UvA-DARE (Digital Academic Repository)

Strengthening teachers in their role to identify and address bullying among students in elementary schools

van Verseveld, M.

Publication date
2021

Citation for published version (APA):
van Verseveld, M. (2021). *Strengthening teachers in their role to identify and address bullying among students in elementary schools*. [Thesis, externally prepared, Universiteit van Amsterdam]. Kenniscentrum Onderwijs en Opvoeding / HvA Publicaties.

CHAPTER 5

Effects of Implementing Multiple Components in a School-Wide Antibullying Program: A Cluster-Randomized Controlled Trial in Elementary Schools

ABSTRACT

This study investigates the effectiveness of the PRIMA antibullying program for elementary education using a cluster-randomized trial with two experimental conditions (with and without student lessons) and a control group. Students of 31 schools participated in the study (N= 3,135; M_{age}=10 yrs). Multilevel regression analyses demonstrated positive effects of the program on peer-reported victimization and reinforcing behavior. Implementing multiple program components was related to stronger program effects. The results provide partial experimental evidence for the beneficial effects of combining student lessons and teacher training in antibullying programs. Future experimental research is needed to investigate other approaches that reduce not only peer-reported victimization, but also self-perceived bullying and victimization.

INTRODUCTION

Bullying is still a common problem in schools, directly involving many students (Jansen et al., 2012; Mitsopoulou, & Giovazolias, 2015; Zych et al., 2015). Bullying is commonly characterized as repeated and intentional aggressive behavior against a victim who cannot readily defend themselves (Olweus, 1993). Bullying is considered a group process in which students can be involved as a victim, bully, reinforcer, outsider, or defender (Salmivalli et al., 1996). Victimized students often develop psychosocial problems, such as low self-esteem, anxiety, and depression (Reijntjes et al., 2010), and these adverse effects can endure into adulthood (Lund et al. 2008).

Bullying behavior already emerges in early elementary school and negatively influences children's socio-emotional development (Jansen et al., 2012). Gender roles and age influence bullying behavior, with higher bullying scores for boys and younger children (Ladd et al., 2017; Mitsopoulou, & Giovazolias, 2015). Especially from grade 3 onwards, students can understand and reflect on the different aspects of bullying behavior, such as the power imbalance and the intention to harm another (Monks & Smith, 2010; Vlachou et al., 2011). In addition, several studies found that the trajectories of peer victimization can differ, with some children being bullied heavily throughout the K-12 school period, while others are bullied for a shorter period or to a less severe extent (Ladd et al., 2017; Nylund et al., 2007; Zhou et al., 2020). These outcomes indicate the need for early bullying interventions in elementary school and into the years of early adolescence and the need for tailored and selective interventions for specific groups of students in addition to more general and universal interventions (Garandeau & Salmivalli, 2019).

Numerous school-based antibullying intervention programs have been developed and evaluated. Meta-analytic reviews have reported small to moderate effect sizes at child level, indicating that antibullying programs reduce bullying and victimization rates in elementary schools, although there is variation in outcomes (Gaffney et al., 2019; Jiménez-Barbero et al., 2016). Many of these programs take a socio-ecological perspective in targeting the many factors that influence bullying behaviors such as the schools' policies and procedures, school's physical environment (supervision by staff, safe places or places of frequent incidents), social environment (school climate/ethos) and engagement with parents, family and the wider community (Axford et al., 2020; Huitsing et al., 2020a; Limber et al., 2018). Programs with a combination of universal components, targeting all school actors, and selective components, targeting students who are involved in bullying situations, have been related to most successful reductions in bullying rates (Ansary et al., 2015; Gaffney et al., 2019). In their meta-analysis, Ttofi and Farrington (2011) found correlational evidence for the effectiveness of specific program components, such as disciplinary methods for bullies, teacher staff training, and parent

meetings. Although these correlational findings suggest that some specific program components may be effective in reducing bullying, there is still a lack of experimental evidence for these components' effectiveness. We need to gain more insight into the effects of antibullying programs and their specific components to open the 'black box'.

Two issues seem especially interesting to investigate: the added value of the student curriculum and the effect of adequate implementation of an antibullying program's component. A student curriculum is a core component in many successful antibullying programs (see Ansary et al., 2015). Some student curricula focus on the development of students' socio-emotional skills. One of the aims of the Steps to Respect program, for example, is to strengthen students to recognize the various forms of bullying behavior and the negative consequences of bullying and aspires to train students in empathy, emotion regulation, and conflict resolution skills (Brown et al., 2011; Frey et al., 2005; Low et al., 2014). Students learn a variety of social skills and coping skills (e.g., assertiveness, emotion management) to deal with bullying and social situations, which, in turn, helps to prevent bullying. This program showed reductions in bullying and positive effects on bullying prevention factors (Brown et al., 2011). Other programs, such as the KiVa program, include a student curriculum aiming to influence the group dynamics of bullying by creating a strong antibullying norm in the classroom and by empowering students to stop the bullying by targeting outsiders to no longer ignore the bullying when it occurs (Salmivalli et al., 1996; Veenstra et al., 2014). The KiVa program has effectively reduced victimization and bullying (see, for example, Kärnä et al., 2011; Salmivalli et al., 2011; Yang & Salmivalli, 2015). Several other studies have also shown that endorsing a strong antibullying norm is associated with less bullying in the classroom (Marachi et al., 2007; Troop-Gordon & Ladd, 2015; Veenstra et al., 2014). Student lessons are considered vital because this component affects all students directly and may influence students' norms in the classroom.

In addition to strengthening students through lessons in the classroom, most effective school-wide programs also include various other components that support teachers and other staff members. Teachers are key figures in implementing the core components of antibullying programs, and therefore, teacher training is crucial (Craig et al., 2011). Successful antibullying programs also depend on teachers and staff to create and maintain antibullying norms, model positive and prosocial behavior, and encourage students to contribute to a positive class- and school climate. Some programs focus, therefore, also on creating awareness among teachers and staff members to identify bullying and to respond adequately when bullying behavior occurs (van Verseveld et al., 2019). Systematic assessments of bullying behavior could help teachers identify bullying because it often happens when adults are not present, and students are reluctant to report bullying (Demaray et al., 2013; Fekkes et al., 2005; Wachs et al., 2019). Since bullying is considered a group process, it is vital to use a multi-informant instrument that measures

bullying and victimization for all students, and other students' roles in bullying behavior (Huitsing & Veenstra, 2012).

Many programs are complex and consist of a mixture of class components and various other, school-wide, and teacher/staff-focused components. Therefore, it is imperative to gain insight into the effectiveness of specific elements. Relatedly, the adequate implementation of individual components is an essential factor for their success (Ttofi & Farrington, 2011). Programs in which multiple individual program components are adequately implemented are more effective than school-wide programs in which the individual components are implemented with less fidelity (Domitrovich et al., 2008; Durlak & DuPre, 2008). Recent studies have shown a wide variation in the level of implementation of antibullying programs (Axford et al., 2020; Orobio de Castro et al., 2018), suggesting that this 'stacking' of various program components is demanding for school professionals. Program implementation may be affected by the complexity of implementing the many components of a school-wide program, a lack of support and resources given by the school management to teachers, and a high workload and low teacher motivation to implement such a program (Haataja et al., 2015; Hall, 2017; Kallestad & Olweus, 2003; Orobio de Castro et al., 2018; Salmivalli et al., 2005). Therefore, investigating the added value of implementing various components is necessary for the evaluation and future development of antibullying programs.

PRIMA Antibullying Program

PRIMA (VeiligheidNL, n.d.) is a Dutch multi-component antibullying program for elementary education, based initially on the Olweus Bullying Prevention Program (Olweus, 1993). Based on the socio-ecological model (Hong & Espelage, 2012), PRIMA focuses on three levels in the school: the individual child, the classroom, and the school. In addition, parents are informed about the preventive antibullying policy and are involved when a bullying situation concerns their child. PRIMA's primary goal is to ensure a prosocial and safe school climate in which students treat each other with respect.

A national antibullying committee has accredited an earlier version of PRIMA (Orobio de Castro et al., 2018). This study reported that after one year of implementation, PRIMA was effective in reducing bullying and victimization. Also, students reported lower levels of depressive symptoms compared to students in control schools.

The PRIMA program was extended in 2017 by including new knowledge and tools related to the group process of bullying and creating a positive group norm in the classroom (Huitsing, & Veenstra, 2012; Salmivalli et al., 1996). The program was also adapted to better meet the needs of teachers and staff members by providing more information and strategies to support them in identifying and addressing bullying behavior.

Students of all elementary school grades receive a student lesson curriculum, including six weekly 45-minute lessons at the beginning of the school year and two lessons to

refresh their knowledge and skills for the remaining year. The lessons pursue to prevent and address bullying together with students by focusing on the following three goals: 1) making students aware of the negative consequences of bullying and their role in bullying situations; 2) strengthening positive antibullying norms in the classroom and teaching students strategies to stand up against a bullying norm; and 3) increasing prosocial skills and promoting positive relations among students. The lessons consist of assignments on class, small-group, and individual level and are supported by videos. The repetition of the student lessons in consecutive school years, adapted to the developmental changes across the K-12 years, aims to create a continuous learning curve in which students become familiar with the program norms and values (Craig et al., 2011; Kärnä et al., 2013).

All school professionals follow a 2-hour e-learning module independently. Teachers are taught different skills that enable them to identify, prevent, and reduce bullying adequately. The e-learning aims to 1) increase teachers' and staff members' awareness and responsiveness to bullying; 2) encourage them to model prosocial and positive behavior; 3) help them create and maintain antibullying norms actively; 4) and support them in the implementation process of the program. Teachers are instructed on how to implement the monitor instrument and recognize the risk factors of victimized students. Methods to respond to bullying include strategies at the class level (e.g., guidelines to create an antibullying norm in the classroom and a positive climate), and strategies at the individual and parent level (e.g., guidelines to talk with victims, bullies, and their parents). In line with the e-learning module, teachers participate in a face-to-face training session from a certified PRIMA-coach to practice strategies that have been introduced in the e-learning module using fictional cases of bullying or bullying situations that occur in teachers' classes. Finally, the training aims to further support staff members in implementing the PRIMA components by discussing school-specific facilitators and barriers.

The monitor aims to provide teachers insight into students' roles in bullying behavior, students' perception of the classroom climate, their social status, and mutual friendships, using a multi-informant approach. The monitor also aims to provide teachers with information about individual students directly involved in bullying situations or students at risk. Teachers in grades 3-6 receive a report twice a year, following from the administration of the monitor (i.e., a digital and interactive questionnaire) for 3-6 grade students. The results of the monitor are shared confidentially with the teacher and not in the classroom. According to the monitoring results, teachers in grades 3-6 also receive digital protocols for students involved in bullying situations. In this way, teachers are provided with tools to tackle the specific situation for students who are being victimized, for students who are bullying other students, or for students who are at-risk for being victimized (e.g., lonely and rejected children).

All school professionals receive access to digital protocols providing guidelines on how to deal with specific bullying situations (see Van Verseveld et al., 2020). Therefore, the protocols support teachers to intervene more effectively in bullying situations that teachers themselves considered to be difficult.

The student curriculum, e-learning module, face-to-face training, and the monitor report are preventive, universal components for all students and teachers. The protocols for specific bullying situations and the protocols following from the monitor results are selective, curative, components, and focus on students who are (at risk of being) involved in bullying.

The Present Study

This study aims to evaluate the effectiveness of the new PRIMA program. Students received the PRIMA program in one experimental condition, including the lessons for students and teacher training (hereafter: PRIMA-L$^+$), whereas students in the other experimental condition received the PRIMA program, including only the teacher training and without the student lessons (PRIMA-L$^-$). As the primary outcomes, we used self-reported and peer-reported bullying and victimization. We hypothesized a stronger decrease in bullying and victimization in PRIMA-L$^+$ schools than PRIMA-L$^-$ schools (H1) as our primary research question. In addition, we evaluated the effects of stacking the universal program components of PRIMA across conditions. We hypothesized a stronger decrease in bullying and victimization when teachers implemented more universal program components (H2). In an explorative fashion, we investigated the effects of both experimental conditions and the effects of stacking universal program components on the roles of reinforcers, outsiders, and defenders.

5

METHOD

Sampling and Design

We conducted a power analysis based on self- and peer-reported victimization as an outcome measure. Since the prevalence of bullying in grades 3-6 varies between 21% and 35% in western countries (Chester et al., 2015; National Center for Educational Statistics, 2019), we estimated that a minimum of 33 classes per condition was needed with a minimum of 25 students per class (assuming a response of 80%; alpha = .05, two-sided, power = .80, ICC = .032) to demonstrate a decrease of 30% of victims (i.e., from 25% to 17.5%) between the two experimental groups and the control group. With this sample size, a small effect (Cohen's d = 0.20) can be demonstrated for primary and secondary measures with adequate power.

A cluster-randomized controlled trial was applied with a pretest and posttest and an 1:1:1 allocation ratio, comparing two experimental conditions with a control group. We selected 354 elementary schools from a database of all Dutch elementary schools and assessed them for eligibility to participate in this study based on the following inclusion criteria: (a) the schools contained more than 50 students; (b) schools were not already using an antibullying prevention program; (c) schools were not participating in any other study in this area; (d) schools were willing to receive additional information about the study. A total of 173 schools met the inclusion criteria. After stratifying schools by school size, the number of special needs students in the school, and the urbanization level of the school's location, schools were randomly assigned to one of the two experimental conditions or the control group. In the PRIMA-L$^+$ condition, schools received all PRIMA program core components, including the student lessons. In the PRIMA-L$^-$ condition, schools received all PRMA core components, except for the student lessons. This design makes it possible to determine the effect of a teacher approach (PRIMA-L$^-$ vs. control) and the additional effect of the student curriculum (PRIMA-L$^+$ vs. control).

After allocation to research conditions, letters were sent to schools from January to July 2017 to invite them to participate in the study. Intervention schools received free access to PRIMA, free coaching of a certified coach during the trial, and monetary compensation of €700. Control schools received a free PRIMA pilot without a certified coach after the trial (i.e., waitlist condition) and €1,000. We informed schools about their assignment (intervention arms or control arm) in September 2017. One of the researchers remained blind to school allocation and led the assessment of the study's outcomes. Figure 1 provides an overview of the school and participant enrollment in the trial. A total of 31 schools participated in the study: nine schools in the PRIMA-L$^+$ condition, ten schools in the PRIMA-L$^-$ condition, and twelve schools in the control group.

Participants

The 31 participating schools included 174 classes representing 4,285 students in grades 3-6 who were eligible for participation in the study. Parents gave written permission for the participation of a total of 3,135 students (73.2% of the initial sample, M_{age} = 10.00, SD = 1.21). We also obtained active, informed consent from the teachers to participate in the trial. In all groups, an approximately equal percentage of students participated in the trial (PRIMA-L$^+$ condition: 70.7%; PRIMA-L$^-$ condition: 79.5%; and control condition: 69.1%). The PRIMA-L$^+$ condition comprised 873 students (M_{age} = 9.97, SD = 1.23), the PRIMA-L$^-$ condition had 982 students (M_{age} = 10.05, SD = 1.17), and the control condition contained 1,389 students (M_{age} = 9.98, SD = 1.21). Of the 3,135 students, 52.4% were girls, 46.8% boys, and 0.8% had missing data on this variable (see Table 1). Most students (75.9%) had a western background, and a smaller proportion had a non-western background (22.8%). Of 1.3% of the students, this information was missing.
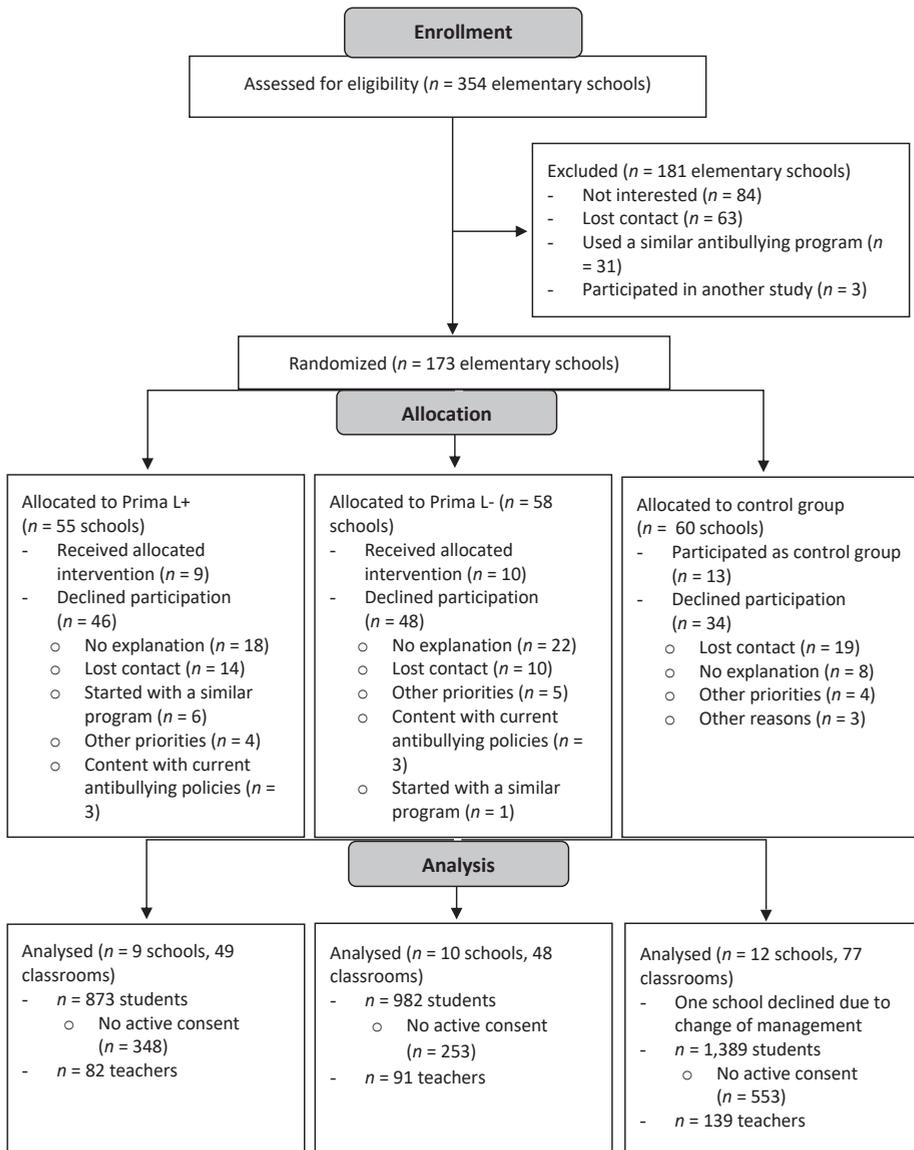
Figure 1. Flow chart of school enrollment in the study

Students between conditions did not differ significantly in age, $F(2, 3093)= 1.43$, $p =$ .241 or gender, $\chi (2)= 1.11$, $p = .574$. However, there was a significant difference in ethnicity, $\chi (2) = 17.60$, $p < .001$, with a smaller proportion of students with a non-western background in the control group. For peer-reports, conditions differed significantly, χ (2) = 8.57, $p = .014$, with slightly less attrition in the control schools (7.0%) compared

Table 1 *Participant Characteristics at Pretest by Condition (N = 3,244)*

	PRIMA-L$^+$		PRIMA-L$^-$		Control group	
	N	%	N	%	N	%
Number of schools	9	29.0	10	32.3	12	38.7
Number of classes	49	25.9	48	27.7	77	45.2
Number of students	873	27.8	982	31.3	1,389	40.8
Sex						
Boy	402	46.0	472	48.1	595	42.8
Girl	464	53.2	499	50.8	679	48.9
Missing	7	0.8	11	1.1	115	8.3
Ethnicity						
Western	643	73.7	715	72.8	1,023	73.7
Non-western	216	24.7	253	25.8	244	17.6
Missing	14	1.6	14	1.4	122	8.8

to the experimental schools (9.6%). A non-response analysis indicated there was no selective attrition. Students who did not participate in the post-test did not differ significantly in any pretest outcome measures compared to students who did participate. The 174 classes represented 312 teachers, 82 teachers in the PRIMA-L$^+$ condition, 91 in the PRIMA-L$^-$ condition, and 139 in the control condition. We controlled for this variable in our analysis. Attrition at the posttest was not different for the three conditions on self-report measures.

Procedures

We visited each participating school to explain the data collection procedure at the start of the school year in September 2017. Data were collected at the pretest in October-November 2017 and the posttest in March-April 2018. During each wave of data collection, students completed two online questionnaires during school hours. Two researchers instructed students on how to complete the questionnaire and ensured students' privacy during the administration. Researchers also explained that students' answers would remain confidential within the classroom.

After the administration of the pretest, schools received access to the program materials. The research team instructed school teachers and principals to deliver the program as follows: 1) Consulting the monitor report, and participating in the e-learning and face-to-face training (November/December 2017); 2) Delivering student curriculum for PRIMA-L$^+$ schools (December 2017/January 2018); 3) Implementing protocols for students (at risk of being) involved in bullying situations or for specific bullying situations if required. Students in the PRIMA-L$^+$ were exposed to the program directly through the student curriculum, while students in the PRIMA-L$^-$ schools were exposed only indirectly

through the teacher being exposed to teacher-focused components. Control schools offered 'care as usual', which means that they implemented nationally established antibullying guidelines, such as monitoring students' wellbeing at school, having an antibullying coordinator, and having a social safety policy. Control schools were interviewed by telephone at the beginning and end of the data collection period to monitor whether they were running a school-wide antibullying program. None of the 13 control schools carried out an antibullying program during the trial period. Ethical approval was granted by the Faculty of Social and Behavioral Sciences' ethical board at the University of [anonymized] (nr. 2016-CDE-8008). The trial has been registered in the ISRCTN register (file number 15425978).

Measures

As formulated in the Revised Bully/Victim Questionnaire of Olweus (BVQ, Olweus, 1996), a definition of bullying was presented in the questionnaire, emphasizing the repetitive and intentional nature of bullying and the imbalance of power between the bully and victim. A description of the different forms of bullying was also given, including overt forms of bullying (e.g., verbal, physical, threatening), covert forms (e.g., social exclusion, gossiping), and digital bullying (e.g., on social media, internet).

We measured bullying and victimization with both self and peer reports. Self-reports are the standard for prevalence estimation and measurement of change (Olweus, 2013). Using self-reports, we can measure how children experience bullying/victimization themselves. Peer-reports are valuable since a multi-informant approach gives a more refined opportunity to measure how observed bullying occurs in a classroom (Kärnä et al., 2011).

Self-reported victims and bullies. We used the global item from the revised Olweus Bully/Victim Questionnaire (OBVQ, Olweus, 1996) to measure self-reported victimization: "How often have you been bullied at school in the last couple of months?". Students answered on a 5-point scale (0 = *not at all*, 1 = *once or twice,* 2 = *two or three times a month,* 3 = about once a week, 4 = *several times a week*).

Self-reported bullying was measured by asking students whether they had engaged in a series of behaviors often associated with bullying in the last couple of months. Students responded to items on a 5-point scale (0 = *not at all*, 1 = *once or twice,* 2 = *two or three times a month,* 3 = about once a week, 4 = *several times a week*). These eight items were based on the OBVQ (Olweus, 1996) and had an internal consistency of α = .882 at the pretest. We have chosen to measure self-reported bullying more subtly through eight related behaviors because children often underreport their bullying behavior due to self-protecting mechanisms (Košir et al., 2019).

Peer-reported victims and bullies. Two single items, based on the Participant Roles Questionnaire (Kärnä et al., 2013; Salmivalli et al., 1996), were used to identify peer-reported victimization and bullying. Students were asked to nominate students who were being bullied in the past couple of months from a list of classmates: *"Which children are being bullied by other children?",* and to nominate students who bullied other children: *"Which classmates bully other children?".* Students could nominate an unlimited number of classmates for each item or nominate no one. To prevent a systematic nomination bias of classmates on top of the list, the order of student names was randomized. Received peer nominations were totalized and divided by the number of classmates responding, resulting in a proportion score ranging from 0.00 to 1.00 for each student on each item.

Peer-reported reinforcers, outsiders, and defenders. Also based on the Participant Roles Questionnaire (Kärnä et al., 2013; Salmivalli et al., 1996), three single items were used to identify students' participant roles in bullying situations concerning the past couple of months; reinforcers of bullies: *"Which classmates reinforce bullies, for example, by laughing or giggling when someone gets bullied?"*; outsider: *"Which classmates do nothing when someone gets bullied, for example, they walk away or act like they did not see the bullying?"*; and defenders of victims: *"Which classmates help children that are being bullied, for example, by comforting, supporting, or defending them?".* Similar to the procedure for peer-reported bullies and victims, students could nominate an unlimited number of classmates or no one. The list of names was randomized, and proportion scores were calculated for each role.

Stacking of program components. To investigate the effects of stacking components, we calculated and dichotomized each program component's implementation level. First, teachers were asked to indicate the degree to which they implemented each part (e.g., lesson 1, lesson 2, etc.) of each PRIMA component (e.g., student lessons) separately on a four-point scale: 0 = *not at all;* 1 = *less than 50%;* 2 = *more than 50%;* 3 = *completely.* We subsequently dichotomized the scores to indicate whether students (or their teachers) were sufficiently exposed to each program component. Regarding the universal program components, we considered an implementation of at least 50% of the components to be a successful implementation of student lessons, e-learning, and the monitor report. The face-to-face training was completed when teachers indicated that they attended the full training session. Concerning the selective components, the protocols for specific bullying situations, and the protocols for students directly involved were used when teachers indicated to have consulted at least one of the protocols for both types of protocols separately.

We determined the PRIMA program's universal components' implementation level by adding the dichotomized variables of student lessons, monitor-report, e-learning, and face-to-face training together into an aggregated implementation score. This resulted in the following scores: 0 = *no components implemented;* 1 = *one component;* 2 = *two components;* 3 = *three components,* or 4 = *four components.* We included these components as they are universal; the use of the selective components (i.e., protocols resulting from the monitor and the protocols for specific situations) heavily depends on specific bullying incidents at school. Therefore, the implementation of selective components is highly context-specific, and its interpretation is, therefore, less straightforward.

Program dosage. In addition to a dichotomous measure of implementation, we also used a continuous measure by calculating the number of hours performed for implementing the different components. Using teachers' reports on the extent to which they have implemented each component, we have estimated the average time spent on each program component. This procedure resulted in a possible program dosage ranging from 0-9 hours (i.e., student lessons: 0-4.5 hours; monitor report: 0-0.5 hour; e-learning: 0-2 hours; and face-to-face training: 0-2 hours).

Demographic information. Students reported their date of birth, gender, grade level, and ethnicity. Ethnicity was measured by asking what the student considered his or her background with the possibility to tick multiple options (e.g., Dutch and Moroccan). We then dichotomized students into 'western' or 'non-western' background, based on the criteria of the Dutch Central Statistical Office (CBS, n.d.).

Statistical Analysis

We used multilevel modeling with SPSS version 25 (IBM Corp. Released, 2017). Three-level hierarchical models were fitted, representing students nested in classrooms, and classrooms nested within schools. We controlled for differences in baseline levels by adding the pretest scores of the variable of interest. Also, ethnicity (i.e., western or non-western), gender, and age (grand-mean centered) were included in all models, as these are well-known covariates (see Salmivalli & Voeten, 2004; Vervoort et al., 2010). We explored possible interaction effects of both PRIMA conditions with ethnicity, gender, or age. Finally, we controlled for differences across conditions on school size, urbanization level, and the number of students with special needs with dummy-coded school-level variables, distinguishing between large schools (> 500 students), urban schools (large and medium cities), and high level of students with special needs (above the national average of 9.31% students with learning difficulties and/or emotional-behavioral problems but without an indicated disability or health care need (Smeets et al., 2007).

To investigate the effects of both PRIMA conditions on bullying behavior and students' roles in bullying situations, we tested a model including all 3,155 students (i.e., intention-to-treat analysis). This analysis provides an estimate of the program effects in general school practices with varying program implementation levels to establish typical effects in educational practice. We also analyzed the data to include only those students who were sufficiently exposed to at least one of the universal PRIMA components (i.e., a received-intervention analysis). To investigate whether classes that implemented one, two, three, or four universal program components showed more positive results, we compared these subgroups with classes that implemented zero universal program components, including the control group. In addition, we investigated whether program dosage was related to program effects.

RESULTS

Table 2 provides an overview of pre- and posttest scores and prevalence changes in mean proportion scores for students' roles in bullying situations. Pretest scores showed that 14.3% of the students ($n = 452$) reported being victimized at least twice a month. For self-reported bullying behaviors, 9.5% ($n = 261$) of the students scored an average of 6 or higher on performing different aggressive behaviors (e.g., kicking or excluding a peer) in the past three months. Self-reported victims declined in all three groups, and the largest decline was observed in PRIMA-L$^+$ schools. Also, self-reported bullies increased in all groups, again with the most considerable change in the PRIMA-L$^+$ schools. The proportion

Table 2 *Pre- and Posttest Scores and Changes in Mean (Proportion) Scores (SD) for Students' Roles in Bullying Situations by Condition (Intention to Treat Analysis)*

	PRIMA-L$^+$			PRIMA-L$^-$			Control		
	T1	T2	CS	T1	T2	CS	T1	T2	CS
Victims									
Self-report	0.635 (1.210)	0.450 (1.018)	-0.185	0.567 (1.127)	0.450 (1.013)	-0.117	0.580 (1.122)	0.440 (0.985)	-0.140
Peer-report	0.051 (0.085)	0.041 (0.082)	-0.010	0.036 (0.079)	0.041 (0.104)	+0.005	0.049 (0.097)	0.059 (0.133)	+0.010
Bullies									
Self-report	2.029 (3.041)	2.223 (2.989)	+0.194	1.788 (2.935)	1.957 (2.989)	+0.169	2.059 (3.052)	2.212 (3.104)	+0.153
Peer-report	0.063 (0.100)	0..078 (0.187)	+0.015	0.046 (0.097)	0.056 (0.1114)	+0.010	0.061 (0.108)	0.087 (0.170)	+0.026
Reinforcers									
Peer-report	0.064 (0.083)	0.075 (0.153)	+0.011	0.054 (0.073)	0.066 (0.102)	+0.012	0.068 (0.087)	0.086 (0.132)	+0.018
Outsiders									
Peer-report	0.074 (0.062)	0.124 (0.142)	+0.050	0.054 (0.054)	0.071 (0.082)	+0.017	0.079 (0.070)	0.086 (0.107)	+0.007
Defenders									
Peer-report	0.208 (0.126)	0.314 (0.317)	+0.106	0.212 (0.108)	0.264 (0.152)	+0.052	0.191 (0.111)	0.296 (0.291)	+0.105

Note. Victims self-report $N = 2,774$; Bullies self-report $N = 2,473$, Peer-report $N = 2,767$. CS = change score (computed as T2-T1)

of the number of nominations for victims decreased significantly from 0.635 to 0.450. The number of nominations, therefore, decreased by 20% in this group. Peer-reported bullies increased in all three groups, with the largest increase in the control group.

Regarding students' roles in bullying situations, the majority of the students received nominations for the role of defender ($M = 0.20$), followed by outsider ($M = 0.07$), and reinforcer ($M = 0.06$) at the pretest (see Table 2). The prevalence of the number of peer-reported defenders, outsiders, and reinforcers increased across all three groups.

Outcomes of the Intention-to-Treat Analysis

We found a statistically significant difference between PRIMA-L$^+$ schools and the control schools on peer-reported victims (see Table 3). PRIMA-L$^+$ schools had a significantly lower number of peer-reported victims than control schools, $b = -0.029$, $SE = 0.010$, $p = .008$, 95% CI [-0.05; -0.01]. Using the differences in adjusted mean proportion scores of the PRIMA-L$^+$ schools and control schools, we observed a small effect of PRIMA-L$^+$ on reducing peer-reported bullying ($d = -0.17$).

Table 3 *Estimates for Intention-to-Treat Intervention Effects of PRIMA on Students' Roles in Bullying Situations*

	Victims				Bullies				Reinforcers		Outsiders		Defenders	
	Self-report		Peer-report		Self-report		Peer-report		Peer-report		Peer-report		Peer-report	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Baseline														
Intercept	0.995***	0.175	0.021	0.024	0.048	0.554	0.006	0.052	0.005	0.047	0.131**	0.046	0.106	0.156
Fixed effects														
PRIMA-L$^+$	-0.029	0.068	-0.029**	0.010	-0.009	0.228	0.008	0.039	0.007	0.036	0.032	0,039	0.024	0.154
PRIMA-L$^-$	0.012	0.066	-0.012	0.009	-0.206	0.219	-0.029	0.038	-0.025	0.035	-0.045	0.038	-0.144	0.151
Pretest score	0.375***	0.015	0.846***	0.019	0.536***	0.017	0.949***	0.023	0.871***	0.025	0.329***	0.028	0.763***	0.027
Boy	0.014	0.034	-0.005	0.003	0.586***	0.099	0.016***	0.004	0.025***	0.004	-0.026***	0.003	-0.040***	0.005
Age	-0.068***	0.016	-0.000	0.002	0.086	0.050	-0.000	0.004	0.001	0.003	-0.006*	0.003	0.004	0.005
Non-western	-0.009	0.047	-0.005	0.004	0.076	0.139	0.017**	0.006	0.014***	0.005	-0.003	0.004	-0.004	0,007
Large schools	-0.084	0.056	-0.009	0.008	-0.183	0.189	-0.025	0.032	-0.030	0.030	-0.010	0.033	-0.009	0.128
Urban area	-0.039	0.057	0.015	0.008	0.174	0.189	0.041	0.032	0.032	0.030	0.081*	0,033	0.202	0.130
Special needs students	-0.036	0.056	0.007	0.008	-0.111	0.187	0.038	0.032	0.032	0.030	-0.005	0,032	0.052	0.127
Random effects														
Group level	0.015	0.007	0.001	0.000	0.222	0.074	0.010	0.002	0.010	0.002	0.031	0.004	0.315	0.041
School level	0.010	0.007	0.000	0.000	0.119	0.078	0.005	0.003	0.004	0,002	0.001	0.002	0.054	0.037
Δ AICª	-1,005.86***		+1,249.58***		-2,692.13***		+1,376.08***		+1,122.38***		-74.98***		+661.23***	

Note. Victims self-report $N = 2,774$; Bullies self-report $N = 2,473$, Peer-report $N = 2,767$.
ªAkaike Information Criteria.
* $p < .05$. ** $p < .01$. *** $p < .001$.

The intention-to-treat analysis showed no significant differences in the number of self-reported victims and bullies and the proportion scores for the roles of reinforcers, outsiders and, defenders between both PRIMA conditions and the control condition. No significant interaction effects were found for PRIMA with gender, age, or ethnicity.

Implementation Level of PRIMA Components

Teachers in the PRIMA-L$^+$ schools implemented universal and selective program components more intensively than teachers in the PRIMA-L$^-$ schools (see Figure 2). In both conditions, the PRIMA monitor report and protocols were used in most classes (n = 56 and n = 55, respectively). The protocols for specific bullying situations were least consulted in both conditions (n = 32). Teachers in the PRIMA-L$^+$ schools delivered an average of 5.32 (SD = 2.86) hours of the program, while teachers in the PRIMA-L$^-$ schools delivered an average of 2.10 hours (SD = 1.67) of the program.
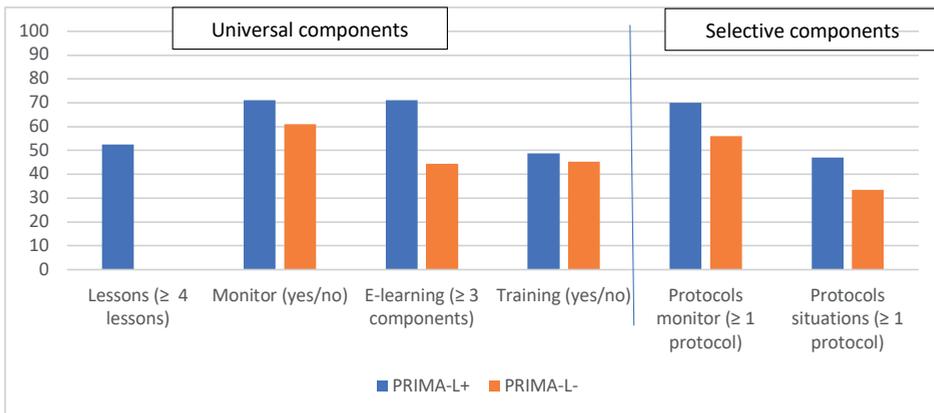

Figure 2. Implementation Level of PRIMA Components and Number of Exposed Students. *Note.* The student lessons were only offered to the PRIMA-L+ schools.

Table 4 provides an overview of the number of universal components implemented by teachers. The majority of the teachers (49 in 26 classes of 548 students) in interventions schools (i.e., PRIMA-L$^+$ and PRIMA-L$^-$) implemented two universal components. Most teachers carried out a combination of a training component (i.e., e-learning or face-to-face training) and the monitor report. One-fifth of the teachers implemented none of these components, indicating that none of these students were (in)directly exposed to the universal PRIMA components.

Outcomes of the Received Intervention Analysis

Removing students who did not receive the universal PRIMA components from the analyses showed similar patterns in pre- and posttest scores and changes in mean (proportion) scores compared to descriptive analyses, including all students (see Table 5). The proportion of the number of nominations for victims decreased by 28% in the PRIMA-L$^+$ group. As expected, effects in the PRIMA-L$^+$ schools were stronger (see Table 6). Compared to the control group, the mean proportion scores for peer-reported victimization declined in PRIMA-L$^+$ condition, b = -0.034, SE = 0.011, p = .005, 95% CI [-0.06; -0.01], Cohen's d = -0.17.

Table 4 *Number of Universal Components Implemented by Teachers*

	Number of classes (n)	Number of teachers (n)	Number of students (n)
Zero components implemented[*]	19	37	343
One component implemented	15	32	303
E-learning	3	6	57
Training	3	5	63
Student lessons	0	0	0
Monitor report	9	21	183
Two components implemented	26	49	548
Student lessons + e-learning	1	1	29
Student lessons + training	2	5	39
Student lessons + monitor report	1	1	25
E-learning + training	4	9	96
E-learning + monitor report	12	22	233
Training + monitor report	6	11	126
Three components implemented	21	39	439
Student lessons + e-learning + training	1	2	9
Student lessons + e-learning + monitor report	5	9	115
Student lessons + training + monitor report	2	3	19
E-learning + training + monitor report	13	25	296
Four components implemented[**]	10	21	222

[*]The 'zero component implemented' category excludes 1,280 control-group students in 73 classes and 139 teachers.
[**]Lessons + teacher e-learning + teacher training + monitor report.

5

Contrary to the intention-to-treat analysis, we found a significant decrease in the mean proportion scores for peer-reported reinforcing behavior in the PRIMA-L[+] schools compared to control schools, $b = -0.0354$, $SE = 0.016$, $p = .044$, 95% CI [-0.068; -0.001], Cohen's $d = -0.11$.

To conclude, the received intervention analysis demonstrated the decline in peer-reported victims more convincingly for the PRIMA-L[+] schools and revealed additional effects for the PRIMA-L[+] school on peer-reported reinforcers, compared to the intention-to-treat analysis. There were no significant differences in the number of self-reported victims and bullies and the number of outsiders and defenders between PRIMA conditions and the control condition. Also, no significant interaction effects were found for gender, age, or ethnicity.

Table 5 *Pre- and Posttest Scores and Changes in Mean (Proportion) Scores (SD) for Students' Roles in Bullying Situations by Condition (Received Intervention Analysis)*

	PRIMA-L+			PRIMA-L-			Control		
	T1	T2	CS	T1	T2	CS	T1	T2	CS
Victims									
Self-report	0.625 (1.199)	0.450 (1.022)	-0.175	0.502 (1.063)	0.400 (0.976)	-0.102	0.580 (1.122)	0.440 (0.985)	-0.140
Peer-report	0.054 (0.087)	0.039 (0.082)	-0.015	0.033 (0.082)	0.034 (0.095)	+0.001	0.049 (0.097)	0.059 (0.133)	+0.010
Bullies									
Self-report	2.061 (3.003)	2.226 (3.070)	+0.165	1.843 (3.054)	2.057 (3.127)	+0.241	2.059 (3.052)	2.212 (3.104)	+0.153
Peer-report	0.064 (0.103)	0.069 (0.138)	+0.005	0.048 (0.105)	0.053 (0.110)	+0.005	0.061 (0.108)	0.087 (0.170)	+0.026
Reinforcers									
Peer-report	0.062 (0.083)	0.064 (0.101)	+0.002	0.054 (0.075)	0.059 (0.089)	+0.005	0.068 (0.087)	0.086 (0.132)	+0.018
Outsiders									
Peer-report	0.074 (0.064)	0.126 (0.133)	+0.052	0.050 (0.053)	0.062 (0.081)	+0.012	0.079 (0.070)	0.086 (0.107)	+0.007
Defenders									
Peer-report	0.198 (0.119)	0.291 (0.202)	+0.093	0.206 (0.098)	0.255 (0.145)	+0.049	0.191 (0.111)	0.296 (0.291)	+0.105

Note. Victims self-report $N = 2,316$; Bullies self-report $N = 2,072$, Peer-report $N = 2,309$. CS = change score (computed as T2-T1)

Table 6 *Estimates for Received-Intervention Effects of PRIMA on Students' Roles in Bullying Situations*

	Victims				Bullies				Reinforcers		Outsiders		Defenders	
	Self-report		Peer-report		Self-report		Peer-report		Peer-report		Peer-report		Peer-report	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Baseline														
Intercept	0.987***	0.193	0.040	0.026	0.168	0.636	0.040	0.036	0.019	0.032	0.118*	0.046	0.127	0.138
Fixed effects														
PRIMA-L+	-0.031	0.072	-0.034**	0.011	-0.117	0.252	-0.033	0.017	-0.035*	0.016	0.020	0.037	-0.093	0.129
PRIMA-L-	-0.017	0.075	-0.018	0.011	-0.222	0.258	-0.033	0.018	-0.035	0.017	-0.045	0.041	-0.144	0.142
Pretest score	0.379***	0.017	0.832***	0.020	0.552***	0.019	0.935***	0.022	0.828***	0.023	0.311***	0.030	0.743***	0.028
Boy	-0.004	0.036	-0.005	0.003	0.621***	0.112	0.011*	0.004	0.022***	0.003	-0.026***	0.003	-0.036***	0.005
Age	-0.067	0.018	-0.002	0.002	0.079	0.059	-0.002	0.003	0.001	0.003	-0.005	0.003	0.002	0.005
Non-western	0.002	0.051	-0.003	0.005	0.254	0.159	0.017**	0.006	0.011*	0.005	-0.005	0.003	-0.012	0.007
Large schools	-0.082	0.061	-0.013	0.009	-0.267	0.213	-0.005	0.015	-0.018	0.014	0.017	0.032	0.094	0.111
Urban area	-0.042	0.061	0.016	0.009	0.097	0.211	0.021	0.014	0.019	0.014	0.080*	0.031	0.181	0.110
Special needs students	-0.039	0.061	0.009	0.009	-0.141	0.213	0.005	0.015	0.004	0.014	-0.029	0.031	-0.049	0.109
Random effects														
Group level	0.017	0.008	0.001	0.000	0.249	0.093	0.003	0.001	0.004	0.001	0.030	0.004	0.329	0.045
School level	0.010	0.008	0.000	0.000	0.136	0.089	0.001	0.000	0.000	0.000	0.000	0.000	0.007	0.018
Δ AICa	-799.62***		+1,091.82***		-2,253.07***		+1,404.73***		+1,147.87***		-47.76*		+577.30***	

Note. Self-report $N = 2,316$; Bullies self-report $N = 2,072$, Peer-report $N = 2,309$.
aAkaike Information Criteria.
* $p < .05$; ** $p < .01$; *** $p < .001$.

Effects of Stacking Universal Program Components

The degree of implementation varied in both experimental conditions. Dividing schools into subgroups of different levels of implementation (i.e., implementation of zero, one, two, three, or four universal components) showed that proportion scores for peer-reported victimization and reinforcing behavior significantly decreased when multiple components were delivered (see Table 7).

Table 7 *Estimates for Effects of Stacking Preventive Components on Students' Roles in Bullying Situations*

	Victims		Bullies		Reinforcers	Outsiders	Defenders
	Self-report	Peer-report	Self-report	Peer-report	Peer-report	Peer-report	Peer-report
Baseline							
Intercept	1.008***	0.024	0.064	0.021	0.027	0.133**	0.136
Fixed effects							
1 component (vs. 0)	-0.014	-0.011	-0.336	-0.050	-0.048	-0.048	-0.152
2 components (vs. 0)	0.007	-0.018	0.053	-0.051	-0.058*	-0.027	-0.139
3 components (vs. 0)	-0.046	-0.021	-0.202	-0.049	-0.068*	-0.025	-0.147
4 components (vs. 0)	-0.084	-0.049**	-0.334	-0.065	-0.055	0.019	-0.151
Pretest score	0.374***	0.848***	0.539***	0.937***	0.868***	0.333***	0.760***
Boy	0.014	-0.005	0.598***	0.016***	0.025***	-0.026***	-0.041***
Age	-0.068***	-0.000	0.083	0.000	0.001	-0.006*	0.004
Non-western	-0.009	-0.004	0.069	0.017**	0.014	-0.003	-0.005
Large schools	-0.092	-0.011	-0.155	-0.034	-0.043	-0.012	-0.023
Urban area	-0.037	0.015	0.184	0.043	0.034	0.083*	0.210
Special needs students	-0.045	0.004	-0.109	0.037	0.030	0.002	0.053
Random effects							
Group level	0.016	0.001	0.239	0.010	0.010	0.032	0.324
School level	0.008	0.000	0.116	0.004	0.003	0.001	0.044
Δ AIC[a]	-1,000.46***	+1,240.49***	-2,692.83***	+1,369.73***	+1,118.05***	-83.98*	+658.64***

Note. Victims self-report N = 2,774; Bullies self-report N = 2,473, Peer-report N = 2,767.
[a]Akaike Information Criteria.
* $p < .05$; ** $p < .01$; *** $p < .001$

5

Compared to classes where zero components had been implemented (including control schools), the proportion of peer-reported victims showed a significant decline in classes where all components were executed, b = -0.049, SE = 0.015, p = .001, 95% CI [-0.08; -0.02], Cohen's d = 0.07. We found similar results for the number of hours that teachers invested in the PRIMA program. The more hours teachers devoted to the program, a decline in mean proportion scores of

peer-reported victims was observed in classes, b = -0.005, SE = 0.001, p = .001, 95% CI [-0.01; -0.00]. In contrast to the stacking analysis, a significant decrease in proportion scores for peer-reported bullies was also revealed for every hour that teachers invested in PRIMA, b = -0.009, SE = 0.004, p = .041, 95% CI [-0.02; -0.00]. No significant rela-

tions were found between program stacking or dosage and self-reported victimization and bullying.

Furthermore, we found positive effects for stacking program components on the mean proportion scores for peer-reported reinforcers. Students in classes where two or three program component were implemented showed significant decreases in mean scores for peer-reported reinforcers compared to students in classes where no components were implemented, b = -0.06, SE = 0.03, p = .040, 95% CI [-0.11; -0.00], d = -0.11; b = -0.07, SE = 0.03, p = .032, 95% CI [-0.13; -0.01], d = -0.18, respectively. The implementation of four components did not further enhance this effect. Also, for the number of hours that teachers invested in the PRIMA program, a decline in mean proportion scores of peer-reported victims was observed in classes, b = -0.009, SE = 0.004, p = .001, 95% CI [-0.01; -0.00]. We found no statistically significant relation between stacking components or the number of hours of program delivery and the number of peer-reported outsiders and defenders. Also, no significant interaction effects were found on gender, age, or ethnicity.

DISCUSSION

This study provides experimental evidence for the assumed relationship between antibullying program components and bullying behavior using a design with two experimental groups and a control group. The results of our study indicate that it is specifically the PRIMA-L$^+$ program, including both student and teacher components, which is effective in reducing peer-reported bullying in grades 3-6 in elementary school. Since bullying emerges in primary education, this study provides positive indications that children in this age range are susceptible to antibullying programs' positive effects.

We found evidence for the effectiveness of PRIMA-L$^+$ to reduce the number of peer-reported victims and reinforcers. However, we did not find any significant declines in bullying and victimization for PRIMA-L$^-$, indicating that the student lessons are a crucial component. We did not find any differences either in self-reported victims and bullies between PRIMA- and control schools. The reduction in self-reported victims was the highest in the PRIMA-L$^+$ condition but did not reach statistical significance. Therefore, our first hypothesis that schools with PRIMA more effectively reduce bullying and victimization than control schools is partially supported by our findings but is not fully confirmed.

Furthermore, we found stronger effects when teachers implemented multiple program components in their classes. Classrooms, where all four universal components were implemented, showed the only statistically significant reduction in the number of peer-reported victims. This finding highlights that full implementation of the multi-

component antibullying program is crucial in achieving optimal results. Again, we did not find similarly positive results for self-reported victimization and bullying, and therefore, our study has found positive, but partial evidence for our hypothesis that the implementation of more program components is related to stronger program effects.

Our results highlight the importance of supporting both students and teachers to decrease bullying in schools. This finding is in line with the correlational outcomes of Ttofi and Farrington's (2011) meta-analysis. Our findings also underline the importance of a school-wide approach in antibullying programs, indicating that bullying is a complex phenomenon that needs to be addressed at multiple levels in the school (Hong & Espelage, 2012), including individual students, the classroom, the teacher, and school.

We found different results for self- and peer-reported bullying behavior. Self-reports are considered to reflect students' subjective perceptions of being bullied, whereas peer-reports are considered to reflect observed bullying behavior and students' reputations in the class (Kärnä et al., 2011). Several studies have reported a modest agreement between the two perspectives (Branson & Cornell, 2009; Demaray et al., 2013). Self-reports are considered the standard to measure bullying and victimization, while peer-reports are considered to be sensitive to reputation-bias (Olweus, 2003). However, as peers are often present at most bullying incidents (Salmivalli, 2014), peer-reports reflect multiple students' observations on the behaviors of each classmate. Therefore, such multiple-informant peer-reports of bullying and victimization are viewed as valid and reliable (Branson & Cornell, 2009; Ladd & Kochenderfer-Ladd, 2002). Besides, peer-reports are less influenced by bullies and victims' possible reluctance to report bullying (Branson & Cornell, 2009; Cornell et al., 2006). Seen from this perspective, peer- and self-reported bullying behavior are the proverbial two sides of the same coin. Few experimental studies have investigated effects on both self- and peer-reported bullying and victimization. The findings from our study on peer-reports are in line with the findings of Kärnä et al. (2011), where also stronger effects of the KiVa intervention were found for peer-reported victimization compared to self-reports.

Unlike previous research that showed effects on self-reported victimization (Gaffney et al., 2019; Kärnä et al., 2011), our findings show that the number of self-reported victims declined in all conditions. A possible explanation for this result is a decrease in self-perceived victimization because of the recent implementation of antibullying guidelines for all schools in the Netherlands, assuming that this policy has affected children's subjective perceptions. Another possible explanation is that peers in the classroom may be the first to observe a change in bullying behavior or reputations, whereas PRIMA may only result in delayed effects for victims' subjective experience. Possibly, bullying behavior has to stop before victims' experience improvement from their point of view. More experimental research into the different perspectives on self- and peer reports of victims is needed to study this matter.

5

We also explored the effects of PRIMA on other roles in bullying situations. Our results indicated a significant decrease in the mean proportion scores of reinforcers in PRIMA-L[+] schools compared to control schools. A notable finding was that even though the student lessons explicitly target all students to respond when bullying occurs, we did not find an increase in the number of defenders in the PRIMA-L[+] schools. This finding is not in line with the meta-analytical review of Polanin et al. (2012), who reported small to medium effect sizes for bullying prevention programs on defending behavior. A possible explanation for our deviant finding is that the number of peer-reported defenders was already relatively high at the pretest in all three conditions (i.e., proportion scores ranging from 0.19-0.21), which left little room for further improvement. Another possible explanation is that the proportion scores for the defender's role remained stable from pretest to post-test in all schools, while the number of victims declined in the intervention schools. However, the group process of bullying is complex, as previous research showed that victims and bullies are defended by their ingroup members, suggesting that the defender role can be controversial for students (Huitsing & Veenstra, 2012). In this study, we also found some combinations of negative roles (i.e., high proportion scores for bully-victims, reinforcer-victims, defender-bullies), indicating that individual students' roles can be controversial (see also, Ladd et al., 2017). Investigating which students defend which classmates, and whether these roles change as a result of the intervention, requires further investigation in future experimental research.

An unexpected finding concerned the differential effects and levels of implementation of PRIMA-L[+] schools versus PRIMA-L[-] schools. Our results show that teachers in PRIMA-L[+] schools implemented the program components more intensively than teachers in PRIMA-L[-] schools. A possible explanation for this finding is that delivering the student lessons in the PRIMA-L[+] condition had a stimulating effect on teachers, which subsequently supported the other components' implementation. Sainio et al. (2020) found similar results for the KiVa student curriculum. This finding suggests that a universal curriculum component with traditional lessons for the regular class may lay the foundation for teachers and students for optimal implementation of school-wide antibullying programs with various other components.

Strengths and Limitations

Our study's strength is that we conducted a cluster randomized controlled trial design, which enhances the internal validity of our experimental study (Farrington & Welsh, 2005). Furthermore, we conducted a priori randomization to prevent self-selection of schools. We then included schools with similar motivation levels in all conditions as a representative educational setting, supporting both the internal and external validity of our experimental findings (Astor et al., 2010).

Another strength of this study is that we used multiple informants to examine victimization and bullying. Because of the complexity of bullying, several scholars addressed the desirability to assess bullying using multiple informants (Branson & Cornell, 2009; Griffin & Gross, 2004). Using both self- and peer-reports enabled us to examine two distinct constructs: self-perceptions, relating to subjective individual experiences, and peer-perceptions, characterizing social representations in a group. In line with previous studies (Branson & Cornell, 2009; Cole et al., 2006), our results showed that self-reporting scores are lower than peer-reports. Victims in elementary school may be reluctant to report bullying in fear of the bully's reprisals, or because they are concerned that their reports are dismissed by their teacher (Fekkes et al., 2005; Newman & Murray, 2005). Students who bully others may be reluctant to report the bullying behavior in fear of social disapproval of their peers (Branson & Cornell, 2009).

A limitation of this study is that both experimental groups implemented the program with varying levels, which complicated the evaluation of the student lessons' specific contribution. Due to a low variation related to program implementation in the PRIMA-L⁻ group, we could not control statistically for the difference in the degree of program implementation in our models. Teachers and school management may not have had enough time to implement all components in addition to their existing curriculum and low motivation of staff to implement a school-wide program (Orobio de Castro et al., 2018). Future experimental research should measures factors that may influence the degree and quality of program implementation, such as perceptions on leadership and program effectiveness (Domitrovich et al., 2008; Durlak & DuPré, 2008; Orobio de Castro et al., 2018; Veenstra et al., 2014).

Furthermore, this study's scope was restricted to outcomes measured in grades 3-6, while the PRIMA program targeted all school students. Assessing bullying among younger children requires different, developmentally appropriate methods, such as using pictures of all children in the class to nominate classmates or observations and recordings of bullying situations (Alsaker & Nägele, 2008). We strongly recommend including such methods in future experimental research to investigate school-wide antibullying programs' effectiveness across all grades, since several studies show that bullying starts early in primary education.

Finally, the findings of this study should be understood within the Dutch cultural context. Since 2015, new guidelines were initiated by the Ministry of Education, which may have motivated both intervention and control schools to evaluate their policies and to implement a new program.

Practical Implications

Our results indicate that a school-wide antibullying program with multiple components, strengthening both students and teachers, effectively reduces peer-reported peer

victimization. An important practical implication is that schools could benefit from evidence-based antibullying programs, and more attention is needed to increase awareness among school management and teachers to select and implement these programs adequately. Implementing a school-wide program requires support and time from all professionals in the school (Durlak & DuPre, 2008; Gaffney et al., 2019; Orobio de Castro et al., 2018), and school management and teachers need to be facilitated in resources and time to implement a school-wide program appropriately.

More attention is needed in experimental research and educational practice for self- and peer-reported victimized students who remain victimized despite the implementation of an antibullying program. There is a growing body of literature indicating that not all students benefit from a universal school-wide antibullying program (Huitsing et al., 2020b; Kaufman et al., 2018) and that victimized students may experience high feelings of distress (Garandeau & Salmivalli, 2019; Huitsing et al., 2020b). These findings indicate the need for more research and future development of programs for this group of vulnerable students.