Educational Research and Evaluation
2004, Vol. 10, No. 3, pp. 205–237

Effects of a Cooperative Learning Program on the Elaborations of Students Working in Dyads

Karen Krol[1], Jeroen Janssen[1], Simon Veenman[1], and Jos van der Linden[2,†]
[1]University of Nijmegen, The Netherlands, and [2]Utrecht University, The Netherlands

ABSTRACT

In this study, the effects of a school improvement program on cooperative learning (CL) with respect to the elaborations of 6th-grade students working in mixed-ability and mixed-sex dyads on 2 cooperative tasks were examined. A posttest-only design with a control group was used to investigate the provision and receipt of elaborations within the dyads and the performance of the dyads working on cooperative mathematics and language tasks. Treatment dyads were found to exchange significantly more high-level elaborations during the language task than the control dyads. The treatment dyads also tended to produce higher performance scores on the 2 tasks.

INTRODUCTION

During the last decades, cooperative learning (CL) has gained ground as an instructional strategy to promote learning and achievement. A large body of research clearly shows the use of CL to enhance student performance and social development (Cohen, 1994; Johnson & Johnson, 1999; Slavin, 1996). Given these positive effects, the interest in the use of CL methods within the classroom has also grown. CL clearly fits with current conceptions of learning

†Unfortunately Jos van der Linden died in February 2003. We deeply regret the loss of our co-author.

Address correspondence to: Karen Krol, Department of Educational Sciences, University of Nijmegen, Spinoza-Hall 4.28, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands. Tel.: +31 24 3612501. Fax: +31 24 3615978. E-mail: k.krol@ped.kun.nl

Manuscript submitted: September 27, 2002.
Accepted for publication: March 26, 2003.

as a social, cultural, and interpersonal constructive process governed as much by social and situational factors as by cognitive factors (Salomon & Perkins, 1998; Shuell, 1996). In addition, our current knowledge-driven society calls for competent people capable of working together in teams to solve complex problems.

In the present study, CL is defined as students working together to achieve common goals that are thus important to not only the individual group member but also the other members of the group. Students see that they can only reach their learning goals when all of the other students in the learning group attain their goals (Johnson & Johnson, 1999). As a consequence, the students in a CL situation are responsible for both their own learning and helping others learn as well. In light of the above, structuring cooperation within the classroom is more than just seating a number of students together and telling them to help each other complete a task. Johnson and Johnson (1999) have argued that, in order for a lesson to be cooperative, five essential elements must be included: (a) positive interdependence (i.e., the perception of group members that they must work together to accomplish a common goal), (b) individual accountability (i.e., group members being held responsible for their contributions to the achievement of the goal), (c) face-to-face interaction (i.e., group members meeting face to face to promote one another's work), (d) the development of social or small-group skills (i.e., acquisition of the interpersonal skills needed to work in a group), and (e) group processing (i.e., group reflection on the collaborative effort and decisions to improve effectiveness). The structuring of positive interdependence and individual accountability can be considered the most important elements related to the effectiveness of CL (Slavin, 1995).

Notwithstanding the substantial amount of research demonstrating the positive effects of CL, the instructional strategy is not commonly used in Dutch education (Veenman, Kenter, & Post, 2000). In Dutch elementary school classrooms, the emphasis is still on whole-class instruction and individualistic learning. That is, with traditional whole-class instruction, the children learn mainly individually and the exchange of ideas, information, and opinions does not occur very frequently. Teachers are also not familiar with CL as a manner of structuring the interactions within the classroom. This is partly due to a lack of explicit teacher training but also to teachers not being accustomed to the use of instructional strategies that give students greater responsibility for their own learning. To familiarize teachers with the potential of CL as an instructional strategy, a Dutch school improvement program for

CL was developed.[1] The program has its theoretical roots in the literatures on CL, school improvement, staff development, the professional development of teachers, and peer learning. The 1st-year evaluation results show teachers to be capable of implementing the desired CL instructional behaviors, such as the structuring of positive interdependence and individual accountability, the promotion of social skills, and the evaluation of group processes (Krol, Veenman, & Voeten, 2002). These results are encouraging because it is important that teachers are able to create a context in which students can cooperate. However, these results only concern the behavior of the teachers. To determine whether CL is effective requires that the students engage in productive interactions. By productive interactions we mean interactions that contribute to the successful performance on cooperative tasks. In other words, one can only speak of effective school improvement when positive results are observed at not only the level of the teacher but also the level of the student. The aim of the present study is thus to examine the interactions of students working in dyads and to determine whether dyads from the classrooms of teachers who participated in the school improvement program on CL perform better on a cooperative task than dyads from the classrooms of teachers who did not participate in the program.

PEER LEARNING

Although the basic elements of CL as described in the introduction above are necessary to structure the interactions between students, such interactions do not automatically lead to higher levels of achievement. Research has shown the quality of the interactions between those students working together to largely determine the outcomes of the CL process (Kneser & Ploetzner, 2001; Mercer, 1995; Webb & Farivar, 1999; Webb, Troper, & Fall, 1995).

There are several ways in which students can learn from each other within a CL setting. The sociocultural perspective of Vygotsky (1978) and the sociocognitive conflict perspective of Piaget (1926) constitute the basis for most studies of peer learning. From the sociocultural perspective inspired by

[1]The program was developed by the Department of Educational Sciences at the University of Nijmegen in collaboration with the national Christian Pedagogical Study Center, the regional Educational Service Centers for Arnhem and Nijmegen, and the Educational Faculty at the Teacher Education College Arnhem and Nijmegen.

the work of Vygotsky (1978), learning is construed as socially constructed during interactions and activities with others. Individuals exchange ideas, information, perspectives, attitudes, and opinions; students can model their patterns of reasoning, thinking strategies, and problem-solving skills on those of their more skilled peers; and as a result, students in CL settings can internalize knowledge, meanings, and skills from each other and cooperatively build new knowledge and insights. Vygotsky (1978) used the concept "zone of proximal development" for the difference between what a student can do on his or her own and what he or she can do in cooperation with a more knowledgeable other. The zone of proximal development is created in the course of social interaction (Hogan & Tudge, 1999). And, as part of this social interaction, the more competent person must be capable of comprehending the thoughts of others and communicating comprehensibly.

From the developmental perspective of Piaget (1926), sociocognitive conflict provides an important impulse. Students can experience a conflict between their own ideas and the ideas of others and, in order to resolve this cognitive conflict, the students must explain their viewpoints to each other. That is, students can learn by not only hearing the explanations of others but also providing their own explanations (De Lisi & Golbeck, 1999). For both Vygotsky (1978) and Piaget (1926), explanation is an important element. Students can learn from the explanations of more capable peers and peers with different ideas or perspectives but also from their own explanations, which require them to explicate and sometimes restructure their own ideas. The provision of explanations is thus an important part of the interaction between more and less capable peers and the resolution of cognitive conflicts.

When students work together, it is important that they verbalize their thoughts. Such verbalization elicits elaborative cognitive processes that typically produce reflection, awareness, (re)organization, differentiation, fine-tuning, and the expansion of knowledge (Van Boxtel, 2000). Teasley (1995) investigated the role of verbalization in children's peer cooperation by studying collaboration and talk as independent variables. Dyads produced higher performance outcomes than students working alone, but this effect was not due to the cooperation per se but the cooperation increasing the likelihood of engagement in the types of talk that support learning. When working with a partner, forms of elaborated talk are more prevalent than when working alone (and talking aloud to oneself).

Other research on helping behavior in small group work shows students to clearly learn more from the provision of elaborated help to others and less from the receipt of low-level elaboration from others. In the studies by Webb and Farivar (1994, 1999), students only learned more from the provision than the receipt of help when the explanations they provided contained some form of elaboration. This is presumably because the provider of elaborated help often reorganizes or clarifies the material and such cognitive restructuring may help the explainer him/herself understand the material more adequately. The explainer may also discover gaps in his or her knowledge and notice that his or her own knowledge does not always match the knowledge of others. By acknowledging this and attempting to neutralize any differences, the explainer can develop new perspectives and construct new knowledge (Fuchs, Fuchs, Bentz, Phillips, & Hamlett, 1994; King, Staffieri, & Adelgais, 1998).

In addition to the provision of help, Webb and Farivar (1994, 1999) also studied the receipt of help and found explanations *in response to requests for help* to benefit the receiver the most. Students were also found to often benefit most from explanations provided by their peers, who are often more aware of what students do not understand than teachers, clearly focus on the main aspect of the problem, and typically provide explanations that are easy to understand. Other students can also sometimes attune their assistance in the zone of proximal development more finely than teachers. The student receiving the help may play a more active role in this process and in applying the help received, he or she can check whether or not he or she understands the help provided. Finally, research has shown that the receipt of non-elaborated help (e.g., the direct answer to a problem) does not correlate strongly with achievement, presumably because such information does not call for cognitive restructuring.

King (1999) has argued that when teacher guidance is lacking, students working in small groups generally interact at a very basic level and appear to be focused on finding the right answers. A correlation has been found between the types of questions asked by students and the nature of the answers that they receive with higher order questions leading to higher level answers (King et al., 1998). Asking thought-provoking questions promotes high-level discussion, which has been found to result in high-level learning; that is, the question triggers an elaborated explanation, which can positively influence the performance of both the student providing the help and the student receiving the help (King, 1999).

The present study borrows from the Piagetian and Vygotskian perspectives on development stressing the importance of social interaction for learning based on the active reconciliation of different perspectives within the cooperative group and the importance of studying learning as a social process. From the cognitive elaboration perspectives of Webb and Farivar (1999) and King (1999), it borrows the view that students in small groups have to verbalize their inner thoughts and learn more by providing elaborated help and learn less by receiving less elaborated help.

RESEARCH QUESTIONS

The school improvement program on CL studied here was aimed at guiding teachers in the development of the instructional behaviors needed to conduct a cooperative lesson. Through participation in 10 workshops and coaching during the 1st year of implementation, the target teachers were helped to implement the essential features of CL as defined by Johnson and Johnson (1999). The school improvement program was expected to not only affect the teachers' instructional behaviors but also the nature of the students' participation in small group work. When the teachers integrate CL activities into their lessons, they are assumed to encourage positive interactions among students via the promotion of positive interdependence, individual accountability, face-to-face interaction, the development of social or small-group skills, and group processing. With the teacher as mediator, we thus expected the CL school improvement program to produce more productive interactions among students working in small groups.

The following research questions were considered:

1. Do the dyads from the classrooms of teachers who participated in the CL school improvement program provide and receive more elaboration when working on a cooperative task than the dyads from the classrooms of teachers who did not participate in the program?
2. Do the dyads from the classrooms of teachers who participated in the CL school improvement program perform better on a cooperative task than the dyads from the classrooms of teachers who did not participate in the program?
3. How do the students of those teachers who participated in the CL school improvement program perceive working on a cooperative task when compared to the students of those teachers who did not participate in the program?

METHOD AND INSTRUMENTATION

Participants

Subjects were 40 sixth-grade students from seven elementary schools involved in the school improvement program on CL in the east and south of The Netherlands. All of the subjects' parents had consented to their participation. During the recruitment phase of the study, the selected schools agreed to actively participate in the school improvement program on CL and to implement CL activities in their classrooms. Prior to the study, the schools and students had had little or no experience with working in cooperative groups. The schools and classes were comparable with regard to location, school size, school enrollment, and an interest in CL methods. Four of the schools agreed to start with the program in the fall of 1999 and constituted the treatment group. The other three schools agreed to start with the program in the fall of 2001 and constituted the control group, which was not exposed to the training of CL methods during the course of the study.

Prior to the pairing of the students, the teachers were asked to divide the students in their classes into three ability levels for mathematics and language arts: low, medium, and high. This list was then compared to the national achievement test scores (Cito) for the mathematics and reading comprehension of the students after administration of the test in the spring of the same school year. The teachers' judgments corresponded to the national achievement test scores for the students. Within each school, six students from grade 6 were next selected to make three dyads (with the exception of one control school where only four students were selected to make two dyads). Based on the students' ability levels, two different groups of dyads were formed. In the first group, a low-ability student was paired with a medium-ability student; in the second group, a medium-ability student was paired with a high-ability student. This pairing was based on the assumption that the ability levels of the students should be different in order to generate help-seeking and helping behaviors but not too different in order to still make it possible for the students to work in their "zone of proximal development." Based on the findings of a study by Webb (1984), who found the achievement and interaction patterns for boys and girls to be identical when working in mixed-sex groups, each dyad consisted of one girl and one boy. The results of the pairing made by the researchers were subsequently checked by the teachers to exclude dyads of students who could not get along with each other. Based on this check, three dyads were omitted and replaced by newly

Table 1. Distribution of 20 Dyads According to Ability Level.

Ability level	Treatment group		Control group	
	Math task	Language task	Math task	Language task
Level 1: low-medium dyad	6	6	5	5
Level 2: medium-high dyad	6	6	3	3

formed dyads. The distribution of the dyads according to ability level is shown in Table 1.

Analyses of variance were used to confirm the differences between the three levels of ability as measured by the nationally standardized Cito test administered in the sixth grade. The differences in the ability levels for mathematics and language indeed proved statistically significant: $F(2, 37) = 47.4$ and $F(2, 37) = 28.3$, $p < .01$, respectively. The mean scores, which could range from 1 (= highest) to 5 (= lowest), were as follows: Mathematics high-ability students ($n = 9$) $M = 1.33$ ($SD = .50$), medium-ability ($n = 20$) $M = 2.75$ ($SD = .64$), and low-ability ($n = 11$) $M = 4.00$ ($SD = .63$); language high-ability students ($n = 9$) $M = 1.33$ ($SD = .50$), medium-ability ($n = 20$) $M = 2.95$ ($SD = .83$), and low-ability ($n = 11$) $M = 3.73$ ($SD = .65$). The pairing procedure thus resulted in dyads of students with clearly different ability levels.

An independent samples *t* test was conducted to identify any initial differences in the mathematics and reading comprehension test scores for the treatment versus control groups. For the mathematics test, the mean score for the treatment group was 2.63 ($SD = 1.10$) and the mean score for the control group was 3.00 ($SD = 1.16$). For the reading comprehension test, the mean score for the treatment group was 2.58 ($SD = 1.18$) and the mean score for the control group was 3.13 ($SD = .96$). These differences were not statistically significant (mathematics: $t(38) = -1.04$, $p = .31$; reading comprehension, $t(38) = -1.53$, $p = .13$, respectively).

Design

A posttest-only design with a treatment group and a control group was used. Both groups were tested after the treatment group teachers had participated in 8 of the 10 workshops constituting part of the CL school improvement program. Given that the school improvement program was already in progress at the start of the present study, it was not possible to conduct a pretest. Given the intensity of the school improvement program, it was also assumed that a pretest would be too much of a burden on the teachers. At the start of the program, the teachers were observed in their classrooms, achievement tests were administered, and extensive questionnaires were administered to both the teachers and the students in order to examine their attitudes towards CL. Some of the results of the school improvement program are published elsewhere (Krol et al., 2002).

Procedures

All of the dyads were asked to solve a mathematics task and a language task. One of the two researchers brought two students from their classrooms into a room where the materials were already set up. On the first morning, half of the dyads solved the mathematics task, and half the language task. On the second morning, the remaining mathematics or language task was completed. The order in which the dyads had to solve the tasks was randomized in order to control for a possible sequence effect. At the start of the session, one of the researchers provided a brief description of the task, how the answers should be recorded, and explicitly stated that the students should cooperate on the task, which had to be solved within 30 min. The instructions were the same for all of the dyads. All of the sessions were video and audio recorded. The researchers also took notes on the behavior of the students during the task. After completion of the task, the students were administered three short questionnaires in order to assess their individual perceptions of the two tasks and the manner in which he or she worked together with the other student. All of the sessions for the 20 dyads were transcribed.

Tasks and Materials

Both of the tasks required formal reasoning and discussion, were developed to be challenging for sixth graders, and did not include topics discussed previously in the classroom. During the development phase, a sample of three dyads from schools not involved in the study (seventh and eighth graders) provided feedback on the adequacy of the materials. On the basis of this pilot work, the wording for the two tasks was slightly revised.

Mathematics Task

In order to promote logical reasoning, a balance beam task was used. In the balance beam task, students must predict which side of the beam will go up or down when various configurations of weights and distances are set up. This task has been used successfully in earlier experiments by Siegler (1976), Phelps and Damon (1989), and Tudge (1992). In contrast to the work by these authors, the balance beam used in our study was not manipulative (i.e., did not involve a real balance beam with removable pegs on each side of the fulcrum or a capacity to really tip to the left, the right, or remain balanced). Our study made use of a paper-and-pencil task with drawings of a balance beam involving different configurations of weights and distances from the fulcrum (cf. Ros, 1994).

Students were given a 15-page booklet and worksheets with 15 problems to solve. The first five worksheets pictured a scale with weights on it, and the students had to indicate whether the scale was balanced or which side would go up or down. The first five problems were the simplest involving basic weight and distance problems. Equal weights at an equal distance (balance), unequal weights at an equal distance (greater weight), and equal weights at an unequal distance (greater distance) proved fairly easy for most of the sixth graders. For the first five problems, feedback on the solution to the problem was provided in the booklet for the students to then compare their solution to the one in the booklet. The first five problems were intended to highlight the importance of different weights and different distances from the fulcrum and thereby familiarize the students with how to work with a balance beam. After completion of these problems, the students were asked to cooperatively solve 10 more problems with the weights and pegs varied in a more complicated manner. The solutions for the last 10 problems required formal reasoning (Phelps & Damon, 1989). In order to share the materials and work cooperatively, each dyad received one booklet and one worksheet with the problems to be solved. After discussion of a problem, one of the students wrote the proposed solution down.

Language Task

In order to promote collaborative reasoning, a reading comprehension task was developed based on the story about some children sailing around an unknown pirate island to discover its characteristics. Two texts were written. In the first, a boy and a girl sailed clockwise around the island; in the second, a boy and a girl sailed counter-clockwise around the island. Each boat thus approached the island from a different perspective. The texts the students received contained different information on the island, so the students had to share information in order to obtain a clear picture of the island. Looking at each other's text was not allowed in order to encourage the verbal exchange of information and discussion. The texts were similar with regard to length and difficulty. The development of the language task was based on the principles of reciprocal teaching, which is a method used to teach reading comprehension skills (Brown & Palincsar, 1989). Students are taught to formulate questions, clarify unclear passages of text, summarize the essential elements of a passage, and predict the ending of a story.

The task consisted of two assignments and seven comprehension problems. The assignments involved reading the text individually and exchanging information about the main points in the text. The purpose of the assignments was to allow the students in the dyads to discover that the two texts contained similar information about the island but also unique information. The students also had to discover that the boy and girl in the two texts sailed in different directions around the island. The first comprehension problem required the students to find those passages in which different information was provided about the island, to discuss these differences, and to write the differences down. The other comprehension problems involved: the explanation of the meaning of a complex word mentioned in one of the texts and explained in the other text; placement of the right information on the map; invention of a suitable name for places on the island after discussion of the descriptions of those places; finding the route for a specific place to organize a party for the participants in the sail camp; and prediction of how the story ends.

Feedback was only provided in the booklet after the second assignment and the first comprehension problem, in order to reassure the students that the two texts indeed differed at some points from each other and that the boy and girl indeed sailed clockwise versus counter-clockwise around the island. This information was needed to solve the remainder of the problems.

Verbal Interaction Categories

The Coding Scheme

The framework used in the present study rests on the assumption that learning can be described in terms of individual cognitive activities and in terms of social processes (Salomon & Perkins, 1998; Shuell, 1996). The methodological framework is inspired by the work of Webb and Farivar (1999), King (1999), Mercer, Wegerif, and Dawes (1999), and Kumpulainen and Mutanen (1999). Three analytic dimensions can be distinguished, namely the cognitive, the affective, and a regulative dimension (Veldhuis-Diermanse, 2002;

Vermunt, 1992). The cognitive dimension refers to the manner in which a student approaches and processes the learning units. Given that the success of working together may depend on affective elements in addition to cognitive elements, the coding scheme also included positive and negative affective elements, which refer to whether the students speak positively or negatively about each other and their respective contributions to a learning task. The regulative dimension refers to metacognitive statements intended to help regulate the necessary cognitive activities, such as the planning of the execution of the learning task, monitoring of learning progress, and the diagnosis of difficulties. Those verbal interactions that did not reflect one of these dimensions were coded as "non-task-related remarks." Although we agree that the affective and regulative dimensions of working cooperatively are also important, the emphasis in the present study is on the cognitive dimension.

The *cognitive dimension* contained 14 verbal interaction categories divided as follows: three categories pertaining to the posing of questions (i.e., factual questions, comprehension questions asking for elaboration, and questions asking for verification), three categories pertaining to the provision of help during the interactions (i.e., answers only, explanations with procedural elaboration – information on how to do something – and explanations with argumentative elaboration); two categories pertaining to the input of new ideas (i.e., presentation of new ideas without elaboration and presentation of new ideas with elaboration); two categories pertaining to references to previously discussed ideas (i.e., elaboration of previously discussed ideas and evaluation of ideas without further elaboration); and four categories pertaining to accepting or rejecting ideas (i.e., acceptance without further elaboration, acceptance with further elaboration, rejection without further elaboration, and rejection with further elaboration). The *affective dimension* contained two categories pertaining to the process of cooperation (i.e., positive vs. negative emotional reaction). The *regulative dimension* contained three categories pertaining to the execution of the learning task (i.e., the planning of the task; evaluation of the group process; and instructing the other student). In all, our coding scheme contained 19 verbal interaction categories.

Unit of Analysis

In order to code the verbal interactions of the dyads, the verbal interaction was first divided into conversational turns defined as a change in speaker. A single turn sometimes contains more than one utterance, and the utterance was the basic unit of analysis. An utterance is distinguished from another utterance via a ''perceptible pause,'' comma, or period and has a singular communicative function (Van Boxtel, Van der Linden, & Kanselaar, 2000). An utterance can vary in length from a single word (''No'') to an extended monologue and each utterance was assigned to one of the categories within the present coding scheme.

The scores for a given student were the number of utterances falling into the verbal categories pertaining to the different dimensions of the verbal interaction coding scheme. The unit of analysis for all of the subsequent analyses was the dyad. This unit of analysis was adopted because the knowledge building that occurs during dyadic interactions can be viewed as largely interdependent; that is, the questions and responses of one partner are, to a great extent, elicited or stimulated by the questions and statements of the other partner (King et al., 1998).

Coding of the Transcripts

Prior to the coding of the transcripts of the videotapes and audiotapes, two researchers went through a training program of about 40 hr. The training program involved the formulation of rules for coding, learning to apply the computer program Multiple Episode Protocol Analysis (MEPA) developed by Erkens (2001) to code transcribed verbal interactions, and the coding of three transcripts from a sample of three dyads from schools not involved in the study. The interrater agreement was based on nine transcripts randomly selected from the treatment and control groups (23% of all transcripts). The percentage agreement was found to be 94%. The Cohen's Kappa was .92. Each transcript was coded in its entirety by one of the two trained coders. The transcripts were randomly assigned to these coders.

Data Analysis

Three levels of elaboration were distinguished to assess the students' elaborations on the problems: high-level elaboration, medium-level elaboration, and low-level elaboration. This classification is based on the work of Webb, Nemer, Chizhik, and Sugrue (1998). *High-level elaboration* included seven categories from the cognitive dimension in our study, namely: comprehension questions asking for elaboration, explanations with procedural elaboration, explanations with argumentative elaboration, presentation of new ideas with further elaboration, elaboration of previously discussed ideas, acceptance with further elaboration, and rejection with further elaboration.

Medium-level elaboration also included seven categories from the cognitive dimension: factual questions, verification questions, answers only, presentation of new ideas without further elaboration, evaluation of ideas without further elaboration, acceptance without further elaboration, and rejection without further elaboration. In the studies by Webb et al. (1998), *low-level elaboration* was defined as listening or watching without making any substantive verbal contribution or inquiry. Given that our coding scheme did not include categories referring to listening or "no response," low-level elaboration was not included in the present study. The affective and regulative dimensions of interaction were also not considered in the present analysis of the levels of elaboration.

For the analyses of the levels of elaboration, the unit of analysis was the dyad. For each dyad, the frequencies of high-level and medium-level elaboration were calculated by summing the relevant codes. Subsequently, the percentages high-level and medium-level elaboration were calculated for each dyad.

Performance Scores

The items from the mathematics and language tasks were used to assess the performances of the dyads. For mathematics, the score was the sum of the points awarded for each correct answer on the balance beam task. The first 5 correct answers were assigned 4 points and the last 10 correct answers 3 points, which produced a maximum score of 50 points.

For language, a different procedure was followed because most of the items in this task did not require a correct answer but justifiable arguments. Each adequate or justifiable answer was scored along a 10-point scale. For complex problems (such as filling in the map of the island), the dyads could earn 10 points; for less complex problems (such as summarizing the contents of the text), they could earn 3 points. The scores on this task thus varied depending on the difficulty of the question, the adequateness of the answer, and the elaborateness of the answer or arguments provided. A maximum score of 50 could be obtained. To check the objectivity of the scoring, two researchers independently scored 10 randomly selected completed language tasks (i.e., 50% of the total number of tasks). The percentage agreement was found to be 92%. The Cohen's Kappa reached .87. For the performance variables, the unit of analysis was also the dyad because the problems in the mathematics and language tasks were solved jointly and the solution written down on a single worksheet.

Measurement of Student Evaluations

To gain insight into the perceptions and attitudes of the students towards the learning tasks, the experience of working together, and their willingness to work together, three short Likert-scale questionnaires were administered directly after completion of the mathematics and language tasks. The first questionnaire examined the students' perceptions of the mathematics and language tasks. The six items composing this questionnaire were rated along a 5-point scale and addressed the difficulty and attractiveness of the task along with the time needed to complete the task. Some of the items were based on the work of Dale (1994) and Meloth and Deering (1994). Given that the items did not form a homogeneous scale, the results will be presented per item.

The second questionnaire was administered to explore the students' perceptions of the manner in which they worked together to solve the problems in the mathematics and language tasks. Seven items concerning the manner in which the students reached mutual agreement, listened to each other, and helped each other were rated along a 5-point scale. The Cronbach's reliability coefficient for the mathematics task was .60 and for the language task .65. In order to obtain a score for the scale "quality of cooperation," the ratings for the seven items were averaged.

The third questionnaire was administered to examine the willingness of the students to work together in small groups in the future. The items addressed the willingness of a student to work together with all students and were rated along a 4-point scale. The Cronbach's reliability coefficient was .80. A higher score on these three scales reflected more positive perceptions and attitudes. For the student perception and attitude data, the unit of analysis was the individual student as each student individually filled out the short questionnaires.

Treatment

Rationale

Studies of the adoption and dissemination of CL methods in elementary and middle schools show the implementation of such methods within the classroom to be a complex and difficult process. Successful implementation of CL methods largely depends on the teacher really understanding what CL is and a capacity to apply CL methods appropriately and with insight.

In our school improvement program, CL is presented as a philosophical and practical approach for changing classroom processes to provide students with more active learning experiences and thereby create a more supportive social environment for students and teachers (cf. Stevens & Slavin, 1995). The main elements of the school improvement program are briefly described below (for more details see Krol et al., 2002).

The aim of the school improvement program is to promote the use of CL in a constructive, appropriate, and integrated manner. In order to enable teachers to master the conceptual framework and actual procedures, they are trained for a period of 2 consecutive years. The teachers also receive support in the form of coaching. Extended support is important for long-term maintenance of the use of CL and to institutionalize CL as a standard instructional practice within the school. For this reason, expert coaching was undertaken to assist teachers in the application of what they learned in the workshops within their classrooms during the 1st year of the program. Coaching as a form of in-class support can also help teachers improve their instructional effectiveness by providing them with feedback and stimulating them to be more reflective. More generally, coaching can help teachers implement CL via the provision of technical support, assistance, and companionship; it can also promote executive control via reflective feedback and discussion (Joyce & Showers, 1995). During the 2 years of the program, the implementation process was therefore supported by a school-based change team composed of the school principal, the vice-principal, and an expert teacher demonstrating clear enthusiasm for the use of CL in his or her classroom. This change team shared responsibility for further implementation of CL within the school by organizing and facilitating the functioning of teaching teams and providing in-class help and support in the form of peer coaching, for instance.

Staff Development Sessions

During the 1st year of implementation, the teachers received six half-day training sessions (i.e., workshops) on the fundamentals of CL, followed by four half-day training sessions during the 2nd year. The training sessions were distributed throughout the school year and addressed the following topics: the nature of CL, the teacher's role in CL, the basic elements needed for CL, research supporting the use of CL, the assessment and evaluation of group work, and effective interaction patterns for CL groups. Each training session was structured as follows: opening, review of the main topics from previous workshops, team-building activity, exchange of experiences related to the use of CL methods within the classroom, presentation of new CL materials, review and discussion of the CL methods to be used in the workshop, discussion of the application of the newly learned cooperative methods within the classroom, and conclusion. During the 2nd year of implementation, the teachers received four half-day training sessions on integrating CL methods with direct instruction, establishing productive interactions in the cooperative work group, reciprocal teaching, paired reading, and supervision. In addition, a special workshop was arranged for the teachers of the lowest grades of elementary school (K-2) to discuss the use of CL methods with kindergartners and young children.

During the workshops, the teachers worked together in heterogeneous cooperative groups using several of the CL structures described by Kagan (1994). After explanation of the rationale behind a particular CL structure and the various steps involved, the teachers are asked to apply the relevant structure in order to directly experience its practical value. CL was introduced during the first workshop and used as the only instructional strategy thereafter. Peer communication and learning were also attended to as CL is clearly mediated by the quality of the interactions within the group. One workshop was specifically devoted to the provision and receipt of help. This special workshop was based on the studies of Webb and Farivar (1994, 1999). In addition, the interactions within the CL groups during the workshops are discussed using the distinctions made by Mercer (1995); that is, in terms of disputational talk, cumulative talk, and exploratory talk.

Each workshop lasted 3 hr. Following each workshop, the teachers were asked to put what they had learned into actual practice. There was also an opportunity to discuss the CL classroom experiences of the teachers during each workshop, and background information was provided on the topics considered in the workshop in the form of a manual distributed to the teachers after each workshop.

RESULTS

Quantitative Analysis of Verbal Interactions

The first research question was whether the dyads from the classrooms of teachers who participated in the CL school improvement program (treatment dyads) provided and received more elaborations while working on the mathematics and language tasks than the dyads of the teachers who did not participate in the program (control dyads). Inspection of the frequency distributions and box-plot scores showed one control dyad to score more than two standard deviations above the control group mean while working on the language task and to therefore constitute an outlier. The results for this dyad, which consisted of a medium- and a high-ability student working together, were therefore omitted from any further analyses of the levels of elaboration for the language task. For the sake of completeness, the results are presented *with* and *without* this outlier.

In Table 2, the mean percentages for the high and medium levels of elaboration, as well as for the affective and regulative categories from the coding scheme, are presented. These percentages were calculated for each dyad by dividing the dyad's score for the relevant categories by the total number of utterances. Note that the percentages do not add up to 100 because some of the utterances (e.g., non-task-related remarks or reading aloud) do not fall into one of the coding categories.

The differences between the treatment and control dyads were examined using *t* tests for independent samples. A significance level of 5% was used in

Table 2. Interaction Differences Between Treatment and Control Dyads While Working on the Mathematics and Language Tasks.

	Treatment dyads ($N = 12$)		Control dyads ($N = 8$)				
Task	*M*	*SD*	*M*	*SD*	*t*	*p*	*ES*
Mathematics							
Cognitive utterances							
High-level elaboration	14.07	3.86	14.24	3.22	– .04	.48	– .02
Medium-level elaboration	59.55	8.82	65.69	9.36	– 1.49	.08	– .68
Affective utterances	2.10	1.26	1.29	1.15	1.45	.16	.66
Regulative utterances	6.54	2.84	5.33	3.43	.86	.20	.39
Language							
Cognitive utterances							
High-level elaboration	6.26	4.07	2.81[a]	1.96	2.09	.03*	.70
Medium-level elaboration	65.01	4.59	67.71	10.25	– .70	.25	– .37
Affective utterances[b]	.84	1.00	.28	.43	– 1.48	.07	.68
Regulative utterances[b]	7.48	2.90	8.73	4.57	– .39	.35	– .34

Note. [a]$n = 7$.
[b]These variables did not meet the assumption of a normal distribution; the nonparametric Mann-Whitney U test was therefore used. The reported values are the corresponding *Z*-values.
$^{*}p < .05$.

all of the statistical tests (one-tailed). Effect sizes (*ES*s) were calculated by dividing the difference in the mean scores for the treatment versus control dyads by the pooled standard deviations for the scores of the treatment versus control dyads at posttest. For a posttest-only design with treatment and control groups, the *ES* is defined as the normalized difference between a trained and an untrained comparison group (Carlson & Schmidt, 1999).

The findings in Table 2 show about 75% of the utterances of the dyads while working cooperatively on the mathematics and language tasks to be cognitive statements (i.e., a combination of high-level and medium-level elaborations). Affective and regulative utterances occurred much less frequently (i.e., 6% and 8%, respectively). Most of the cognitive utterances for the mathematics and language tasks were characterized as medium-level elaborations (62%); high-level elaborations occurred less frequently (14%).

The data displayed in Table 2 also show some differences in the verbal interactions of the treatment versus control dyads. While working on the language task, the treatment dyads exchanged significantly more high-level elaborations than the control dyads ($p < .05$). The relatively high effect size for this cognitive dimension ($ES = .70$) was in favor of the treatment dyads. However, with the *aforementioned outlier* included in the analysis, the difference between treatment versus control dyads was not significant, $M_{control} = 3.98$, $SD = 3.77$, $t = 1.26$, $p = .11$, $ES = .58$. The effect size for the affective dimension was also in favor of the treatment dyads ($ES = .68$). For the regulative dimension, a difference between the two groups was found in favor of the control dyads ($ES = -.34$). For the mathematics task, no statistically significant differences were found between the treatment versus control groups with respect to the exchange of high-level elaborations. No statistical significant differences were found for the affective or regulative dimensions. The effect sizes for the affective and regulative dimensions while working on the mathematics task nevertheless showed moderate effects in favor of the treatment dyads (*ES*s = .66 and .39, respectively).

Additional analyses revealed significant differences for the language task between Level-1 ability dyads (a low-ability student combined with a medium-ability student) and Level-2 ability dyads (a medium-ability student combined with a high-ability student). Level-2 ability dyads exchanged significantly more high-level elaborations ($M = 7.14$, $SD = 3.72$) while working on the language task than Level-1 dyads ($M = 2.84$, $SD = 1.61$), $t(11) = -3.20$, $p < .01$. Conversely, the Level-2 ability dyads ($M = 62.59$, $SD = 7.01$) exchanged significantly fewer medium-level elaborations while working on the language task than the Level-1 dyads ($M = 68.97$, $SD = 6.38$), $t(18) = 2.13$, $p < .05$. No statistical significant differences were found for the mathematics task.

Qualitative Analyses of the Verbal Interactions

The effects of the CL school improvement program on the elaborations of the students were also explored by comparing the videotaped interactions of the treatment versus control dyads. Qualitative analyses can clearly supplement quantitative analyses by illustrating the processes involved.

In Table 3, a sample interaction from a Level-2 treatment dyad is shown. The dyad is solving a problem from the language task and trying to define the word "helm." In Table 4, a sample interaction from a Level-2 control dyad working on the same problem is shown.

The most striking difference between the two sequences is the difference in their lengths. Compared to the control dyad, the treatment dyad engaged in a much longer discussion before agreement was reached on the answer to be written down. The treatment dyad also conducted a more balanced discussion than the control dyad. Both of the students in the treatment dyad contributed thoughts and ideas to the discussion and also put the same amount of effort into the cooperation. Both of the students in the treatment dyad contributed to the discussion by reading relevant parts of the text aloud (e.g., lines 1, 3–4, and 15–16), asking questions (e.g., lines 7 and 10), providing elaborated explanations (lines 20–22), and providing unelaborated explanations (e.g., lines 2–3, 17, and 18). In contrast, most of the effort in the control dyad comes from one high-ability student, Annemarie (student 1), who tries to establish a common ground for solving the problem by reading relevant parts of the text aloud (lines 1 and 9–10). She also provides unelaborated explanations (lines 4–5 and 12–13) and tries to engage Rik (student 2) in the discussion by asking several questions (lines 7 and 13). However, this does not appear to work as Rik simply refers to his text without any further explanation (line 8) and simply accepts Annemarie's proposal to write down the answer (line 14). In our opinion, Rik's contribution to the discussion is rather limited.

Performance Scores

The second research question was whether the treatment dyads performed better than the control dyads while working cooperatively on the mathematics and language tasks. In Table 5, the findings for the two groups are presented with a maximum possible score of 50 (see Method and Instrumentation section).

Table 3. Interaction of a Treatment Dyad Working on the Fourth Language Problem.

Student 1: Nicolien[a] (medium-ability) Student 2: Dolf (high-ability)	/ short pause // pause longer than 3 s
1.	2: *What do the following words mean. Helm.*
2.	1: Yeah, oooh//Oh/Take a look at the text. Helm, ok. //I think, it's a/handle, the helm,
3.	that's what's here, right? *Mike pulled on the helm, reduced his speed, and steered the boat at a snail's*
4.	*pace around the island.*
5.	2: I hardly see anything like that.
6.	1: Oh. It's the word in bold.//For me, in the third paragraph //
7.	2: The third paragraph?
8.	1: Yeah.
9.	2: It's not in mine./Mine *does* have navi navigate.
10.	1: Navigate?/
11.	Mine doesn't have that./So, we'll have to use my text.
12.	2: Now wow./
13.	1: So, you'll have to listen, okay?
14.	2: Okay.
15.	1: *Mike pulled on the helm, reduced his speed, and steered the boat at a snail's pace around the island*
16.	*counter-clockwise.*
17.	2: Could be a kind of brake.
18.	1: Could be a handle.
19.	2: Yeah, or a brake, and a steering wheel, or something like that.
20.	1: Yeah./A boat has something like that, and if you make such a trip, yeah, a boat with a motor, and
21.	then you pull like that and then it brakes as well, or/travels slower. /
22.	Oh yes, it takes a turn like this (talk through one another)

(*continued*)

Table 3. (*continued*).

Student 1: Nicolien[a] (medium-ability) Student 2: Dolf (high-ability)	/ short pause // pause longer than 3 s
23.	2: A brake and a steering wheel, really.
24.	1: Yeah. Shall we write that down?
25.	2: Yeah, a brake and a steering wheel. /
26.	1: For example, a kind of handle, or steering wheel.
27.	2: Yeah, that should be good. /
28.	1: (written) Handle or steering wheel//which you with which you can brake, or with which . . .
29.	2: You can steer.
30.	1: Yeah, or steering with which you can brake or . . .
31.	2: Or steer.

Note. [a]To preserve the anonymity of the students, their names have been changed.

Table 4. Interaction of a Control Dyad Working on the Fourth Language Problem.

Student 1: Annemarie (high-ability)	/ short pause
Student 2: Rik (medium-ability)	// pause longer than 3 s

1: *What do the following words means. Helm, board, navig/*Helm, I think uh that that it is a motor of a . . .
2. It was right here in (unintelligible) //
1. Yeah, a sailboat. It doesn't usually have a motor, I think. It always has some sort of sails. So they can maybe uh/pull the sails down, if they put them up, then they then they go they faster. //
2. Yeah. Board.//
1. Do you think so, too?/What do you think?//Helm is printed in bold in mine. /
2. Not in mine. Yeah, I don't have that word./
1. Hmmm. *Mike pulled on the helm, reduced his speed, and steered the boat at a snail's pace around the island counter-clockwise.*
2. Yeah, that's possible. /
1: Actually both are. Or the motor. It depends on whether the things are gone or not.// But it's a sailboat. I think they pull the sails down./Do it?
2. Yeah. //

Table 5. Mathematics and Language Task Scores for Treatment Versus Control Dyads.

	Treatment dyads ($N = 12$)		Control dyads ($N = 8$)				
Performance scores	*M*	*SD*	*M*	*SD*	*t*	*p*	*ES*
Mathematics task	34.83	4.53	31.13	5.14	1.71	.05	.77
Language task	19.58	8.04	18.50	9.59	.27	.39	.12

The data displayed in Table 5 show the treatment dyads to attain higher performance scores than the control dyads on both tasks. For the mathematics task, the difference between the groups approached significance ($p = .05$). The effect size for the mathematics task also showed a relatively large effect in favor of the treatment dyads ($ES = .77$). For the language task, a much smaller effect in favor of the treatment dyads was found ($ES = .12$).

Additional analyses revealed a significant correlation between the frequency of high-level elaborations provided by the dyads while working on the language task and their performance score on the language task, $r = .41$, $p < .05$. A positive but nonsignificant correlation was found for the mathematics task, $r = .09$, $p = .35$. No significant correlations were found between the frequency of medium-level elaborations and the performance scores for either the mathematics or language tasks. In fact, both of the correlations were negative: $r = -.34$, $p = .07$, and $r = -.20$, $p = .20$, respectively.

As mentioned above, Level-2 dyads exchanged significantly more high-level elaborations than Level-1 dyads while cooperating on the language task. Level-2 dyads also performed better on the language task than the Level-1 dyads, $t(18) = -1.84$, $p < .05$ while the mean score for the Level-2 dyads was 22.8 ($SD = 6.3$), the mean score for the Level-1 dyads was 16.2 ($SD = 9.1$).

Student Perceptions

The last research question addressed any differences in the perceptions and attitudes of the treatment versus control group students towards the characteristics of the mathematics and language tasks, the experience of cooperating on the these tasks, and their willingness to cooperate in small groups in the future.

Table 6 shows the findings for the students' perceptions of the characteristics of the mathematics and language tasks with the differences

Table 6. Student Perceptions of the Characteristics of the Mathematics and Language Tasks.

	Treatment group ($N = 24$)		Control group ($N = 16$)					
	M	*SD*	*M*	*SD*	*t*	*df*	*p*	*ES*
Mathematics task								
Liked the task	4.42	.65	4.53	.74	− .51	37	.61	− .16
Understood the task	4.33	.87	4.87	.35	− 2.67	33	.01*	− .75
Found the task difficult	2.21	1.10	1.79	1.19	1.11	36	.28	.31
Conduct a new, similar task	3.75	.99	2.93	1.53	1.84	21	.08	.67
Too many questions	1.42	.72	1.40	.91	.06	37	.95	.03
Too little time	1.50	1.25	1.07	.26	1.64	26	.11	.44
Language task								
Liked the task	4.13	.61	4.25	.68	− .60	38	.55	− .19
Understood the task	4.00	.95	4.38	.72	− 1.33	37	.19	− .44
Found the task difficult	3.21	1.25	3.13	1.63	.17	26	.86	.06
Conduct a new, similar task	3.04	1.27	3.06	1.34	− .05	38	.96	− .02
Too many questions	2.38	1.13	1.69	.79	2.10	38	.04*	.68
Too little time	2.79	1.25	2.81	1.47	− .05	38	.96	− .01

Note. Mean scores along a 5-point scale ranging from 1 (= *highly disagree*) to 5 (= *highly agree*).
$^{*}p < .05$.

between the treatment and control groups examined using a two-tailed *t* test for independent samples. With regard to the mathematics task, the students liked the task, understood it well, did not find it difficult, would like to perform a similar task in the future, did not think that the task had too many questions, and thought they had enough time to complete the task. Only one significant difference in the perceptions of the students was found: The control group students reported a better understanding of the mathematics task than the treatment group students, $t(33) = -2.67$, $p < .05$, $ES = -.75$. The effect size for the readiness to conduct a similar task in the future was in favor of the treatment group students ($ES = .67$).

With regard to the language task, the students liked the task and understood it quite well but nevertheless found the task relatively difficult with too many questions and too little time to answer them. The treatment group students indicated that the task had too many questions significantly more often than the control group students, $t(38) = 2.10$, $p < .05$, $ES = .68$. The effect size for

Table 7. Student Perceptions of the Cooperation on the Mathematics and Language Tasks and Their Willingness to Cooperate in the Future.

	Treatment group ($N=24$)		Control group ($N=16$)				
	M	*SD*	*M*	*SD*	*t*	*p*	*ES*
Quality of cooperation math task	4.45	.43	4.58[a]	.33	−1.03	.31	−.34
Quality of cooperation language task	4.41	.42	4.41	.35	−.02	.98	−.01
Willingness to cooperate	3.24	.56	3.07	.53	.99	.33	.31

Note. Mean scores along a scale ranging from 1 (= *low quality cooperation*) to 5 (= *high quality cooperation*) and, for the Willingness to cooperate scale, from 1 (= *not at all willing to cooperate*) to 4 (= *very willing to cooperate*). [a]$n = 15$.

understanding the language task was in favor of the control group students ($ES = -.44$).

The findings in Table 7 show both the treatment and control group students to consider their cooperation on the mathematics and language tasks to be high quality. Both groups also express a willingness to cooperate on similar tasks in the near future. No significant differences were found between the two groups. The effect sizes nevertheless showed the control group students to rate the quality of their cooperation on the mathematics task moderately higher than the treatment group students. In contrast, the effect size for willingness to cooperate together in the future showed greater readiness on the part of the treatment group students than on the part of the control group students.

DISCUSSION AND CONCLUSION

In the present study, the effects of a CL school improvement program on the interactions of students working in dyads were examined. The program was aimed at guiding the teachers in the development of the instructional behaviors necessary to conduct a cooperative lesson. It is assumed that the program should not only affect the instructional behaviors of the teachers but also the quality of the student interactions while working in small groups.

The first research question was whether the dyads from the classrooms of teachers who had participated in the school improvement program on CL provided and received more elaborations while working on problems than the dyads from the classrooms of teachers who had not participated in the program. The results show moderately positive effects on the interactions of the treatment dyads. In addition, significant differences between the tasks were found. For the language task, statistically significant differences were found between the treatment and control dyads in favor of the treatment dyads. Treatment dyads exchanged more high-level elaborations than the control dyads ($ES = .70$). For the mathematics task, no statistical significant differences were found between the groups ($ES = -.02$). The differences in the results for the two tasks are in line with the findings of a review by Cohen (1994) who concluded that the characteristics of a task can influence the quality of peer interactions. The students in the present study provided more elaborations during the language task than during the mathematics task. This is probably due to the fact that the language task can be characterized as an open-ended, ill-structured task (Cohen, 1994). To solve this task, the students were clearly dependent on each other and needed to reach joint agreement as there was often more than one manner to solve a particular problem. The mathematics task, in contrast, can be characterized as a well-structured task with a single correct answer. The students were not really interdependent as it was possible to solve the task individually (although this was not the instruction), and the task clearly evoked less elaboration than the language task.

Additional analyses showed a significant difference between the Level-1 and Level-2 ability dyads on the language task. The Level-2 ability dyads (i.e., a medium-ability student combined with a high-ability student) exchanged significantly more high-level elaborations than the Level-1 dyads (i.e., a low-ability student combined with a medium-ability student). These findings are in line with the results of a study by Webb et al. (1998), who investigated the effects of group ability composition on group processes and outcomes and found ability levels and group composition to indeed influence the quality of the group discussion while working together to solve a problem. Webb et al. (1998) found the groups with at least one above-average student to produce more accurate and high-quality answers and explanations than groups without such an above-average student.

The second research question was whether the treatment dyads performed better on the two tasks than the control dyads. The results showed this to be partly the case. On average, the treatment dyads performed better than the control dyads. For the mathematics task, the difference between the two groups approached significance ($p = .05$), while the effect size showed a clear difference in favor of the treatment dyads ($ES = .77$). For the language task, however, the differences between the groups were quite small ($ES = .12$). Although the treatment dyads exchanged more high-level elaborations during the language task, they did not perform better on this task. This is contrary to our expectation that more high-level elaborations would result in higher performance scores. One possible explanation is that it was more difficult to attain a higher score on the language task because there was often more than one answer to the problem and the students therefore had to discuss the alternatives to reach agreement. Each student also had only part of the information necessary to solve the problem. This means that the input of both students was necessary to search for a possible solution; considerable time was needed to reach agreement, and the students may have had too little time to complete the entire task as a result. The questionnaire data from the students confirmed that they found the language task to be quite difficult and also had too little time to complete the entire task.

The third research question was whether the students in the treatment group perceived working cooperatively differently than the students in the control group. The first questionnaire pertained to the students' perceptions of the characteristics of the tasks. For the mathematics task, only one statistically significant difference between the two groups was found: The students in the control group reported a better understanding of the mathematics task than the students in the treatment group. Despite their report of a better understanding of the mathematics task, however, the students in the control group did not actually perform better on the mathematics task than the students in the treatment group. For the language task, only one significant difference between the groups was again found: The students in the treatment group indicated that the task had too many questions more often than the students in the control group.

The second questionnaire pertained to the students' experience of cooperating on the tasks. The students in the control group reported a more positive cooperation experience than the students in the treatment group, but only for the mathematics task. One explanation for this unexpected result is that the students in the control group may understand cooperation to mean something different than the students in the treatment group and that the students in the control group may have been less critical of their cooperation than the students in the treatment group who experienced what it takes to work together on many more occasions and were therefore more aware of the criteria that cooperation needs to meet for it to be of good quality. For the language task, no differences were found between the groups. The students in both groups found their cooperation to be high quality.

The third questionnaire pertained to the willingness of the students to cooperate in small groups in the future. The students in the treatment group showed a greater willingness to cooperate on a task in the near future than the students in the control group ($ES = .31$).

Although the focus of the present study was mainly on the cognitive activities of student dyads while working together on a problem-solving task, the affective and regulative dimensions were also distinguished within the coding scheme. The results show the students to generally not spend much time on the regulation of their activities; about 75% of the utterances of the dyads were classified as cognitive statements and only 8% as regulative. The students do not appear to be very aware of the importance of regulating their problem-solving activities. Only 6% of the utterances were classified as affective. In other words, the students were very task oriented and either did not have or apply the skills needed to regulate their activities (e.g., orienting, planning, monitoring, and checking of their progress) and did not encourage each other with affective statements.

In interpreting the results of this study, some possible limitations should be kept in mind. First, the data were collected on a single occasion (posttest-only design), which may limit the representativeness of the interaction data. The absence of pretest data also means that any posttest differences between the groups may be due to treatment effects or initial differences between the groups (Cook & Campbell, 1979). In future studies, data should therefore be collected at the beginning of the study and on multiple occasions thereafter in order to examine the stability of student interactive behavior and make it easier to interpret the effects of the treatment on student behavior. Second, the students were studied outside the classroom for recording ease. This setting does not resemble the normal classroom situation. The important role of the teacher during CL activities (e.g., for monitoring and intervention as needed) is therefore not taken into account. Although complicated, future studies should examine peer interactions within the natural classroom context under the guidance of the teacher. Third, the mixed-sex group composition of the dyads in the present study was undertaken to neutralize any sex differences, which means that same-sex dyads may have produced more variable interaction patterns. When Webb (1984) investigated the effects of sex composition on achievement and interaction patterns for groups of four students, she found the achievement and interaction results to clearly depend on the ratio of girls to boys within the group. The achievement of the girls and boys was nearly identical in groups with two girls and two boys but in majority-girl groups and majority-boy groups, the boys consistently showed higher achievement than the girls. The girls and boys also showed similar interaction patterns in groups with two girls and two boys but not in groups with one or the other majority. Fourth, we did not examine the degree to which the teachers in the treatment group implemented the essential features of the CL school improvement program. Therefore, no conclusions can be drawn with regard to the degree to which the students were actually prepared to provide and receive elaborated explanations. Findings on the degree of implementation will be published in the near future.

In general, it can be concluded that the school improvement program on CL positively affected the interactions of the student dyads. These effects are quite encouraging in light of the fact that only 1 of the 10 workshops was devoted to the quality of peer discussion while working in small groups. As already mentioned, the main aim of the school improvement program was to guide the teachers in the development of the instructional behaviors necessary to conduct cooperative lessons. Given that the implementation of CL in the classroom is a complex matter, it is understandable that the teacher initially focuses on his or her own teaching practice and pays less attention to the interactions of the students working in small groups. This assumption is supported by the research literature on teacher concerns and the role that these concerns play in the innovation process. For example, studies based on the Concerns Based Adoption Model (see Hord, Rutherford, Huling-Austin, & Hall, 1987; Van den Berg & Ros, 1999) show teachers to have different types of concerns depending on the stage of implementation for an innovation. The feelings and concerns of teachers shift from largely self-concerns during the adoption phase (e.g., concerns about personal ability), to task-concerns during the implementation phase (e.g., concerns about the actual performance of the task), and to other-concerns during the institutionalization phase (e.g., concerns about cooperation with colleagues, further progress, and implications for students). In the present study, the teachers may have been struggling with various self-concerns and, until these concerns are addressed, there is little room for task- or other-concerns. That is, concerns about the quality of peer interactions may only appear when basic concerns about the proper implementation of the relevant instructional skills for CL have been overcome.

At this moment, the teachers in our research project have used CL in their classrooms for 3 consecutive years. Only after years of intensive training to broaden their didactic repertoire to include CL do the teachers appear to be ready to deepen their knowledge of actual student interactions in order to promote student performance within cooperative groups. In our own future research, we therefore plan to develop a follow-up program to train students on the skills needed for productive cooperation, such as effective communication and high-level elaboration. And suffice it to say that the insights provided by the present study will certainly help us in this endeavor.

REFERENCES

Brown, A.L., & Palincsar, A.S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 393–451). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carlson, K.D., & Schmidt, F.L. (1999). Impact of experimental design on effect size: Findings from the research literature. *Journal of Applied Psychology*, *84*(6), 851–862.

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

Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analyses issues for field settings*. Chicago: Rand McNally.

Dale, H. (1994). Collaborative writing interactions in one ninth-grade classroom. *Journal of Educational Research*, *87*(6), 334–344.

De Lisi, R., & Golbeck, S.L. (1999). Implications of Piagetian theory for peer learning. In A.M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 3–37). Mahwah, NJ: Lawrence Erlbaum Associates.

Erkens, G. (2001). *Multiple Episode Protocol Analysis* (Version 4.6.4) [Computer software]. Utrecht, The Netherlands: University of Utrecht.

Fuchs, L.S., Fuchs, D., Bentz, J., Phillips, N.B., & Hamlett, C.L. (1994). The nature of student interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal*, *31*(1), 75–103.

Hogan, D.M., & Tudge, J.R.H. (1999). Implications of Vygotsky's theory for peer learning. In A.M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 39–65). Mahwah, NJ: Lawrence Erlbaum Associates.

Hord, S.M., Rutherford, W.L., Huling-Austin, L., & Hall, G.E. (1987). *Taking charge of change*. Alexandria, VA: Association for Supervision and Curriculum Development.

Johnson, D.W., & Johnson, R.T. (1999). *Learning together and alone: Cooperative, competitive, and individualistic learning* (5th ed.). Boston: Allyn and Bacon.

Joyce, B., & Showers, B. (1995). *Student achievement through staff development* (2nd ed.). New York/London: Longman.

Kagan, S. (1994). *Cooperative learning* (10th ed.). San Juan Capistrano, CA: Kagan Cooperative Learning.

King, A. (1999). Discourse patterns for mediating peer learning. In A.M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 87–115). Mahwah, NJ: Lawrence Erlbaum Associates.

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structural tutorial interaction to scaffold peer learning. *Journal of Educational Psychology*, *90*(1), 134–152.

Kneser, C., & Ploetzner, R. (2001). Collaboration on the basis of complementary domain knowledge: Observed dialogue structures and their relation to learning success. *Learning and Instruction*, *11*(1), 53–83.

Krol, K., Veenman, S., & Voeten, M. (2002). Toward a more cooperative classroom: Observations of teachers' instructional behaviors. *Journal of Classroom Interaction*, *37*(2), 37–46.

Kumpulainen, K., & Mutanen, M. (1999). The situated dynamics of peer group interaction: An introduction to an analytic framework. *Learning and Instruction*, *9*(5), 449–473.

Meloth, M.S., & Deering, P.D. (1994). Task talk and task awareness under different cooperative learning conditions. *American Educational Research Journal*, *31*(1), 138–165.

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners*. Clevedon, UK: Multilingual Matters.

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.

Phelps, E., & Damon, W. (1989). Problem solving with equals: Peer collaboration as a context for learning mathematics and spatial concepts. *Journal of Educational Psychology*, *81*(4), 639–646.

Piaget, J. (1926). *Judgement and reasoning in the child*. New York: Harcourt.

Ros, A.A. (1994). *Analyse en effecten van uitleg door leerlingen en leerkrachten* [Explanations by students and teachers: Analyses and effects]. Groningen, The Netherlands: RION, Instituut voor Onderwijsonderzoek.

Salomon, G., & Perkins, D.N. (1998). Individual and social aspects of learning. In P.D. Pearson & A. Iran-Nejad (Eds.), *Review of research in education* (Vol. 23, pp. 1–24). Washington, DC: American Educational Research Association.

Shuell, T.J. (1996). Teaching and learning in a classroom context. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 726–764). New York: Macmillan.

Siegler, R.S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*(4), 481–520.

Slavin, R.E. (1995). *Cooperative learning: Theory, research and practice* (2nd ed.). Boston: Allyn and Bacon.

Slavin, R.E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, *21*(1), 43–69.

Stevens, R.J., & Slavin, R.E. (1995). The cooperative elementary school: Effects on students' achievement, attitudes, and social relationship. *American Educational Research Journal*, *32*(2), 321–351.

Teasley, S.D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology*, *31*(2), 207–220.

Tudge, J.R.H. (1992). Processess and consequences of peer collaboration: A Vygotskian analysis. *Child Development*, *63*(6), 1364–1379.

Van Boxtel, C. (2000). *Collaborative concept learning: Collaborative learning tasks, student interaction, and the learning of physics concepts.* Doctoral dissertation, Utrecht University, The Netherlands.

Van Boxtel, C., Van der Linden, J., & Kanselaar, G. (2000). The use of textbooks as a tool during collaborative physics learning. *Journal of Experimental Education*, *69*(1), 57–76.

Van den Berg, R., & Ros, A. (1999). The permanent importance of the subjective reality of teachers during educational innovations: A concerns-based approach. *American Educational Research Journal*, *36*(4), 879–906.

Veenman, S., Kenter, B., & Post, K. (2000). Cooperative learning in Dutch primary classrooms. *Educational Studies*, *26*(3), 281–302.

Veldhuis-Diermanse, A.E. (2002). *CSCLearning? Participation, learning activities and knowledge construction in computer-supported collaborative learning in higher education.* Doctoral dissertation, Wageningen University, The Netherlands.

Vermunt, J.D.H.M. (1992). *Leerstijlen en het sturen van leerprocessen in het hoger onderwijs* [Learning styles and the regulation of learning processes in higher education]. Doctoral dissertation, Tilburg University, The Netherlands.

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Webb, N.M. (1984). Sex differences in interaction and achievement in cooperative small groups. *Journal of Educational Psychology*, *76*(1), 33–44.

Webb, N.M., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal*, *31*(2), 369–395.

Webb, N.M., & Farivar, S. (1999). Developing productive group interaction in middle school mathematics. In A.M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 117–149). Mahwah, NJ: Lawrence Erlbaum Associates.

Webb, N.M., Nemer, K.M., Chizhik, A.W., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal*, *35*(4), 607–651.

Webb, N.M., Troper, J.D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology*, *87*(3), 406–423.