology*, *85*, 37–56. https://doi.org/10.1016/j.jsp.2020.12.002

Garandeau, C. F., Laninga-Wijnen, L., & Salmivalli, C. (2021). Effects of the KiVa anti-bullying program on affective and cognitive empathy in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *51*, 515–529.

Garandeau, C. F., Vermande, M. M., Reijntjes, A. H. A., & Aarts, E. (2022). Classroom bullying norms and peer status: Effects on victim-oriented and bully-oriented defending. *International Journal of Behavioral*

DEFENDING AND PEER STATUS

JOURNAL OF
Research on Adolescence SRA Society for Research on Adolescence
www.s-r-a.org

17

Development, *46*(5), 401–410. https://doi.org/10.1177/0165025419894722

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, *144*, 517–546.

Hawley, P. H. (2014). Evolution, prosocial behavior, and altruism: A roadmap for understanding where the proximate meets the ultimate. In L. Padilla-Walker & G. Carlo (Eds.), *Prosocial development: A multidimensional approach* (pp. 43–69). Oxford University Press.

Hine, B. A. (2017). Identifying the male prosocial niche: The gendertyping of prosocial behaviour across childhood and adolescence. *European Journal of Developmental Psychology*, *14*, 206–220. https://doi.org/10.1080/17405629.2016.1186009

Hodges, E. V. E., & Perry, D. G. (1999). Personal and interpersonal antecedents and consequences of victimization by peers. *Journal of Personality and Social Psychology*, *76*, 677–685.

Huitsing, G., Snijders, T. A., Van Duijn, M. A., & Veenstra, R. (2014). Victims, bullies, and their defenders: A longitudinal study of the coevolution of positive and negative networks. *Development and Psychopathology*, *26*(3), 645–659.

Isaacs, J., Voeten, M., & Salmivalli, C. (2013). Gender-specific or common classroom norms? Examining the contextual moderators of the risk for victimization. *Social Development*, *22*(3), 555–579.

Jones, L. M., Mitchell, K. J., & Turner, H. A. (2015). Victim reports of bystander reactions to in person and online peer harassment: A national survey of adolescents. *Journal of Youth and Adolescence*, *44*, 2308–2320.

Kärnä, A., Voeten, M., Little, T. D., Alanen, E., Poskiparta, E., & Salmivalli, C. (2013). Effectiveness of the KiVa anti-bullying program: Grades 1–3 and 7–9. *Journal of Educational Psychology*, *105*, 535–551. https://doi.org/10.1037/a0030417

Kaufman, T. M. L., Laninga-Wijnen, L., & Lodder, G. (2022). Are victims of bullying primarily social outcasts? Person-group dissimilarities in relational, socio-behavioral and physical characteristics as predictors of victimization. *Child Development*, *93*, 1458–1474.

Kim, M. J., Lee, J. W., & Kim, D. M. (2013). The effects of affective empathy, self-efficacy and social responsibility on defending behaviors in bullying: Focused on the moderating effects of perceived popularity. *Journal of Asia Pacific Counseling*, *3*(2), 139–150.

Kline, R. (2005). *Principles and practice of structural equation modeling*. Guilford Press.

LaFontana, K. M., & Cillessen, A. H. N. (2010). Developmental changes in the priority of perceived status in childhood and adolescence. *Social Development*, *19*, 130–147.

Lambe, L. J., & Craig, W. M. (2020). Peer defending as a multidimensional behavior: Development and validation of the defending behaviors scale. *Journal of School Psychology*, *78*, 38–53.

Laninga-Wijnen, L., Gremmen, M. C., Dijkstra, J. K., Veenstra, R., Vollebergh, W. A. M., & Harakeh, Z. (2019). The role of academic status norms in friendship selection and influence processes related to academic achievement. *Developmental Psychology*, *55*(2), 337–350. https://doi.org/10.1037/dev0000611

Laninga-Wijnen, L., Van den Berg, Y., Garandeau, C. G., Mulder, S., & Orobio de Castro, B. (2022). Does being defended relate to decreases in victimization and improved psychological adjustment among victims? *Journal of Educational Psychology*, *115*(2), 363–377.

Laninga-Wijnen, L., Van den Berg, Y. H. M., Mainhard, T., & Cillessen, A. H. N. (2021). The role of aggressive peer norms in elementary school children's perceptions of classroom peer climate and school adjustment. *Journal of Youth and Adolescence*, *50*, 1582–1600. https://doi.org/10.1007/s10964-021-01432-0

Laninga-Wijnen, L., & Veenstra, R. (2021). Peer similarity in adolescent social networks: Types of selection, influence, and factors contribution to openness to peer influence. In B. Halpern-Felsher (Ed.), *The encyclopedia of child and adolescent health* (pp. 1–9). Elsevier.

Laursen, B., & Veenstra, R. (2021). Toward understanding the functions of peer influence: A summary and synthesis of recent empirical research. *Journal of Research on Adolescence*, *31*, 889–907. https://doi.org/10.1111/jora.12606

Lucas-Molina, B., Perez-Albeniz, A., Fonseca-Pedrero, E., & Gimenez-Dası, M. (2018). Bullying, defending, and outsider behaviors: The moderating role of social status and gender in their relationship with empathy. *Scandinavian Journal of Psychology*, *59*, 473–482. https://doi.org/10.1111/sjop.12453

Ma, T. L., Meter, D. J., Chen, W. T., & Lee, Y. (2019). Defending behavior of peer victimization in school and cyber context during childhood and adolescence: A meta-analytic review of individual and peer-relational characteristics. *Psychological Bulletin*, *145*, 891–928. https://doi.org/10.1037/bul0000205

Malamut, S. T., Trach, J., Garandeau, C. F., & Salmivalli, C. (2021). Examining the potential mental health costs of defending victims of bullying: A longitudinal analysis. *Research on Child and Adolescent Psychopathology*, *49*, 1197–1210.

Malamut, S. T., Trach, J., Garandeau, C. F., & Salmivalli, C. (2022). Does defending victimized peers put youth at risk of being victimized? *Child Development*, *94*, 380–394. https://doi.org/10.1111/cdev.13866

McAndrew, F. T. (2019). *Costly signaling theory*. In T. Shackelford & V. Weekes-Shackelford (Eds.), *Encyclopedia of evolutionary psychological science* (pp. 1–8). Springer.

Meter, D. J., & Card, N. A. (2015). Effects of defending: The longitudinal relations among peer-perceived defending of victimized peers, victimization, and liking. *Social Development*, *24*(4), 734–747.

Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation modeling. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 488–508). SAGE Publications Inc.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Nickerson, A. B., Aloe, A. M., Livingston, J. A., & Feeley, T. H. (2014). Measurement of the bystander intervention model for bullying and sexual harassment. *Journal of Adolescence*, *37*, 391–400. https://doi.org/10.1016/j.adolescence.2014.03.003

Nielson, M. G., Padilla-Walker, L., & Holmes, E. K. (2017). How do men and women help? Validation of a multidimensional measure of prosocial behavior. *Journal of Adolescence*, *56*, 91–106. https://doi.org/10.1016/j.adolescence.2017.02.006

Olthof, T., Goossens, F. A., Vermande, M. M., Aleva, E. A., & Van der Meulen, M. (2011). Bullying as strategic behavior: Relations with desired and acquired dominance in the peer group. *Journal of School Psychology*, *49*, 339–359.

Olweus, D. (1996). Bully/victim problems at school: facts and effective intervention. *Reclaiming children and youth: Journal of Emotional and Behavioral Problems*, *5*(1), 15–22.

Peeters, M., Laninga-Wijnen, L., & Veenstra, R. (2021). Differences in adolescents' alcohol use and smoking behavior between educational tracks: Do popularity norms matter? *Journal of Youth and Adolescence*, *50*, 1884–1895.

Peets, K., Pöyhönen, V., Juvonen, J., & Salmivalli, C. (2015). Classroom norms of bullying alter the degree to which children defend in response to their affective empathy and power. *Developmental Psychology*, *51*(7), 913–920. https://doi.org/10.1037/a0039287

Perkins, H. W., Craig, D. W., & Perkins, J. M. (2011). Using social norms to reduce bullying: A research intervention among adolescents in five middle schools. *Group Processes & Intergroup Relations*, *14*(5), 703–722. https://doi.org/10.1177/1368430210398004

Pfister, R., & Janczyk, M. (2013). Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. *Advances in Cognitive Psychology*, *9*(2), 74–80. https://doi.org/10.2478/v10053-008-0133-x

Poorthuis, A. M. G., Slagt, M., van Aken, M. A. G., Denissen, J. J. A., & Thomaes, S. (2021). Narcissism and popularity among peers: A cross-transition longitudinal study. *Self and Identity*, *20*, 282–296. https://doi.org/10.1080/15298868.2019.1609575

Pouwels, J. L., van Noorden, T. H. J., & Caravita, S. C. S. (2019). Defending victims of bullying in the classroom: The role of moral responsibility and social costs. *Journal of Experimental Social Psychology*, *84*, 103–831. https://doi.org/10.1016/j.jesp.2019.103831

Pöyhönen, V., Juvonen, J., & Salmivalli, C. (2010). What does it take to stand up for the victim of bullying? The interplay between personal and social factors. *Merrill-Palmer Quarterly-Journal of Developmental Psychology*, *56*(2), 143–163.

Pozzoli, T., & Gini, G. (2013). Why do bystanders of bullying help or not? A multidimensional model. *The Journal of Early Adolescence*, *33*, 315–340. https://doi.org/10.1177/0272431612440172

Pozzoli, T., & Gini, G. (2021). Longitudinal relations between students' social status and their roles in bullying: The mediating role of self-perceived social status. *Journal of School Violence*, *20*, 76–88. https://doi.org/10.1080/15388220.2020.1850462

Pronk, J., Olthof, T., Aleva, E. A., Van der Meulen, M., Vermande, M. M., & Goossens, F. A. (2020). Longitudinal associations between adolescents' bullying-related indirect defending, outsider behavior, and peer-group status. *Journal of Research on Adolescence*, *30*, 87–99. https://doi.org/10.1111/jora.12450

Pronk, J., Olthof, T., Goossens, F. A., & Krabbendam, L. (2019). Differences in adolescents' motivations for indirect, direct, and hybrid peer defending. *Social Development*, *28*(2), 414–429. https://doi.org/10.1111/sode.12348

Reijntjes, A., Vermande, M. M., Olthof, T., Goossens, F. A., Aleva, E. A., & van der Meulen, M. (2016). Defending victimized peers: Opposing the bully, supporting the victim, or both? *Aggressive Behavior*, *42*, 585–597. https://doi.org/10.1002/ab.21653

Rigby, K., & Slee, P. T. (1991). Bullying among Australian school children: Reported behavior and attitudes toward victims. *Journal of Social Psychology*, *131*, 615–627.

Saarento, S., Boulton, A., & Salmivalli, C. (2015). Reducing bullying and victimization: Student- and classroom-level mechanisms of change. *Journal of Abnormal Child Psychology*, *43*, 61–76.

Sainio, M., Veenstra, R., Huitsing, G., & Salmivalli, C. (2011). Victims and their defenders: A dyadic approach. *International Journal of Behavioral Development*, *35*, 144–151.

Salmivalli, C., Laninga-Wijnen, L., Malamut, S. T., & Garandeau, C. F. (2021). Bullying prevention in adolescence: Solutions and new challenges from the past decade. *Journal of Research on Adolescence*, *31*, 1023–1046.

Salmivalli, C., & Voeten, M. (2004). Connections between attitudes, group norms, and behaviors associated with bullying in schools. *International Journal of Behavioral Development*, *28*, 246–258.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage.

Shaw, M. (1981). *Group dynamics* (3rd ed.). McGraw-Hill.

Smith, E. A., & Bliege Bird, R. L. B. (2000). Turtle hunting and tombstone opening: Public generosity as costly signaling. *Evolution and Human Behavior*, *21*, 245–261.

Spadafora, N., Marini, Z. A., & Volk, A. A. (2020). Should I defend or should I go? An adaptive, qualitative examination of the personal costs and benefits associated with bullying intervention. *Canadian Journal of School Psychology*, *35*(1), 23–40.

Strindberg, J., Horton, P., & Thornberg, R. (2020). The fear of being singled out: pupils' perspectives on victimisation and bystanding in bullying situations. *British Journal of Sociology of Education*, *41*, 942–957. https://doi.org/10.1080/01425692.2020.1789846

van der Ploeg, R., Kretschmer, T., Salmivalli, C., & Veenstra, R. (2017). Defending victims: What does it take to intervene in bullying and how is it rewarded by peers? *Journal of School Psychology*, *65*, 1–10.

Wright, J. C., Giammarino, M., & Parad, H. W. (1986). Social status in small groups: Individual-group similarity and the social "misfit". *Journal of Personality and Social Psychology*, *50*, 523–536. https://doi.org/10.1037/0022-3514.50.3

Yun, H. Y., & Graham, S. (2018). Defending victims of bullying in early adolescence: A multilevel analysis. *Journal of Youth and Adolescence*, *47*, 1926–1937.