Journal of Clinical Child & Adolescent Psychology, 00(00), 1–15, 2016

ISSN: 1537-4416 print/1537-4424 online
DOI: 10.1080/15374416.2016.1233498

Routledge
Taylor & Francis Group

Dynamic Change of Aggressive Behavior and Victimization Among Adolescents: Effectiveness of the ViSC Program

Takuya Yanagida and Dagmar Strohmeier

School of Medical Engineering and Applied Social Sciences, University of Applied Sciences Upper Austria

Christiane Spiel

Department of Applied Psychology: Work, Education, Economy, University of Vienna

The present study reports a high-quality evaluation of the ViSC Social Competence Program, which was implemented large scale in Austria. A rigorous test of program effectiveness has been performed to investigate the dynamic change of aggressive behavior and victimization and to ensure a high level of statistical conclusion validity. A cluster randomized control study was applied to examine program effectiveness regarding aggressive behavior and victimization. In sum, 1,377 adolescents (48.5% girls, M_{age} = 11.7) enrolled in 13 schools participated in the program; 665 adolescents (45.2% girls, M_{age} = 11.6) enrolled in 5 schools were in the control group. Data were collected with Internet-based questionnaires at pre- and posttest with several validated scales to capture the full range of the two constructs. To ensure construct validity, a series of invariance tests of the second-order factor models were performed. To test program effectiveness, a multiple group bivariate latent change score model was applied. Evidence for a dynamic change of aggressive behavior and victimization was found. As predicted, the pretest levels and the change scores of aggressive behavior and victimization were associated. Moreover, higher levels of pretest values predicted more change. The program was effective in reducing victimization but not aggressive behavior. Gender did not moderate the results. Results are important for national rollout and cross-national dissemination of the program. However, further research is needed to investigate the underlying mechanisms of the intervention effects.

INTRODUCTION

Large-scale country-comparative studies consistently show that a substantial number of adolescents around the world are involved in aggressive behavior and bullying in schools (Craig & Harel, 2004; Currie et al., 2008, 2012). As a result, several European countries started to tackle these problems by large-scale implementing whole-school prevention programs (e.g., Finland; see Kärnä et al., 2011). In line with an ecological understanding of development (Bronfenbrenner, 1979; Hawley & Williford, 2015; Swearer & Espelage, 2004), whole-school programs target not only individuals but also the class (e.g., peers and bystanders) and the school level (e.g., teachers and school policies). Although several meta-analyses (e.g., Ttofi & Farrington, 2011; Wilson & Lipsey, 2007) indicate that whole-school programs are effective, high-quality evaluation studies of whole-school programs that have been implemented large scale are rare (exceptions are, e.g., Kärnä et al., 2011; Olweus, 2004). Also, there is a lack of studies that applied rigorous methods to ensure statistical conclusion validity when evaluating program effectiveness. In addition, the vast majority of the existing program evaluation studies treated aggressive behavior and victimization as two separate outcome variables (see Ttofi & Farrington, 2011) and neglected that these phenomena are both concurrently and longitudinally associated (Swearer & Cary, 2003). Moreover, it is important to understand that whole-school programs are heterogeneous, as they comprise a variety of different concrete measures and they are always implemented in a particular national school system. Because the school

Correspondence should be addressed to Takuya Yanagida; Dagmar Strohmeier, School of Medical Engineering and Applied Social Sciences, University of Applied Sciences Upper Austria, Garnisonstraße 21, Linz 4020, Austria. E-mail: takuya.yanagida@fh-linz.at; dagmar.strohmeier@fh-linz.at

systems also vary considerably between different European countries, the effectiveness of these approaches to tackle bullying and aggressive behavior cannot be taken for granted.

Thus, the present study advances the literature both methodologically and conceptually by applying the strongest possible statistical approach to test whether a whole-school prevention program that was implemented large scale in Austria dynamically changed aggressive behavior and victimization. The present study is also important for policy and practice given the potential impact of a whole-school program, which is rolled out nationwide, on the development of young people.

Why Statistical Conclusion Validity Matters

For many years there has been a call for program evaluation studies based on the highest methodological standards to ensure their validity (e.g., Farrington & Ttofi, 2009). However, not many studies considered statistical conclusion validity when investigating whole-school program effectiveness (Trochim & Donnelly, 2006). Statistical conclusion validity (T. D. Cook & Campbell, 1979) refers to the appropriate use of statistics to infer the covariance of two variables (i.e., treatment and outcome variable). In program evaluation research, common threats to statistical conclusion validity include unreliability of measures, biased estimates of effects, violation of statistical assumptions, repeated tests for significant relationships, and heterogeneity of units (Maxwell & Delaney, 2004). These threats result from several methodological challenges, for example, missing data, measurement error, or multiple outcome variables. Over the past decades, statistical research has advanced considerably, providing statistical methods to tackle these challenges, which have been applied for the present study. Although 53 program evaluation studies were included in the most comprehensive meta-analysis investigating the effectiveness of school-based programs to reduce bullying (Ttofi & Farrington, 2011), the vast majority of these studies did not fully consider statistical conclusion validity. Only a few program evaluation studies treated missing data according to state-of-the-art statistical approaches, rigorously tested construct validity, adequately treated measurement error, or used multiple outcome variables. However, all these aspects of statistical conclusion validity were not considered in one single evaluation study to date.

Given the potential impact of a whole-school program on the development of young people when rolled out nationwide, there is no doubt that high-quality program evaluation including high statistical conclusion validity is important. Moreover, high-quality program evaluation studies are able to enhance the methodological level of this research field.

Why Dynamic Change Is Important

The present study argues that it is important to simultaneously investigate the change of aggressive behavior and victimization in program evaluation. Aggressive behavior comprises intentional harm doing, whereas bullying is a subcategory of aggressive behavior, which also includes repetition and imbalance of power (Olweus, 1996; Roland & Idsøe, 2001). Both bullying and aggressive behavior are expressed directly and indirectly and include a variety of negative forms, like physical attacks, verbal insults, relational harassments (Olweus, 1996), and offenses via electronic means (Gradinger, Strohmeier, & Spiel, 2009; Smith et al., 2008). Victimization comprises exactly the same phenomena but from the perspective of the victimized adolescents. It is important to capture different forms of aggressive behaviors and victimization when evaluating the effectiveness of a whole-school program, because they are moderately associated and they usually co-occur (Gradinger et al., 2009). Moreover, research demonstrated that there is a perpetrator/victim continuum (Swearer & Cary, 2003), meaning that aggressive behavior and victimization are both concurrently and longitudinally associated. Therefore, it is plausible to assume that these two phenomena also change simultaneously due to an intervention. Conceptually, it is therefore a more accurate view on intervention effects to consider the dynamic change of aggressive behavior and victimization. More specifically, dynamic change comprises three important aspects: (a) the covariation of the change scores of aggressive behavior and victimization, (b) the impact of the prescores of the same construct, and (c) the impact of the prescores of the other construct. Therefore, it is crucial to account for all three aspects within one statistical model to control for effects on one construct while testing for the other. It is surprising that the dynamic change has never been taken into account when evaluating the effectiveness of an antibullying program. Instead, in studies to date, aggressive behavior and victimization were treated as independent outcome variables. Therefore, the reliable estimation of program effects is questionable.

The ViSC Program

The ViSC Social Competence Program (Strohmeier, Hoffmann, Schiller, Stefanek, & Spiel, 2012) has been developed and implemented in Austria, a country with comparatively high rates of aggressive behavior and bullying (Craig & Harel, 2004; Currie et al., 2008, 2012) and an educational system that is characterized by resistance to change and low commitment (Spiel & Strohmeier, 2011). The ViSC program is designed for secondary schools; these are fifth to eighth grades in Austria, when students are 11 to 14 years old. The program has been developed for these grades because prevalence rates of bullying and victimization are related to school transitions (Pellegrini & Long, 2002), which take place between Grades 4 and 5 in Austria and have been shown to be highest among 13-year-olds (Currie et al., 2012). Between 2008 and 2013, the program was promoted in Austria as one measure of the Austrian national strategy to prevent violence in schools. The Austrian national strategy comprised six activity domains including prevention/intervention and evaluation/research (for more details, see Spiel & Strohmeier, 2011). Therefore, it was possible to implement the ViSC program in Austria and

to conduct a cluster-randomized evaluation study. Based on a socioecological perspective on development (Bronfenbrenner, 1979; Swearer & Espelage, 2004), the program defines the prevention of aggressive behavior and victimization as a whole-school task, and the initial implementation of the program lasts 1 school year. Thus, the program aims not only to change the behavior of the students (e.g., individual and class level) but also to foster knowledge and competences among teachers to be able to initiate several change processes on the school level. During program implementation, several routines on the school level (e.g., playground supervision or the handling of acute bullying cases) are targeted. The program is implemented via several in-school teacher trainings and a class project for students (Strohmeier et al., 2012b).

The main goal at the school level is to create a change process during which as many teachers as possible are committed to work "together against violence" (Spiel & Strohmeier, 2011). Thus, at the school level teachers are the main change agents. Therefore, the program aims to foster the shared responsibility among them, which in turn implies that as many teachers as possible in the school have worked out a common understanding of aggressive behavior and bullying, agreed on procedures on how to tackle acute cases, and jointly implemented preventive measures (Schultes, Stefanek, van de Schoot, Strohmeier, & Spiel, 2014). The change processes on school level are promoted via a series of in-school training sessions in which all teachers are invited to participate. This in-school training is organized at least three times during the whole school year by so-called ViSC coaches—school psychologists or other multipliers who attended a 1-year training offered by the program developers and researchers.

At the class level, a 13-unit class project is implemented by teachers in their classes. The fully manualized class project subsequently trains a broad spectrum of competencies among students considered important for the prevention of aggressive behavior and victimization. The class units are highly structured and contain single worksheets, group activities, interactive games, and a summary sheet with main messages.

To begin, in many classes there is a rather big group of students who neither feel responsible for what happens around them nor intervene in critical situations (Craig, Pepler, & Atlas, 2000). Therefore, all students are trained (a) to feel responsible when something negative is going on and (b) to react in a way that is likely to improve the situation.

Second, in secondary school there is still a group of adolescents who are not able to manage their negative emotions in a nonaggressive way (Salmivalli & Nieminen, 2002). Therefore, all students are trained (a) to recognize their own emotions and the emotions of others and (b) to cope with these emotions in a positive, nonaggressive way. Third, it is necessary to empower students who might get victimized easily so that they are not attacked because of their nonassertive behavior (Veenstra et al., 2007). Thus, all students are trained on how to best react when getting picked on by others. Besides reacting in an assertive way, students are advised to tell somebody (ideally an adult) if somebody is victimized in the class.

These competences are trained during the 13 units by applying a large variety of didactic methods (e.g., role-plays, small-group work, or whole-class discussions). All units comprise single worksheets, group activities, interactive games, and a summary sheet (for more details regarding the units, see Strohmeier et al, 2012b).

At the individual level, the program assumes that aggressive behavior has two underlying functions (Card & Little, 2006). Reactive aggression is theoretically grounded in the frustration-aggression model; instrumental aggression has its roots in social cognitive learning theory describing a planned behavior controlled by external rewards (Vitaro, Brendgen, & Barker, 2006). Although anger is the central emotion for reactive aggression, instrumental aggression is characterized by positive or neutral emotions. In addition, the program acknowledges that victimized students are heterogeneous (Yang & Salmivalli, 2013). Whereas victims of bullying usually cannot easily defend themselves, reactive aggressive students are also often victims and thus labelled as bully-victims. Important to note, the program acknowledges that aggressive behavior and victimization co-occurs and aims to change these behaviors simultaneously. The knowledge regarding the individual-level mechanisms of reactive and instrumental aggressive behavior is important for the teachers who are trained (a) to recognize and differentiate bullies, victims and bully-victims and (b) to conduct structured conversations with them. These state-of-the-art conversations are based on the approach suggested by Roland and Vaaland (2006), and they differ between bullies, victims, and bully-victims.

The Present Study

The present study is the first investigation of the effectiveness of the ViSC program regarding aggressive behavior and victimization. The goal of the evaluation study was to meet high research standards including an analytic strategy based on state-of-the-art methodology to ensure statistical conclusion validity (T. D. Cook & Campbell, 1979). To begin with, aggressive behavior and victimization were each measured with three validated scales to capture the full range of the two constructs. In addition, the analytic strategy includes missing data imputation, testing measurement invariance, and structural equation modelling.

We hypothesized that the ViSC program yields substantial reduction in aggressive behavior and victimization controlling for age, pretest scores of aggressive behavior, and victimization, and we explored whether gender moderates the effectiveness of the program. Regarding age, it is well researched that age covaries with aggressive behavior and bullying and that antibullying programs are more effective for younger students (Yeager, Fong, Lee, & Espelage, 2015). As for pretest scores of aggressive behavior and victimization, it is conceivable that

change in these constructs is negatively related to the initial level, that is, the higher the initial level of aggressive behavior and victimization, the stronger the decrease in these constructs. That is, program effectiveness was investigated, accounting for the dynamic change of aggressive behavior and victimization among youth. Accounting for the dynamic change is important because aggressive behavior and victimization are moderately associated in both cross-sectional and longitudinal studies (C. Cook, Williams, Guerra, Kim, & Sadek, 2010). Thus, it is likely that these two behaviors might also change simultaneously, and changes in one variable (e.g., aggressive behavior) might be misattributed to program effectiveness when not controlling for the changes in the other variable (e.g., victimization). In contrary to other evaluation studies of antibullying programs we are aware of, the present study is the first that investigated the dynamic change of aggressive behavior and victimization.

Conceptually, it is a more accurate view on intervention effects when considering the dynamic change of aggressive behavior and victimization than investigating them separately.

To explore whether gender moderates program effectiveness is important for several reasons: First, 29% of boys but 12% of girls were identified as being part of a risk group regarding aggressive behavior and victimization in a national representative study conducted in Austria (Strohmeier, Gradinger, Schabmann, & Spiel, 2012). Moreover, boys are overrepresented in the small but highly maladapted group of bully-victims (Yang & Salmivalli, 2013). Thus, it is possible that gender moderates program effectiveness, because aggressive behavior and victimization might be differently related among boys than girls.

METHOD

Design and Procedure

In December 2008, all secondary schools located in the capital city of Austria were invited to participate in the program. Out of all 155 secondary schools located in Vienna, 34 schools applied for participation, from which 26 schools fulfilled the necessary requirements (e.g., they were willing to participate in the evaluation study). A cluster randomized controlled study design was applied, that is, 13 schools were randomly assigned to the intervention group, and five of the 13 remaining schools agreed to serve as control schools. Thus, eight schools declined participation in the study after being selected as control schools; they did not receive the program. In the 13 intervention schools, the program was implemented between September 2009 and June 2010 with a cascaded train-the-trainer model. That is, scientists trained ViSC coaches, ViSC coaches trained teachers, and teachers trained their students (see Strohmeier et al., 2012b). To keep the data collection month constant, we collected the pretest in May 2009, when adolescents were at the end of Grade 5, and the posttest in May 2010, when adolescents were at the end of Grade 6. In Austria, the class compositions are fairly constant in secondary schools. Thus, very few children left the classes after the pretest data collection, and very few newcomers entered the classes either at the beginning or during the Grade 6 school year.

After the study was accepted by the local school council and the school principals, active parental consent was obtained. In general, parental consent rates are very high in Austria, ranging between 85% and 100%. Because participation in the longitudinal study was a requirement at the school level to be chosen for the cluster-randomized trial, the parental consent was greater than 90% in all 18 schools. In total, 71.1% of students were present at the day of data collection and participated in the study. Data were collected through Internet-based questionnaires, which were completed during 1 regular school hr in the school's computer lab under the supervision of two trained research assistants. Prior to data collection students were assured that their answers would be kept confidential. To avoid any systematic order effect, items within scales were counterbalanced across participants.

Program Implementation Fidelity

The quality of program implementation was carefully monitored in the intervention schools during the whole school year. Schultes and colleagues (2014) assessed several indicators of implementation fidelity and participant responsiveness and demonstrated that they are related with proximal program outcomes among teachers. For instance, teachers' self-efficacy to stop aggressive behavior among students was significantly more enhanced in schools where the ViSC program had been implemented with high fidelity. Furthermore, only teachers with high participant responsiveness significantly changed their strategies when being asked to handle hypothetical bullying situations.

Participants

In sum, 2,042 students (1,377 in the intervention group, 665 in the control group) participated in at least one occasion of measurement and were included in the current study. At Wave 1 (pretest), the sample comprised 1,639 students (47.6% girls) from 103 classrooms (50 fifth-grade classes, 51 sixth-grade classes, and 2 seventh-grade classes) with a mean age of 11.7 years (SD = 0.9, Min = 10, Max = 15). According to a commonly used classification system (see Stefanek, Strohmeier, van de Schoot, & Spiel, 2011), 46.4% of the students were nonimmigrant Austrians, 20.2% were immigrants from countries of the former Yugoslavia, 14.3% were immigrants from Turkey, and 19.1% were immigrants from other countries. Intervention and control groups were compared on pretest demographic characteristics using a series of Pearson's chi-square tests and a two-sample t

test. There was no statistically significant result indicating that the two groups were fairly comparable.

Missing Data

In total, 20.69% of data were missing, resulting from two main missing data patterns: Students who participated at pretest only (n = 515) and students who participated at posttest only (n = 403). The remaining 56 students had a general missing data pattern with missing value on single items. The mean percentage of missing values across the 53 variables ranged between 0.00% and 25.60%.

A series of two-sample Wilcoxon tests with Bonferroni-Holm correction for multiple comparisons was conducted for attrition analysis. Results showed no differences between students who dropped out of the study (effect sizes ranged between r = .01 and r = .08) or students who dropped in the study (effect sizes ranged between r = .00 and r = .07) after pretest and students with complete data in any variable used in the present study.

Multiple imputation (Rubin, 1987) under the missing at random assumption was used to deal with missing data. For more details of the imputation process, see the appendix. Calculations were performed in R (R Core Team, 2014) using the mice package (van Buuren & Groothuis-Oudshoorn, 2011).

Measures

Aggressive behavior and victimization were each measured with three scales: (a) Bullying Perpetration/Victimization, (b) Physical Aggression/Victimization, and (c) Relational Aggression/Victimization.

Bullying Perpetration and Bullying Victimization

Self-reported bullying and victimization were measured by items developed for the PISA 2009 study in Austria (Strohmeier et al., 2012a). The Bullying scale as well as the Victimization scale consists of a global item and three specific items covering different forms of bullying. In the global item, students were asked, "How often have you insulted or hurt other students during the last 2 months?" and "How often have others insulted or hurt you in the last 2 months?" The three specific items were similar to the global ones, except that they described specific forms of bullying and victimization. Cronbach's alpha coefficients were .82/.83 (pretest/posttest) for the Bullying Perpetration scale and .81/.82 (pretest/posttest) for the Bullying Victimization scale.

Physical Aggression and Physical Victimization

The peer nomination measure developed by Crick and Grotpeter (1995) was modified into a self-report questionnaire and comprised three items, for example, "How often have you hit one or more classmates during the last 2 months?" and "How often have you been hit by one or more classmates during the last 2 months?" Cronbach's alpha coefficients were .79/.79 (pretest/posttest) for the Physical Aggression scale and .74/.76 (pretest/posttest) for the Physical Victimization scale.

Relational Aggression and Relational Victimization

These five items were also adapted from the peer nomination measure originally developed by Crick and Grotpeter (1995), for example, "Some kids leave other kids out on purpose when it's time to play or do an activity. How often have you done that during the last 2 months?" and "How often during the last 2 months have you been excluded from play or another activity by one or more classmates?" Cronbach's alpha coefficients were .83/.87 (pretest/posttest) for the Relational Aggression scale and .82/.81 (pretest/posttest) for the Relational Victimization scale.

Answers to all questions were given on a 5-point response scale: 0 (*not at all*), 1 (*once or twice*), 2 (*two or three times a month*), 3 (*once a week*), and 4 (*nearly every day*).

Measurement Model and Measurement Invariance

A measurement model for aggressive behavior and victimization that represents the hierarchical relations among the investigated constructs was established (see Figure 1). We modeled that the higher order factors aggressive behavior and victimization account for the relations among bullying perpetration/victimization, physical aggression/victimization, and relational aggression/victimization, resulting in a second-order factor model. Furthermore, we took into account that the items are not continuous but ordered-categorical. Thus, our subsequent analyses are based on a second-order factor model with ordered-categorical indicators (see Bovaird & Koziol, 2012).

Measurement models of aggressive behavior and victimization were tested for measurement invariance between girls and boys within control and intervention group across pre- and posttest. A series of multiple group confirmatory factor analysis was conducted in Mplus version 7.11 (Muthén & Muthén, 1998–2012) to test hierarchical series of models to establish strong measurement invariance for the first- and second-order models (Chen, Sousa, & West, 2005). Model specification and identification were based on Millsap and Yun-Tein (2004) using theta parameterization and a robust weighted least squares estimator. To evaluate whether the assumption of invariance is tenable, change in comparative fit index (CFI) and root mean square error of approximation (RMSEA) were considered. It has been suggested that a change in CFI more than .01 (Cheung & Rensvold, 2002) and a change in RMSEA more than .01 (Chen, 2007) indicate a meaningful decrease in model fit, making the invariance assumption not reasonable.

The path diagram for the unstandardized solution of the measurement model assuming strong measurement

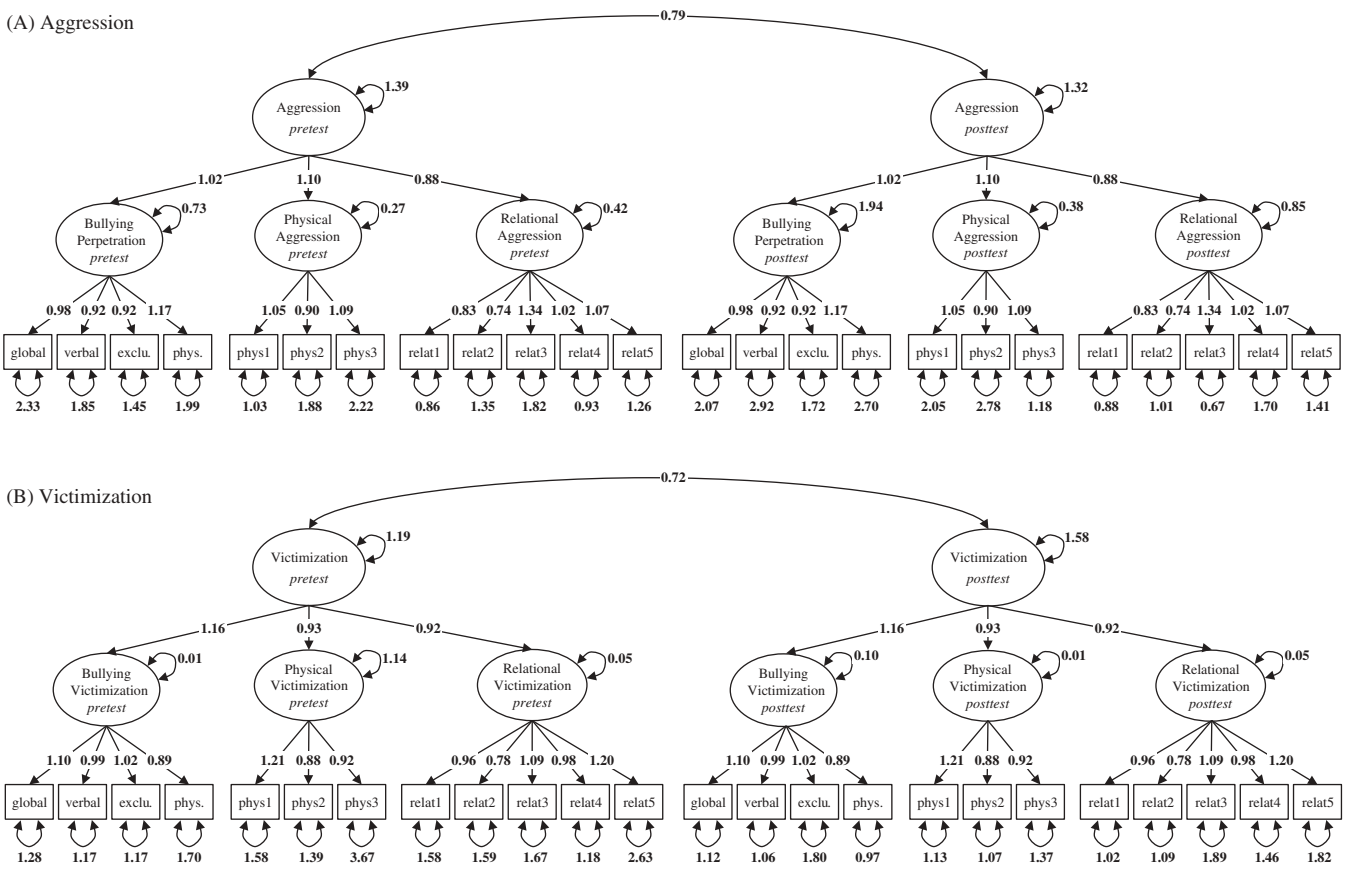

FIGURE 1 Results of the second-order measurement model for aggressive behavior (Panel A) and victimization (Panel B): Unstandardized solution. *Note*: Residual covariances stemming from the correlation of indicator-specific variance over time are not shown in the path diagram; effect coding (Little, Slegers, & Card, 2006) was used for model identification. global = global bullying/victimization; verbal = verbal bullying/victimization; exclu. = exclusion bullying/victimization; phys. = physical bullying/victimization; phys1-phys3 = Item 1–3 of physical aggression/victimization; relat1-relat5 = Item 1–5 of relational aggression/victimization.

invariance (Model 3b) is depicted in Figure 1, Panel A for aggressive behavior and Panel B for victimization.

Results showed no meaningful decrease in model fit between the hierarchically nested models for aggressive behavior and victimization. Thus, all first- and second-order factor loadings, threshold of the measured variables, and first-order factors were invariant between girls and boys within control and intervention group across pre- and posttest. Detailed results for testing measurement invariance can be obtained from the first author upon request.

Analytic Strategy

A structural equation modeling approach was used to test the main hypotheses of the study (see Little, 2013). More specifically, we applied a bivariate multiple group latent change score (LCS) model (McArdle, 1988) comparing intervention and control group (Model 1) and comparing four groups: (a) intervention group girls, (b) intervention group boys, (c) control group girls, and (d) control group boys (Model 2).

The basic idea of the LCS model is that the observed score Y at time t can be decomposed into the score Y at time t – 1 plus a difference score ΔY, where ΔY represents latent change from time t – 1 to time t:

$$Y_t = Y_{t-1} + \Delta Y. \quad (1)$$

From Equation 1 we can see that

$$\Delta Y = Y_t - Y_{t-1}. \quad (2)$$

The latent change score ΔY represents interindividual differences in intraindividual change from t – 1 to time t on Y. A positive sign indicates an increase, whereas a negative sign indicates a decrease over time. Because the latent change score ΔY is part of the model, it can be statistically controlled for the pretest score and additional covariates.

In the main analysis of the present study, we investigate not observed variables but common factors based on

measurement models. That is, the latent change score from pretest to posttest on aggressive behavior and victimization are corrected for measurement error.

To account for the covariation between aggressive behavior and victimization, we applied a bivariate latent change score model. Thus, we statistically controlled the latent change for pretest scores of both aggressive behavior and victimization. In addition, we incorporated mean centered age as covariate.

Intervention effect was investigated comparing intervention and control group; that is, the difference between the change in the intervention and the control group represents the intervention effect. In Model 1, we compared intervention and control group for the whole sample in order to investigate overall intervention effects. In Model 2, we compared intervention and control group within the subgroup of girls and boys in order to investigate intervention effects for girls and boys separately. Moderation effect of gender on program effectiveness was investigated comparing the intervention effects between girls and boys.

Statistical analyses were conducted using Mplus version 7.11 (Muthén & Muthén, 1998–2012). Models were estimated using robust weighted least squares estimator.

RESULTS

Descriptive Statistics

Response Category Proportions and Intraclass Correlations

Response category proportions for all scale items are reported for pre- and posttest in Table 1. As expected, the item response distribution was highly positive skewed with the largest proportion of students reporting the absence of aggressive behavior and victimization incidents. However, a substantial proportion of students reported being involved in aggressive behavior and/or victimization once or twice within the last 2 months at pretest (range = .105–.365) and at posttest (range = .092–.372).

The intraclass correlations (ICCs) indicated that, on average, 4.8% (*SD* = 14%, *Min* = 2.1%, *Max* = 7.6%) of the overall variance in the scale items is attributable to the clustering of students within classes. Thus, the standard errors were adjusted for the design effects (see Snijders & Bosker, 2012) in all analyses of the present study. However, we did not account for the ICCs on the school level because the ICCs on school level after accounting for the class level were very low (*M* = .021, *SD* = 0.010, *Min* = .004, *Max* = .041), resulting in a negligible impact of the design effect on standard error estimation.

Baseline Equivalence

Latent means of aggressive behavior and victimization were tested for differences between control and intervention group at pretest. We found no mean differences for aggressive behavior ($\Delta\hat{\alpha}$ = 0.171, p = .130, latent d = 0.131) or victimization ($\Delta\hat{\alpha}$ = 0.170, p = .057, latent d = 0.148).

Because we investigate intervention effects in girls and boys, we also tested latent mean differences within girls and boys separately. As for girls, we found no mean differences for aggressive behavior ($\Delta\hat{\alpha}$ = −0.100, p = .506, latent d = 0.080) or victimization ($\Delta\hat{\alpha}$ = −0.170, p = .146, latent d = 0.148). Likewise, there were also no mean differences for aggressive behavior ($\Delta\hat{\alpha}$ = 0.256, p = .052, latent d = 0.195) or victimization ($\Delta\hat{\alpha}$ = 0.156, p = .181, latent d = 0.131) in boys. These results indicate that control and intervention group are equivalent at baseline in the whole sample as well as in the subgroups of girls and boys separately.

Overall Intervention Effects on Aggressive Behavior and Victimization

In the main analysis, the bivariate multiple group LCS model investigating intervention effects on aggressive behavior and victimization controlling for age was analyzed in the whole sample (Model 1). Results revealed a good model fit, $\chi^2(2469) = 3513.566$, $p < .05$, CFI = .969, RMSEA = .020.

As expected, aggressive behavior and victimization at pretest were correlated ($\hat{\psi}$ = 1.044, $p < .001$). In addition, the change of aggressive behavior and the change victimization were also correlated in the control group ($\hat{\psi}$ = 0.925, $p < .001$) and in the intervention group ($\hat{\psi}$ = 0.869, $p < .001$). The mean change of aggressive behavior and victimization in the intervention and control group, as well as the mean difference between these groups (i.e., intervention effect), are shown in Table 2.

Results indicate that aggressive behavior did not change in the intervention group ($\hat{\alpha}$ = −0.145, p = .364) or in the control group ($\hat{\alpha}$ = 0.089, p = .253). The mean difference between intervention and control group did not reach statistical significance ($\Delta\hat{\alpha}$ = −0.233, p = .253) indicating no intervention effect for aggressive behavior. Although the effect size of the latent mean difference is not negligible (latent d = 0.185). Victimization decreased in the intervention group ($\hat{\alpha}$ = −0.938, $p < .05$) but did not change in the control group ($\hat{\alpha}$ = −0.076, p = .397). The mean difference between intervention and control was statistically significant ($\Delta\hat{\alpha}$ = −0.862, $p < .05$) with an effect size of latent d = 0.725.

Intervention Effects on Aggressive Behavior and Victimization for Girls and Boys

For testing intervention effects on aggressive behavior and victimization in girls and boys, the bivariate multiple group LCS model controlling for pretest scores and age was analyzed (Model 2). Results revealed a good model fit, $\chi^2(5022) = 6169.712$, $p < .05$, CFI = .969, RMSEA = .021. Figure 2 shows the path diagram for the LCS model without the measurement models, Panel A for girls and Panel B for boys.

TABLE 1
Response Category Proportions for Bullying Perpetration and Bullying Victimization, Physical Aggression, and Physical Victimization, and Relational Aggression and Relational Victimization Scale for Pre- and Posttest

	Pretest					*Posttest*				
Variable	*Not at All*	*Once or Twice*	*Two or Three Times a Month*	*Once a Week*	*Nearly Every Day*	*Not at All*	*Once or Twice*	*Two or Three Times a Month*	*Once a Week*	*Nearly Every Day*
Bullying Perpetration (ICC class level: *M* = .041, *SD* = .005, *Min* = .033, *Max* = .046; ICC school level: *M* = .024, *SD* = .010, *Min* = .013, *Max* = .041)										
global	.504	.349	.068	.033	.046	.476	.372	.047	.042	.064
verbal	.545	.325	.049	.034	.047	.572	.304	.046	.028	.051
physical	.737	.193	.021	.023	.027	.782	.140	.028	.021	.030
exclusion	.772	.152	.021	.023	.031	.787	.131	.026	.021	.035
Physical Aggression (ICC class level: *M* = .041, *SD* = .009, *Min* = .030, *Max* = .054; ICC school level: *M* = .027, *SD* = .010, *Min* = .014, *Max* = .040)										
hit	.640	.228	.048	.032	.052	.605	.258	.055	.030	.052
push	.514	.351	.046	.040	.049	.501	.346	.058	.038	.057
kick	.789	.136	.028	.026	.020	.784	.126	.032	.029	.029
Relational Aggression (ICC class level: *M* = .028, *SD* = .010, *Min* = .009, *Max* = .045; ICC school level: *M* = .021, *SD* = .011, *Min* = .006, *Max* = .037)										
exclude	.638	.271	.035	.023	.033	.626	.268	.042	.031	.034
together	.636	.272	.038	.022	.031	.645	.246	.046	.031	.033
lies	.806	.123	.029	.013	.029	.795	.126	.029	.025	.024
like	.829	.105	.028	.015	.022	.823	.108	.026	.017	.026
mean	.789	.146	.023	.015	.027	.744	.170	.042	.023	.021
Bullying Victimization (ICC class level: *M* = .034, *SD* = .012, *Min* = .013, *Max* = .057; ICC school level: *M* = .013, SD = .006, *Min* = .004, *Max* = .024)										
global	.424	.365	.077	.050	.084	.482	.349	.059	.039	.071
verbal	.430	.356	.071	.054	.089	.489	.356	.044	.040	.070
physical	.673	.220	.040	.026	.040	.748	.176	.025	.023	.028
exclusion	.693	.197	.028	.032	.049	.779	.149	.022	.023	.027
Physical Victimization (ICC class level: *M* = .037, *SD* = .007, *Min* = .028, *Max* = .046; ICC school level: *M* = .024, *SD* = .010, *Min* = .014, *Max* = .040)										
hit	.740	.163	.025	.025	.047	.803	.119	.024	.027	.027
push	.511	.302	.057	.045	.084	.604	.268	.039	.033	.057
kick	.767	.140	.025	.031	.037	.839	.096	.020	.020	.025
Relational Victimization (ICC class level: *M* = .023, *SD* = .006, *Min* = .014, *Max* = .032; ICC school level: *M* = .021, *SD* = .010, *Min* = .007, *Max* = .036)										
exclude	.664	.200	.041	.034	.061	.720	.188	.034	.021	.038
together	.603	.256	.052	.039	.049	.608	.278	.052	.026	.035
lies	.601	.235	.058	.052	.055	.595	.267	.055	.032	.052
like	.790	.117	.025	.027	.041	.843	.092	.023	.019	.024
mean	.659	.211	.042	.040	.048	.682	.221	.033	.025	.039

Note: Imputed data. *N* = 2,042. Intraclass correlation coefficients (ICCs) were computed using variance components of a three-level multilevel model estimated in Mplus (Muthén & Muthén, 1998–2012).

TABLE 2
Results of the Bivariate LCS Model for Girls and Boys: Mean Change of Aggressive Behavior and Victimization for Control and Intervention Groups

	Mean Change Aggressive Behavior				*Mean Change Victimization*			
Model	*Unstandardized Estimate 95% CI*	*SE*	*p-value*	*Effect size latent* d	*Unstandardized Estimate 95% CI*	*SE*	*p-value*	*Effect size latent* d
Mode 1: Whole sample								
Intervention Group (IG)	-0.145 [-0.457, 0.167]	0.159	.364		**-0.938** [-1.279, -0.597]	0.174	<.001	
Control Group (CG)	0.089 [-0.095, 0.273]	0.094	.346		-0.076 [-0.252, 0.100]	0.090	.397	
IG – CG	-0.233 [-0.703, 0.363]	0.204	.253	0.185	**-0.862** [-1.701, -0.369]	0.221	<.001	0.725
Model 2: Girls and Boys								
Girls								
Intervention Group (IG)	-0.163 [-0.616, 0.290]	0.231	.482		**-0.80**2 [-1.186, -0.418]	0.196	<.001	
Control Group (CG)	**0.285** [0.036, 0.534]	0.127	.025		0.016 [-0.229, 0.261]	0.125	.900	
IG – CG	-0.448 [-1.016, 0.120]	0.290	.122	0.442	**-0.818** [-1.333, -0.303]	0.263	<.05	0.775
Boys								
Intervention Group (IG)	**-0.503** [-0.840, -0.166]	0.172	<.01		**-0.962** [-1.387, -0.537]	0.217	<.001	
Control Group (CG)	-0.332 [-0.761, 0.097]	0.219	.128		0.073 [-0.407, 0.553]	0.245	.766	
IG – CG	-0.170 [-0.703, 0.363]	0.272	.531	0.128	**-1.035** [-1.701, -0.369]	0.340	<.01	0.765
Program Effectiveness x Gender Girls (IG – CG) – Boys (IG – CG)	-0.278 [-1.072, 0.516]	0.405	.493		0.217 [-0.608, 1.042]	0.421	.607	

Note. 95% CI = 95% confidence interval; Statistically significant results at α = .05 are boldface; effect size latent *d* according to Hancock (2001).

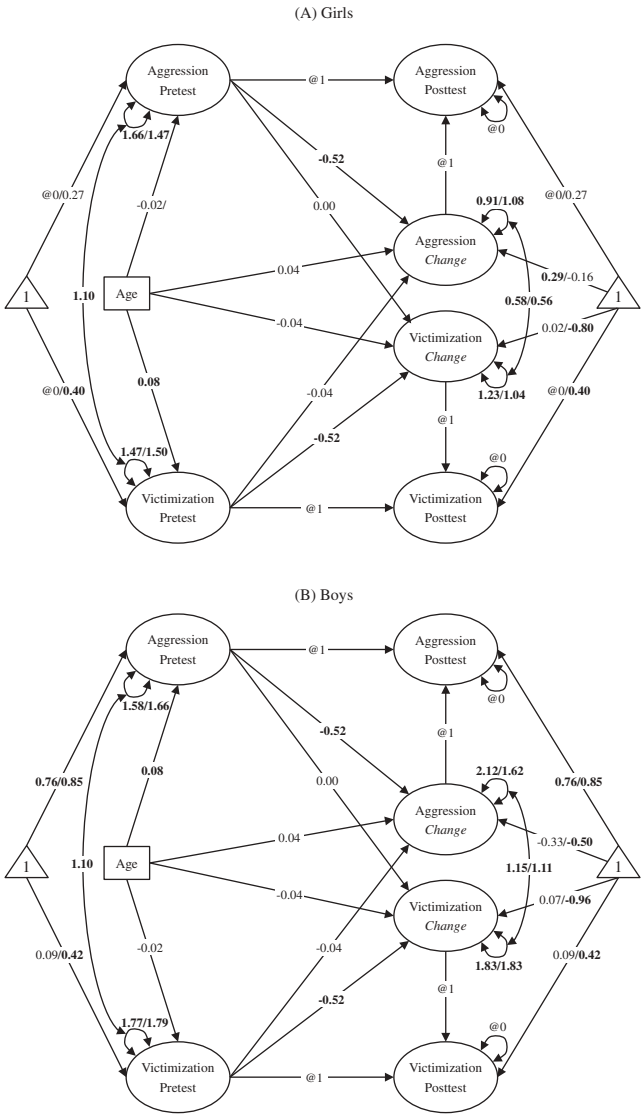

FIGURE 2 Structural model of the bivariate latent change score model: Unstandardized solution for intervention/control group for girls (Panel A) and boys (Panel B). *Note*: Statistically significant results at α = .05 are shown in bold. Parameter constraints are indicated with @. The triangle with the number 1 in the center represent the intercept of a regression equation (see Little, 2013), that is, latent mean of the factor.

The mean change of aggressive behavior and victimization in the intervention and control group as well as the mean difference between these groups (i.e., intervention effect) for girls and boys are shown in Table 2.

As for girls, results indicate that aggressive behavior did not change in the intervention group ($\hat{\alpha} = -0.163$, $p = .482$) but increased in the control group ($\hat{\alpha} = 0.285$, $p < .05$). However, the mean difference between intervention and control group did not reach statistical significance ($\Delta\hat{\alpha} = -0.448$, $p = .122$), although the effect size of the latent mean difference was substantial (latent $d = 0.442$). Victimization decreased in the intervention group ($\hat{\alpha} = -0.802$, $p < .05$) but did not change in the control group ($\hat{\alpha} = 0.016$, $p = .900$). The mean difference between intervention and control was statistically significant ($\Delta\hat{\alpha} = -0.818$, $p < .05$) with an effect size of latent $d = 0.775$.

As for boys, results revealed that aggressive behavior decreased in the intervention group ($\hat{\alpha} = -0.503$, $p < .05$), whereas aggressive behavior did not change in the control group ($\hat{\alpha} = -0.332$, $p = .128$). However, the mean difference between intervention and control group did not reach statistical significance ($\Delta\hat{\alpha} = -0.170$ $p = .531$), although the effect size of the latent mean difference is not negligible (latent $d = 0.128$). Victimization, on the other hand, decreased in the intervention group ($\hat{\alpha} = -0.962$, $p < .05$) but did not change in the control group ($\hat{\alpha} = 0.073$, $p = .766$). The mean difference between intervention and control was statistically significant ($\Delta\hat{\alpha} = -1.035$, $p < .05$) with an effect size of latent $d = 0.765$.

There were no moderation effects of gender on program effectiveness, neither on aggressive behavior ($\Delta\hat{\alpha} = -0.278$, $p = .493$) nor on victimization ($\Delta\hat{\alpha} = 0.217$, $p = .607$).

Dynamic Change of Aggressive Behavior and Victimization

Based on Model 1, the results of the effects of covariates in the whole sample on the mean change of aggressive behavior and victimization are shown in Table 3.

The mean change of aggressive behavior was related to the pretest score of aggressive behavior ($\hat{\beta} = -0.518$, $p < .05$) but not to the pretest score of victimization ($\hat{\beta} = -0.037$, $p = .454$), indicating the higher aggressive behavior at pretest, the stronger the decrease in aggressive behavior. The mean change of victimization was not related to the pretest score of aggressive behavior ($\hat{\beta} = 0.002$, $p = .972$) but to the pretest score of victimization ($\hat{\beta} = -0.520$, $p < .05$), indicating the higher victimization at pretest, the stronger the decrease in victimization. Age was related to neither the mean change of aggressive behavior ($\hat{\beta} = 0.042$, $p = .343$) nor the mean change of victimization ($\hat{\beta} = -0.041$, $p = .325$).

DISCUSSION

International research shows that aggressive behavior and bullying are severe problems in schools around the word (Craig & Harel, 2004; Currie et al., 2008, 2012). In the last decades, various prevention programs have been developed, implemented, and evaluated in several countries (Ttofi & Farrington, 2011). In contrary to most of the available programs that are designed to prevent bullying, a subtype of aggressive behavior, the ViSC program aims to prevent a broad spectrum of aggressive behavior (including bullying) and victimization.

The present study is the first longitudinal cluster randomized control group design study to investigate the effectiveness of the ViSC program regarding aggressive behavior and victimization.

TABLE 3
Results of the Bivariate LCS model: Regression of Mean Change of Aggressive Behavior and Victimization on Aggressive Behavior at Pretest, Victimization at Pretest, and Age

	Mean Change Aggressive Behavior				*Mean Change Victimization*			
Predictor	*Unstandardized Estimate 95% CI*	*SE*	p-*value*	*Standardized Estimate*	*Unstandardized Estimate 95% CI*	*SE*	p-*value*	*Standardized Estimate*
Aggressive Behavior at pretest	**–0.518** [–0.630, –0.406]	0.057	<.001	–0.565	0.002 [–0.112, 0.116]	0.058	.972	0.002
Victimization at pretest	–0.037 [–0.131, 0.059]	0.049	.454	–0.038	**–0.520** [–0.634, 0.406]	0.058	<.001	–0.488
Age	0.042 [–0.044, 0.128]	0.044	.343	0.029	–0.041 [–0.123, 0.041]	0.042	.325	–0.026

Note. 95% CI = 95% confidence interval. Statistically significant results at α = .05 are boldface.

We tested the effectiveness of the program over 1 year regarding the reduction of aggressive behavior and victimization both for the whole sample and separately for boys and girls. We took into account the dynamic change of aggressive behavior and victimization and investigated pretest scores and age as possible moderators of change. To ensure statistical conclusion validity, a highly sophisticated statistical approach was applied.

Program Effectiveness on Aggressive Behavior and Victimization

In line with our assumption that aggressive behavior and victimization change in a dynamic way, the present analyses revealed a substantial positive association between the change of aggressive behavior and the change of victimization. This result is important, as it gives a more accurate view on intervention effects compared with investigating the two constructs seperately. Furthermore, the pretest values of the two constructs were statistically significant predictors of the change variables, indicating that youth with initial high levels of aggressive behavior and victimization change more in the respective variable compared with initially low-involved youth. Taking these important longitudinal associations between constructs into account is important so as not to overestimate or misattribute possible program effects.

Thus, considering these dynamic changes of aggressive behavior and victimization, results indicate that the ViSC program is effective in reducing victimization in the whole sample (latent $d = 0.725$), as well as girls (latent $d = 0.775$) and boys (latent $d = 0.765$) controlling for age. Important to note, the effect size of the latent mean difference was substantial, indicating that the large-scale implementation of a systemic whole-school prevention program with a broad understanding of aggressive behavior and victimization is valuable. However, results indicated that the ViSC program was not effective in reducing aggressive behavior, although the change was in the desired direction. The effect size indicates a practically relevant reduction of aggressive behavior in the intervention group compared to the control group for girls (latent $d = 0.442$), although the effect did not reach statistical significance.

What could be the reasons that the program had such a large impact on victimization but a nonsignificant and much lower impact on aggressive behavior?

There might be several reasons for this result. To begin with, the ViSC program was implemented based on a cascaded train-the-trainer model. That is, we did not work directly with the schools; we trained ViSC coaches who worked with teachers who worked with their students. As a result, the implementation quality varied between schools (Schultes et al., 2014). The ViSC program contains universal and indicated actions. While the universal actions compose in-school trainings and the class project and target all teachers and students, the indicated actions consist of a structured way of conversations that the teachers should realize with students who have been identified as victims, bullies, or bully-victims. Important to note, these conversations differ for these three groups of youth. During the in-school trainings, teachers are trained to distinguish reactive and proactive aggression as underlying motives in perpetrators and are advised to implement the talks in a different way depending on the motive system of the perpetrator. Our previous analyses give reason to assume that overall the universal actions were implemented with moderate to high fidelity in all participating schools (for details, see Schultes et al., 2014). However we did not gather any data regarding the implementation quality of the indicated actions. Thus, we are not aware how well teachers were able to distinguish reactive and proactive aggression and how many talks with bullies and bully-victims have been realized in the participating schools. This lack of knowledge is unfortunate, as it can be assumed that these talks have the highest potential to sustainably change aggressive behavior in subgroups of aggressive youth (Roland & Vaaland, 2006). Thus, it is possible that these conversations were not realized well enough in the participating schools and that the universal

actions that foster responsibility, empathy, social-emotional competencies, and school climate more generally were too weak to have an impact on aggressive children.

Our analyses revealed that the level of aggressive behavior had an impact on the change score of aggressive behavior. More specifically, the higher aggressive behavior at pretest, the higher the decrease in aggressive behavior. However, we did not investigate other subgroups of aggressive youth, for example, the popular bullies (Garandeau, Lee, & Salmivalli, 2014; Sijtsema, Veenstra, Lindenberg, & Salmivalli, 2009). It is reasonable to assume that the program was more effective for particular groups of aggressive youth (e.g., bully-victims), which should be investigated in future studies (Yang & Salmivalli, 2015). To apply person-oriented approaches for data analysis in future studies is also important in order to be able to test Risk × Intervention effects, as it was shown that a similar whole-school program that was implemented in Finland was particular beneficial among youth who were most victimized before program implementation (Juvonen, Schacter, Sainio, & Salmivalli, 2016). However, it is important to investigate the psychosocial adaptation of those youth who remain victims even after program implementation, as it is possible that their situation in the class gets worse.

Limitations

Although the present study has a high methodological standard, several limitations should be mentioned. First, we relied on self-assessments only. When applying an intervention program large scale (as was the case in the present study), self-report measures are often chosen because they are easy to apply and are reliable, given that multiple items are used to measure a construct (Yanagida et al., 2016). The strengths and weaknesses of self-report measures in studies about aggressive behavior including bullying have already been discussed extensively in the literature (e.g., Solberg & Olweus, 2003). Likely, aggressive behavior is underestimated using self-reports because perpetrators might not report the "true" frequency of their behavior but might underestimate it. Thus, self-report measures should be interpreted with caution. Second, in the present study the program developers were also the program evaluators. This is a disadvantage because program developers led implementations so that evaluations might be positively biased toward program effectiveness (Malti, Ribeaud, & Eisner, 2011). Thus, independent implementations and evaluations are needed. On the other hand, the program developers did not work directly with the schools but with trained ViSC coaches who implemented the program in schools. Third, the cluster randomized study might be considered a disadvantage from an experimental point of view, because only self-selected volunteer schools were randomized and not all of them were willing to serve in the control group. However, given the complexity and duration of the intervention, including a large amount of teacher trainings and class units, an experimental approach was not feasible.

Implications for Research, Policy, and Practice

To replicate the present findings is an important step toward "evidence-based practice." Currently, the ViSC program has been implemented in Romania (Trip et al., 2015), Cyprus (Solomontos-Kountouri, Gradinger, Yanagida, & Strohmeier, 2016), and Turkey. These cross-national dissemination studies are important to generalize the present findings. Moreover, investigation of the theoretical model and underlying mechanisms of the intervention effects is needed. Furthermore, subgroups of children (e.g., bullies, victims, bully-victims, and uninvolved) should be investigated based on a person-centered analytic approach (see von Eye & Spiel, 2010). Last, side effects of the program (e.g., academic motivation) should be investigated in future studies. Most important, effort is needed to implement evidence-based programs into educational systems (Spiel, Schober, & Strohmeier, 2016; Spiel & Strohmeier, 2011, 2012; Spiel, Wagner, & Strohmeier, 2012). To reach this goal, intensive collaboration and communication between policymakers, practitioners, and researchers are needed.

ACKNOWLEDGMENTS

We are very grateful to the whole ViSC project team consisting of Eva-Maria Schiller, Elisabeth Stefanek, Petra Gradinger, Christoph Burger, Bianca Pollhammer, Katharina Derndarsky, Marie Therese Schultes and Christine Hoffmann for their invaluable work during the intervention study. We also want to thank the ViSC coaches and teachers who implemented the program in the schools and classes. We thank all schools and students who participated in this study. Special thanks to Alexander Robitzsch from the Leibniz Institute for Science and Mathematics Education (IPN) for his advice and helpful comments concerning the multiple imputation of missing data.

FUNDING

The implementation and evaluation of the ViSC program was funded by the Austrian Federal Ministry for Education, Arts and Cultural Affairs (PI: Christiane Spiel) between 2008 and 2011. The data analyses and writing of the present study was funded by the Platform for Intercultural Competences, University of Applied Sciences Upper Austria (PI: Dagmar Strohmeier) between 2012 and 2015.

REFERENCES

Bovaird, J. A., & Koziol, N. A. (2012). Measurement models for ordered categorical indicators in structural equation modelling. In R. H. Hoyle, D. Kaplan, G. Marcoulides, & S. West (Eds.), *Handbook of structural equation modeling* (pp. 495–511). New York, NY: Guilford Press.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Card, N. A., & Little, T. D. (2006). Proactive and reactive aggression in childhood and adolescence: A meta-analysis of differential relations with psychosocial adjustment. *International Journal of Behavioral Development*, *30*, 466–480. doi:10.1177/0165025406071904

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 464–504. doi:10.1080/10705510701301834

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*, 471–492. doi:10.1207/s15328007sem1203_7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 233–255. doi:10.1207/S15328007SEM0902_5

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in missing data procedures. *Psychological Methods*, *6*, 330–351. doi:10.1037/1082-989X.6.4.330

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

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Craig, W., & Harel, Y. (2004). Bullying, physical fighting and victimization. In C. Currie, C. Roberts, A. Morgan, R. Smith, W. Settertobulte, & O. Samdal (Eds.), *Young people's health in context: Health behavior in school-aged children (HBSC) study: International report from the 2001/2002 survey* (pp. 133–144). Copenhagen, Denmark: WHO Regional Office for Europe.

Craig, W., Pepler, D., & Atlas, R. (2000). Observations of bullying in the playground and in the classroom. *School Psychology International*, *21*, 22–36. doi:10.1177/0143034300211002

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development*, *66*, 710–722. doi:10.2307/1131945

Currie, C., Gabhainn, S. N., Godeau, E., Roberts, C., Smith, R., Currie, D., … Barnekow, V. (Eds.). (2008). *Inequalities in young people's health: HBSC international report from the 2005/06 survey. Health policy for children and adolescents, No. 5*. Copenhagen, Denmark: WHO Regional Office for Europe.

Currie, C., Zanotti, C., Morgan, A., Currie, D., De Looze, M., Roberts, C., … Barnekow, V. (Eds.). (2012). *Social determinants of health and well-being among young people: Health behaviour in school-aged children (HBSC) study: International report from the 2009/2010 survey*. Copenhagen, Denmark: WHO Regional Office for Europe.

Farrington, D. P., & Ttofi, M. M. (2009). School-based programs to reduce bullying and victimization. *Campbell Systematic Reviews*, *6*, 1–147. doi:10.4073/csr.2009.6

Garandeau, C. F., Lee, I. A., & Salmivalli, C. (2014). Differential effects of the KiVa anti-bullying program on popular and unpopular bullies. *Journal of Applied Developmental Psychology*, *35*, 44–50. doi:10.1016/j.appdev.2013.10.004

Gradinger, P., Strohmeier, D., & Spiel, C. (2009). Traditional bullying and cyberbullying: Identification of risk groups for adjustment problems. *Journal of Psychology*, *217*, 205–213. doi:10.1027/0044-3409.217.4.205

Graham, J. W. (2009). Missing data analysis: Making it work in the Real World. *Annual Review of Psychology*, *60*, 549–576. doi:10.1146/annurev.psych.58.110405.085530

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206–213. doi:10.1007/s11121-007-0070-9

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, *66*, 373–388. doi:10.1007/BF02294440

Harrell, F. H. (2006). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York, NY: Springer.

Hawley, P. H., & Williford, A. (2015). Articulating the theory of bullying intervention programs: Views from social psychology, social work, and organizational science. *Journal of Applied Developmental Psychology*, *37*, 3–15. doi:10.1016/j.appdev.2014.11.006

Juvonen, J., Schacter, H. L., Sainio, M., & Salmivalli, C. (2016). Can a school-wide bullying prevention program improve the plight of victims? Evidence for Risk × Intervention effects. *Journal of Consulting and Clinical Psychology*, *84*, 334–344. doi:10.1037/ccp0000078

Kärnä, A., Voeten, M., Little, T., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa anti-bullying program; Grades 4-6. *Child Development*, *82*, 311–330. doi:10.1111/j.1467-8624.2010.01557.x

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Little, T. D., Slegers, D. W., & Card, N. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*, 59–72. doi:10.1207/s15328007sem1301_3

Malti, T., Ribeaud, D., & Eisner, M. P. (2011). The effectiveness of two universal preventive interventions in reducing children's externalizing behavior: A cluster randomized controlled trial. *Journal of Clinical Child & Adolescent Psychology*, *40*, 677–692. doi:10.1080/15374416.2011.597084

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective*. Mahwah, NJ: Erlbaum.

McArdle, J. J. (1988). Dynamic but structural equation modelling of repeated measures data. In R. B. Cattell & J. Nesselroade (Eds.), *Handbook of multivariate experimental psychology* (pp. 561–614). New York, NY: Plenum Press.

Mevik, B.-H., & Wehrens, R. (2007). The pls package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, *18*, 1–24. doi:10.18637/jss.v018.i02

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*, 479–515. doi:10.1207/S15327906MBR3903_4

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Olweus, D. (1996). *The revised olweus bully/victim questionnaire*. Bergen, Norway: Research Center for Health Promotion (HEMIL Center), University of Bergen.

Olweus, D. (2004). Bullying at school: Prevalence estimation, a useful evaluation design, and a new national initiative in Norway. *Association for Child Psychology and Psychiatry Occasional Papers*, *23*, 5–17.

Pellegrini, A. D., & Long, J. D. (2002). A longitudinal study of bullying, dominance, and victimization during the transition from primary school through secondary school. *British Journal of Developmental Psychology*, *20*, 259–280. doi:10.1348/026151002166442

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Roland, E., & Idsøe, T. (2001). Aggression and bullying. *Aggressive Behavior*, *27*, 446–462. doi:10.1002/ab.1029

Roland, E., & Vaaland, G. (2006). *ZERO teacher´s guide to the Zero Anti-Bullying Programme*. Stavanger, Norway: Centre for Behavioural Research, University of Stavanger.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley & Sons.

Salmivalli, C., & Nieminen, E. (2002). Proactive and reactive aggression among school bullies, victims, and bully-victims. *Aggressive Behavior*, *28*, 30–44. doi:10.1002/ab.90004

Schultes, M.-T., Stefanek, E., van de Schoot, R., Strohmeier, D., & Spiel, C. (2014). Measuring implementation of a school-based violence prevention program: Fidelity and teachers' responsiveness as predictors of proximal outcomes. *Zeitschrift Für Psychologie/Journal of Psychology*, *222*, 49–57. doi:10.1027/2151-2604/a000165

Sijtsema, J. J., Veenstra, R., Lindenberg, S., & Salmivalli, C. (2009). Empirical test of Bullies' Status Goals: Assessing direct goals, aggression, and prestige. *Aggressive Behavior*, *35*, 57–67. doi:10.1002/ab.20282

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

Snijders, T. A. B., & Bosker, T. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

Solberg, M., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior*, *29*, 239–268. doi:10.1002/ab.10047

Solomontos-Kountouri, O., Gradinger, P., Yanagida, T., & Strohmeier, D. (2016). The implementation and evaluation of the ViSC Program in Cyprus: Challenges of cross-national dissemination and evaluation results. *European Journal of Developmental Psychology*. Advance online publication. doi:10.1080/17405629.2015.1136618

Spiel, C., Schober, B., & Strohmeier, D. (2016). Implementing intervention research into public policy - the "I3-Approach". *Prevention Science*. Advance online publication. doi:10.1007/s11121-016-0638-3

Spiel, C., & Strohmeier, D. (2011). National strategy for violence prevention in Austrian schools and kindergarten: Development and implementation. *International Journal of Behavioral Development*, *35*, 412–418. doi:10.1177/0165025411407458

Spiel, C., & Strohmeier, D. (2012). Evidence-based practice and policy: When researchers, policy makers, and practitioners learn how to work together. *European Journal of Developmental Psychology*, *9*, 150–162. doi:10.1080/17405629.2011.616776

Spiel, C., Wagner, P., & Strohmeier, D. (2012). Violence prevention in Austrian schools: Implementation and evaluation of a national strategy. *International Journal of Conflict and Violence*, *6*, 176–186.

Stefanek, E., Strohmeier, D., van de Schoot, R., & Spiel, C. (2011). Bully-victim behaviour in ethnically diverse classes: Risk and protective factors on the individual and classroom level. *International Journal of Developmental Science*, *5*, 73–84. doi:10.3233/DEV-2011-11073

Strohmeier, D., Gradinger, P., Schabmann, A., & Spiel, C. (2012a). Gewalterfahrungen von Jugendlichen. Prävalenzen und Risikogruppen [Adolescent's experiences of violence]. In F. Eder (Ed.), *PISA 2009. Nationale Zusatzerhebungen [PISA 2009. Additional National Investigations]* (pp. 165–208). Münster, Germany: Waxmann.

Strohmeier, D., Hoffmann, C., Schiller, E.-M., Stefanek, E., & Spiel, C. (2012b). ViSC Social Competence Program. *New Directions for Youth Development*, *133*, 71–84. doi:10.1002/yd.20008

Swearer, S. M., & Cary, P. T. (2003). Perceptions and attitudes towards bullying in middle school youth: A developmental examination across the bully/victim continuum. *Journal of Applied School Psychology*, *19*, 63–79. doi:10.1300/J008v19n02_05

Swearer, S. M., & Espelage, D. L. (2004). Introduction: A social-ecological framework. of bullying among youth. In D. L. Espelage & S. M. Swearer (Eds.), *Bullying in American schools: A social-ecological perspective on prevention and intervention* (pp. 1–12). Mahwah, NJ: Erlbaum.

Trip, S., Bora, C., Sipos-Gug, S., Tocai, I., Gradinger, P., Yanagida, T., & Strohmeier, D. (2015). Bullying prevention in schools by targeting cognitions, emotions and behavior: Evaluating the effectiveness of the REBE-ViSC program. *Journal of Counseling Psychology*, *62*, 732–740. doi:10.1037/cou0000084

Trochim, W. M. K., & Donnelly, J. P. (2006). *The research methods knowledge base* (3rd ed.). Cincinnati, OH: Thomson.

Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic meta-analytic review. *Journal of Experimental Criminology*, *7*, 27–56. doi:10.1007/s11292-010-9109-1

van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: Chapman & Hall/CRC Press.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1–67.

Veenstra, R., Lindenberg, S., Zijlstra, B. J. H., 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*, 1843–1854. doi:10.1111/j.1467-8624.2007.01102.x

Vitaro, F., Brendgen, M., & Barker, E. D. (2006). Subtypes of aggressive behaviors: A developmental perspective. *International Journal of Behavioral Development*, *30*, 12–19. doi:10.1177/0165025406059968

von Eye, A., & Spiel, C. (Eds.). (2010). *Person-oriented Research. Zeitschrift für Psychologie/Journal of Psychology*. Göttingen, Germany: Hogrefe.

Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine*, *33*, 130–143. doi:10.1016/j.amepre.2007.04.011

Yanagida, T., Gradinger, P., Strohmeier, D., Solomontos-Kountouri, O., Trip, S., & Bora, C. (2016). Cross-national prevalence of traditional bullying, traditional victimization, cyberbullying and cyber-victimization: Comparing single-item and multiple-item approaches of measurement. *International Journal of Developmental Science*, *10*, 21–32. doi:10.3233/DEV-150173

Yang, A., & Salmivalli, C. (2013). Different forms of bullying and victimization: Bully-victims versus bullies and victims. *European Journal of Developmental Psychology*, *10*, 723–738. doi:10.1080/17405629.2013.793596

Yang, A., & Salmivalli, C. (2015). Effectiveness of the KiVa antibullying programme on bully-victims, bullies and victims. *Educational Research*, *57*(1), 80–90. doi:10.1080/00131881.2014.983724

Yeager, D. S., Fong, C. J., Lee, H. Y., & Espelage, D. L. (2015). Declines in efficacy of anti-bullying programs among older adolescents: Theory and a three-level meta-analysis. *Journal of Applied Developmental Psychology*, *37*, 36–51. doi:10.1016/j.appdev.2014.11.005

APPENDIX: Missing Data Imputation

An important goal of missing data imputation is to capture the missing at random mechanism. Thus, we adopted an inclusive analysis strategy that incorporates not only variables used in the present study but also potential auxiliary variables into the missing data handling procedure (Collins, Schafer, & Kam, 2001). In total, up to 373 main effects, 34,500 interaction terms, and 182 quadratic terms were considered in the imputation process. In addition, 34 scale cluster means and scale item cluster means were included to account for the hierarchical data structure (Graham, 2009). Based on the fully conditional specification, an imputation model for each incomplete variable was specified. The imputation process involved a number of steps. First, the imputation model was specified by selecting variables based on a minimal correlation criteria (r = .15 for main effects and r = .25 for interaction effects) resulting in up to 1,000 predictor variables. Next, to reduce the number of predictor variables, a partial least

square (Mevik & Wehrens, 2007) dimension reduction was conducted, resulting in up to 30 component scores. Last, partial least square component scores were used to predict missing values based on the predictive mean matching algorithm (van Buuren, 2012) using Tukey's tricube weighting function (Harrell, 2006) for metric and ordered-categorical variables and polytomous logistic regression for nominal variables. This imputation process was carried out for each incomplete variable in the data set completing one iteration. Graphical diagnostics using trace plots suggested that convergence was reached after 300 burn-in iterations. However, we took the conservative tack saving the first data set at the 501th iteration and saved additional data set every 50th iteration thereafter. At least 20 imputed data sets are recommended for most situation (Graham, Olchowski, & Gilreath, 2007). To follow a conservative tack, we generated 50 imputed data sets. The imputation process was conducted separately for the intervention and control group and combined afterward to preserve interactions between the grouping variable and other variables (Graham, 2009).