Young children working together

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

Published in:
Learning and Instruction

DOI:
10.1016/j.learninstruc.2020.101308

Publication date:
2020

Learning and Instruction 67 (2020) 101308

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Young children working together. Cooperative learning effects on group work of children in Grade 1 of primary education

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

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

ARTICLE INFO

Keywords:
Cooperative learning
Group work behaviour
Primary education
Observations
Multilevel analysis

ABSTRACT

It was examined whether cooperative learning within the Success for All (SfA) program led to improved group work behaviour of Grade 1 pupils. 168 pupils of six SfA schools and 144 pupils of four control schools participated. Positive and negative group work behaviour was observed during a group task, taking into account socioemotional ethos, group participation, and type of dialogue. Longitudinal multilevel analysis was used for the sequence of observed 20-s time intervals. SfA groups showed more positive and less negative group work behaviour compared to control groups, whilst controlling for several group characteristics. Results suggest that negative group work behaviour increased gradually during the whole task in control groups, while in SfA groups it increased only towards the end of the task. The findings indicate that cooperative learning may lead to improved group work behaviour of young pupils (6–7 years old).

1. Introduction

In recent decades a considerable amount of research has been conducted on the effects of cooperative learning on student achievement (e.g., Sharan, Ackerman, & Hertz-Lazarowitz, 1979; Slavin, 1983; Slavin, Hurley, & Chamberlain, 2003), including a number of meta-analyses (Kyndt et al., 2013; Roseth, Johnson, & Johnson, 2008). These studies provide convincing evidence that cooperative learning can lead to improvement of cognitive student achievement, amounting up to medium sized effects (Kyndt et al., 2013; Roseth et al., 2008).

However, cooperative learning is not only a means of improving cognitive student achievement. It also has shown considerable potential to facilitate social-emotional outcomes (Slavin et al., 2003). Moreover, learning to cooperate is in itself an important educational goal. Collaborative skills and teamwork have an important place in all sets of 21st-century skills, as they are expected to become even more important in the future because of social and economic developments (Ananiadou & Claro, 2009; Binkley et al., 2012). In order to become successful inside and outside school children need to learn how to cooperate with others, which makes skills for working together an educational outcome in its own right (Barron, 2003; Blatchford, Baines, Rubie-Davies, Bassett, & Chowne, 2006).

The aim of the present study is to investigate whether cooperative learning integrated in day-to-day classroom practices impacts group work behaviour of young-aged primary school pupils in a setting outside the classroom. The study contributes in particular to the existing literature because of its focus on a) pupils' group work behaviour rather than cognitive achievement, and b) the young age group (6–7 years).

1.1. Pupils' group work behaviour

It is well recognized that placing pupils in groups and telling them to work together will not necessarily promote cooperation. In fact, groups often struggle to make it work (Gillies, 2014). In the context of Grade 1 pupils performing a group task, skilled group work involves: 1) dialogue between pupils that is of high level and on task, 2) participation of all group members, and 3) a positive socioemotional ethos within the group.

High-level and on-task dialogue. In skilled group work, group members engage in peer dialogue that is about the topic of the task or about planning and monitoring the group process to conduct the task. Talk that is of high level goes beyond simply sharing information, it requires that pupils make a serious effort to explain their ideas to others and ask their group members to make the same effort. This would involve talk that aims to make reasoning explicit to others, e.g., giving arguments, explanations, and justifications (Baines, Rubie-Davies, & Blatchford, 2009; Blatchford et al., 2006). Mercer and others (e.g., Mercer, 1996; Mercer, Wegerif, & Dawes, 1999) called this high-level talk in peer

* Corresponding author. University of Groningen, Faculty of Behavioural and Social Sciences, Department of Educational Sciences, Grote Rozenstraat 3, 9712 TG, Groningen, the Netherlands.
E-mail addresses: m.a.veldman@rug.nl (M.A. Veldman), s.doolaard@rug.nl (S. Doolaard), r.j.bosker@rug.nl (R.J. Bosker), t.a.b.snijders@rug.nl (T.A.B. Snijders).

https://doi.org/10.1016/j.learninstruc.2020.101308
Received 24 July 2019; Received in revised form 10 February 2020; Accepted 15 February 2020

interaction exploratory talk. In exploratory talk, partners engage critically but constructively with each other's ideas. Knowledge is made publicly accountable and reasoning is visible in the talk (Mercer et al., 1999).

Participation of all group members. The participation of all group members is needed when pupils are working together. Every group member should contribute to the group process, which means that problems such as free riding and social loafing have to be overcome (Karau & Williams, 1993; Latané, Williams, & Harkins, 1979; Slavin et al., 2003). Joint attention of all group members needs to be recruited and maintained (Barron, 2003).

Positive socioemotional group ethos. Furthermore, there is a social-emotional dimension of group work: appropriate group work depends on trust, sensitivity, and respect among pupils (Baines, Blatchford, & Kutnick, 2008; Baines et al., 2009; Galton & Hargreaves, 2009; Kutnick, Ota, & Berdondini, 2008). Negative social-emotional behaviour, such as insulting and domineering behaviour, has a negative impact on group functioning. By contrast, pupils need to develop prosocial behaviours necessary for cooperation, such as being able to promote and seek help.

In successful groups, group members give room to each other's contributions and monitor how unfolding contributions relate to the group goal (Barron, 2003). It follows that after they have made all relevant information explicit, have listened to each other's views, and have checked for agreement, they can take joint action to complete a group task (see also Mercer & Littleton, 2007).

1.2. Cooperative learning to improve pupils' group work behaviour

Cooperative learning refers to the instructional use of small groups in which pupils work together with the goal to increase their learning (Johnson & Johnson, 1999; Slavin et al., 2003), which might involve pupils' cognitive as well as social-emotional learning. Through cooperative learning pupils get a lot of opportunities to practice their group work skills. For instance, during cooperative learning all pupils get the chance to use various speech elements: ask questions, provide answers, give suggestions, and critically reflect on each other's ideas (Gillies, 2003). Cooperative learning techniques, such as *Think-Pair-Share*, provide structure and guidance to group work (Kagan, 1989). During cooperative learning activities, positive interdependence between group members is promoted. Positive interdependence exists when pupils perceive that they are linked together and their actions promote the achievement of joint goals. It promotes a situation in which pupils work together to maximize the learning of all group members, provide mutual support, share resources, and celebrate joint success (Johnson & Johnson, 1999, 2009). Group rewards may be used to motivate groups. The theoretical rationale for the use of group rewards is that if pupils value the success of the group, they will encourage and help one another to achieve the desired goal (Slavin et al., 2003).

Gillies (2004) showed that providing pupils with the opportunity to work cooperatively together on a regular basis in structured cooperative learning groups encourages them to develop social behaviours that promote participation in the group's activities. Blatchford et al. (2006) also found that group work behaviour within the classroom can be supported by a cooperative learning program. This was indicated by less negative behaviours among group members, increased joint involvement, and more high-level talk during group work. Likewise, the study of Tolmie et al. (2010) showed that pupils' collaborative skills improved by implementing a group work skills training followed by structured group work within the classroom.

Most studies on cooperative learning were focused on older children. Even though children in the aforementioned studies were in primary education, they were aged eight years or older. However, positive effects were also found with pupils in the early years of primary education. In the study of Kutnick et al. (2008) the participating pupils were five to seven years old. The young pupils of their experimental classes participated in group work more often, were more likely to stay on task, and were more likely to communicate effectively with their peers during lessons than the pupils in control classes. Similar results for this age group were found in the study of Kutnick and Berdondini (2009). Altogether, in all aforementioned studies a positive impact of a cooperative learning approach on primary school pupils' group work behaviour was shown.

1.3. Cooperative learning implementation in the early grades

In spite of findings emphasizing the importance of cooperative learning, in general primary school practice, pupils work mostly *in* groups, but not *as* groups (Baines et al., 2008; Galton & Hargreaves, 2009; Veenman, Kenter, & Post, 2000). Moreover, several studies indicate that teachers have difficulties implementing cooperative learning in the classroom. For instance, teachers have time management issues and problems with preparing pupils to work together (Gillies & Boyle, 2010). The amount of preparation and effort required of teachers to implement cooperative learning in the classroom should not be underestimated (Blatchford et al., 2006; Kutnick et al., 2008). However, the majority of teachers are not trained to use cooperative learning in their daily classroom practice (Blatchford et al., 2006; Kutnick, Blatchford, & Baines, 2002).

Cooperative learning implementation may be particularly challenging in the lower grades of primary education as more teacher guidance is needed (Battistich & Watson, 2003). As also noted by Kutnick et al. (2008), most studies focused on older children and only a few studies included pupils who are in the first grades of primary education. This might be because it is often suggested that young pupils are not able to work and learn together, but this idea contradicts with research findings on cooperative learning in the early grades. There is some evidence that young pupils are well able to participate in teacher guided cooperative learning activities (Kutnick & Berdondini, 2009; Kutnick et al., 2008).

1.4. Research questions and hypotheses

We investigated the impact of cooperative learning on group work behaviour among young primary school pupils (6- and 7-year-olds) in a group task outside the classroom setting. Cooperative learning was part of a broader intervention: the comprehensive school reform program Success for All (SfA). Its implementation is described below. The main research question to be answered in the current study was whether cooperative learning leads to improved group work behaviour, i.e., more positive group work behaviour and less negative group work behaviour. Furthermore, we explored how pupils' group work behaviour developed over time of the group task, and whether this differed between the intervention and the control condition. The aim of the SfA intervention is to improve pupils' cognitive as well as social-emotional learning. Cooperative learning should promote pupils' achievement by increasing active learning and having pupils learn from each other. It also aims to teach pupils how to work together and improve pupils' group work behaviour. Therefore, in general, it may be expected that involvement in the cooperative learning condition leads to improved pupils' group work. It is hypothesized that involvement in the cooperative learning condition gives pupils the skills needed to avoid negative and sustain positive group work behaviour, that is: 1) skills to sustain a positive socioemotional group ethos and not block the group

process, 2) to involve all group members and deal with the difficulties caused by children who are dominating, isolated, or withdrawn, and 3) increase the likelihood of peer dialogue in which group members are involved in on-task and high-level talk.

2. Method

2.1. Intervention

Success for All (SfA) is a comprehensive school reform program that makes use of cooperative learning in primary education classrooms to improve cognitive as well as social-emotional student outcomes. Success for All has been shown to be effective in increasing student achievement 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. Besides cooperative learning, an engaging reading and language curriculum, tutoring, and parental involvement are core elements of the program.

Instruction in the daily 90-min SfA lessons is characterized by scripted lesson plans that make extensive use of cooperative learning in pairs and small groups. The cooperative learning in SfA includes training for pupils in group work skills. First, cooperative learning used in SfA lessons involves explicit instruction in the required cooperative behaviours, such as *'explain your ideas and ask each other questions'*, *'everyone participates'*, and *'help each other'*. Furthermore, SfA lessons involve various cooperative learning methods, such as *Numbered Heads* and *Think-Pair-Share*. In SfA lessons, teachers use rewards aimed at motivating children. Good teamwork is rewarded with a points system and certificates and successes are celebrated, for instance with cheers.

Teachers have an active role in implementing the cooperative learning approach in the classroom, including serving as a modelling example in how to behave during group work. Teachers are provided with the required materials and manuals and with professional training in how to teach SfA lessons and in the use of cooperative learning. SfA is described in more detail in Slavin, Madden, Chambers, and Haxby (2009).

2.2. Design

A quasi-experimental design was used. At the end of the school year 2015–2016 and school year 2017–2018, Grade 1 pupils (6- and 7-year-olds) performed a group task in small groups of four pupils. The group processes during the group task between the intervention group and the control group were compared using an observation instrument. When the group task took place, the pupils in the intervention group had been following SfA lessons for one school year.

In school year 2015–2016 the SfA program was implemented in Grade 1 of the schools for one year, except for one pilot class in which it was implemented for the second year in Grade 1. In school year 2017–2018 the program was implemented for three years in Grade 1 of the schools. In Grade 1, the Dutch SfA program was integrated in an established Dutch reading curriculum. The control schools followed the same reading curriculum. Teachers in control schools were not given any instructions concerning teaching methods or cooperative learning implementation.

2.3. Sample

Schools were assigned to the intervention group or the control group in accordance with agreements among the school boards, the municipality, and the research and development team of Success for All. One of the control schools chose not to participate in the data collection of the first school year, but was involved in the second year of data collection of the current study. The SfA schools and control schools were all located in the north of the Netherlands and in the same neighbourhoods. All SfA schools and control schools were situated in neighbourhoods of the city 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 pupils with low socio-economic backgrounds, meaning that their parents/legal guardians completed less than two years of secondary school. In school year 2015–2016, percentages ranged from 7% to 18% for SfA schools and from 6% to 27% for control schools (Dienst Uitvoering Onderwijs, 2015). 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 (2015–2016: $t(7) = 1.01, p = .35$; 2017–2018: $t(8) = -.95, p = .37$). This suggests comparability of SfA and control schools.

Groups of four pupils were selected from every class of the participating schools. Selection of pupils was based on active informed parental consent for making video observations in research on SfA. Subsequently, the teachers were asked to compose groups of four pupils. Teachers were asked to form groups heterogeneously, with language achievement as the most important determining factor. By asking the teachers to compose heterogeneous groups, all groups were mixed so there were no groups of only high-achievers or only low-achievers. The number of groups ranged from one to five groups per class. In most classes there were two or three groups per class. Due to missing data for covariates, five groups were excluded from the analyses. In total, this study involved 168 pupils of 18 classes from six intervention schools and 144 pupils of 15 classes from four control schools. Hence, the intervention condition involved 42 groups of four pupils and the control condition involved 36 groups of four pupils.

2.4. Procedure

Ethical approval for this study was obtained from the Ethics Committee Pedagogical and Educational Sciences from the University of Groningen. The group task was performed at the end of the school year 2015–2016 and at the end of school year 2017–2018. Groups of four pupils were taken out of the classroom in order to perform the task in a separate room. By taking the pupils out of the classroom, a controlled setting without distractions was created. All groups of four pupils, from both the intervention and the control group, performed exactly the same task. It should be noted that the SfA pupils did not execute the task in their own team (the small groups in which the pupils work together on a daily basis) and the group task used in the study is not used in the SfA program. Neither the intervention group nor the control group was familiar with the cooperative learning strategy used in this study.

2.4.1. Group task

The group task was designed specifically for this study. Within the task, the test leader told a story that ended with a social problem. We deliberately chose a social problem, as for instance moral or religious problems could be associated with stronger emotions. The task involved an open-ended question lacking a single right answer, as it is often suggested that open-ended tasks are more effective in facilitating productive interaction than more closed tasks (Cohen, 1994; Slavin et al., 2003).

First, the pupils had to think for themselves about a solution for the problem. Second, the test leader asked them to write down (or draw)

Table 1
Descriptive statistics of observation scores at the group level.

	Total (n = 78)	SfA group (n = 42)	Control group (n = 36)
	M (SD)	*M (SD)*	*M (SD)*
Socioemotional ethos: Maintenance			
Maintenance	.105 (.15)	.121 (.13)	.086 (.14)
No Maintenance	.895 (.15)	.879 (.13	.914 (.14)
Socioemotional ethos: Blocking			
Blocking	.032 (.09)	.018 (.05)	.049 (.12)
No blocking	.968 (.09)	.982 (.05)	.951 (.12)
Group participation			
All on task	.470 (.35)	.530 (.38)**	.402 (.30)
All on task, some passive	.266 (.32)	.312 (.34)	.211 (.29)
Some uninvolved, passive	.043 (.10)	.013 (.05)	.077(.14)**
Some uninvolved, active	.176 (.19)	.136 (.17)	.220 (.21)
All off task	.023 (.07)	.006 (.03)	.043 (.10)*
On task, Split	.023 (.09)	.003 (.02)	.047 (.12)*
Type of pupil-pupil dialogue			
Coll. discussion: Inferential talk	.069 (.10)	.089 (.12)*	.046 (.08)
Coll. discussion: Text-based	.007 (.02)	.002 (.01)	.015 (.04)
Metagroup talk	.160 (.18)	.204 (.18)*	.109 (.17)
Sharing Information	.344 (.19)	.281 (.15)	.419 (.20)**
Procedural talk	.209 (.18)	.231 (.17)	.182 (.18)
Reading-out task (asking)	.021 (.06)	.025 (.08)	.017 (.04)
Disputational talk	.028 (.07)	.017 (.06)	.041 (.09)
Off task	.118 (.15)	.111 (.14)	.134 (.15)
No talk	.042 (.11)	.046 (.12)	.038 (.08)

**p < .01 *p < .05.

three solutions for solving the problem within their own part of a cooperative learning placemat. The placemat, printed on a large piece of paper (A2 size), was divided into five fields: one part for every child to write down his or her solution(s) and one part in the middle where they could write down their final solution as a group. By giving pupils appropriate time to think first, it was assumed that the quality of the pupils' responses improved and that they all could contribute to the discussion. After all pupils had thought about the problem individually and had written something down in their own field of the cooperative learning placemat, the test leader asked them to discuss what would be the best solution for the problem. Finally, they had to write down, as a group, their cooperative solution in the center of the placemat. When the pupils were finished, the test leader asked them why they had chosen their solution and evaluated the group work process with the pupils. In total, the session took about 15–30 min per group.

2.4.2. Video observations

To observe the group processes in detail, video observations were used. Cameras of two tablets were directed towards the group of four pupils at different places in the room. By using two cameras all the pupils' faces were visible on the videos. Furthermore, an audio recorder was placed on the table to make sure that the talk of all pupils could be heard. Pupils spent some time in the room before the group discussion started to get used to the equipment. Moreover, the same equipment had been used in the classrooms earlier that school year, which made the pupils somewhat accustomed to the technical devices. While filming the tablet screens turned black to avoid distraction.

2.5. Instruments

2.5.1. Observation instrument group work behaviour

Videos were analyzed using an observation instrument. The group process was observed from the moment the test leader gave the sign to start the discussion until the moment the test leader started to evaluate the final group solution. The duration of this process differed per group, from slightly over 1 min to over 15 min. Because the process of group work depends on how the children react to each other, the unit of analysis was the group rather than the individual child.

Three category sets of Blatchford et al.'s (2006) observation instrument (see also Baines et al., 2009) were adopted: *socioemotional ethos, group participation,* and *type of pupil-pupil dialogue.* We decided not to use the category set *discourse topic*, concerning whether discussion was sustained or changed, because we included all time intervals in the analysis allowing us to identify how group processes developed over the time of task, e.g., whether specific talk was sustained or changed over time. Following the original observation instrument, *socioemotional ethos, group participation,* and *type of pupil-pupil dialogue* were coded on playback for every 20 s of group interaction. Full definitions of the used categories are described in Appendix A.

The categories in the sets *group participation* and *type of pupil-pupil dialogue* were mutually exclusive, i.e., in both category sets exactly one

Table 2
Descriptive statistics of dependent variables.

	Total (n = 1616)	SfA group (n = 908)	Control group (n = 708)
Positive group work behaviour	*M (SD)*	*M (SD)*	*M (SD)*
	2.06 (.84)	2.21 (.79)	1.87 (.88)
	VAR (SE)	*VAR (SE)*	*VAR (SE)*
Level-three variance (classes)	.104 (.044)	.065 (.038)	.065 (.062)
Level-two variance (groups)	.117 (.031)	.065 (.026)	.174 (.064)
Level-one variance (time)	.510 (.018)	.487 (.023)	.539 (.029)
Negative group work behaviour	*M (SD)*	*M (SD)*	*M (SD)*
	1.35 (.83)	1.23 (.76)	1.51 (.79)
	VAR (SE)	*VAR (SE)*	*VAR (SE)*
Level-three variance (classes)	.079 (.036)	.055 (.036)	.062 (.050)
Level-two variance (groups)	.097 (.027)	.080 (.030)	.112 (.046)
Level-one variance (time)	.527 (.019)	.448 (.022)	.628 (.034)

Table 3
Descriptive statistics of control variables.

	Total (n = 78)	SfA group (n = 42)	Control group (n = 36)
Oral language skills	*M (SD)*	*M (SD)*	*M (SD)*
Group mean	19.53 (3.30)	19.69 (3.14)	19.18 (3.34)
Group sd.	3.62 (1.65)	3.77 (1.81)	3.48 (1.33)
Social competence	*M (SD)*	*M (SD)*	*M (SD)*
Group mean	104.58 (9.19)	103.71 (8.79)	105.58 (9.65)
Group sd.	9.99 (5.70)	10.51 (6.38)	9.39 (4.81)
Gender (0: boy, 1: girl)	*M (SD)*	*M (SD)*	*M (SD)*
Group mean	.47 (.40)	.52 (.40)	0.42 (.36)
Age	*M (SD)*	*M (SD)*	*M (SD)*
Group mean	7.21 (.22)	7.19 (.22)	7.25 (.23)
Group sd.	.40 (.20)	.39 (.20)	.41 (.18)
Within-group correlation between social competence and oral language skills	.22 (.59)	.31 (.53)	.11 (.64)

of the categories had to be coded as 1 for any given 20-s interval. In the category set *socioemotional ethos* of the original observation instrument the observer could code either or both of two categories, i.e., maintenance and/or blocking of the group process. For the aim of analysis, we split this set into two category sets also with mutually exclusive categories. This means that we created two category sets for socioemotional ethos: 1) *socioemotional ethos: maintenance* (maintenance = 1, no maintenance = 0), and 2) *socioemotional ethos: blocking* (blocking = 1, no blocking = 0). Correlations between scores are shown in Appendix B.

For both conditions descriptive statistics of scores per category averaged over the total duration of the task per group are shown in Table 1. To give more insight in the group processes, differences between the intervention and control group of the average scores over the whole task for each category were analyzed using multilevel analyses (Table 1: models with two levels: groups in classes). SfA pupils showed more collaborative inferential talk ($p < 0.05$) and more metagroup talk ($p < 0.05$) compared to the control group. SfA pupils showed also more on-task behaviour ($p < 0.01$). The pupils in the control group showed more off-task behaviour ($p < 0.05$) and used more sharing information ($p < 0.01$).

2.5.2. Dependent variables: Positive and negative group work behaviour

Based on the observation categories of Blatchford et al.'s (2006) instrument, two dependent variables were constructed. The first variable combined the positive group behaviours: 1) *socioemotional ethos*: maintenance; 2) *socioemotional ethos*: no blocking; 3) *group participation*: All on task and All on task, some passive; 4) *type of pupil-pupil dialogue*: Collaborative discussion (inferential talk and text-based talk) and Metagroup talk.

The second variable combined the negative group behaviours: 1) *socioemotional ethos*: no maintenance; 2) *socioemotional ethos*: blocking; 3) *group participation*: Some uninvolved, passive, Some uninvolved, active, and All off task; 4) *type of pupil-pupil dialogue*: Off task and Disputational talk.

The constructed variables, positive and negative group work behaviour, are the sum of the scores on the above mentioned categories at the level of the 20-s time sequences. Hence, both variables range from 0 to 4. Skewness and Kurtosis for positive group work behaviour were 0.15 and -.13 and for negative group work behaviour .91 and .34, which is no reason for concern. We also checked normality of residuals in the multilevel analyses. Means and variance components of the dependent variables at the various levels (time intervals in groups in classes) are given in Table 2.

2.5.3. Interrater reliability

To make the comparison between the intervention group and the control group more reliable, all videos were coded by the same researcher. A second observer, also a researcher and an experienced primary school teacher, coded approximately 15% of the videos of the first school year of data collection to check the reliability of the observation instrument. Cohen's kappa's for the dependent variables were above .60, indicating substantial agreement (Landis & Koch, 1977): $\kappa = .71$ for positive group work behaviour and $\kappa = .64$ for negative group work behaviour.

2.5.4. Control variables

We controlled for several group characteristics taking into account social competence scores, scores on an oral language test, age, and gender of pupils. To control for gender composition, the mean of gender was used (0: boy, 1: girl). To control for differences in group composition regarding age, social competence, and oral language skills, both the mean of a group and the within-group standard deviation, as a measure of heterogeneity, were taken into account.

Social competence. The teachers scored social competence of every pupil in their class using the Social Competence Observation List (Joosten, 2006). This questionnaire consists of 26 questions (scale 1–5 per item). The validity and reliability of this instrument were assessed as sufficient by the Dutch National Committee of Tests and Testing (COTAN, 2008). Cronbach's alpha was .93 for Grade 1 pupils in the study of Joosten (2006).

Oral language skills. Pupils' oral language skills were tested, using a test of which the reliability and validity have been assessed as good (COTAN, 2007; Verhoeven & Vermeer, 2006). Cronbach's alpha was .91 for the test items for Grade 1 pupils in the study of Verhoeven and Vermeer (2006). In the test, pupils were asked to tell a story based on pictures that were shown to them. Two stories with different pictures were used in the test. The completeness of the stories was scored on a scale from 1 to 32 using a checklist.

Relation between oral language skills and social competence. The correlation between oral language skills and social competence within a group was also taken into account as a control variable. It was expected that when the composition of the group was such that higher oral language skills of pupils were combined within the group with lower social competence, the level of group dialogue might be worse than would be expected of pupils with average levels of language and social competence skills.

There were some missing data for the control variables. Of the 336 pupils participating in the group task, there were nine missing scores on the oral language skills test and one missing for age. Of the pupils participating in the first school year, there were four missing scores for social competence. Hence, for a few groups, covariates (which all were within group measures) were calculated using only the available student cases. In the second school year of data collection of this study, teachers of one school did not fill in the social competence questionnaire, leading to missing social competence scores for 20 pupils who did the task together in five groups. As mentioned before, these five groups were excluded from analyses.

Tested with multilevel analyses (models with two levels: groups in classes), there were no significant differences between the SfA groups and the control groups with respect to the control variables. Descriptive statistics of control variables are shown in Table 3.

Table 4
Multilevel models of positive group work behaviour.

	Model 1		Model 2		Model 3		Model 4	
Fixed Part	β	SE	β	SE	β	SE	β	SE
Intercept	2.337	0.140	1.196	2.282	0.898	2.178	0.735	2.179
Relative time	−0.475**	0.082	−0.474**	0.082	−0.473**	0.082	0.382	0.383
Duration task	−0.001	0.005	0.000	0.005	−0.000	0.005	−0.000	0.005
Group mean social competence			0.001	0.008	0.002	0.007	0.002	0.007
Group sd. social competence			−0.009	0.012	−0.010	0.011	−0.010	0.011
Group mean age			0.154	0.286	0.166	0.274	0.166	0.274
Group sd. age			−0.560	0.287	−0.510	0.278	−0.509	0.278
Group mean gender			0.246	0.133	0.209	0.129	0.209	0.129
Group mean oral language skills			0.015	0.020	0.006	0.019	0.006	0.019
Group sd. oral language skills			−0.040	0.032	−0.048	0.031	−0.048	0.031
Group correlation social competence and oral language skills			0.037	0.096	−0.029	0.094	−0.028	0.094
SfA					0.455**	0.138	0.357	0.184
Relative time squared							−0.789*	0.344
SfA * relative time							0.621	0.511
SfA * relative time squared							−0.642	0.460
Random Part	Variance	SE	Variance	SE	Variance	SE	Variance	SE
Level-three variance								
Intercept	0.106	0.045	0.133	0.050	0.082	0.037	0.082	0.037
Relative time slope	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Covariance	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Level-two variance								
Intercept	0.185	0.055	0.161	0.051	0.156	0.049	0.159	0.049
Relative time slope	−0.116	0.056	−0.109	0.055	−0.103	0.053	−0.109	0.054
Covariance	0.222	0.082	0.221	0.082	0.220	0.081	0.231	0.083
Level-one variance								
Intercept	0.456	0.019	0.455	0.019	0.455	0.019	0.447	0.019
Dummy last part * intercept	0.036	0.022	0.038	0.023	0.038	0.023	0.038	0.022
−2*loglikelihood	3720.008		3569.700		3558.934		3531.893	

n = 1616 time intervals in 78 groups 33 classes.
**p < .01 *p < .05.

2.6. Analyses

Multilevel analyses were used to test the effect of SfA on positive and negative group work behaviour, treating the repeated measures obtained using the 20-s time intervals during the group task as longitudinal data. The longitudinal data were unbalanced: the number of 20-s intervals per group ranged from 4 to 47. Multilevel analysis is an appropriate method for analyzing unbalanced longitudinal data (Snijders & Bosker, 2012). In addition, the data set had a hierarchical character, with the nested levels of time interval, group, class, and school. Predictor variables were defined at several of these levels. This supported the decision to use multilevel analysis. Positive and negative group work behaviour were analyzed separately. Analyses were performed using the program MLwiN version 3.0 (Rasbash et al., 2000). Because the number of schools in this study (n = 10) was limited, and preliminary analyses showed that there was no variance at the school level, this level was not taken into account in the model. Hence, the multilevel models concerned three levels: time intervals, i.e., measurement occasions (level 1), groups (level 2), and classes (level 3). In total, there were 1616 time intervals in 78 groups in 33 classes.

Total duration of the task, and hence the number of measurement occasions, differed between the groups. As the result of explorations (see Appendix B), it was decided to include relative time, defined as absolute time divided by total duration, as predictor variable. Thus, in each group, relative time runs from 0 to 1. Because we included relative time, we decided to include the duration of the task as a covariate in the models. The duration of the task varied per group from slightly over 1 min to over 15 min. The inclusion of it as a covariate was permitted because there was no significant difference between the SfA group and the control group in the duration of the task (Mann-Whitney U = 664.5, n_1 = 42, n_2 = 36, p = .36). Furthermore, it was decided to specify a random slope for relative time at the second level (groups) and third level (classes). A random slope effect of time at the third level did not improve the models significantly. However, because condition (SfA vs. control group) is a variable defined at the third level, models with a random slope of time at both the second and third level were used in further analysis (cf. chapter 6 of Snijders & Bosker, 2012). There was more variability in both positive and negative group work behaviour after the 20th time interval than before; therefore, we modelled this heteroscedasticity using a dummy variable, called *Dummy last part*, being random at level 1.

Models were fitted in the following steps. The first models include variables representing time: duration of the task (total number of intervals per group) and relative time (interval number divided by the total number of intervals). The second step includes the covariates. A backward selection procedure was planned for the covariates, starting with the models including all covariates with fixed effects. As covariates are correlated we decided to keep all covariates in the models. Covariates were not centered in the analyses. Third, condition (1: SfA, 0: control) was added to the models and interactions between time and condition were tested. A squared term of relative time was added to the models with interactions of time and condition, because of its

Table 5
Multilevel models of negative group work behaviour.

Fixed Part	Model 1		Model 2		Model 3		Model 4	
	β	SE	β	SE	β	SE	β	SE
Intercept	0.807	0.124	0.672	2.110	0.909	2.047	0.852	2.049
Relative time	0.588**	0.094	0.590**	0.094	0.595**	0.093	0.906*	0.386
Duration task	0.008	0.005	0.008	0.005	0.009	0.005	0.00	0.005
Group mean social competence			−0.004	0.007	−0.005	0.006	−0.005	0.006
Group sd. social competence			0.007	0.011	0.009	0.010	0.009	0.010
Group mean age			0.053	0.266	0.029	0.259	0.024	0.259
Group sd. age			0.348	0.269	0.316	0.264	0.312	0.264
Group mean gender			−0.226	0.124	−0.197	0.122	−0.200	0.122
Group mean oral language skills			0.001	0.018	0.006	0.018	0.006	0.018
Group sd. oral language skills			0.015	0.030	0.027	0.030	0.026	0.030
Group correlation social competence and oral language skills			0.013	0.090	0.056	0.089	0.058	0.089
SfA					−0.326*	0.123	−0.047	0.171
Relative time squared							−0.148	0.343
SfA * relative time							−1.112*	0.514
SfA * relative time squared							0.794	0.458
Random Part	**Variance**	**SE**	**Variance**	**SE**	**Variance**	**SE**	**Variance**	**SE**
Level-three variance								
Intercept	0.041	0.043	0.060	0.047	0.054	0.045	0.051	0.045
Relative time slope	0.033	0.046	0.023	0.048	0.002	0.049	0.002	0.049
Covariance	0.012	0.080	0.010	0.080	0.002	0.080	0.011	0.080
Level-two variance								
Intercept	0.168	0.055	0.158	0.053	0.157	0.053	0.156	0.053
Relative time slope	−0.152	0.073	−0.146	0.072	−0.144	0.072	−0.140	0.071
Covariance	0.347	0.128	0.352	0.129	0.360	0.131	0.347	0.128
Level-one variance								
Intercept	0.423	0.018	0.423	0.018	0.423	0.018	0.422	0.018
Dummy last part * intercept	0.095	0.025	0.096	0.025	0.095	0.025	0.096	0.025
−2*loglikelihood	3735.593		3539.274		3532.451		3525.461	

n = 1615 time intervals in 78 groups 33 classes.
**p < .01 *p < .05.

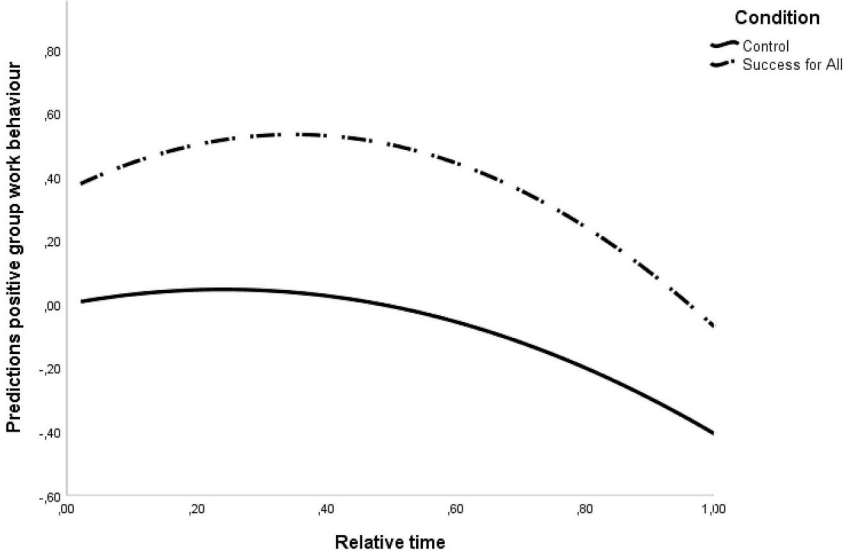

Fig. 1. Interaction effect of time and condition: predictions for positive group work behaviour.

significance. Assumptions were checked for the final multilevel models. Residuals were approximately normally distributed and variance of residuals was evenly distributed.

2.6.1. Power analysis

We did a power analysis using the software Optimal Design Plus (Spybrook et al., 2011), with a two-level design (groups in classes), i.e., not taking into account the time over task. There were 33 participating

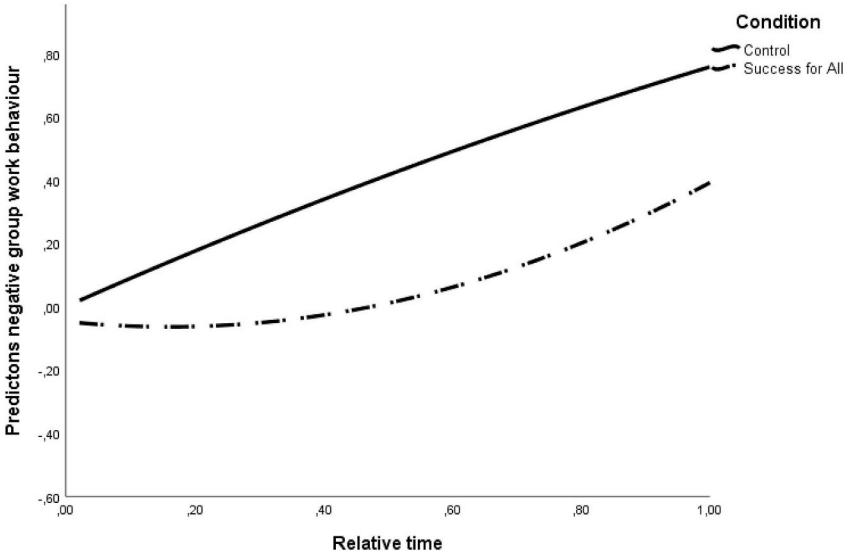

Fig. 2. Interaction effect of time and condition: predictions for negative group work behaviour.

classes in the study and we assumed an intraclass correlation of ρ = .10. Depending on whether there are two or three groups per class, the power ranges between .50 and .63 to determine a medium effect size of .50 with a two-sided significance level of .05. This indicates that the power is of medium size; but it was impossible in practice to recruit more classes.

2.6.2. Effect sizes

We took the following approach to calculate effect sizes of the intervention. As for intervention effects we are interested in general differences in group work behaviour between SfA and control groups, we did not take into account how group processes developed over time of task. Therefore, we calculated effect sizes by dividing regression coefficients of condition by the square root of the sum of level-three (classes) and level-two variances (groups) of the models that included condition and all covariates but no random slopes. This is analogous to Cohen's *d*. These multilevel models are shown in Appendix D.

3. Results

3.1. Models including time and inclusion of covariates

In Tables 4 and 5, the estimates for the multilevel models of positive and negative group work behaviour are given. Models 1 included time variables. In general, positive group work behaviour decreased over time of the task. Negative group work behaviour increased over time. Duration of task was significant for negative group work behaviour ($p < .05$), meaning that the longer the task lasts, the more negative group work behaviour was shown.

Residual variance for time at level 1 is not constant, meaning that for later time intervals (after the 20th time interval) there is more variability between time intervals with respect to positive and negative group work behaviour (see Appendix B for an elaboration of the modelled heteroscedasticity). This increase in variance suggests greater irregularity toward the end of the group work, with bursts of negative group work behaviour (note that the heteroscedasticity is especially strong for the negative group work behaviour).

All covariates were added to the models: the group mean and group standard deviation of social competence, age, and oral language skills, the group mean of gender, and the correlation between social competence and oral language skills. None of the covariates was found to have significant effect at the level of .05.

3.2. Models including condition and interaction effects between time and condition

Condition was added to the models and it was subsequently tested whether relative time interacted with condition. A significant main effect of condition was found. SfA groups showed more positive group work behaviour ($p < .01$) and less negative group work behaviour ($p < .05$) compared to the control groups. Not taking into account developments over time of task, but only general differences between SfA groups and control groups, the effect size of the intervention for positive group work behaviour is 1.05 (ES = .451/√(.080 + .105)) and for negative group work behaviour -.87 (ES = -.343/√(.059 + .097)).

To explore differences between SfA groups and control groups over time of task, we added interactions between time and condition for positive and negative group work behaviour. As shown in Fig. 1, positive group work behaviour followed trends of a similar inverse U-shape in both conditions, with the SfA groups maintaining a consistent advantage compared to the control groups throughout the duration of the group task. The model was significantly improved ($\chi^2 = 27.04$, $df = 3$, $p < .01$) by adding quadratic time and interactions between time and quadratic time multiplied by condition. This improvement is due to adding the squared term of relative time ($p < .01$). The interaction terms between time and condition were not significant for positive group work behaviour.

For negative group work behaviour, adding quadratic time and interactions between time and quadratic time multiplied by condition did not improve the model significantly ($\chi^2 = 6.99$, $df = 3$, $p = .07$). However, the interaction between time and condition ($p < .05$) was significant when considered on its own. This suggests there may be differences between SfA groups and control groups in the development of negative behaviours over time of task. In control groups negative group work behaviour increased gradually during the whole task, while in SfA groups negative group work behaviour was more or less constant initially, but increased towards the end of the task (see Fig. 2).

The lack of significance in these explorative tests with three degrees of freedom is no reason to put in doubt the differences between SfA and control groups found for Model 3. The tests including also the linear and quadratic trends consider any shape differences between the two curves, thereby diluting the power to detect differences between the two conditions.

4. Discussion

The aim of the current study was to investigate the effect of cooperative learning in the Success for All (SfA) program on group work behaviour of pupils in Grade 1 in primary education. At the end of the school year, Grade 1 pupils (6- and 7-year-olds) performed a group task in groups of four children. At that moment, pupils in the intervention group had been following SfA lessons including cooperative learning on a daily basis for one whole school year. Multilevel analyses were used to test the effect of cooperative learning in SfA on positive and negative group work behaviour, treating the repeated measures obtained using the 20-s time intervals during the group task as longitudinal data.

The SfA groups showed more positive group work behaviour and less negative group work behaviour compared to the control group whilst controlling for several group characteristics, supporting the hypothesis that changes in children's behaviour as a result of involvement in the SfA program would lead to improved group work. Effect sizes were large: ES = 1.05 for positive and ES = -.87 for negative group work behaviour. The findings of this study are in line with other studies that found evidence for the improvement of group work behaviour of young-aged primary school pupils by cooperative learning (Blatchford et al., 2006; Gillies, 2004; Kutnick et al., 2008; Tolmie et al., 2010).

We also explored how pupils' group work behaviour developed over time of the group task and whether this differed between the SfA groups and the control groups. In general, positive group work behaviour decreased towards the end of the task, while negative group work behaviour increased towards the end of the task. Although this is not a surprising finding, it is important for the implementation of group work among pupils of this young age. It leads to the suggestion that young primary school pupils should be introduced to group work activities in restricted periods of time, which can be gradually increased under guidance of the teacher.

Moreover, results suggest that there may be differences between the SfA groups and the control groups in how negative group work behaviour changes over time of task. In SfA groups the level of negative group work behaviour was approximately constant at first, and increased only towards the end of the task, whereas in control groups negative group work behaviour increased gradually during the whole task. Future research is needed to explore further how group work behaviour develops over time. Such research might explore different phases in group work and investigate whether the time development of positive and negative group work behaviours is affected by cooperative learning. Perhaps cooperative learning methods used in SfA lessons give pupils specific tools to structure their group work over time. The results of our experiment suggest they may have learned tools to avoid negative group work behaviour, but that this is effective only for a limited period of time.

Important limitations of the current study are the relatively small sample size (78 groups in 33 classes), the absence of random assignment of schools, and the lack of a pretest of group work behaviour. Power of statistical analyses was of medium size, but it was impossible in practice to recruit more classes. The sample size can be regarded as justifiable since it is known to be difficult to conduct such complex intervention studies in practice. With respect to the non-random assignment of schools, we note that no significant differences were found between SfA and control schools on various measures. Conducting a pretest was impossible, because before the intervention pupils were in different classes in Kindergarten. The unit of analysis was a group of four pupils who were familiar with each other from the classroom, and these groups could not have existed earlier. Given the lack of a pretest, some support for the comparative results still is given by the absence of significant differences between the intervention groups and the control groups in several group characteristics, such as oral language skills.

Based on the observation categories used in Blatchford et al.'s (2006) observation instrument, two variables were constructed: one variable that combined positive behaviours and one that combined negative behaviours, taking into account socioemotional ethos, group participation, and type of pupil-pupil dialogue in the group process. This provided a method that includes the most important aspects of group work. However, additional research is needed to explore in more detail the effects of cooperative learning on particular aspects of group work. For instance, in this study SfA pupils showed significant more collaborative inferential talk, but not more collaborative text-based talk (as shown in Table 1). Both types of pupil-pupil dialogue are part of skilled group work, but thinking beyond the text of the story may be an important learned skill within the SfA program.

In the current study, we focused on pupils' group work behaviour during a group task. We did not investigate the outcomes of pupils' group work, i.e., the group solutions for the social problem stated in the group task. It would be relevant for future research to study how specific aspects of pupils' group work behaviour are related to group work outcomes. We expect that for instance more collaborative inferential talk will lead to more thoughtful solutions, as this is expected to lead to a more focused discussion about why one solution would be better than another. However, it should be noted that the task used by us was an open-ended group task not having one right answer, so several solutions for the problem were possible.

Others studies found evidence for improvement of group work behaviour of pupils by cooperative learning inside classroom settings (Blatchford et al., 2006; Gillies, 2004; Kutnick et al., 2008; Tolmie et al., 2010). Using a controlled setting for conducting the group task outside the classroom allowed us to investigate whether learned group work behaviours were transferable to a situation without teacher guidance. The results indicate that the hoped-for benefits of a cooperative learning approach on pupils group work behaviour can transfer to a setting outside the classroom. However, as we did not examine the interactions in groups under normal classroom conditions, further research needs to be done to provide more information about the effectiveness (for both cognitive as well as behavourial outcomes) of group work within the teams in SfA lessons. The amount of preparation and effort required of teachers to implement cooperative learning in the classroom should not be underestimated (Blatchford et al., 2006; Kutnick et al., 2008). Hence, future research should take into account challenges teachers may experience when implementing cooperative learning in their classrooms, in particular with regard to cooperative learning implementation in the lower grades of primary education.

The current study contributes to the existing research literature about cooperative learning mainly because of its focus on a) the young age group of Grade 1 pupils, and b) group work behaviour rather than cognitive achievement. Working together consists of many different aspects and is a highly complex process that even many adults may struggle with. Because of this, one may think that young children are not able to work and learn together. However, there is evidence that young children are capable of participating in group work activities

(Kutnick & Berdondini, 2009; Kutnick et al., 2008). As learning to work together is an important educational goal (Barron, 2003; Blatchford et al., 2006), the general practice in which pupils mostly work *in* groups, but not *as* groups should be questioned (Baines et al., 2008; Galton & Hargreaves, 2009; Veenman et al., 2000). The results of the current study show that group work of young children might be improved by cooperative learning. This emphasizes the importance of implementing cooperative learning in the early grades of primary education.

Funding

This research was supported by a grant from the Netherlands Organisation for Scientific Research (NWO). Grant number: 405-15-731.

CRediT authorship contribution statement

M.A. Veldman: Writing - original draft, Conceptualization. **S. Doolaard:** Conceptualization. **R.J. Bosker:** Conceptualization. **T.A.B. Snijders:** Conceptualization.

Appendix A

The full definitions of the categories of the observation instrument (based on Blatchford et al., 2006, p. 264–265, see also; Baines et al., 2009).

Observation categories are related to groups. Coding every 20 s.		
Socioemotional ethos	For this category set, the observer could code either or both categories, if they occurred once or more during the 20-s sequence.	***Group maintenance:*** Speech or gestures that sustain the interaction and function of the group, that draw others in, for example, by saying "That's a good suggestion" or "Do you agree?" ***Group blocking:*** A child in the group makes an explicit–active attempt to block progress by refusal to participate or cooperate by saying, for example, "I'm not doing this" or by ridiculing another pupil.
Group participation	For this mutually exclusive category set, the observer coded the prominent form of interaction within the 20-s sequence.	***All actively involved and on task:*** When all group members contribute and actively listen to others and all are equally engaged in the work and the interaction is on task. ***All involved and on task, some passive:*** As above but one or some are actively listening but not contributing. ***Some uninvolved and appear off task, passive:*** Some are involved in the group interaction, but 1 or more children are passively disengaged and plainly not attending. ***Some uninvolved and actively off task:*** Some are involved in the group interaction but 1, 2, or 3 children are actively engaged in something else not related to the task. ***All off task:*** None of the pupils are on task and all are clearly off task. ***All involved and on task, split:*** When the group splits into two pairs. Both groups are on task and actively involved.
Type of pupil-pupil dialogue	For this mutually exclusive category set, categories were ordered when talk was on task such that an occurrence of on-task inferential collaborative discussion superseded all other categories (unless all talk was off task) followed by text-based collaborative, metagroup, sharing information, and so on.	***Collaborative discussion:*** Dialogue in which group members engage in talk that aims to make reasoning explicit to others and involves giving explanations, counterarguments, justifications, conditionals, and result statements. This was divided into two types: ***Collaborative inferential talk*** involves reasoning that draws on evidence or ideas that go beyond the text of the story of the task. This superseded ***collaborative text-based talk***, which was more limited in that it was reasoning based on information given in the story ***Metagroup talk:*** Organizing and planning within the group; talk about the group rather than about the task at hand (e.g., children may discuss group roles, how the group should proceed, etc.). ***Sharing information:*** Sharing members' ideas, opinions, suggestions, and different knowledge but with little effort to explore the ideas further or to investigate the evidence. Reasons for opinions or decisions are not given. ***Procedural talk:*** When pupils talk about preparation of the task and materials, e.g., spelling out words for others. ***Disputational talk:*** Speakers are concerned with defending their own idea–opinion at the expense of any attempt at reasoning a solution or compromise. ***Asking the test leader:*** Because in the present study no reading task or written instructions were used (only oral instructions given by the test leader), we replaced the category *reading-out task* of the original observation instrument, which involved reading out the task instructions by the pupils before beginning the discussion with the category asking the test leader a question (about what to do or about the story). ***Off-task talk:*** When the talk is not related to the task or topic of discussion. ***No talk:*** When there was no talk during the 20-s time interval none of the above categories was scored.

Appendix B

Table B1
Spearman correlations between cooperative behaviours

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.
Socioemotional ethos: Maintenance																		
1. Maintenance																		
2. No maintenance	−1.00																	
Socioemotional ethos: Blocking																		
3. Blocking	-.003	.003																
4. No blocking	.003	-.003	−1.00															
Group participation																		
5. All on task	-.110**	.110**	-.113**	.113**														
6. All on task some passive	.173**	-.173**	-.062*	.062*	-.564**													
7. Some uninvolved passive	-.027	.027	-.035	.035	-.164**	-.111**												
8. Some uninvolved active	-.010	.010	.184**	-.184**	-.448**	-.308**	-.089**											
9. All off task	-.042	.042	.011	-.011	-.138**	-.099**	-.028	-.079**										
10. On task Split	-.044	.044	.080**	-.080**	-.078**	-.079**	-.023	-.064*	-.020									
Type of pupil-pupil dialogue																		
11. Collaborative discussion inferential talk	-.035	.035	-.017	.017	.102**	-.016	-.037	-.075**	-.030	-.036								
12. Collaborative discussion text based	-.033	.033	.016	-.016	.002	-.031	.020	.033	-.015	-.012	.047							
13. Metagroup talk	.191**	-.191**	-.027	.027	.174**	.016	-.062*	-.168**	-.051*	-.006	-.109**	-.044						
14. Sharing Information	.052*	-.052*	-.057*	.057*	-.001	.116**	.045	-.154**	-.105**	.108**	-.187**	-.064*	-.300**					
15. Disputational talk	-.042	.042	.236**	-.236**	-.077**	.036	.026	.061*	-.030	-.024	-.053*	-.018	-.085**	-.124**				
16. Reading out task (asking)	.040	-.040	.015	-.015	.059*	-.069**	.003	-.006	.033	-.025	-.055*	-.019	-.088**	-.129**	-.037			
17. Procedural talk	-.132**	.132**	-.039	.039	.049*	.006	-.016	-.018	-.084**	-.020	-.149**	-.051*	-.239**	-.349**	-.099**	-.103**		
18. Off task	-.020	.020	.047	-.047	-.309**	-.159**	-.051*	.464**	.353**	-.051*	-.113**	-.039	-.172**	-.260**	-.075**	-.001	-.201**	
19. No talk	-.022	.022	.041	-.041	-.019	.008	.141**	-.035	-.014	-.027	-.059*	-.020	-.095**	-.132**	-.022	.189**	-.058*	-.084**

**p < .01 *p < .05.

Appendix C

Modelling time in the multilevel models

Total duration of the task, and hence the number of measurement occasions, differed between the groups. This implies that time could be measured in two different ways: as absolute time, which is the sequence number of the time interval, and as relative time, defined as absolute time divided by total duration. The first compares the group work behaviour processes between the groups for the same number of seconds elapsed, the second for the same fraction of total task elapsed. Absolute time runs from 0 to the number of 20-s intervals per group, which ranges from 4 to 47 (depending on the group); relative time runs in each group from 0 to 1. In an explorative phase, without taking into account which were the intervention and which the control classes, a choice was made between these two time variables, between a linear or non-linear effect of time, and between a homoscedastic or heteroscedastic model. For non-linear models, polynomials as well as spline functions were considered (cf. chapter 15 of Snijders & Bosker, 2012). These choices were made on grounds of statistical fit as well as interpretability.

As the result of explorations, it was decided to include relative time as predictor variable, without spline functions. A squared term of relative time was only added to the models with interactions of time and condition, because in that case it proved to be statistically significant. There was more variability in both positive and negative group work behaviour after the 20th time interval than before; therefore, we modelled this heteroscedasticity using a dummy variable, called *Dummy last part*, being random at level 1. A linear dependence of the level-one residual variance on a variable can be expressed by the formula $\sigma_0^2 + 2\sigma_{01}\, x_{1ij}$ (Snijders & Bosker, 2012, p. 120). As x_1 is a dummy variable with values 0 for time intervals ≤ 20 and 1 for time intervals > 20, the residual variance is σ_0^2 for the units up to time interval 20 and $\sigma_0^2 + 2\sigma_{01}$ after the 20th time interval. Thus, the level 1 variance for the first 20 time intervals is .45 for positive group work behaviour and .43 for negative group work behaviour. After the 20th time interval, the level 1 variance is .53 (0.447 + 2 * 0.040) for positive group work behaviour and .61 (0.430 + 2 * 0.090) for negative group work behaviour. This means that after 400 seconds had elapsed, for both kinds of behaviour there was a greater over-time variability within the groups.

Appendix D

Table D1
Multilevel models without random slopes

	Positive group work behaviour		Negative group work behaviour	
Fixed part	β	SE	β	SE
Intercept	0.920	2.168	0.922	2.053
Relative time	−0.468**	0.060	0.592**	0.061
Duration task	0.000	0.005	0.008	0.005
Group mean social competence	0.002	0.007	−0.004	0.006
Group sd. social competence	−0.010	0.011	0.011	0.010
Group mean age	0.164	0.273	0.027	0.259
Group sd. age	−0.506	0.277	0.307	0.264
Group mean gender	0.208	0.128	−0.188	0.122
Group mean oral language skills	0.006	0.019	0.004	0.018
Group sd. oral language skills	−0.045	0.031	0.019	0.030
Group correlation social competence and oral language skills	−0.021	0.093	0.031	0.089
SfA	0.451**	0.136	−0.343*	0.124
Random Part	Variance	SE	Variance	SE
Level-three variance	0.080	0.037	0.059	0.030
Level-two variance	0.105	0.028	0.097	0.026
Level-one variance	0.491	0.018	0.497	0.018
−2*loglikelihood	3578.476		3586.296	

**p < .01 *p < .05.

References

Ananiadou, K., & Claro, M. (2009). *21st century skills and competences for new millennium learners in OECD countries.* Paris: OECD.

Baines, E., Blatchford, P., & Kutnick, P. (2008). Pupil grouping for learning: Developing a social pedagogy of the classroom. In R. M. Gillies, A. Ashman, & J. Terwel (Eds.). *The teacher's role in implementing cooperative learning in the classroom* (pp. 56–72). New York: Springer.

Baines, E., Rubie-Davies, C., & Blatchford, P. (2009). Improving pupil group work interaction and dialogue in primary classrooms: Results from a year-long intervention study. *Cambridge Journal of Education, 39*(1), 95–117.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*, 307–359.

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). London: RoutledgeFalmer.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., et al. (2012). Defining twenty-first century skills. In P. Griffin, & E. Care (Eds.). *Assessment and teaching of 21st century skills. Methods and approach* (pp. 17–66). Dordrecht: Springer Netherlands.

Blatchford, P., Baines, E., Rubie-Davies, C., Bassett, P., & Chowne, A. (2006). The effect of a new approach to group work on pupil-pupil and teacher-pupil interactions. *Journal of Educational Psychology, 98*(4), 750–765.

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.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1–35.

COTAN (2007). *Taaltoets Alle Kinderen [language test all children], TAK, 2006. CED-groep.* Retrieved October 3, 2019 from https://www.cotandocumentatie.nl/beoordelingen/b/13480/taaltoets-alle-kinderen/.

COTAN (2008). *Sociale competentie observatie lijst [social competence observation list], SCOL, 2006. CED-groep.* Retrieved October 3, 2019 from https://www.cotandocumentatie.nl/beoordelingen/b/13591/sociale-competentie-observatie-lijst/.

Dienst Uitvoering Onderwijs. (2015). *Leerlingen bo naar gewicht en leeftijd 2015-2016 [Primary school students by social advantage and age 2015-2016].* Retrieved October 2, 2019 from URL: 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].* Retrieved October 2, 2019 from URL: https://duo.nl/open_onderwijsdata/databestanden/po/leerlingen-po/bo-sbo/po-owsoort-leeftijd.jsp.

Galton, M., & Hargreaves, L. (2009). Group work: Still a neglected art? *Cambridge Journal of Education, 39*(1), 1–6.

Gillies, R. M. (2003). Structuring cooperative group work in classrooms. *International Journal of Educational Research, 39*(1–2), 35–49.
Gillies, R. M. (2004). The effects of communication training on teachers' and students' verbal behaviours during cooperative learning. *International Journal of Educational Research, 41*(3), 257–279.
Gillies, R. M. (2014). Cooperative learning: Developments in research. *International Journal of Educational Psychology, 3*(2), 125–140.
Gillies, R. M., & Boyle, M. (2010). Teachers' reflections on cooperative learning: Issues of implementation. *Teaching and Teacher Education, 26*(4), 933–940.
Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory Into Practice, 38*(2), 67–73.
Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365–379.
Joosten, F. (2006). *Verantwoording van de Sociale competentie observatie lijst (SCOL) [justification of the social competence observation list (SCOL)].* Rotterdam: CED-groep.
Kagan, S. (1989). The structural approach to cooperative learning. *Educational Leadership, 47*(4), 12–15.
Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*(4), 681.
Kutnick, P., & Berdondini, L. (2009). Can the enhancement of group working in classrooms provide a basis for effective communication in support of school-based cognitive achievement in classrooms of young learners? *Cambridge Journal of Education, 39*(1), 71–94.
Kutnick, P., Blatchford, P., & Baines, E. (2002). Pupil groupings in primary school classrooms: Sites for learning and social pedagogy? *British Educational Research Journal, 28*(2), 187–206.
Kutnick, P., Ota, C., & Berdondini, L. (2008). Improving the effects of group working in classrooms with young school-aged children: Facilitating attainment, interaction and classroom activity. *Learning and Instruction, 18*(1), 83–95.
Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review, 10*(2), 133–149.
Landis, J. R., & Koch, G. G. (1977). *The measurement of observer agreement for categorical data.* Biometrics.
Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*(6), 822–832.
Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction, 6*(4), 359–377.
Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach.* London: Routledge.
Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal, 25*(1), 95–111.
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*New York: MDRC.
Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., & Healy, M. (2000). *A user's guide to MLwiN.* London: Centre for Multilevel Modeling.
Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin, 134*(2), 223.
Sharan, S., Ackerman, Z., & Hertz-Lazarowitz, R. (1979). Academic achievement of elementary school children in small-group versus whole-class instruction. *The Journal of Experimental Education, 48*(2), 125–129.
Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin, 94*(3), 429–445.
Slavin, R., Hurley, E. A., & Chamberlain, A. (2003). Cooperative learning and achievement: Theory and research. In W. M. Reynolds, & G. E. Miller (Vol. Eds.), *Handbook of psychology: Educational psychology: Vol. 7*, (pp. 177–198). New York: Wiley.
Slavin, R. E., Madden, N. A., Chambers, B., & Haxby, B. (2009). *2 million children: Success for All.* Thousand Oaks, California: Corwin Press.
Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.
Spybrook, J., Bloom, H., Congdon, R., Hill, C., Martinez, A., & Raudenbush, S. (2011). *Optimal design plus empirical evidence: Documentation for the "Optimal Design" software.* Ann Arbor: Western Michigan University.
Tolmie, A. K., Topping, K. J., Christie, D., Donaldson, C., Howe, C., Jessiman, E., ... Thurston, A. (2010). Social effects of collaborative learning in primary schools. *Learning and Instruction, 20*(3), 177–191.
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.
Veenman, S., Kenter, B., & Post, K. (2000). Cooperative learning in Dutch primary classrooms. *Educational Studies, 26*(3), 281–302.
Verhoeven, L., & Vermeer, A. (2006). *Verantwoording Taaltoets Alle Kinderen (TAK) [justification of the language test for all children].* Arnhem: Cito.
van der Werff, E., & Kloosterman, K. (2016). *Minima in Groningen [Low-income households in Groningen].* Groningen: Onderzoek en Statistiek Groningen.