Promoting students' social behavior in primary education through Success for All lessons

Veldman, M. A.; Hingstman, M.; Doolaard, S.; Snijders, T. A. B.; Bosker, R. J.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Promoting students' social behavior in primary education through Success for All lessons

M.A. Veldman [a,*], M. Hingstman [a], S. Doolaard [a], T.A.B. Snijders [b], R.J. Bosker [a]

[a] *Department of Educational Sciences, University of Groningen, Faculty of Behavioral and Social Sciences, Groningen, the Netherlands*
[b] *Department of Sociology, University of Groningen, Faculty of Behavioral and Social Sciences, Groningen, the Netherlands*

ARTICLE INFO

ABSTRACT

Success for All (SfA) is a comprehensive school reform program with a strong emphasis on cooperative learning that aims to improve students' social emotional learning alongside students' cognitive learning. In the present study it was examined whether SfA led to improved students' social behavior in Grade 1–3 of primary education. Peer sociometric data was collected for 974 students aged 6–9. Using multivariate multilevel analysis we found no significant effect of SfA on students' pro- and antisocial behavior over time. However, a significant interaction effect was found showing that antisocial behavior of students from disadvantaged backgrounds decreased in the intervention condition in Grade 2. This is a promising finding given that the SfA program was especially developed for schools serving large numbers of disadvantaged students. Implications of the study are described.

1. Introduction

Social emotional learning (SEL) interventions aim to help children to develop their social emotional skills, such as managing emotions and maintaining positive relationships (Weissberg, Resnik, Payton, & O'Brien, 2003). SEL at school may be especially important for children from less advantaged backgrounds, as they are more often growing up under conditions of adversity (Nix et al., 2016; OECD, 2015). Generally, it has been found that SEL interventions are a promising approach to enhance success at school and in life of students from diverse family backgrounds (Taylor, Oberle, Durlak, & Weissberg, 2017). The aim of the current study is to investigate the effect of Success for All lessons, in which SEL strategies are embedded, on prosocial and antisocial behavior of 6–9 years old students in the early elementary grades and to explore whether there is a differential intervention effect for children from disadvantaged backgrounds.

1.1. Enhancing social emotional learning

SEL refers to the process through which students acquire and effectively apply the knowledge, attitudes and skills necessary to recognize and manage emotions, solve problems effectively, and establish positive relationships with others (Collaborative for Academic, Social, and Emotional Learning CASEL, 2005). In SEL interventions, social emotional skills are taught, modeled, practiced and applied to diverse situations so that students use them as part of their daily repertoire of behaviors. In turn, these achieved SEL competencies should provide a foundation for better school adjustment and academic performance as reflected in more positive social behavior, and also in improved grades, fewer conduct problems, and less emotional distress (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Taylor et al., 2017).

Several meta-analyses (Durlak et al., 2011; Goldberg et al., 2019; Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2016; Sklad, Diekstra, Ritter, Ben, & Gravesteijn, 2012; Taylor et al., 2017) showed that SEL programs can have significant and meaningful effects on social behavioral student outcomes. Positive social behavior of students refers to getting along with others (Durlak et al., 2011). SEL programs can have both promotional and preventive impact (Taylor et al., 2017). Desired social behavioral outcomes can be both an increase of prosocial behavior, i.e., behaviors intended to help other people and expressing concerns about them, and a reduction or prevention of antisocial behavior, that is disruptive behaviors, hurting someone on purpose or showing other externalizing problem behavior (Sklad et al., 2012). Important findings of the meta-analyses are that SEL interventions can

be incorporated into daily educational practices and that the programs can be effectively conducted by classroom teachers. However, it was also found that in practice SEL programs show a considerable variety in effectiveness.

According to Durlak et al. (2011), and emphasized by Sklad et al. (2012) and Taylor et al. (2017), SEL interventions should follow the acronym "SAFE" to be effective: using a connected and coordinated set of activities (*Sequenced*), using active forms of learning (*Active*), using dedicated time or a specific program element that focuses on developing SEL competencies (*Focused*), and identifying specific SEL competencies that are tried to develop within the intervention, i.e., targeting specific SEL skills (*Explicit*). For consistent delivery of programs, scripted curricula and manuals can be helpful (Catalano et al., 2003; Greenberg et al., 2003), in particular when they are based on the SAFE principles (Sklad et al., 2012).

Although SEL interventions have the potential to positively impact all students (Taylor et al., 2017), some results suggest stronger effects for students from disadvantaged backgrounds (e.g., Hawkins, Catalano, Kosterman, Abbott, & Hill, 1999; Raver et al., 2009). Students from disadvantaged backgrounds, which can be indicated by parental educational level, parental income, and parental occupations (Dietrichson, Bøg, Filges, & Klint Jørgensen, 2017), may be in a greater need of SEL at school as they appear to have fewer resources in many domains (Dietrichson et al., 2017; Nix et al., 2016; OECD, 2015). Hence, although universal SEL interventions should work to enhance social behavior of all children in the classroom, more insight in the degree to which these programs affect behavior of children who vary in background is needed.

1.2. Success for All

Success for All (SfA) is a comprehensive school reform program that aims to change schools, especially those serving large numbers of disadvantaged students and/or in high-poverty areas, with a relentless investment in several improvements to ensure success for all students (Slavin & Madden, 2013). SfA has been shown to be effective in increasing student achievement, especially reading performance, in the US and the UK (Borman et al., 2007; Quint, Zhu, Balu, Rappaport, & DeLaurentis, 2015; Tracey, Chambers, Slavin, Hanley, & Cheung, 2014), and is currently being adapted for Dutch educational practice.

Daily ninety-minute language and reading lessons with extensive use of cooperative learning are the core of the SfA program. Instruction in SfA lessons is based on scripted lesson plans; teachers receive materials and manuals containing descriptions of every single SfA lesson. The lessons cover a wide range of activities, for example partner reading or story summarization. Several cooperative learning techniques, such as Think-Pair-Share and Numbered Heads Together (Kagan, 1989), are involved. Teachers explicitly instruct students in the required cooperative behaviors, e.g., "active listening" and "helping each other". During the lessons, students are seated together in heterogeneous groups of approximately four students in order to provide all class members opportunities to learn together. To stimulate prosocial behavior and foster effective teamwork, social behavior reinforcement strategies are used. For example, when all students of a team managed to complete a task well, their success is celebrated by using a team cheer.

In addition to the daily ninety-minute lessons, struggling readers can receive tutoring in one-on-one or small group settings. Furthermore, SfA combines the reading and language curriculum with (a) Solutions Teams that address non-instructional issues that may affect learning, such as partnerships with parents, and (b) an emphasis on continuous improvement, in which the SfA program facilitator plays a key role by coaching and assisting teachers in their implementation of the SfA program (Quint et al., 2015; see for a more detailed description of the SfA program also Slavin, Madden, Chambers, & Haxby, 2009).

1.3. Promotion of social behavior by Success for All

SfA's first goal is to promote academic achievement, but it also aims to improve students' social emotional skills (e.g., Slavin et al., 2009). Several SEL strategies are integrated in SfA lessons. Altogether, SfA lessons are expected to adhere to the findings mentioned above from research on effective SEL programs. SfA lessons follow the SAFE acronym specified by Durlak et al. (2011). Throughout the school year, SfA lessons use a sequenced set of activities, building up from more simple to more complex activities (*Sequenced*). For example, the SfA lessons in the beginning of the school year focus on practicing active listening in pairs, and throughout the school year they build up towards more complex cooperative learning activities. A prerequisite within the classroom for effective (social emotional) learning is effective classroom management. To create a safe and positive classroom climate, clear rules, and stable routines need to be communicated, established, and maintained (Kunter, Baumert, & Köller, 2007; Korpershoek et al., 2016; Hutchings, Martin-Forbes, Daley, & Williams, 2013). Proactive classroom management strategies, such as the use of gestures, energizers, and randomizations when asking questions, are integrated in SfA lessons. Extensive use of cooperative learning techniques provides a lot of opportunities for active student learning (*Active*). A great strength of cooperative learning is that it provides teachers with many opportunities to instruct children in their social emotional skills when such instruction is immediately relevant (see also Battistich & Watson, 2003). Attention is also paid to the development of social skills in SfA lessons through the use of reinforcement strategies to promote social behavior (*Focused*). It has been found that modeling and rewarding prosocial behaviors can be effective for preventing problem behavior and stimulating prosocial behavior (Korpershoek et al., 2016). Specific SEL skills are targeted in SfA lessons through explicit instruction in cooperative behaviors (*Explicit*). Explicit instruction in cooperative behaviors is also often emphasized in the cooperative learning literature, as placing students in groups and expecting them to work together will not necessarily promote cooperation (Goodwin, 1999; Johnson & Johnson, 2009; Slavin, Hurley, & Chamberlain, 2003).

Several studies showed that teachers experienced improvement of students' behavior since SfA was introduced in their classrooms (Harris, Hopkins, Youngman, & Wordsworth, 2001; Hopkins, Youngman, Harris, & Wordsworth, 1999; Quint et al., 2015; Tracey et al., 2014). Teachers attributed students' behavioral changes mostly to the rapid pace, classroom management routines, and, in particular to the cooperative learning component of the program. Teachers noted that cooperative learning in SfA keeps students engaged and makes them responsible for each other. In several studies it was observed that students' engagement was high during SfA lessons (Datnow & Castellano, 2000; Ross & Smith, 1994). Furthermore, teachers reported that students became more tolerant and calmer in interactions with their peers, and that these developed social skills are transferable to contexts outside the SfA lessons (Harris et al., 2001). In the study of Skindrud and Gersten (2006), teachers felt SfA contributed to better social behavioral outcomes due to its emphasis on cooperative learning and social reinforcement strategies.

1.4. Present study

This study aims to find out whether there is empirical evidence for effects of the SfA program on social behavior of young students over the course of two school years, using peer sociometric measurements of students' behavior. Peers are privy to aspects of children's social worlds that teachers may not and they thus can provide unique information about children's social behavior (Rabiner, Godwin, & Dodge, 2016). In addition, in the peer sociometric method scores are based on multiple classmates instead of only one informant; the multiplicity of informants may be expected to give a balanced result.

The main research question to be answered in the current study is *(1) is there an effect of the SfA program on pro- and antisocial behavior of*

students in the early grades of primary education? It was hypothesized that involvement in the SfA program would lead to more prosocial behavior and less antisocial behavior of young school-aged children. A second explorative research question is *(2) is there a differential intervention effect for students from disadvantaged backgrounds (indicated by a low level of parental education)?* The SfA program is especially developed for schools serving large numbers of students from disadvantaged backgrounds. Children from less advantaged backgrounds may be in a greater need of explicitly taught SEL skills at school. Therefore, it was explored whether there was a differential intervention effect depending on students' backgrounds.

2. Method

2.1. Intervention

At the moment of the current study, the SfA program was implemented in six schools in the north of the Netherlands. Similar to the American SfA program, the programs consists of daily ninety-minute SfA lessons with a strong emphasis on cooperative learning. The developers of the Dutch program aim to adhere to the American program, but as the program was in an early development phase, there were some differences. The Dutch SfA program did not yet involve a stand-alone version of the *Getting Along Together* (GAT) program. GAT involves lessons in the first weeks of the school year and weekly activities aiming at the development of social emotional skills. However, assignments based on the GAT program were integrated in the Dutch SfA lessons. Another difference with the American program was that cross-grade ability grouping, i.e., placing students in different groups during SfA lessons based on their reading and language performance, was not yet possible at the moment of the current study. In the Dutch program students received SfA instruction in their homeroom classes. Nonetheless, the "SAFE" elements mentioned above, expected to promote students' social behavior, were all included in the Dutch SfA lessons.

At the moment of the current study, Dutch SfA lessons had only been developed for the lower grades (Grade 1–3). At the moment of data collection, SfA lessons were implemented for the second year in Grade 1 and for the first year in Grade 2 and Grade 3. Although the aim was to expand the Dutch SfA program every school year by one grade, three schools decided not to implement SfA in Grade 3 in the school year it was planned. Those schools decided to first refine the implementation of the program in the lower grades, i.e., Grade 1 and Grade 2, in which the SfA lessons were already implemented at that time. At the moment of the data collection of Grade 3, these schools form the follow up condition.

Teachers are supported in their implementation of SfA through the detailed and scripted manuals for each lesson. Furthermore, SfA provides professional training in how to teach SfA lessons and in the use of cooperative learning in these lessons. The focus of the initial training course, which took place right before the start of the school year, was on explaining teaching methods and materials. During the school year, program developers visited each classroom and had meetings per grade level with SfA teachers of all schools.

2.2. Design

The current study has a quasi-experimental design with an intervention and a control group in the first year of data collection, and an extended intervention condition, follow up condition, and control condition in the second year of data collection. Schools were assigned to the intervention group or the control group in accordance with agreements among the school boards, the municipality, and the Dutch SfA research and development team. In the first school year (2016–2017), data were collected in Grade 1 and Grade 2. In the second school year (2017–2018), data were collected in Grade 3. To investigate whether students' behavior improved, there were two waves of data collection in each school year: in October, in the last three weeks before the first

holiday in the school year, and in May, in the last three weeks before the last holiday in the school year. In the first year of the study, i.e., in Grade 1 and Grade 2, two conditions can be distinguished: SfA classes and control classes. In the second year of the study, i.e., in Grade 3, three conditions can be distinguished: children who were not involved in the intervention (control condition), children who were involved in the intervention condition for three years (extended intervention condition), and children who were involved in the intervention during two years in Grade 1 and Grade 2, but not in the third year in Grade 3 (follow up condition). Teachers in control schools were not given any instructions concerning teaching methods or SEL implementation.

2.3. Sample

All SfA schools and control schools were situated in one city, in the neighborhoods with the highest percentages of low-income households, ranging from 11% to 22% (van der Werff & Kloosterman, 2016). Schools in both conditions served relatively large numbers of students from disadvantaged backgrounds, i.e., parents or legal guardians completed less than two years of secondary school. In school year 2016–2017, percentages ranged from 8% to 20% for SfA schools and from 8% to 23% for control schools (Dienst Uitvoering Onderwijs, 2016). In school year 2017–2018, percentages ranged from 8% to 27% for SfA schools and from 9% to 20% for control schools (Dienst Uitvoering Onderwijs, 2017). Differences between SfA and control schools were not significant (2016-2017: $t(7) = .23$, $p = .83$; 2017-2018: $t(7) = -0.60$, $p = .57$). This suggests comparability of SfA and control schools.

In the first school year data were collected in Grade 1 and Grade 2 classes in six SfA schools and three control schools. In this school year, there were a total of 17 SfA classes (9 Grade 1 classes and 8 Grade 2 classes) and 11 control classes (5 Grade 1 classes and 6 Grade 2 classes). In the second school year, data were collected in two schools that represent the extended intervention condition, three schools that represent the follow up condition, and three control schools. There were three classes in the extended SfA condition and seven control classes. The follow up condition included five classes that used to work with the SfA program in Grade 2, but did not implement the SfA program any more in Grade 3.

The total sample of students with available sociometric scores at two measurement occasions consisted of 327 Grade 1 students (50% girls) and 333 Grade 2 students (49% girls) in the first school year of data collection. In the second school year, the sample consisted of 314 Grade 3 students (50% girls). Ages of the participants ranged from 6 to 9 years. At the second measurement in the first school year, Grade 1 students were on average 7.0 years old (SD = 5.5 months), Grade 2 students were on average 8.0 years old (SD = 5.9 months). At the second measurement in the second school year, Grade 3 students were on average 9.0 years old (SD = 5.8 months).

2.4. Procedure

Active informed consent for participation in this research was obtained from the parents or legal guardians. Procedures were approved by the ethics committee of the Department of Pedagogical and Educational Sciences of the University of Groningen. All children with parental consent to participate filled in a questionnaire assessing pro- and antisocial behavior. The questionnaire was administered in a one-on-one setting by trained test leaders: undergraduate students who were not familiar to the children. During the training, the procedure of administering the questionnaire and the role of the test leaders was discussed. While at the schools, the test leaders could always contact one of the researchers for questions and advice. Before the questionnaire administration, the homeroom teacher told his/her students they were going to fill in a questionnaire and introduced the test leader. The test leaders read the questionnaire out loud to the children. Test leaders took great care that children were at ease and understood the questions. The

Table 1
Number of students with sociometric scores and number of nominating students.

	Grade 1		Grade 2		Grade 3	
Measurement moment	T1	T2	T1	T2	T1	T2
Number of students with sociometric scores	322	316	326	325	307	309
Number of nominating students	281	291	289	298	282	282

children were instructed to provide their own responses to the questions and they were ensured that their answers would not be shared with their classmates. In all schools it was quite common to ask similar peer nomination questions approximately two times each school year to give insight in students' relationships in the classroom. Filling in the questionnaire took approximately 15 minutes per child. Preliminary multi-level analyses per measurement occasion showed no test leaders effect on the outcome variables.

2.5. Instruments and variables

2.5.1. Pro- and antisocial behavior

Peer measures of prosocial behavior were obtained using the questions: *'Who says and doesfriendly things?'* and *'Who helps other children?'*. Measures of antisocial behavior were obtained using the questions: *'Who says and does mean things?'* and *'Who quarrels a lot?'*. Per question, the children were given a roster with the names of their classmates and they could select as many, or as few, classmates as they wanted. Students could not name themselves or children outside of their class. Classmates appeared in random order in the roster. Before the questions about social behavior were asked, all classmates passed by one by one in other questions about working and playing together in the questionnaire.

The sociometric method used to assess students' social behavior was also used in several other studies (e.g., Bierman et al., 2010; Carlo, Koller, Eisenberg, Da Silva, & Frohlich, 1996; Deković & Janssens, 1992; Jansma, Opdenakker, & van der Werf, 2017; van Lier, Vuijk, & Crijnen, 2005). Peer nominations received were totaled and divided by the number of nominating classmates (for students who were a nominator themselves: number of nominating students minus 1), resulting in proportion scores per student. Note that nominating students are those who filled in the questionnaire, and students who could be nominated are all students who were in the classrooms at the time of the study. The total sample of this study consists of the students who could be nominated: those are the students with available sociometric scores. The numbers of nominating students and students with sociometric scores differed, because all students in the classroom could be nominated by their classmates while some students did not fill in the questionnaire

themselves due to missing parental consent. Percentages of nominating classmates, i.e., the students with parental consent to fill in the questionnaire, on which students' sociometric scores were based, ranged from 77% to 100% per class. In Table 1, the numbers of nominating and students with sociometric scores are shown per grade and per time point.

To reduce skewness of the scores, both for pro- and antisocial behavior the sums of the two items were transformed by taking the square root. To give an example of the final measurements, when a student was nominated by four classmates at the question 'who says and does nice/friendly things?' and by five classmates at the question 'who helps other children?' in a class with 23 nominating children (and the student was not a nominator him or herself), the student received a score of $\sqrt{((4/23) + (5/23))} = .63$ for prosocial behavior. Descriptives of the dependent variables are shown in Table 2.

2.5.2. Reliability

The Spearman-Brown formula was used for reliability analysis, as the scales of prosocial behavior and antisocial behavior were both based on two items (Eisinga, Grotenhuis, & Pelzer, 2013). Spearman-Brown coefficients for prosocial behavior (.84–.95) and antisocial behavior (.93–.94) were good for all four measurement points.

2.5.3. Missing data

After the first measurement, eleven students left their school and five students entered in Grade 1. Thus, there are missing scores for 16 Grade 1 students on one of the two measurements occasions. In Grade 2, there were 15 missing scores, because eight students left and seven students entered the schools. There are 12 students with missing scores on one of the two measurements in the Grade 3 sample, because five students left their school and seven students entered the schools after the third measurement.

Four Grade 1 students and two Grade 2 students switched schools or classes within our sample between measurement one and two. In the second year of data collection, one student switched to another class in our data between the two measurement moments. The data of these students were not taken into account in the data analyses.

2.5.4. Student background

Student background (non-disadvantaged background: 0, disadvantaged background: 1) was taken into account in the analyses as a covariate and it was investigated whether there was a differential intervention effect of student background. In the Netherlands, schools receive additional funding for students from disadvantaged backgrounds. This is indicated based on the educational level of the parents or legal guardians, meaning that of students from disadvantaged

Table 2
Descriptives of pro- and antisocial behavior for the different measurement occasions.

Prosocial behavior

	Measurement 1						Measurement 2					
	SfA				Control		SfA				Control	
	M(SD)	min-max			M(SD)	min-max	M(SD)	min-max			M(SD)	min-max
Grade 1	.48 (.21)	0-1.20			.49 (.23)	0-1.10	.67 (.33)	0-1.13			.63 (.22)	0-1.37
Grade 2	.55 (.21)	0-1.15			.64 (.27)	0-1.26	.52 (.24)	0-1.21			.70 (.30)	0-1.28
	SfA extended		SfA follow up		Control		SfA extended		SfA follow up		Control	
	M(SD)	min-max	M(SD)	min-max	M(SD)	min-max	M(SD)	min-max	M(SD)	min-max	M(SD)	min-max
Grade 3	.50 (.28)	0-1.06	.83 (.36)	0-1.39	.88 (.34)	0-1.45	.48 (.25)	0-1.00	.59 (.27)	0-1.21	.69 (.32)	0-1.29

Antisocial behavior

	Measurement 1						Measurement 2					
	SfA				Control		SfA				Control	
	M(SD)	min-max			M(SD)	min-max	M(SD)	min-max			M(SD)	min-max
Grade 1	.28 (.25)	0-1.26			.28 (.28)	0-1.00	.40 (.36)	0-1.35			.38 (.30)	0-1.29
Grade 2	.34 (.32)	0-1.35			.38 (.38)	0-1.37	.34 (.36)	0-1.40			.40 (.40)	0-1.40
	SfA extended		SfA follow up		Control		SfA extended		SfA follow up		Control	
	M(SD)	min-max	M(SD)	min-max	M(SD)	min-max	M(SD)	min-max	M(SD)	min-max	M(SD)	min-max
Grade 3	.39 (.35)	0-1.29	.40 (.38)	0-1.38	.48 (.42)	0-1.40	.32 (.33)	0-1.26	.33 (.37)	0-1.26	.43 (.44)	0-1.41

Table 3
Descriptives of covariates.

Grade 1			SfA	Control
			M(SD)	*M(SD)*
Age (months)			83.7 (5.9)	85.1 (4.7)
			N(%)	*N(%)*
Gender (girl)			214 (51.7)	108 (45.0)
Student background (disadvantaged)			66 (15.9)	40 (16.7)

Grade 2			SfA	Control
			M(SD)	*M(SD)*
Age (months)			95.3 (5.6)	98.1 (5.9)
			N(%)	*N(%)*
Gender (girl)			178 (45.4)	146 (53.3)
Student background (disadvantaged)			42 (10.7)	44 (16.1)

Grade 3	SfA extended	SfA follow up	Control
	M(SD)	*M(SD)*	*M(SD)*
Age (months)	107.6 (5.2)	107.1 (5.5)	110.1 (6.1)
	N(%)	*N(%)*	*N(%)*
Gender (girl)	56 (40.0)	118 (50.0)	140 (55.6)
Student background (disadvantaged)	22 (15.7)	16 (6.8)	38 (15.1)

backgrounds, parents or legal guardians completed less than two years of secondary school. This is similar to not entering ISCED level 3, which is typically entered by students aged between 14 and 16 years old (Eurydice, 2018).

2.5.5. Covariates

The age of the student in months at the second measurement was taken into account as covariate (centered around the grand mean in analyses). Furthermore, gender (boy: 0, girl: 1) was taken into account as covariate because research has shown that peers, especially girls, are more likely to nominate girls for prosocial behavior and nominate boys for antisocial behavior (Warden & Mackinnon, 2003; Warden, Cheyne, Christie, Fitzpatrick, & Reid, 2003). Descriptives of the students' age, gender and background per grade and condition are shown in Table 3.

2.6. Analyses

To take into account the nested structure of the data, we conducted multilevel analyses (Snijders & Bosker, 2012) using the MLwiN software (Rasbash et al., 2000). Analyses were conducted per grade. Although most of the students of Grade 2 and Grade 3 were the same, we decided to conduct analyses per grade because of changes in how they were distributed over the classrooms. Furthermore, in Grade 2 two conditions were distinguished, while in Grade 3 three conditions were distinguished. The effect of SfA was tested by the interaction term of SfA and time, to take into account students' behavioral scores at the beginning of the school year. In the analysis of Grade 1 and Grade 2, the SfA condition was compared to the control condition. In the analysis of Grade 3, the SfA extended condition and the follow up condition were compared to the control condition.

As the dependent variables pro- and antisocial behavior were correlated and observed for the same individuals, we used multivariate multilevel analyses[1] (cf. chapter 16, Snijders & Bosker, 2012). To model the two measurement occasions, we used two dummy variables in the

multivariate model, and therefore, the models do not contain a constant term (cf. chapter 15, Snijders & Bosker, 2012). Dummy 1 is 1 for the first measurement moment and 0 for the second measurement moment, and Dummy 2 is 0 for the first measurement moment and 1 for the second measurement moment. Hence, to represent the two measurement occasions and multivariate data, four levels were used. The first level is that of the dependent variables. The other three levels are the measurements (level 2) nested in students (level 3) nested in classrooms (level 4).

Ideally, a five-level model should be applied to the data, because of the nested structure of the data of classrooms within schools. However, the school level was not included in the multilevel models. The number of schools in the study was limited ($n = 9$) and, for this type of dependent variables, we expected most variance between students and classrooms and less variance between schools. Preliminary analyses confirmed this; the estimated variance at the school level was zero. This means that observed differences between schools are smaller than what could be expected by chance given the differences between classrooms (see p. 61 in Snijders & Bosker, 2012).

Models were fitted in the following steps. The first models include two dummies for the two measurement occasions with a random slope for classes and for students. Furthermore, the first models include a methodological covariate: the number of nominators per class (not centered in analyses). We take this covariate into account, because in a larger class there is a smaller chance for being nominated as there is a tendency to nominate only a few (e.g., three to four) classmates. Second, all covariates were added to the model as fixed explanatory variables. Third, condition and the interaction effect of time and condition were added to the models to test the effect of the intervention. The interaction term between SfA and time added in models 3 is the effect of interest to answer our first research question. Subsequently, to answer the second research question, the interaction effect between time, condition and student background was added to test whether there was a differential intervention effect for disadvantaged students (models 4). Assumptions were checked for the final multilevel models. We checked variance and normality of residuals: variance of residuals was evenly distributed and although the dependent variables were somewhat skewed, residuals were normally distributed.

3. Results

Results of the multivariate multilevel analyses for Grade 1, Grade 2, and Grade 3 are shown in respectively Tables 4–6. For the sake of clarity, we only show the variances for pro- and antisocial behavior at the two time measurement occasions at the levels of students and classes in the tables. Correlations between pro- and antisocial behavior at the two measurement occasions are shown in Table A1 in the Appendix A. With respect to covariates, we found significant results of age and gender in all grades. In all grades, girls showed more prosocial behavior and less antisocial behavior compared to boys ($p < .01$ for pro- and antisocial behavior in all grades). In Grade 1, older students showed more antisocial behavior ($p = .03$). In Grade 2, older students showed less prosocial behavior ($p = .02$) as well as in Grade 3 ($p = .01$). We found significant effects of student background in Grade 2: students from disadvantaged backgrounds showed less prosocial behavior and more antisocial behavior ($p < .01$). In Grade 3, students from disadvantaged backgrounds showed more antisocial behavior ($p < .01$).

3.1. Intervention effects

The effect of interest is the interaction effect of the intervention over time, by using time dummy 2, which is coded 1 for the second measurement occasion, contrasting this to the behavior scores at the beginning of the school year. Taking into account the scores at the beginning of the school year, in none of the grades a significant effect of the intervention between the two measurement occasions was found. In

---

[1] Because the number of classes was limited and to enable model convergence, we made some restrictions to the random part of the models at the classroom level. The following covariances to be estimated are forced to be equal (T1 refers to time point 1, T2 to time point 2, pro to the dependent variable prosocial behavior, and anti to the dependent variable antisocial behavior): antiT1proT1 = antiT2proT2, proT2antiT1 = antiT2proT1, proT1-proT1 = proT2proT2, antiT1antiT1 = antiT2antiT2.

Table 4
Multivariate multilevel models pro- and antisocial behavior Grade 1.

Grade 1	Model 1: time				Model 2: covariates				Model 3: condition x time				Model 4: condition x time x background			
	Prosocial		Antisocial		Prosocial		Antisocial		Prosocial		Antisocial		Prosocial		Antisocial	
Fixed Part	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Measurement 1	0.839*	0.116	0.655*	0.077	0.805*	0.114	0.742*	0.078	0.815*	0.130	0.732*	0.081	0.808*	0.131	0.733*	0.081
Measurement 2	1.015*	0.119	0.769*	0.080	0.982*	0.117	0.856*	0.081	0.929*	0.134	0.815*	0.085	0.933*	0.135	0.823*	0.086
Number of nominators in a class	−0.015*	0.005	−0.017*	0.003	−0.015*	0.005	−0.016*	0.003	−0.015*	0.005	−0.016*	0.003	−0.015*	0.005	−0.016*	0.003
Student background (disadvantaged)					−0.048	0.028	0.024	0.034	−0.048	0.028	0.024	0.034	−0.007	0.049	0.004	0.060
Gender (girl)					0.078*	0.020	−0.219	0.025	0.078*	0.020	−0.221*	0.025	0.077*	0.020	−0.218	0.025*
Age (centered around grand mean)					0.003	0.002	0.005*	0.002	0.003	0.002	0.005*	0.002	0.003	0.002	0.005	0.002*
Condition									−0.019	0.090	0.019	0.048	−0.009	0.091	0.012	0.050
Condition x time									0.100	0.140	0.049	0.072	0.090	0.140	0.033	0.073
Condition x background													−0.057	0.049	−0.042	0.059
Background x time													−0.053	0.062	0.043	0.075
Condition x time x background													0.052	0.062	0.095	0.075
Random Part	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE
Level Classrooms																
Measurement 1	0.024	0.007	0.004	0.002	0.023	0.007	0.005	0.002	0.024	0.007	0.005	0.002	0.024	0.007	0.005	0.002
Measurement 2	0.024	0.007	0.004	0.002	0.023	0.007	0.005	0.002	0.024	0.007	0.005	0.002	0.024	0.007	0.005	0.002
Level Students																
Measurement 1	0.037	0.003	0.072	0.006	0.037	0.003	0.057	0.005	0.037	0.003	0.057	0.005	0.037	0.003	0.057	0.005
Measurement 2	0.051	0.004	0.114	0.009	0.046	0.004	0.089	0.007	0.046	0.004	0.089	0.007	0.046	0.004	0.089	0.007
−2*loglikelihood:	−261.399				−344.835				−347.095				−351.984			

n = 637 measurements in 327 students 14 in classes in model 1 and *n* = 628 measurements in 320 students 14 in classes in due to missing data in models 2–4.

*$p < .05$.

Table 5
Multivariate multilevel models pro- and antisocial behavior Grade 2.

Grade 2	Model 1: time				Model 2: covariates				Model 3: condition x time				Model 4: condition x time x background			
	Prosocial		Antisocial		Prosocial		Antisocial		Prosocial		Antisocial		Prosocial		Antisocial	
Fixed Part	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Measurement 1	0.795*	0.136	0.463*	0.090	0.686*	0.146	0.682*	0.079	0.672*	0.137	0.679*	0.080	0.676*	0.137	0.681*	0.082
Measurement 2	0.806*	0.140	0.477*	0.093	0.696*	0.151	0.697*	0.082	0.727*	0.141	0.712*	0.083	0.721*	0.141	0.700*	0.085
Number of nominators in a class	−0.009	0.006	−0.005	0.004	−0.007	0.007	−0.009*	0.003	−0.004	0.006	−0.008*	0.003	−0.004	0.006	−0.008*	0.003
Student background (disadvantaged)					−0.106*	0.033	0.172*	0.052	−0.106*	0.033	0.170*	0.052	−0.140*	0.051	0.165*	0.076
Gender (girl)					0.161*	0.022	−0.300*	0.035	0.160*	0.022	−0.301*	0.035	0.160*	0.022	−0.300*	0.035
Age (centered around grand mean)					−0.004*	0.002	0.005	0.003	−0.005*	0.002	0.005	0.003	−0.005*	0.002	0.005	0.003
Condition									−0.079	0.079	−0.026	0.042	−0.079	0.079	−0.024	0.045
Condition x time									−0.080	0.091	−0.033	0.037	−0.087	0.090	−0.013	0.039
Condition x background													0.053	0.044	0.088	0.050
Background x time													−0.022	0.071	−0.023	0.106
Condition x time x background													0.099	0.062	−0.141*	0.070
Random Part	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE
Level Classrooms																
Measurement 1	0.019	0.006	0.003	0.003	0.021	0.007	0.001	0.002	0.019	0.006	0.001	0.002	0.018	0.006	0.001	0.002
Measurement 2	0.019	0.006	0.003	0.003	0.021	0.007	0.001	0.002	0.019	0.006	0.001	0.002	0.018	0.006	0.001	0.002
Level Students																
Measurement 1	0.052	0.004	0.123	0.010	0.044	0.004	0.101	0.008	0.044	0.004	0.100	0.008	0.044	0.004	0.101	0.008
Measurement 2	0.056	0.004	0.147	0.012	0.050	0.004	0.121	0.010	0.050	0.004	0.121	0.010	0.049	0.004	0.121	0.010
−2*loglikelihood:	−169.566				−269.077				−273.586				−291.822			

$n = 651$ measurements in 333 students 14 in classes in model 1 and $n = 636$ measurements in 325 students 14 in classes in due to missing data in models 2-4.

* $p < .05$.

Table 6
Multivariate multilevel models pro- and antisocial behavior Grade 3.

Grade 3	Model 1: time				Model 2: covariates				Model 3: condition x time				Model 4: condition x time x background			
	Prosocial		Antisocial		Prosocial		Antisocial		Prosocial		Antisocial		Prosocial		Antisocial	
Fixed Part	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Measurement 1	1.322*	0.191	0.873*	0.121	1.221*	0.192	1.039*	0.119	1.233*	0.183	1.055*	0.108	1.227*	0.184	1.026*	0.110
Measurement 2	1.120*	0.189	0.803*	0.121	1.019*	0.190	0.968*	0.118	0.989*	0.182	0.976*	0.107	0.991*	0.183	0.949*	0.109
Number of nominators in a class	−0.025*	0.010	−0.021*	0.006	−0.023*	0.010	−0.024*	0.006	−0.019	0.010	−0.023*	0.005	−0.019	0.010	−0.023*	0.005
Student background (disadvantaged)					−0.094*	0.038	0.189*	0.060	−0.091*	0.038	0.185*	0.060	−0.061	0.056	0.278*	0.088
Gender (girl)					0.156*	0.025	−0.275*	0.040	0.155*	0.025	−0.273*	0.040	0.156*	0.026	−0.270*	0.040
Age (centered around grand mean)					−0.006*	0.002	0.004	0.003	−0.006*	0.002	0.004	0.003	−0.006*	0.002	0.004	0.003
Condition (SfA follow up)									−0.023	0.140	0.003	0.084	−0.017	0.141	0.026	0.084
Condition (SfA extended)									−0.384*	0.157	−0.097	0.094	−0.370*	0.157	−0.083	0.095
Condition (SfA follow up) x time									−0.019	0.130	0.016	0.044	−0.031	0.130	0.019	0.044
Condition (SfA extended) x time									0.250	0.153	0.012	0.051	0.235	0.153	0.026	0.052
background x time													−0.047	0.044	−0.004	0.049
Condition (SfA follow up) x background													−0.027	0.099	−0.237	0.155
Condition (SfA extended) x background													−0.084	0.093	−0.091	0.145
Condition (SfA follow up) x time x background													0.107	0.081	−0.063	0.093
Condition (SfA extended) x time x background													0.087	0.073	−0.081	0.082
Random Part	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE	Variance	SE
Level Classrooms																
Measurement 1	0.053	0.016	0.013	0.007	0.053	0.016	0.011	0.006	0.046	0.014	0.012	0.006	0.046	0.014	0.011	0.006
Measurement 2	0.053	0.016	0.013	0.007	0.053	0.016	0.011	0.006	0.046	0.014	0.012	0.006	0.046	0.014	0.011	0.006
Level Students																
Measurement 1	0.055	0.005	0.140	0.012	0.049	0.004	0.122	0.010	0.049	0.004	0.122	0.010	0.049	0.004	0.122	0.010
Measurement 2	0.067	0.006	0.154	0.013	0.059	0.005	0.135	0.011	0.059	0.005	0.135	0.011	0.059	0.005	0.134	0.011
−2*loglikelihood:	−136.422				−203.571				−224.793				−234.398			

n = 609 measurements in 307 students 15 in classes in model 1 and n = 607 measurements in 306 students 15 in classes in due to missing data in models 2–4.
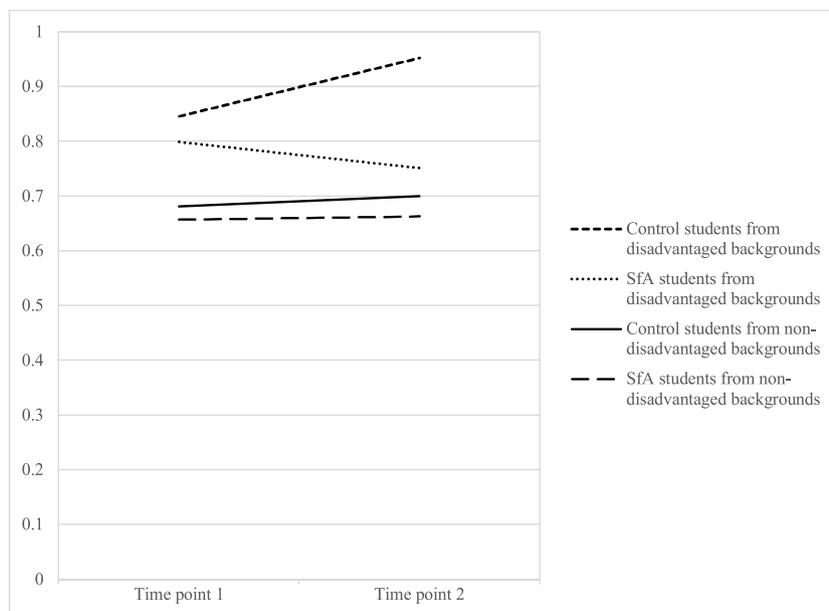
* $p < .05$.

Fig. 1. Interaction between time, condition and student background in Grade 2 for antisocial behavior.

Grade 3, the students in the extended intervention condition showed significantly less prosocial behavior than the control group on the first measurement occasion ($p = .01$). However, we did not find an effect of the (extended) intervention over time in this or other grades.

3.2. Differential intervention effects for students from disadvantaged backgrounds

Including an interaction effect between condition, time and student background led to a significant model improvement for Grade 2 ($\chi^2 = 18.24$, $df = 6$, $p < .01$). As shown in Fig. 1, antisocial behavior of students from disadvantaged backgrounds in the SfA condition decreased, while antisocial behavior of students from disadvantaged backgrounds in the control condition increased. Antisocial behavior of students from non-disadvantaged backgrounds slightly increased in the SfA condition as well as in the control condition (interaction effect: $p = .04$). The interaction effect was not significant for prosocial behavior ($p = .11$) in Grade 2. We also did not find significant interaction effects between time, condition and student background in Grade 1 and Grade 3.

4. Discussion

The aim of this study was to investigate the effect of the SfA program on pro- and antisocial behavior of 6–9 years old students in Grade 1, Grade 2 and Grade 3 of primary education. To investigate whether students' social behavior improved, peer sociometric measurements of students' behavior were collected twice a year over the course of two school years. In the first year of the study, we collected data of Grade 1 and Grade 2 students. At that moment, students of the intervention group had followed SfA lessons on a daily basis for one or two whole school year(s). In the second year of the study, we studied the behavior of Grade 3 students in an extended intervention condition, follow up condition, and control condition. Taking into account students' scores at the beginning of the school year, we found no significant effects of the SfA program on students' social behavior. Thus, based on the findings of the current study, our hypothesis cannot be confirmed: it cannot be concluded that the SfA program leads to increased prosocial behavior or decreased antisocial behavior of students in the early grades of primary education.

Furthermore, we explored whether there was a differential intervention effect for students from disadvantaged backgrounds. In Grade 2,

antisocial behavior of students from disadvantaged backgrounds in the SfA condition decreased, while antisocial behavior of students from disadvantaged backgrounds in the control condition increased. Antisocial behavior of students from non-disadvantaged backgrounds slightly increased in the SfA condition as well as in the control condition. This shows a positive effect of the intervention for the group of students from disadvantaged backgrounds: a promising finding given that the SfA program was especially developed for schools serving many disadvantaged students. SEL at school may be especially important for these children, because they are more often growing up under conditions of adversity (Nix et al., 2016; OECD, 2015). However, we should interpret the current finding with caution, as we only found this in Grade 2 and not in Grade 1 or Grade 3. Moreover, although researchers agree that parental education, income, and occupations are the main indicators of students' socio-economic background (Dietrichson et al., 2017), we only had access to information about parents' educational level. Based on this information we could identify students from disadvantaged backgrounds, indicated by whose parents had only completed less than two years of secondary school. In future research, preferably a more thorough measure of student background is used.

A significant initial difference between the extended intervention group and the control group in Grade 3 was found: on the first measurement occasion the students of the extended intervention condition scored lower on prosocial behavior than the students of the control group. This in spite of the fact that the Grade 3 students in the extended condition had been following the SfA program for already two years at that moment. The quasi-experimental design of this study with two repeated measures implies that effects of condition are tested by interactions of condition by time; these tests were presented above. The main effect of condition on the first measurement moment is necessary in the model for the purpose of controlling for initial differences. Its significance is not of particular interest for the hypothesis tests.

The quasi-experimental design is a limitation of the current study. However, SfA schools and control schools were situated in the same neighborhoods and no significant differences were found with regard to the percentage of students from disadvantaged backgrounds, suggesting comparability of SfA and control schools. Another limitation that should be noted is that the study only included nine schools, and consequently, the number of classrooms was limited (Grade 1 $n = 17$, Grade 2 $n = 11$, Grade 3 $n = 15$). Preferably, more schools and classrooms should be included for more precise predictions, although it is known to be

difficult to conduct large-scale intervention studies in practice.

Another limitation is that measurements of pro- and antisocial behavior were based on multiple informant measures from classmates using only two items. Classmates experience each other's behavior from nearby, and the use of the average of all classmates is likely to give a balanced and reliable impression, but the class might also be collectively misguided. We suppose this will be the case only in a small minority of cases, if at all. This way of measuring students' pro- and antisocial behavior by using peer sociometric scores is quite often used (e.g., Carlo et al., 1996; Deković & Janssens, 1992), but using it as an outcome measure to study the effect of an intervention is not very common. There are many intervention studies that used peer nominations for measuring students' behavior (Bierman & Furman, 1984; Garandeau, Lee, & Salmivalli, 2014; Kärnä et al., 2011, 2013; Salmivalli, Kaukiainen, & Voeten, 2005), however, most studies focus on bullying and victimization; which is more extreme behavior than the students' behavior we studied. To our knowledge, only in a few other studies intervention effects on pro- and/or antisocial behavior were investigated the same way. Although only positive effects on antisocial behavior were found, these studies show that effects can be found using these measurements methods (Bierman et al., 2010; Jansma et al., 2017; van Lier et al., 2005).

A disadvantage of using peer reports to measure students' social behavior is that such a measure requires substantial changes in behavior to reveal intervention effects. Peer reports are known to be notably resistant to change, because of stability of reputations (Juvonen, Nishina, & Graham, 2001; Salmivalli et al., 2005). A change in students' behavior does not immediately lead to a change in reputation. The sociometric scores are based on perceptions of peers, which might be biased. There might also be differences between children in their perceptions about behavior of others, for instance about what it entails to do something friendly for another. Moreover, children tend to nominate a few classmates who stand out relative to classroom norms, even though the questions allowed for unlimited nominations of classmates (children could nominate as few or as many children as they want). Therefore, this method also reflects social comparison processes. This makes these sociometric measures less sensitive to universal intervention effects that are designed to change classroom behavioral norms (Bierman et al., 2010). Previous researchers also have suggested the possibility that an intervention caused more reporting of antisocial behavior. This so-called sensitization effect is the phenomenon that students become better in recognizing antisocial behavior, because of raised awareness caused by the intervention (Frey et al., 2005; Kärnä et al., 2011; Smith, Ananiadou, & Cowie, 2003). A sensitization effect could have contributed to the null findings in the present study.

It may be that the SfA program led to behavioral changes that are visible and observable in the classroom, but not measurable in the way we did it. In earlier studies (Harris et al., 2001; Hopkins et al., 1999; Quint et al., 2015), teachers reported that students' behavior improved markedly since SfA was introduced. Classroom observations showed that students' engagement was high during SfA lessons (Datnow & Castellano, 2000; Ross & Smith, 1994). Teachers reported that students' peer interactions improved and that developed social skills were transferable to contexts outside SfA lessons (Harris et al., 2001). Hence, although no significant main effect of the SfA program was found in the present study, it would be premature to conclude that SfA is ineffective in improving students' social behavior. The SfA program is a highly complex intervention and the results might be explained by the early phase of the development and implementation of the Dutch SfA program. Due to the early development phase of the Dutch program, cross-grade ability grouping and the GAT program targeting students' social emotional skills were not yet implemented. However, SEL practices following the "SAFE" principles (Durlak et al., 2011) were embedded in the Dutch SfA lessons. In the present study no implementation measure was included in the analyses, hence, a lack of effects could be due to implementation problems rather than program

Table A1
Correlations between dependent variables at two time points (models 4).

	Grade 1	Grade 2	Grade 3
Classroom level			
Corr(Antisocial measurement 1/Prosocial measurement 1)	0.53	1.12	1.05
Corr(Prosocial measurement 2/Prosocial measurement 1)	−0.24	0.29	0.48
Corr(Prosocial measurement 2/Antisocial measurement 1)	−0.56	0.17	0.70
Corr(Antisocial measurement 2/Prosocial measurement 1)	−0.56	0.17	0.70
Corr(Antisocial measurement 2/Antisocial measurement 1)	−0.45	−0.18	0.84
Corr(Antisocial measurement 2/Prosocial measurement 2)	0.53	1.12	1.05
Student level			
Corr(Antisocial measurement 1/Prosocial measurement 1)	−0.03	−0.38	−0.55
Corr(Prosocial measurement 2/Prosocial measurement 1)	0.55	0.66	0.73
Corr(Prosocial measurement 2/Antisocial measurement 1)	−0.33	−0.41	−0.47
Corr(Antisocial measurement 2/Prosocial measurement 1)	−0.09	−0.42	−0.51
Corr(Antisocial measurement 2/Antisocial measurement 1)	0.64	0.81	0.84
Corr(Antisocial measurement 2/Prosocial measurement 2)	−0.39	−0.48	−0.48

Correlations of the dependent variables pro- and antisocial behavior at the two measurement moments of the fourth models of the multilevel analyses are shown above. Some of the correlation coeffiencts are greater than 1.0. This is similar to the possibility to find negative variances in the multilevel models by using MLwiN (Goldstein, 2011). MLwiN provides approximations for the observed variances and correlations of the observations as sums of covariance matrices for the various levels, without requiring that these matrices are positive definite. The correlations greater than 1 can be interpreted as the result of random variability, and point to the large uncertainty associated with these estimates. Indeed, the correlations at the classroom level do not show any clear pattern. At the student level, we did find a clear pattern of correlations. All correlations between pro- and antisocial behavior are negative. Positive correlations were found between prosocial behavior at the two different measurement occasions as well as for antisocial behavior. Stronger correlations were found in the higher grades, which suggests that students' behavior becomes more stable after Grade 1.

limitations. Multi-year studies of SfA have found better outcomes with each successive year of program implementation (Chamberlain, Daniels, Madden, & Slavin, 2007). Thus, it is expected that the Dutch SfA program and the implementation of the program will improve in future years. The decision of the three schools to not expand the implementation to the third grade in the school year it was planned reflects the demandingness of the implementation of the program.

Although in the current study we did not find positive effects of SfA on the social behavioral measures, with the exception of the decrease in antisocial behavior of students from disadvantaged backgrounds in Grade 2, other research has shown that SEL practices can have several positive outcomes. Therefore, for practice we would recommend to give more priority to SEL in daily educational practice. It could be that SEL becomes snowed under by all other challenges that schools are facing. Embedding SEL practices in a comprehensive school reform program, such as SfA, has the potential to produce multiple benefits. However, implementation of comprehensive school reform programs can be demanding. Here, we note that prescriptiveness of curricula can be both a strength and a limitation. Script-based curricula are emphasized for consistent delivery of the program. At the same time, the programs require extensive training and strong support for high quality implementation and consequently may lead to reduced teacher autonomy. Future studies that examine outcomes of the SfA program should

preferably include multiple measures of implementation and effects to provide more insight in whether and how SfA improves students' social behavior.

Acknowledgments

Appendix A. Correlations between pro- and antisocial behavior

References

Battistich, V., & Watson, M. (2003). Fostering social development in preschool and the early elementary grades through co-operative classroom activities. In R. Gillies, & A. Ashman (Eds.), *Co-operative learning: The social and intellectual outcomes of learning in groups* (pp. 19–35). Routledge Falmer.

Bierman, K. L., & Furman, W. (1984). The effects of social skills training and peer involvement on the social adjustment of preadolescents. *Child Development, 55*(1), 151–162. https://doi.org/10.2307/1129841

Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., McMahon, R. J., et al. (2010). The effects of a multiyear universal social-emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology, 78*(2), 156–168.

Borman, G. D., Slavin, R. E., Cheung, A. C., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal, 44*(3), 701–731. https://doi.org/10.3102/0002831207306743

Carlo, G., Koller, S. H., Eisenberg, N., Da Silva, M. S., & Frohlich, C. B. (1996). A cross-national study on the relations among prosocial moral reasoning, gender role orientations, and prosocial behaviors. *Developmental Psychology, 32*(2), 231–240. https://doi.org/10.1037/0012-1649.32.2.231

Catalano, R. F., Mazza, J. J., Harachi, T. W., Abbott, R. D., Haggerty, K. P., & Fleming, C. B. (2003). Raising healthy children through enhancing social development in an elementary school: Results after 1.5 years. *Journal of School Psychology, 41*(2), 143–164. https://doi.org/10.1016/s0022-4405(03)00031-1

Chamberlain, A., Daniels, C., Madden, N. A., & Slavin, R. E. (2007). A randomized evaluation of the Success for All middle school reading program. *Middle Grades Research Journal, 2*(1), 1–22.

Collaborative for Academic, Social, and Emotional Learning [CASEL]. (2005). *Safe and sound: An educational leader's guide to evidence-based social and emotional learning programs* (Illinois edition).

Datnow, A., & Castellano, M. (2000). *Inside look center for research on the education of students placed at risk*. Johns Hopkins University.

Deković, M., & Janssens, J. M. (1992). Parents' child-rearing style and child's sociometric status. *Developmental Psychology, 28*(5), 925–932. https://doi.org/10.1037/0012-1649.28.5.925

Dienst Uitvoering Onderwijs. (2016). *Leerlingen bo naar gewicht en leeftijd 2016–2017. [Primary school students by social advantage and age 2016–2017]*. https://duo.nl/open_onderwijsdata/databestanden/po/leerlingen-po/bo-sbo/po-owsoort-leeftijd.jsp.

Dienst Uitvoering Onderwijs. (2017). *Leerlingen bo naar gewicht en leeftijd 2017–2018. [Primary school students by social advantage and age 2017–2018]*. https://duo.nl/open_onderwijsdata/databestanden/po/leerlingen-po/bo-sbo/po-owsoort-leeftijd.jsp.

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A. M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research, 87*(2), 243–282. https://doi.org/10.3102/0034654316687036

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Eisinga, R., Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health, 58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Eurydice. (2018). *The structure of the European education systems 2018/19: Schematic diagrams. Eurydice facts and figures*. Publications Office of the European Union.

Frey, K. S., Hirschstein, M. K., Snell, J. L., Edstrom, L. V. S., MacKenzie, E. P., & Broderick, C. J. (2005). Reducing playground bullying and supporting beliefs: An experimental trial of the steps to respect program. *Developmental Psychology, 41*(3), 479–491. https://doi.org/10.1037/0012-1649.41.3.479

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*(1), 44–50. https://doi.org/10.1016/j.appdev.2013.10.004

Goldberg, J. M., Sklad, M., Elfrink, T. R., Schreurs, K. M., Bohlmeijer, E. T., & Clarke, A. M. (2019). Effectiveness of interventions adopting a whole school approach to enhancing social and emotional development: A meta-analysis. *European Journal of Psychology of Education, 34*(4), 755–782. https://doi.org/10.1007/s10212-018-0406-9

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Chichester: Wiley. https://doi.org/10.1002/9780470973394

Goodwin, M. W. (1999). Cooperative learning and social skills: What skills to teach and how to teach them. *Intervention in School and Clinic, 35*(1), 29–33. https://doi.org/10.1177/105345129903500105

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., et al. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *The American Psychologist, 58*(6-7), 466–474. https://doi.org/10.1037/0003-066x.58.6-7.466

Harris, A., Hopkins, D., Youngman, M., & Wordsworth, J. (2001). The implementation and impact of Success for All in English schools. In R. E. Slavin, & N. A. Madden (Eds.), *Success for All: Research and reform in elementary education* (pp. 81–92). Erlbaum.

Hawkins, J. D., Catalano, R. F., Kosterman, R., Abbott, R., & Hill, K. G. (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Archives of Pediatrics & Adolescent Medicine, 153*(3), 226–234. https://doi.org/10.1001/archpedi.153.3.226

Hopkins, D., Youngman, M., Harris, A., & Wordsworth, J. (1999). Evaluation of the initial effects and implementation of Success for All in England. *Journal of Research in Reading, 22*(3), 257–270. https://doi.org/10.1111/1467-9817.00089

Hutchings, J., Martin-Forbes, P., Daley, D., & Williams, M. E. (2013). A randomized controlled trial of the impact of a teacher classroom management program on the classroom behavior of children with and without behavior problems. *Journal of School Psychology, 51*(5), 571–585. https://doi.org/10.1016/j.jsp.2013.08.001

Jansma, D. J., Opdenakker, M. C. J. L., & van der Werf, M. P. C. (2017). *Promoting agreeableness in middle childhood: Effects of a class-based intervention* (Doctoral Dissertation). University of Groningen.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365–379. https://doi.org/10.3102/0013189x09339057

Juvonen, J., Nishina, A., & Graham, S. (2001). Self views versus peer perceptions of victims status among early adolescents. In J. Juvonen, & S. Graham (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 105–124). Guilford.

Kagan, S. (1989). The structural approach to cooperative learning. *Educational Leadership, 47*(4), 12–15.

Kärnä, A., Voeten, M., Little, T. D., Alanen, E., Poskiparta, E., & Salmivalli, C. (2013). Effectiveness of the KiVa Antibullying Program: Grades 1–3 and 7–9. *Journal of Educational Psychology, 105*(2), 535–551. https://doi.org/10.1037/a0030417

Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa antibullying program: Grades 4–6. *Child Development, 82*(1), 311–330. https://doi.org/10.1111/j.1467-8624.2010.01557.x

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research, 86*(3), 643–680. https://doi.org/10.3102/0034654315626799

Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction, 17*(5), 494–509. https://doi.org/10.1016/j.learninstruc.2007.09.002

Nix, R. L., Bierman, K. L., Heinrichs, B. S., Gest, S. D., Welsh, J. A., & Domitrovich, C. E. (2016). The randomized controlled trial of Head Start REDI: Sustained effects on developmental trajectories of social–emotional functioning. *Journal of Consulting and Clinical Psychology, 84*(4), 310–322. https://doi.org/10.1016/j.learninstruc.2007.09.002

OECD. (2015). *Skills for social progress: The power of social and emotional skills*. OECD. https://doi.org/10.1787/9789264226159

Quint, J., Zhu, P., Balu, R., Rappaport, S., & DeLaurentis, M. (2015). *Scaling up the success for all model of school reform: Final report from the investing in innovation (i3) evaluation*. MDRC. https://doi.org/10.2139/ssrn.2477948

Rabiner, D. L., Godwin, J., & Dodge, K. A. (2016). Predicting academic achievement and attainment: The contribution of early academic skills, attention difficulties, and social competence. *School Psychology Review, 45*(2), 250–267. https://doi.org/10.17105/spr45-2.250-267

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., & Healy, M. (2000). *A user's guide to MLwiN*. Centre for Multilevel Modeling.

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M. W., & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology, 77*(2), 302–316.

Ross, S. M., & Smith, L. J. (1994). Effects of the Success for All model on kindergarten through second-grade reading achievement, teachers' adjustment, and classroom-school climate at an inner-city school. *The Elementary School Journal, 95*(2), 121–138. https://doi.org/10.1086/461794

Salmivalli, C., Kaukiainen, A., & Voeten, M. (2005). Anti-bullying intervention: Implementation and outcome. *The British Journal of Educational Psychology, 75*(3), 465–487. https://doi.org/10.1348/000709905x26011

Skindrud, K., & Gersten, R. (2006). An evaluation of two contrasting approaches for improving reading achievement in a large urban district. *The Elementary School Journal, 106*(5), 389–408. https://doi.org/10.1086/505437

Sklad, M., Diekstra, R., Ritter, M. D., Ben, J., & Gravesteijn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools, 49*(9), 892–909. https://doi.org/10.1002/pits.21641

Slavin, R. E., & Madden, N. A. (2013). Success for All at 27: New developments in whole-school reform. *Journal of Education for Students Placed at Risk, 18*(3-4), 169–176. https://doi.org/10.1080/10824669.2013.862095

Slavin, R. E., Madden, N. A., Chambers, B., & Haxby, B. (2009). *2 million children: Success for All.* Corwin Press.

Slavin, R., Hurley, E. A., & Chamberlain, A. (2003). Cooperative learning and achievement: Theory and research. In W. M. Reynolds, & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (Vol. 7, pp. 177–198). Wiley.

Smith, P. K., Ananiadou, K., & Cowie, H. (2003). Interventions to reduce school bullying. *The Canadian Journal of Psychiatry, 48*(9), 295–303. https://doi.org/10.1177/070674370304800905

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development, 88*(4), 1156–1171. https://doi.org/10.1111/cdev.12864

Tracey, L., Chambers, B., Slavin, R. E., Hanley, P., & Cheung, A. (2014). Success for All in England: Results from the third year of a national evaluation. *SAGE Open, 4*(3), 1–10. https://doi.org/10.1177/2158244014547031

van der Werff, E., & Kloosterman, K. (2016). *Minima in Groningen. [Low-income households in Groningen].* Onderzoek en Statistiek Groningen.

van Lier, P. A., Vuijk, P., & Crijnen, A. A. (2005). Understanding mechanisms of change in the development of antisocial behavior: The impact of a universal intervention. *Journal of Abnormal Child Psychology, 33*(5), 521–535. https://doi.org/10.1007/s10802-005-6735-7

Warden, D., & Mackinnon, S. (2003). Prosocial children, bullies and victims: An investigation of their sociometric status, empathy and social problem-solving strategies. *The British Journal of Developmental Psychology, 21*(3), 367–385. https://doi.org/10.1348/026151003322277757

Warden, D., Cheyne, B., Christie, D., Fitzpatrick, H., & Reid, K. (2003). Assessing children's perceptions of prosocial and antisocial peer behaviour. *Educational Psychology, 23*(5), 547–567. https://doi.org/10.1080/0144341032000123796

Weissberg, R. P., Resnik, H., Payton, J., & O'Brien, M. U. (2003). Evaluating social and emotional learning programs. *Educational Leadership, 60*(6), 46–50.