

Abstract

Bullying has been studied for several decades, but its measurement remains a debated issue. The two main questions are whether measures should include a definition or specific behaviors and whether self- or peer-reports should be used. However, empirical data about the convergence of the different measures and their unique associations with students' psychosocial adjustment are lacking. This was the main goal of this study. A sample of 475 primary students completed the three most common types of bullying measures (self-reported behavior-based, self-reported definition-based, and peer-reported definition-based), and psychosocial adjustment indicators (prosocial behaviors, conduct problems, hyperactivity, emotional symptoms, peer problems, social acceptance, and social rejection). Results indicated that these three types of measures converge to some extent. Our model also revealed that self-reported behavior-based measures were more strongly associated with several students' psychosocial adjustment indicators. Overall, our findings support the idea that researchers and practitioners should select the type of measure that best fits their specific goals and constraints.

Keywords: bullying; victimization; peer-reported; assessment; psychosocial adjustment.

School bullying measurement: Convergence between self-reported behavior-based, definition-based, and peer-reported measures and their unique associations with students' psychosocial adjustment

School bullying is a form of repetitive aggressive behavior that is deliberately inflicted to cause harm and involves an imbalance of power between the author and the victim (Younan, 2019). School bullying has been linked to students' psychosocial adjustment, which refers to students' subjective well-being, psychopathology, and social adaptation (Rodríguez-Fernández et al., 2016; Zych et al., 2019). More specifically, being a victim of bullying (i.e., victimization) is positively associated with several internalizing symptoms, including depression, anxiety (Cook et al., 2010), and social isolation (Moore et al., 2017). Perpetrating bullying is positively associated with several externalizing behaviors, such as attention-deficit/hyperactivity disorder symptoms (Farrington & Baldry, 2010) and conduct problems (Cook et al., 2010), and negatively associated with prosocial behaviors (Volk et al., 2022). In addition, both bullies and victims are less socially accepted and more rejected by their peers (Wiertsema et al., 2023).

However, the measurement of bullying remains a major topic of debate among scholars and is considered the “Achilles’ heel of bullying research” (Volk et al., 2017, p. 36). The two main questions dividing researchers are whether bullying measures should include a definition of bullying, and whether self- or peer-reports should be used (Jia & Mikami, 2018). This ongoing debate is notably reflected in the most commonly used bullying measures, which are based on either a definition of bullying (definition-based; e.g., Inchley et al., 2020) or a description of bullying behaviors without definition (behavior-based; e.g., Joseph & Stockton, 2018) and are either self-reported (e.g., Solberg & Olweus, 2003) or peer-reported (e.g., Salmivalli et al., 1996). While these measures are partially correlated, some studies have suggested that they do not fully converge (Cornell & Brockenbrough, 2013; Green et al.,

2013). Therefore, it is unclear whether the different types of measures capture the same phenomenon. In addition, there is a lack of knowledge about the unique associations between these different bullying measures and students' psychosocial adjustment. As a result, the implications of using one measure rather than another are poorly understood, and practitioners lack clear guidelines to follow when selecting an appropriate scale for their objectives and context. Arguments for and against each type of measure are mainly based on theoretical considerations, and empirical data are lacking, which hinders progress in the ongoing debates about which scales should be preferred depending on the objective of the survey (Tolmatcheff & Veenstra, 2025; Volk et al., 2017; Xie et al., 2022).

Overall, further research is needed to provide empirical insights into the convergence of the most common types of bullying and victimization measures and how they are associated with students' psychological adjustment. This would help define evidence-based recommendations to help researchers and practitioners in selecting bullying measures, depending on the context and specific objectives. This study aimed to examine the convergent validity and associations between the three most common types of bullying measures and several indicators of students' psychosocial adjustment, including anxiety and depression, peer problems, attention-deficit/hyperactivity disorder symptoms, conduct problems, prosocial behaviors, social acceptance, and social rejection.

Bullying Measures With or Without Definition

Definition-based measures first present and define bullying to students before asking them to report how often they have bullied others (i.e., bullying) and been bullied by others (i.e., victimization). It has been argued that including a definition can help researchers and students reach a common understanding of bullying (Vivolo-Kantor et al., 2014; Volk et al., 2017) and distinguish it from other aggressive behaviors (Shaw et al., 2013) or students' own conceptions (Vaillancourt et al., 2008). Including a definition, it is claimed, is the best way to

capture the three characteristics of bullying: intention, repetition, and power imbalance (Vivolo-Kantor et al., 2014; Ybarra et al., 2014).

However, several studies have found that both perpetrators and victims of bullying do not always report bullying or victimization when bullying is explicitly defined (Kert et al., 2010; Lai & Kao, 2018). It has been argued that mentioning the word “bullying” can lead to underreporting due to the associated stigma (Thomas et al., 2014; Vessey et al., 2014). A second issue is the length of the definition, which ranges from 150 to more than 200 words and may be misunderstood by younger students (Hamby & Finkelhor, 2000; Swain, 1998). A final issue is that definition-based bullying measures sometimes rely on a single general item following the definition (e.g., Inchley et al., 2020; Solberg & Olweus, 2003), which increases the risk of random measurement error (Thomas et al., 2014).

Alternatively, some bullying measures do not use a definition and instead ask students to report how often they have experienced or engaged in various bullying behaviors within a given period (Bear et al., 2014; Joseph & Stockton, 2018). Describing specific behaviors may facilitate students’ understanding, as short items are easier to process than long definitions (Kaufman et al., 2020). In addition, specific behaviors may be easier for students to remember and report reliably (Huang & Cornell, 2015). However, bullying measures without a definition have been criticized for focusing on repetition at the expense of power imbalance and intentionality. Consequently, critics have argued that these measures provide a measure of general aggression rather than bullying (Vaillancourt et al., 2008; Ybarra et al., 2014).

Definition-based and behavior-based bullying measures have shown moderate convergence in identifying bullies and victims (Chen & Cheng, 2013; Felix et al., 2011). However, the overall pattern of findings is inconclusive, with some studies reporting different rates using different types of measures (Cook et al., 2009; Modecki et al., 2014) and others reporting similar rates (Huang & Cornell, 2015). On the other hand, similar associations have

been found between bullying and victimization and various indicators of students' psychosocial adjustment across studies, regardless of the presence of a definition (Cook et al., 2010). However, no study has simultaneously examined the unique associations between definition-based and behavior-based bullying measures and students' psychosocial adjustment by including both types of measures in a single model. Table 1 summarizes the theoretical arguments for and against both types of bullying measures.

Table 1

Summary of the Theoretical Strengths and Limitations of Bullying Measures With and Without a Definition of Bullying

With a definition	Without a definition (behavior-based)
• Presents the three characteristics of bullying explicitly: intention, repetition, and power imbalance • Helps researchers and responders reach a common understanding of bullying • Helps distinguish bullying from other aggressive behaviors	• Prevents the stigma attached to the word "bullying" • Relies on specific behavior items, which are easier for students to understand and reliably remember
• Definitions used vary across studies • Sometimes single-item, which increases the risk of random error and does not distinguish between indirect, direct, and cyberbullying • Risk of underreporting due to the stigma attached to the word "bullying" • Definitions are often lengthy and may be misunderstood by younger students	• Behaviors included can vary across studies • The power imbalance characteristic may be neglected, causing a possible overlap with general aggression

Note. Bullying measures can include both a definition and specific behaviors. However, such a combination is less commonly found among bullying measures.

Self-Reported or Peer-Reported Bullying Measures

While some studies rely on self-reported measures of bullying and victimization, others also use peer nominations (e.g., Kärnä et al., 2011). Self-reported scales have been more widely used for several reasons. First, students can be considered the most reliable informants about their own behaviors and experiences (Olweus, 2013; Shaw et al., 2013). Victims may sometimes be the only ones to know the details (e.g., frequency) of the bullying they experience (Thomas et al., 2014). In line with this, adults may need to know the victim's subjective perceptions of the bullying to determine whether an intervention is needed. However, a major shortcoming of self-reported measures is that they are subject to social desirability bias (Cole et al., 2006; Cornell & Brockenbrough, 2013).

Peer-reported bullying measures are typically definition-based and ask students to nominate those who bully others or are bullied from a list of classmates. Since bullying has been conceptualized as a social phenomenon (Salmivalli et al., 1996), peer-reported measures have been praised for providing information about the peer context (Volk et al., 2017). Peer-reported bullying measures prevent social desirability bias (Furlong & Sharkey, 2009; Green et al., 2018). Psychometrically, peer-reported measures have multiple reporters for one item, ensuring reliability and reducing random measurement error (Branson & Cornell, 2009).

However, peer-reported measures also have several limitations. First, nominations may be based on reputations rather than actual behavior (Casper et al., 2015; Hymel et al., 1990) and may be less suitable for capturing behavior change over relatively short periods of time (e.g., in an intervention study design). In addition, student relationships may affect the reliability of the answers, for instance, with students being reluctant to nominate their friends as bullies (Casper et al., 2015). Some students may also perceive the nominations as a form of “snitching” and answer inaccurately. As students are not trained to detect all forms of bullying, they may not consider less visible or indirect bullying (e.g., spreading rumors; Volk

et al., 2017). Peer-reported nominations are often limited to classmates for practical reasons, so that children cannot identify bullies outside their class (Casper et al., 2015; van der Ploeg et al., 2015). Because the validity of peer-reported measures is contingent on the number of respondents, results from small classes or classes with many absent students during data collection may be unreliable (Cillessen & Marks, 2011).

The few studies that have compared different bullying measures have reported correlations between peer-reported and self-reported measures ranging from small ($r < .20$) to moderate ($r = .46$) for bullying and from moderate ($r = .32$) to large ($r = .75$) for victimization (Cornell & Brockenbrough, 2013; Kärnä et al., 2011; Lee & Cornell, 2010). Similarly, the effects of anti-bullying programs have varied depending on whether a peer-reported or self-reported measure is used (van Verseveld et al., 2021). Some studies have found that peer-reported victims are less liked and more rejected by their peers, whereas self-reported victimization is more strongly associated with internalizing problems (Bouman et al., 2012; Košir et al., 2020). Regarding bullying perpetration, peer-reported measures correlate more with social rejection (Caravita et al., 2009; Juvonen et al., 2001; Yun & Espelage, 2024), whereas self-reported measures are more associated with internalizing problems such as depression and anxiety (Bouman et al., 2012; Yun & Espelage, 2024). Overall, although self- and peer-reported bullying measures seem to partially overlap, inconsistencies remain across studies. Therefore, knowledge about the distinct and unique associations of peer-reported and self-reported bullying measures with students' psychosocial adjustment is limited. Table 2 summarizes the main strengths and limitations of both types of reports.

The Present Study

The goal of this study was twofold. The first aim was to assess the convergent validity between the three most commonly used types of bullying and victimization measures: self-reported behavior-based, self-reported definition-based, and peer-reported definition-based

(Xie et al., 2022). The second aim was to provide insights into the unique associations between these three types of bullying measures and several key indicators of students' psychosocial adjustment: anxiety and depression, peer problems, attention-deficit/hyperactivity disorder symptoms, conduct problems, prosocial behaviors, social acceptance, and social rejection. To achieve these goals, we first examined the correlations among these three types of measures of bullying and victimization. We then simultaneously examined the unique associations between the different bullying and victimization measures and the seven students' psychosocial adjustment indicators in a single model.

Table 2

Summary of the Theoretical Strengths and Limitations of Self- and Peer-Reported Bullying Measures

Self-reported bullying measures	Peer-reported bullying measures
• Students are the most reliable informants about their own behaviors or experiences • Only the victims might know some details of the bullying they experience • Adults may need to know students' subjective perceptions to make intervention-related decisions	• Prevent social desirability bias • Reduce random measurement error because they rely on multiple informants • Provide information about the peer context
• Subject to social desirability bias	• Nominations may be based on reputations rather than behaviors • Student relationships may affect reliability • Students might be resistant if they see identifying a bully as "snitching" • Students may not consider less visible or indirect bullying • Often limited to classmates for practical reasons • Results may be unreliable if many absent students or smaller classes

We first assumed that measures assessing the same construct (bullying or victimization) would converge, and that the two self-reported measures (definition-based and behavior-based) would be more strongly correlated because of method-specific shared variance (Hypothesis 1). Because students can be considered the most reliable informants about their own behaviors and experiences (Olweus, 2013; Shaw et al., 2013), we hypothesized that both types of self-reported measures would be more strongly associated with students' psychosocial adjustment than peer-reported measures, except for social acceptance and social rejection, due to common method variance (Hypothesis 2). As definition-based measures are supposed to better capture the three bullying characteristics, which have been found to be associated with greater psychosocial maladjustment (Felix et al., 2011; Kaufman et al., 2020), we hypothesized that definition-based measures would be more strongly associated with students' psychosocial adjustment indicators than behavior-based measures (Hypothesis 3).

Materials and Methods

Participants and Procedure

Data were collected from six schools in French-speaking Belgium prior to program implementation in May 2021. At that time, no bullying-related activities had yet been delivered to students or teachers. The research team conducting this study was independent from the practitioner team responsible for program implementation and was not involved in any stage of the program delivery. All information about school size, official socioeconomic indices (SES), geographical location are available in Table S1. Passive parental consent was obtained through a letter sent by the schools. Sixty students (11%) did not participate because they were absent ($N = 25$), their parents objected to their participation ($N = 29$), or they themselves refused to participate ($N = 6$). Therefore, our final sample consisted of 475 students ($M_{\text{age}} = 10.37$ years; $SD = 1.07$; 54.5% girls).

Students were asked to fill out a confidential questionnaire on school climate and well-being during school hours. A research team member was present to explain the general procedure of the study and answer questions. The researcher read the questionnaire aloud in every classroom except in sixth grade, where students completed the questionnaire independently after receiving general instructions. This study and its procedure obtained the approval of the ethics committee of University of Louvain.

Measures

The behavior-based measures were presented first in the questionnaire to avoid the priming effect of the definition (Huang & Cornell, 2015). Next, the definition of bullying was presented, followed by the definition-based measures, then the peer-reported measures. The psychosocial adjustment indicators were presented at the end of the questionnaire.

Bullying and Victimization

Behavior-Based Bullying and Victimization. We used a 10-item scale validated for French-speaking elementary school students (Galand et al., 2014; Tolmatcheff et al., 2022), including direct (three items: insulting, hitting, scratching or pulling hair), indirect (four items: mocking or teasing, gossiping, excluding, hiding or damaging personal things), and cyber (three items: calling or texting, sending offensive or threatening messages through the Internet, publishing pictures or videos) forms of aggression. These scales have been validated among French-speaking elementary students and have consistently reported good psychometric indices over the years (e.g., Authors, 2014; 2022). For victimization, students had to report how many times they had undergone these specific behaviors since Christmas on a five-point Likert scale ranging from 0 (*0 times*) to 4 (*four times or more*). For bullying, students were asked to report how often they had engaged in each behavior toward another student since Christmas (from *0 times* to *four times or more*). Using Christmas as the time point of reference implied approximately two months (excluding holidays) before data

collection, which is a common timeframe in bullying research (Vivolo-Kantor et al., 2014; Xie et al., 2022). We used a bifactor model to concurrently assess the global construct and the three subdimensions (Morin et al., 2016). To evaluate the fit of our CFA and model, we used the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean squared error of approximation (RMSEA) and its related 90% confidence interval (CI), and the standardized root mean square residual (SRMR). The confirmatory factor analyses (CFA) showed a good fit for both perpetration ($\chi^2 [28] = 36.320, p = .13, CFI = .98, RMSEA = .025, 95\% CI: [0-.046]$, and $SRMR = .034$) and victimization ($\chi^2 [28] = 51.95, p = .004, CFI = .95, RMSEA = .043, 95\% CI: [.024-.06]$, and $SRMR = .037$). We used ordinal Cronbach's alpha (α) based on a polychoric matrix to estimate the reliability of our scales (Gadermann et al., 2012). The ordinal Cronbach's alpha was $\alpha = .86$ for bullying and $\alpha = .82$ for victimization.

Definition-Based Bullying and Victimization. We used the French-validated version of the Olweus Bully/Victim Questionnaire–Revised (OBVQ-r; Kubiszewski et al., 2014; Solberg & Olweus, 2003). After reading the definition of bullying, students were asked to report how many times they had been bullied since the Christmas holidays on a five-point Likert scale ($0 = I \text{ have never suffered from this}; 4 = \text{several times a week}$). We used this reference period to be consistent with the behavior-based measures and provide students with a concrete temporal marker. Then, students were asked to report how many times they had bullied others since the Christmas holidays ($0 = I \text{ have never done this}; 4 = \text{several times a week}$). The definition used in French in this study and the original version in English are provided in Appendix A.

Peer-Reported Bullying and Victimization. These single-item measures were based on the Participant Roles Questionnaire (Kärnä et al., 2011; Salmivalli et al., 1996) and were translated into French for this study. Although the definition of bullying was not repeated a second time, these measures followed the definition-based measures and thus relied on the

previously provided definition. Students received the list of their classmates and were asked to identify those who bullied others in their class. Next, students were asked to nominate classmates who were victimized by others. The number of nominations was unlimited. We summed up the number of nominations received for each student and divided it by the number of present students minus one (because students could not nominate themselves). The final scores ranged from 0 to 1.

Strengths and Difficulties Questionnaire

We used the French-validated version of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997; Capron et al., 2007) to assess five indicators of students' psychosocial adjustment: emotional symptoms (anxiety and depression), peer problems, attention-deficit/hyperactivity disorder symptoms, conduct problems, and prosocial behaviors. Students had to rate each item on a three-point Likert scale ranging from 0 (*not true*) to 3 (*certainly true*). The SDQ has been used in mental health and bullying research (Capron et al., 2007; Shaw et al., 2013), and has been translated and validated in French with satisfactory psychometric properties (Capron et al., 2007).

Prosocial Behaviors. Five items assessed the extent to which students acted with other people in a caring and prosocial way (e.g., "I try to be nice to other people"). CFA results indicated a good fit ($\chi^2 [5] = 5.8, p = .33, CFI = .99, RMSEA = .02, 95\% CI: [.0-.075]$, and SRMR = .025), and the ordinal Cronbach's alpha was $\alpha = .75$.

Attention-Deficit/Hyperactivity. Five items assessed externalized symptoms of attention-deficit/hyperactivity disorder (e.g., "I am restless. I cannot stay still for a long time"). We removed one reversed item ("I think before I do things") that decreased the reliability of the scale. CFA results showed a good fit ($\chi^2 [2] = 5.64, p = .06, CFI = .98, RMSEA = .067, 95\% CI: [0-.136]$, and SRMR = .024). The ordinal Cronbach's alpha was $\alpha = .62$.

Conduct Problems. Five items referred to students' disciplinary issues (e.g., "I get very angry and often lose my temper"). The results of the CFA indicated a good fit ($\chi^2 [3] = 85.42, p = .27, CFI = .98, RMSEA = .024, 95\% CI: [0-.078],$ and $SRMR = .024$). The ordinal Cronbach's alpha was below the conventional threshold ($\alpha = .52$).

Peer Problems. Five items referred to negative relationships with others and social isolation (e.g., "I am usually on my own"). The CFA results indicated an acceptable fit ($\chi^2 [5] = 11.1, p = .049, CFI = .93, RMSEA = .055, 95\% CI: [.003-.1],$ and $SRMR = .038$), and the ordinal Cronbach's alpha was $\alpha = .73$.

Emotional Symptoms (Anxiety and Depression). Five items assessed symptoms of anxiety and depression (e.g., "I worry a lot"). The CFA results indicated a good fit ($\chi^2 [5] = 0.7, p = .984, CFI = 1, RMSEA = 0, 95\% CI: [.0-.0],$ and $SRMR = .007$), and the ordinal Cronbach's alpha was $\alpha = .74$.

Social Acceptance and Social Rejection

Students received the list of their classmates and were asked to nominate those they liked the most and the least (Cillessen & Marks, 2011). They could nominate a maximum of 5 classmates but not themselves. We summed up the number of nominations received for social acceptance and social rejection for each student. We divided it by the number of present students minus one. The final scores ranged from 0 to 1.

Analytical Strategy

We first performed all descriptive analyses based on mean scores using SPSS (version 29). For the SDQ subscales, we then extracted factor scores and used these factor scores in the subsequent analyses to mitigate measurement error and account for the suboptimal reliability of some subscales. Next, we conducted a correlation matrix and a path analysis using Mplus (version 8.4) to estimate all regressions simultaneously within a single model.

The proportion of missing data at the item level ranged from 0.6% to 1.3% for behavior-based victimization, 0.6 to 1.5% for behavior-based bullying, and was 5.5% for the definition-based victimization and bullying single items. Missingness was higher for the SDQ subscales, ranging from 16.0% to 18.1%, likely reflecting their placement at the end of the questionnaire and associated fatigue or time effects (16.2–17.7% for prosocial behaviors, 16.0%–17.5% for attention-deficit/hyperactivity, 16.4%–18.1% for conduct problems, 16.6%–17.7% for peer problems, and 16.4%–17.1% for emotional symptoms). Although the missing completely at random (MCAR) test was significant ($\chi^2 [1079] = 1928.05, p < 0.001$), inspection of the missing data did not indicate strong systematic patterns. Therefore, we used full information maximum likelihood (FIML) estimation to handle missing data under the assumption that data were missing at random (MAR) (Newman, 2014). We used the robust maximum likelihood estimator to deal with the non-normality of our data. Since students were nested in their classrooms, standard errors were adjusted using a sandwich estimator to account for the non-independence of observations on the class level (Muthén & Muthén, 2017). We ran a path model with all six continuous measures of bullying and victimization as independent variables. The SDQ subscales (operationalized using extracted factor scores) and the social rejection and social acceptance measures were considered dependent variables (see Figure 1). Among the independent variables, we allowed for correlations among measures assessing the same construct, that is, between the three victimization measures on the one hand, and between the three bullying measures on the other. In addition, because bullying and victimization are moderately correlated (Wolke & Lereya, 2015), we allowed each victimization measure to covary with its equivalent bullying measure (e.g., behavior-based victimization with behavior-based bullying). Given the number of estimated paths, interpretation focused on the pattern and magnitude of standardized effects rather than on

individual *p*-values. We conducted additional exploratory analyses to examine the influence of gender and age as covariates.

Results

Table 3 shows the descriptive statistics and correlations between variables. The correlation between definition-based and behavior-based measures was moderate to high for victimization ($r = .60$) but was lower for bullying ($r = .44$). The correlations between the behavior-based and peer-reported measures were lower for bullying ($r = .27$) and victimization ($r = .39$). The weakest correlations were between the definition-based and peer-reported measures for both bullying ($r = .15$) and victimization ($r = .27$). Venn diagrams illustrating the overlap between students identified by the different bullying and victimization measures based on commonly used thresholds, are presented in the Supplementary Materials for descriptive purposes only. These cut-off-derived categorical classifications were not used in the analyses.

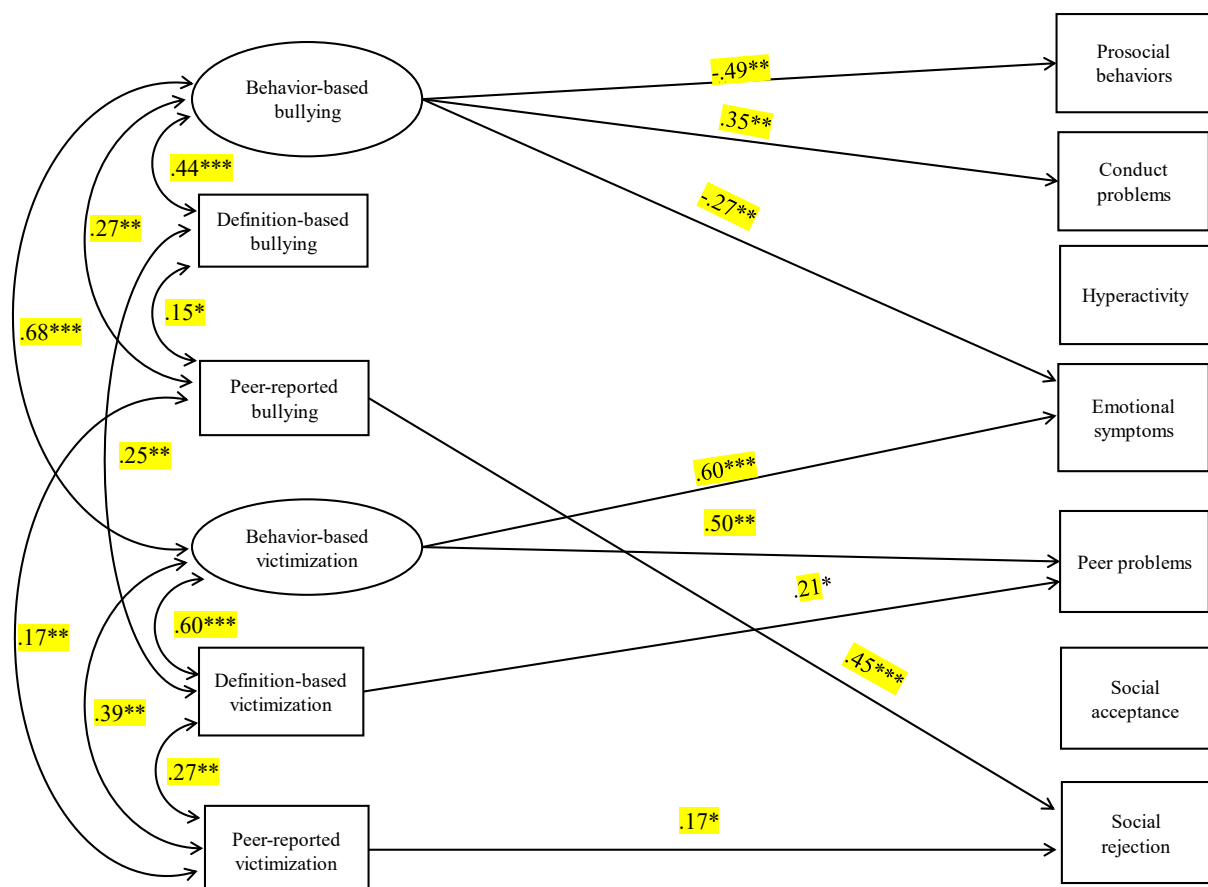
The fit of the main model was good: $\chi^2(325) = 500.851, p < .001$, CFI = .93, RMSEA = .034, 95% CI: .028–.039, and SRMR = .044. Figure 1 shows all significant standardized regression coefficients. The 95% confidence intervals for all predictors, as well as the standard errors and the R^2 values for each outcome, are reported in the Supplementary Materials (Tables S2, S3, S4, S5, S6, S7, S8). Results indicated that self-reported measures were more strongly associated with psychosocial adjustment. More specifically, behavior-based bullying was significantly and negatively associated with prosocial behaviors ($\beta = -.49$) and emotional symptoms of anxiety and depression ($\beta = -.27$), and positively associated with conduct problems ($\beta = .35$). This means that students reporting more frequent bullying behaviors were less prone to report prosocial behaviors and more prone to report other conduct problems. They were also less likely to report emotional symptoms of anxiety and depression. Behavior-based victimization was significantly linked to more anxiety and

emotional symptoms ($\beta = .60$) and more peer problems ($\beta = .50$). Regarding peer-reported measures, social rejection was positively associated with both peer-reported bullying ($\beta = .45$) and peer-reported victimization ($\beta = .17$). In other words, students identified as bullies and, to a lesser extent, as victims by their peers were also more likely to be identified as least liked.

Concerning definition-based measures, victimization was positively associated with peer problems when controlling for the other measures ($\beta = .21$). Students identifying themselves as victims, as measured by definition-based victimization, were also more likely to report peer problems.

Figure 1.

Path Diagram of the Associations Between Bullying Measures, the Strengths and Difficulties Questionnaire subscales, and the Social Acceptance and Rejection subscales



Note. For the sake of clarity, only significant standardized path coefficients are presented.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Sensitivity Analyses

Since gender and age are related to bullying and bystander behaviors, we conducted exploratory sensitivity analysis examining whether the results changed when controlling for these variables in the final model. The results were highly similar to those of the main analyses (see Supplementary Materials Tables S9-S15). Although the inclusion of additional age- and gender-related paths slightly reduced the fit indices ($\chi^2 (361) = 593.122, p < .001$, CFI = .91, RMSEA = .037, 95% CI: .031-.042, SRMR = .045), the overall model fit remained good and the parameter estimates did not change substantially. This reduction in fit may reflect more complex associations involving gender and age that were not fully captured in the model, but further investigation of these associations was beyond the scope of the present study. Overall, these findings suggest that the coefficients reported in Figure 1 are robust to the inclusion of gender and age effects.

Table 3*Descriptive Statistics and Correlations Between Variables*

Variables	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
Bullying and Victimization														
1. Behavior-based bullying	0.32	0.35												
2. Behavior-based victimization	0.53	0.46	.68**											
3. Definition-based bullying	0.14	0.48	.44**	.20**										
4. Definition-based victimization	0.48	0.94	.29**	.60**	.25**									
5. Peer-reported bullying	0.06	0.10	.27**	.19**	.15**	.06								
6. Peer-reported victimization	0.04	0.08	.13**	.39**	.02	.27**	.17**							
Strengths and Difficulties questionnaire (self-reported)														
7. Prosocial behaviors	1.63	0.35	-.36**	-.17**	-.13*	-.09	-.09*	.06						
8. Conduct problems	0.47	0.33	.56**	.53**	.23**	.24**	.21**	.21**	-.20**					
9. Attention deficit/hyperactivity	0.92	0.48	.21**	.19**	.04	0.01	.09	.05	-.09	.27**				
10. Emotional symptoms (anxiety and depression)	0.65	0.48	.15**	.38**	.09*	.21**	-.03	.23**	.01	.25**	.30**			
11. Peer problems	0.36	0.38	.34**	.58**	.16**	.50**	.14	.30**	-.23**	.43**	.11*	.31**		
Social acceptance and rejection (peer-reported)														
12. Social acceptance	0.29	0.02	-.12	-.20**	-.09	-.16**	-.05	-.11	.12*	-.21**	-.07	-.11*	-.26**	
13. Social rejection	0.15	0.02	.25*	.19*	.14	-.10	.51**	.24**	-.06	.21**	.09	-.03	.22**	-.27**

Note. * $p < .05$; ** $p < .01$.

Discussion

Even if bullying is studied for several decades, there is an ongoing debate about whether its measure should include an explicit definition and whether self- or peer-reports should be preferred. The most common types of bullying measures include definition-based self-reports, behavior-based self-reports, and definition-based peer-reports. Although arguments have been made in favor or against each type of measure, they are mainly based on theoretical considerations. Overall, it remains unclear whether these three types of measures capture the same phenomenon and what their unique associations are with students' psychosocial adjustment.

The results of the present study showed that all these three measures were correlated, indicating that they all partially measured a similar construct, supporting our first hypothesis. However, peer-reported scales are more correlated with behaviors-based than with definition-based scales. This result could mean that students seem more inclined to report bullying-related behaviors than to identify with the definition when they are nominated as bullies by their peers. This aligns with another study that found that peer-reported bullies were more reluctant to admit their behavior when presented with a definition (Garandeau et al., 2024). It might be that students are less likely to identify as bullies or victims due to the stigma attached to the label (Thomas et al., 2014; Vessey et al., 2014), their misunderstanding of a long definition (Hamby & Finkelhor, 2000), their minimization of their acts (Monks & Smith, 2006), or their focus only on severe cases of bullying.

The results also indicated that social rejection is associated with peer-reported measures, but not self-reported measures, which is likely due to common method variance (Casper et al., 2015) and partially support our second hypothesis. Surprisingly, we did not find an association between any of the types of measure and social acceptance. Moreover, neither type of self-reported measures was associated with social rejection when peer-reported measures were included in the same model. This underlines how peer nomination is a unique source of information about social dynamics among

students (Kaufman et al., 2021; Veenstra & Huitsing, 2021). Future studies could include popularity in the model as it is more strongly associated with bullying than social acceptance (Wiertsema et al., 2023).

Contrary to our third hypothesis, behavior-based measures were more consistently associated with students' psychosocial adjustment than definition-based measures. Behavior-based bullying was more strongly associated with conduct problems and prosocial behaviors than definition-based bullying. Likewise, behavior-based victimization was more strongly associated with emotional symptoms of anxiety and depression and peer problems than definition-based victimization. These results align with other studies suggesting that providing a definition of bullying does not necessarily better capture maladjusted students (Cook et al., 2010). Moreover, definition-based measures do not seem to better capture all bullying characteristics (Felix et al., 2011; Green et al., 2013). Although behavior-based measures without a definition may capture elements of non-bullying aggression (Andrews et al., 2023; Vivolo-Kantor et al., 2014), studies tend to indicate that definition-based measures also fail to precisely capture all the theoretical components of bullying (Kaufman et al., 2020; Felix et al., 2011). Overall, further progress in the debate on bullying measurement and the validity of different types of bullying measures may require innovative strategies to advance the debate beyond theoretical considerations alone (Tolmatcheff & Veenstra, 2025; Turner et al., 2015). For example, future research could take an ethnographic approach to compare the categorization of students into bullying roles based on observations of social dynamics in the classroom with that derived from different types of bullying measures.

Limitations and Future Perspectives

Our study included three different bullying measures in a single model, allowing us to simultaneously examine the unique associations between each type of measure and students' psychosocial adjustment. Despite this main contribution, this study also has several limitations. First, the conduct problems and attention deficit/hyperactivity subscales showed relatively low internal

consistency, suggesting that these results should be interpreted with caution. However, goodness-of-fit indices indicated a good fit for both subscales, which is more relevant in the context of path modeling (Stanley & Edwards, 2016). Future studies should replicate these findings using scales with better psychometric properties.

Second, although we made sure to include some peer-reported indicators of students' psychosocial adjustment, most were self-reported, which may have inflated the correlations with self-reported bullying measures. Prioritizing students' subjective psychosocial adjustment, which has been extensively studied and consistently shown to influence future life outcomes (Ttofi & Farrington, 2008; Wolke & Lereya, 2015), is arguably relevant. However, including both self-reported and peer-reported measures of similar constructs (e.g., peer-reported and self-reported social rejection) may help disentangle common method variance from the relations of interest in future studies.

Third, while our measures encompassed the three most common types of bullying measures, alternative measures within these types exist. Using different phrasing or including different items may yield different outcomes. In addition, using multiple items can reduce measurement error. The resulting asymmetry between our multiple-item and single-item measures may partially explain why our behavior-based measures showed stronger associations with psychosocial adjustment. Although this is less common, definition-based and peer-reported measures can also rely on multiple items instead of a single item (e.g., Huitsing et al., 2020; Salmivalli et al., 1996). Another possible variation is the use of dyadic peer nominations, which provides insights into relational networks and allows for selective focus on specific ties and roles (e.g., bullies nominated by victims). This approach also prevents nominations from being based on reputations rather than actual behaviors (Veenstra et al., 2007). Furthermore, it could also be argued that, although definition-based measures (either self- or peer-reported) explicitly mention power imbalance, students may not necessarily integrate this characteristic when answering. Therefore, perhaps none of these most commonly used

scales fully capture all characteristics of bullying (Green et al., 2013; Kaufman et al., 2020) and new measures ought to be developed to assess these characteristics (Andrews et al., 2023). For example, some behavior-based measures add separate items about power imbalance (e.g., Felix et al., 2011; Sjögren et al., 2025). Power imbalance could be assessed by asking students whether the bully is more popular or stronger than them (e.g., Kaufman et al., 2020). As power imbalance can arise from countless sources (Finkelhor et al., 2012), future studies should focus on victims' subjective feelings of helplessness. An important challenge would be to avoid items that could be confounded with depression, low self-esteem, or low social skills. Overall, more research is needed to examine the implications of these variations on the convergence of bullying measures and their associations with students' psychosocial adjustment (Turner et al., 2015).

Finally, our sample consisted of elementary school students, which may limit the generalizability of the results to older or younger students. Some developmental effects may influence the way students report bullying. For instance, adolescents may better understand the definition of bullying but might be less inclined to report it using self-reports (Furlong & Sharkey, 2009). Moreover, a convenience sample was used. The schools predominantly had medium-to-high SES, which may have influenced the results. Additionally, although no activities with students had yet been implemented at the time of data collection, it is possible that general information about the program given to teachers had a small impact on students. To increase the external validity of our findings, future research should replicate these results in more diverse and random samples.

Practical Implications

Our results indicate that the three different types of bullying and victimization measures used in this study converge to some extent, suggesting a certain degree of overlap. Additionally, these different measures were associated with several indicators of students' psychosocial adjustment, suggesting that each measure captures, to some extent, a construct related to psychosocial adjustment. Taken together, these findings support the recommendation to use the type of measure

that best fits the purpose and constraints of the survey (Volk et al., 2017). For example, if the goal is to monitor global trends in bullying over time, self-reported definition-based single-item measures are easier to use and may facilitate comparisons across countries in international surveys (e.g., Inchley et al., 2020). On the other hand, behavior-based measures might be more appropriate for practitioners to identify specific forms of bullying or victimization that are more prevalent in a given school and tailor the prevention strategies accordingly (Bear et al., 2014; Kennedy, 2020).

In general, self-reported measures are more accessible and easier to administer as the data collection is more straightforward and raises fewer ethical issues compared to peer nomination (Casper et al., 2015; Furlong & Sharkey, 2009). Self-reported measures might also be more suitable for evaluations of psychological interventions and prevention programs, in which students' subjective perceptions and psychosocial adjustment are critical (Olweus, 2013; Xie et al., 2022). Nevertheless, as peer-reported measures provide unique insights into social interactions within a group, they could be valuable for helping teachers recognize bullying (Kaufman et al., 2021; Veenstra & Huitsing, 2021) and for managing the classroom social dynamics, for instance through the choice of seating arrangement (Hoekstra et al., 2023).

To conclude, this study provides some empirical insights into the convergence of the three most common types of bullying and victimization measures and their unique associations with students' psychosocial adjustment. Our results support the idea that both researchers and practitioners should select the type of measure that best fits their specific goals and constraints (Tolmatcheff & Veenstra, 2025; Volk et al., 2017). Overall, we advocate for a more empirical approach when selecting a specific scale of bullying to inform theoretical discussions about the critical issue of bullying measurement.

References

- Andrews, N. C. Z., Cillessen, A. H. N., Craig, W., Dane, A. V., & Volk, A. A. (2023). Bullying and the Abuse of Power. *International Journal of Bullying Prevention*, 5(3), 261-270.
<https://doi.org/10.1007/s42380-023-00170-0>
- Bear, G., Yang, C., Mantz, L., Pasipanodya, E., Hearn, S., & Boyer, D. (2014). Technical manual for Delaware School Survey: Scales of school climate, bullying victimization, student engagement, and positive, punitive, and social emotional learning techniques. *Delaware Positive Behavior Support (DE-PBS) and School Climate Transformation Projects*.
- Bouman, T., van der Meulen, M., Goossens, F. A., Olthof, T., Vermande, M. M., & Aleva, E. A. (2012). Peer and self-reports of victimization and bullying: Their differential association with internalizing problems and social adjustment. *Journal of School Psychology*, 50(6), 759-774.
<https://doi.org/10.1016/j.jsp.2009.10.001>
- Branson, C. E., & Cornell, D. G. (2009). A Comparison of Self and Peer Reports in the Assessment of Middle School Bullying. *Journal of Applied School Psychology*, 25(1), 5-27.
<https://doi.org/10.1080/15377900802484133>
- Capron, C., Théron, C., & Duyme, M. (2007). Psychometric properties of the French version of the self-report and teacher Strengths and Difficulties Questionnaire (SDQ). *European Journal of Psychological Assessment*, 23(2), 79-88. <https://doi.org/10.1027/1015-5759.23.2.79>
- Caravita, S. C. S., Di Blasio, P., & Salmivalli, C. (2009). Unique and Interactive Effects of Empathy and Social Status on Involvement in Bullying. *Social Development*, 18(1), 140-163.
<https://doi.org/10.1111/j.1467-9507.2008.00465.x>
- Casper, D. M., Meter, D. J., Card, N. A., & Rose, C. A. (2015). Addressing Measurement Issues Related to Bullying Involvement. *School Psychology Review*, 44(4), 353-371.
<https://doi.org/10.17105/spr-15-0036.1>

- Chen, L.-M., & Cheng, Y.-Y. (2013). Prevalence of school bullying among secondary students in Taiwan: Measurements with and without a specific definition of bullying. *School Psychology International*, 34(6), 707-720. <https://doi.org/10.1177/0143034313479694>
- Cillessen, A. H. N., & Marks, P. E. L. (2011). Conceptualizing and measuring popularity. In *Popularity in the peer system*. (pp. 25-56). The Guilford Press.
- Cole, J. C. M., Cornell, D. G., & Sheras, P. (2006). Identification of School Bullies by Survey Methods. *Professional School Counseling*, 9(4), 305-313.
<http://www.jstor.org/stable/42732690>
- Cook, C. R., Williams, K. R., Guerra, N. G., & Kim, T. E. (2009). Variability in the prevalence of bullying and victimization: A cross-national and methodological analysis. In *Handbook of Bullying in Schools* (pp. 347-358). Routledge.
- 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>
- Cornell, D. G., & Brockenbrough, K. (2013). Identification of bullies and victims: A comparison of methods. In *Issues in school violence research* (pp. 63-87). Routledge.
https://doi.org/10.1300/J202v03n02_05
- Farrington, D., & Baldry, A. (2010). Individual risk factors for school bullying. *Journal of Aggression, Conflict and Peace Research*, 2(1), 4-16. <https://doi.org/10.5042/jacpr.2010.0001>
- Felix, E. D., Sharkey, J. D., Green, J. G., Furlong, M. J., & Tanigawa, D. (2011). Getting precise and pragmatic about the assessment of bullying: the development of the California Bullying Victimization Scale. *Aggressive Behavior*, 37(3), 234-247. <https://doi.org/10.1002/ab.20389>
- Finkelhor, D., Turner, H. A., & Hamby, S. (2012). Let's prevent peer victimization, not just bullying. *Child Abuse & Neglect*, 36(4), 271-274. <https://doi.org/10.1016/j.chiabu.2011.12.001>

- Furlong, M. J., & Sharkey, J. D. (2009). Bullying assessment: A call for increased precision of self-reporting procedures. In *Handbook of bullying in schools* (pp. 339-356). Routledge.
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research, and Evaluation*, 17(3).<https://doi.org/10.7275/n560-j767>
- Galand, B., Hospel, V., & Baudoin, N. (2014). Prévenir le harcèlement via les pratiques de classe? Une étude multiniveaux. *Revue québécoise de psychologie*, 35(3), 137-156.
- Garandeau, C. F., Turunen, T., Trach, J., & Salmivalli, C. (2024). Admitting to bullying others or denying it: Differences in children's psychosocial adjustment and implications for intervention. *International Journal of Behavioral Development*, 49(1), 1-11.
<https://doi.org/10.1177/01650254241242690>
- Green, J. G., Felix, E. D., Sharkey, J. D., Furlong, M. J., & Kras, J. E. (2013). Identifying bully victims: definitional versus behavioral approaches. *Psychological Assessment*, 25(2), 651-657. <https://doi.org/10.1037/a0031248>
- Green, J. G., Oblath, R., Felix, E. D., Furlong, M. J., Holt, M. K., & Sharkey, J. D. (2018). Initial evidence for the validity of the California Bullying Victimization Scale (CBVS-R) as a retrospective measure for adults. *Psychological Assessment*, 30(11), 1444-1453.
<https://doi.org/10.1037/pas0000592>
- Hamby, S. L., & Finkelhor, D. (2000). The Victimization of Children: Recommendations for Assessment and Instrument Development. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39(7), 829-840. <https://doi.org/10.1097/00004583-200007000-00011>
- Hoekstra, N. A. H., van den Berg, Y. H. M., Lansu, T. A. M., Mainhard, M. T., & Cillessen, A. H. N. (2023). Teachers' goals and strategies for classroom seating arrangements: A qualitative study. *Teaching and Teacher Education*, 124, 104016.
<https://doi.org/10.1016/j.tate.2023.104016>

- Huang, F. L., & Cornell, D. G. (2015). The impact of definition and question order on the prevalence of bullying victimization using student self-reports. *Psychological Assessment*, 27(4), 1484-1493. <https://doi.org/10.1037/pas0000149>
- Huitsing, G., Lodder, G. M. A., Browne, W. J., Oldenburg, B., Van der Ploeg, R., & Veenstra, R. (2020). A Large-Scale Replication of the Effectiveness of the KiVa Antibullying Program: a Randomized Controlled Trial in the Netherlands. *Prevention Science*, 21(5), 627-638. <https://doi.org/10.1007/s11121-020-01116-4>
- Hymel, S., Rubin, K. H., Rowden, L., & LeMare, L. (1990). Children's peer relationships: longitudinal prediction of internalizing and externalizing problems from middle to late childhood. *Child Development*, 61(6), 2004-2021. <https://doi.org/10.1111/j.1467-8624.1990.tb03582.x>
- Inchley, J., Currie, D., Budisavljevic, S., Torsheim, T., Jåstad, A., Cosma, A., Kelly, C., Arnarsson, Á., & Samdal, O. (2020). Spotlight on adolescent health and well-being. Findings from the 2017/2018 Health Behaviour in School-aged Children (HBSC) survey in Europe and Canada. *International report*, 1, 146.
- Jia, M., & Mikami, A. (2018). Issues in the assessment of bullying: Implications for conceptualizations and future directions. *Aggression and Violent Behavior*, 41, 108-118. <https://doi.org/10.1016/j.avb.2018.05.004>
- Joseph, S., & Stockton, H. (2018). The multidimensional peer victimization scale: a systematic review. *Aggression and Violent Behavior*, 42, 96-114. <https://doi.org/10.1016/j.avb.2018.07.009>
- Juvonen, J., Nishina, A., & Graham, S. (2001). Self-views versus peer perceptions of victim status among early adolescents. In *Peer harassment in school: The plight of the vulnerable and victimized*. (pp. 105-124). The Guilford Press.

- Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa antibullying program: grades 4-6. *Child Development*, 82(1), 311-330. <https://doi.org/10.1111/j.1467-8624.2010.01557.x>
- Kaufman, T. M. L., Huitsing, G., Bloemberg, R., & Veenstra, R. (2021). The Systematic Application of Network Diagnostics to Monitor and Tackle Bullying and Victimization in Schools. *International Journal of Bullying Prevention*, 3(1), 75-87. <https://doi.org/10.1007/s42380-020-00064-5>
- Kaufman, T. M. L., Huitsing, G., & Veenstra, R. (2020). Refining victims' self-reports on bullying: Assessing frequency, intensity, power imbalance, and goal-directedness. *Social Development*, 29(2), 375-390. <https://doi.org/10.1111/sode.12441>
- Kennedy, R. S. (2020). A meta-analysis of the outcomes of bullying prevention programs on subtypes of traditional bullying victimization: Verbal, relational, and physical. *Aggression and Violent Behavior*, 55, 101485. <https://doi.org/10.1016/j.avb.2020.101485>
- Kert, A. S., Coddington, R. S., Tryon, G. S., & Shiyko, M. (2010). Impact of the word "bully" on the reported rate of bullying behavior. *Psychology in the Schools*, 47(2), 193-204. <https://doi.org/10.1002/pits.20464>
- Košir, K., Klasinc, L., Špes, T., Pivec, T., Cankar, G., & Horvat, M. (2020). Predictors of self-reported and peer-reported victimization and bullying behavior in early adolescents: the role of school, classroom, and individual factors. *European Journal of Psychology of Education*, 35(2), 381-402. <https://doi.org/10.1007/s10212-019-00430-y>
- Kubiszewski, V., Fontaine, R., Chasseigne, G., & Rusch, E. (2014). Évaluation du bullying scolaire (harcèlement scolaire) chez les adolescents français: validité de l'adaptation française du questionnaire Agresseur/Victime révisé d'Olweus (1996). *Annales Médico-Psychologiques, Revue Psychiatrique*, 172(4), 261-267. <https://doi.org/10.1016/j.amp.2012.09.018>

- Lai, T., & Kao, G. (2018). Hit, Robbed, and Put Down (but not Bullied): Underreporting of Bullying by Minority and Male Students. *Journal of Youth and Adolescence*, 47(3), 619-635.
<https://doi.org/10.1007/s10964-017-0748-7>
- Lee, T., & Cornell, D. (2010). Concurrent Validity of the Olweus Bully/Victim Questionnaire. *Journal of School Violence*, 9(1), 56-73. <https://doi.org/10.1080/15388220903185613>
- Modecki, K. L., Minchin, J., Harbaugh, A. G., Guerra, N. 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. <https://doi.org/10.1016/j.jadohealth.2014.06.007>
- Moore, S. E., Norman, R. E., Suetani, S., Thomas, H. J., Sly, P. D., & Scott, J. G. (2017). Consequences of bullying victimization in childhood and adolescence: A systematic review and meta-analysis. *World Journal Psychiatry*, 7(1), 60-76.
<https://doi.org/10.5498/wjp.v7.i1.60>
- Morin, A. J., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 116-139. <https://doi.org/10.1080/10705511.2014.961800>
- Muthén, B., & Muthén, L. (2017). Mplus. In *Handbook of item response theory* (pp. 507-518). Chapman and Hall/CRC.
- Newman, D. A. (2014). Missing Data: Five Practical Guidelines: Five Practical Guidelines. *Organizational Research Methods*, 17(4), 372-411.
<https://doi.org/10.1177/1094428114548590>
- Olweus, D. (2013). School bullying: Development and some important challenges. *Annual Review of Clinical Psychology*, 9(1), 751-780. <https://doi.org/10.1146/annurev-clinpsy-050212-185516>
- Rodríguez-Fernández, A., Ramos-Díaz, E., Madariaga, J. M., Arrivillaga, A., & Galende, N. (2016). Steps in the construction and verification of an explanatory model of psychosocial

adjustment. *European Journal of Education and Psychology*, 9(1), 20-28.

<https://doi.org/10.1016/j.ejeps.2015.11.002>

Salmivalli, C., Lagerspetz, K., Björkqvist, K., Österman, K., & Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group.

Aggressive Behavior: Official Journal of the International Society for Research on

Aggression, 22(1), 1-15. [https://doi.org/10.1002/\(SICI\)1098-2337\(1996\)22:1<1::AID-](https://doi.org/10.1002/(SICI)1098-2337(1996)22:1<1::AID-)

[AB1>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1098-2337(1996)22:1<1::AID-AB1>3.0.CO;2-T)

Shaw, T., Dooley, J. J., Cross, D., Zubrick, S. R., & Waters, S. (2013). The Forms of Bullying Scale (FBS): validity and reliability estimates for a measure of bullying victimization and perpetration in adolescence. *Psychological Assessment* 25(4), 1045-1057.

<https://doi.org/10.1037/a0032955>

Sjögren, B., Bjereld, Y., Thornberg, R., Hong, J. S., & Espelage, D. L. (2025). Prevalence Rates of Bullying: A Comparison Between a Definition-Based Scale and a Behavior-Based Scale. *Journal of Interpersonal Violence*, 40(7-8), 1530-1552.

<https://doi.org/10.1177/08862605241262216>

Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 29(3), 239-268. <https://doi.org/10.1002/ab.10047>

Stanley, L. M., & Edwards, M. C. (2016). Reliability and Model Fit. *Educational and Psychological Measurement*, 76(6), 976-985. <https://doi.org/10.1177/0013164416638900>

Swain, J. (1998). What does bullying really mean? *Educational Research*, 40(3), 358-364.

<https://doi.org/10.1080/0013188980400307>

Thomas, H. J., Connor, J. P., & Scott, J. G. (2014). Integrating Traditional Bullying and

Cyberbullying: Challenges of Definition and Measurement in Adolescents – a Review.

Educational Psychology Review, 27(1), 135-152. <https://doi.org/10.1007/s10648-014-9261-7>

- Tolmatcheff, C., Galand, B., Roskam, I., & Veenstra, R. (2022). The effectiveness of moral disengagement and social norms as anti-bullying components: A randomized controlled trial. *Child development*, 93(6), 1873-1888.
- Tolmatcheff, C., & Veenstra, R. (2025). So you want to design an anti-bullying program? Recommendations for evaluating the potential of discrete ingredients and improving program cost-effectiveness. *Aggression and Violent Behavior*, 86, 102110.
<https://doi.org/10.1016/j.avb.2025.102110>
- Ttofi, M. M., & Farrington, D. P. (2008). Bullying: Short-term and long-term effects, and the importance of defiance theory in explanation and prevention. *Victims and Offenders*, 3(2-3), 289-312.
<https://doi.org/10.1080/15564880802143397>
- Turner, H. A., Finkelhor, D., Shattuck, A., Hamby, S., & Mitchell, K. (2015). Beyond bullying: Aggravating elements of peer victimization episodes. *School Psychology Quarterly*, 30(3), 366–384. <https://doi.org/10.1037/spq0000058>
- Vaillancourt, T., McDougall, P., Hymel, S., Krygsman, A., Miller, J., Stiver, K., & Davis, C. (2008). Bullying: Are researchers and children/youth talking about the same thing? *International Journal of Behavioral Development*, 32(6), 486-495.
<https://doi.org/10.1177/0165025408095553>
- van der Ploeg, R., Steglich, C., Salmivalli, C., & Veenstra, R. (2015). The Intensity of Victimization: Associations with Children's Psychosocial Well-Being and Social Standing in the Classroom. *PLoS One*, 10(10), e0141490. <https://doi.org/10.1371/journal.pone.0141490>
- van Verseveld, M. D. A., Fekkes, M., Fekkink, R. G., & Oostdam, R. J. (2021). Effects of Implementing Multiple Components in a School-Wide Antibullying Program: A Randomized Controlled Trial in Elementary Schools. *Child Development*, 92(4), 1605-1623.
<https://doi.org/10.1111/cdev.13529>

- Veenstra, R., & Huitsing, G. (2021). Social Network Approaches to Bullying and Victimization. In *The Wiley Blackwell Handbook of Bullying* (pp. 196-214).
<https://doi.org/10.1002/9781118482650.ch11>
- Veenstra, R., Lindenberg, S., Zijlstra, B. J., De Winter, A. F., Verhulst, F. C., & Ormel, J. (2007). The dyadic nature of bullying and victimization: testing a dual-perspective theory. *Child Development*, 78(6), 1843-1854. <https://doi.org/10.1111/j.1467-8624.2007.01102.x>
- 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(12), 819-843. <https://doi.org/10.1111/josh.12210>
- 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. <https://doi.org/10.1016/j.avb.2014.06.008>
- Volk, A. A., Dane, A. V., & Al-Jbouri, E. (2022). Is adolescent bullying an evolutionary adaptation? A 10-year review. *Educational Psychology Review*, 34(4), 2351-2378.
<https://doi.org/10.1007/s10648-022-09703-3>
- Volk, A. A., Veenstra, R., & Espelage, D. L. (2017). So you want to study bullying? Recommendations to enhance the validity, transparency, and compatibility of bullying research. *Aggression and Violent Behavior*, 36, 34-43.
<https://doi.org/10.1016/j.avb.2017.07.003>
- Wiertsema, M., Vrijen, C., van der Ploeg, R., Sentse, M., & Kretschmer, T. (2023). Bullying perpetration and social status in the peer group: A meta-analysis. *Journal of Adolescence*, 95(1), 34-55. <https://doi.org/10.1002/jad.12109>
- Wolke, D., & Lereya, S. T. (2015). Long-term effects of bullying. *Archives of Disease in Childhood*, 100(9), 879-885.
<https://doi.org/10.1136/archdischild-2014-306667>

- Xie, Z., Man, W., Liu, C., & Fu, X. (2022). A PRISMA-based systematic review of measurements for school bullying. *Adolescent Research Review*, 8(2), 219-259.
<https://doi.org/10.1007/s40894-022-00194-5>
- Ybarra, M. L., Espelage, D. L., & Mitchell, K. J. (2014). Differentiating youth who are bullied from other victims of peer-aggression: the importance of differential power and repetition. *Journal of Adolescent Health*, 55(2), 293-300. <https://doi.org/10.1016/j.jadohealth.2014.02.009>
- Younan, B. (2019). A systematic review of bullying definitions: How definition and format affect study outcome. *Journal of Aggression, Conflict and Peace Research*, 11(2), 109-115.
<https://doi.org/10.1108/JACPR-02-2018-0347>
- Yun, H.-Y., & Espelage, D. (2024). Self-ratings and peer-ratings of bullying perpetrators: Intrapersonal and interpersonal factors that differentiate bully subgroups. *Journal of School Psychology*, 106, 101358.
<https://doi.org/10.1016/j.jsp.2024.101358>
- Zych, I., Farrington, D. P., & Ttofi, M. M. (2019). Protective factors against bullying and cyberbullying: A systematic review of meta-analyses. *Aggression and Violent Behavior*, 45, 4-19. <https://doi.org/10.1016/j.avb.2018.06.008>

Appendix A

Definition presented in the original OBVQ-r (Solberg & Olweus, 2003).

We say a student is being bullied when another student, or several other students:

- say mean and hurtful things or make fun of him or her or call him or her mean and hurtful names;
- completely ignore or leave him or her out from their group of friends or
- leave him or her out of things on purpose;
- hit, kick, push, shove around, or lock him or her inside a room;
- tell lies or spread false rumours about him or her or send mean notes and try to make other students dislike him or her;
- and other hurtful things like that.

When we talk about bullying, these things happen repeatedly, and it is difficult for the student being bullied to defend himself or herself. We also call it bullying, when a student is teased repeatedly in a mean and hurtful way. But we don't call it bullying when the teasing is done in a friendly and playful way. Also, it is not bullying when two students of about equal strength or power argue or fight.

French translation of the OBVQ-r (Kubiszewski et al., 2014).

On considère qu'un élève est victime de bullying quand un ou plusieurs de ses camarades :

- tien(nen)t des propos désobligeants, blessants, menaçants, se moque(nt) de lui, l'insulte(nt) ou lui donne(nt) des surnoms méchants ;
- l'ignore(nt) ou l'exclue(nt) complètement d'un groupe d'élèves, le met(tent) intentionnellement à l'écart ;

- le pousse(nt), le frappe(nt), le percute(nt), le bouscule(nt) ou encore l'enferme(nt) dans une pièce ;
- raconte(nt) des mensonges à son propos, profère(nt) de fausses rumeurs, ou envoie(nt) des mots menaçants et essaie(nt) de faire en sorte de le discréditer aux yeux des autres élèves ;
- et tous autres agissements blessants de ce type.

On parle de bullying quand ces situations surviennent fréquemment et qu'il est difficile pour l'élève qui subit cela de se défendre par lui-même. On parle également de bullying quand un élève est régulièrement taquiné de manière méchante et blessante. Cependant, on ne parle pas de bullying quand les taquineries/moqueries sont amicales ou « pour rire ». De la même façon, on ne parle pas de bullying quand deux élèves de même « force »/« pouvoir » se disputent ou se battent.