# Cooperative learning during math lessons in multi-ethnic elementary schools: counting on each other

Oortwijn, M.B.

# CHAPTER 2

## The impact of teacher stimulation and pupils' ethnicity and prior knowledge on pupils' performance and motivation to cooperate

Abstract

*How can the teacher bring about effective cooperative learning (CL) in multi-ethnic elementary classrooms? To answer this question we hypothesized that teachers who stimulate pupils' helping behavior (experimental condition) boost their performance and CL motivation as compared to pupils whose helping behavior is not stimulated by the teachers (control condition). Subjects were 166 pupils from 10 schools. The results show that national pupils in the experimental condition outperformed pupils in the control condition and teams with low and medium prior math knowledge performed better in the experimental condition. Additionally, immigrant teams with high prior math knowledge in the control condition outperformed their low prior math knowledge counterparts and showed a higher motivation for CL. Our results suggest that, next to teacher stimulation, attention has to be paid to both pupils' background characteristics (ethnicity and prior math knowledge) and the teacher's prior knowledge of CL skills.*

**Key words**: cooperative learning; teacher stimulation; prior knowledge; ethnicity; elementary schools

## 1. Introduction

Cooperative learning is an educational method that has received a great deal of attention in the last decades (Cohen, 1994; DeVries & Slavin, 1978; Gillies, 2004; Salomon & Perkins, 1998). Following Cohen (1994) we define cooperative learning (CL) as an educational setting in which pupils work together in a group small enough that everyone can participate on a collective task that has been clearly assigned, without direct and immediate supervision from the teacher. Ample research has revealed that CL can be effective in enhancing the educational development of students (for meta analyses see Qin, Johnson & Johnson, 1995; Rohrbeck, Ginsburg-Block, Fantuzzo & Miller, 2003). However, CL is not per se more effective than direct teaching methods (see for instance Pollock, Chandler & Sweller, 2002). In this study, we attempt to answer the question: what can teachers do to make CL effective? First, we outline the theoretical background of CL, next we outline the three independent variables we studied (teacher stimulation, prior knowledge, and ethnicity). We conclude with the design of the study and our hypotheses.

### *1.1. Theoretical background of CL: the sociocultural approach*

The history of research into CL traces back to Vygotsky's sociocultural approach (Vygotsky, 1978, first published in 1930), who argued that learning is socially rooted. In his view, interactions with other learners in the social learning context are essential for the learner's development. Whether or not these interactions result in fruitful cognitive development depends on the level of cognitive development of each of the learners in the social learning context and the presence of an expert supervisor. The distance between the developmental age of the most capable learner and the least capable learner should not be too great, but just large enough for the least capable learner to benefit from the most capable learner. Vygotsky called this the zone of proximal developmental (ZPD). Learners have to be supported by an expert supervisor in this learning process, called scaffolding (Valsiner & Van Der Veer, 2000). Vygotsky argued that CL will be successful, provided that the students are working in the ZPD or a supervisor is present.

### *1.2. Teacher stimulation during CL*

In this study we define the supervisor as the teacher. A number of studies have highlighted the importance of teacher stimulation for successful CL (e.g. Gillies & Ashman, 1997; 2000; Webb & Farivar, 1994). Teacher stimulation refers, broadly speaking, to the educational tools the teacher applies to stimulate pupils' performance. Studies by Webb and her colleagues (Webb & Farivar, 1994; Webb, Troper & Fall, 1995) have shown that teachers who encourage pupils to use high-quality helping behavior -defined as helping behavior that includes asking for, providing, and applying explanations- stimulate the pupils' performance. Additionally, Chinn, O'Donnell, and Jinks (2000) reported that teacher stimulation of high-quality helping behavior increases both the cooperation and the performance. Following these researchers, we define teacher stimulation during CL as teachers' ability to stimulate high-quality verbal helping behavior in both individual team members and teams as a whole.

### *1.3. Prior knowledge and teacher stimulation in CL*

A number of studies have revealed that the effectiveness of teacher stimulation interacts with pupil background characteristics. For instance, students with high prior knowledge are more able to maintain focused on the group task and to plan and evaluate their actions (Hmelo, Nagarajan & Day, 2000; O'Donnell & Dansereau, 2000). Puustinen (1998) argued that the quality of teacher behavior is important to compensate for the differences between pupils in prior knowledge. More precisely, Puustinen argued that pupils with low prior knowledge are less able to self-regulate their learning. Consequently, they need more support from the teacher during CL. In accordance with this, Gillies and Ashman (2000) showed that teachers who encourage pupils to use high-quality helping behavior augment the math post-test scores of pupils with low prior knowledge as compared to pupils whose high-quality helping behavior is not encouraged. Additionally, they demonstrated that teachers who stimulate the high-quality helping behavior of pupils with low prior knowledge augment their motivation to cooperate. Pollock et al. (2002) found that novice students who are required to work alone perform better than when they are required to interact. This appears to contradict the results of Gillies and Ashman (2000) and the assertion of Puustinen (1998). Caution must be taken in comparing these two seemingly contrasting results. The subjects in the study by Pollock et al. (2002) were novice *adolescent* students. In contrast, the sample in the studies by Gillies and Ashman (2000) consisted of low ability *elementary aged* pupils. In keeping with Puustinen (1998) it can be argued that elementary aged, low ability pupils are less able to cope with independent problem-solving than adolescent students. This assertion is supported by the study by Veenman and Spaans (2005), who found that 15 year olds were more able to regulate their own learning process than were 13 year olds.

### *1.4. Ethnicity and teacher stimulation in CL*

Ethnicity is a second background characteristic that has been found to interact with the effectiveness of teacher stimulation during CL. There is evidence that pupils in ethnically heterogeneous classrooms whose peer interactions are stimulated by the teacher perform better than pupils whose peer interaction are not stimulated (Kagan & Knight, 1981; Klingner, Vaughn & Schumm, 1998). Webb and Farivar (1994) carried out a study in which the quality of helping behavior that the teacher provided was manipulated: pupils were either encouraged to help their peers (experimental condition) or not encouraged (control condition). Their sample consisted of pupils from multi-ethnic elementary schools, of whom most immigrant pupils had an educational disadvantage. Their study revealed that immigrant pupils in the experimental condition outperformed the immigrant pupils in the control condition. Other studies have reported similar findings (e.g. Calderón, Hertz-Lazarowitz & Slavin, 1998). In addition, reviews by Slavin and Cooper (1999) and Webb and Palincsar (1996) revealed that CL can reduce interracial prejudice and augment the quality of cooperation (see also Cohen, 1994).

## 2. Aim of the study, research considerations, and hypotheses

This study investigates the effect of high-quality helping behavior on pupils' performance and motivation to cooperate. The aim of this study is to corroborate the findings of Webb and Farivar (1994) and Gillies and Ashman (2000) that the stimulation of pupils' high-quality helping behavior augments their math post-test scores. In addition to these studies we investigate how the effectiveness of teacher stimulation during CL interacts with ethnicity and prior knowledge. The effectiveness of teacher stimulation is operationalized here in terms of the pupils' gain in math post-test scores and their motivation for CL. Since the teachers in this study were required to address not only the individual team members during CL, but also the team as a whole, analyses were performed both at the individual and at the team level.

The following hypotheses are investigated: 1) Teachers who stimulate the pupils' use of high-quality helping behavior (experimental condition) boost their math post-test scores as compared to pupils whose high-quality helping behavior is not stimulated (control condition). This holds especially for pupils with low prior knowledge and immigrant pupils; 2) pupils in the experimental condition are more motivated to cooperate than are their counterparts in the control condition, especially pupils with low prior knowledge and immigrant pupils.

Since the CL curriculum is in math, prior knowledge is defined here as prior math ability.

## 3. Method

### *3.1. Sample*

A CL math curriculum of nine lessons was carried out in ten multi-ethnic elementary schools. Letters were sent to 200 schools, of which ten teachers responded positively in a subsequent telephone conversation. Eight classes of the participating schools had little or no prior knowledge of CL skills, teachers of two classes (one in the experimental condition, and one in the control condition) indicated implementing group work frequently, around 80 percent of the time. Classes were randomly assigned to the experimental or the control condition. The total sample consisted of 172 children. Six pupils who did not complete the math exam were dropped from the data set. As illustrated in Table 1, 166 children remained (average age 135.7 months, *SD* = 6.5) - see also paragraph 5 of Chapter 1 of this thesis, entitled: *Overview of the thesis and hypotheses*. 71 Pupils were placed in the control condition (57.7 % male, 42.3 % female), and 95 were placed in the experimental condition (44.5 % male, 55.5 % female). With respect to ethnicity, pupils of whom both parents were of Dutch origin were regarded as national, 'mixed' if one parent was of Dutch origin, and 'immigrant' if both parents were of foreign origin (i.e. not Dutch).

Table 1
*Sample characteristics*

| Condition | *N* | Gender | Ethnicity |
|---|---|---|---|
| Control condition | 71 | 30 female | 16 national |
| | | | 8 mixed |
| | | 41 male | 47 immigrant |
| Experimental condition | 95 | 52 female | 40 national |
| | | | 21 mixed |
| | | 43 male | 34 immigrant |
| **Total** | **166** | | |

'mixed' = one parent is Dutch, 'immigrant' = both parents are of foreign origin

In the control condition, 16 pupils were national, 47 immigrant, and eight had a mixed background. The experimental condition consisted of 40 national pupils, 21 mixed, and 34 immigrant pupils. Because there were relatively few pupils categorized as mixed in both conditions, the ethnicity variable was dichotomized. The mixed pupils were classified as national pupils. Dutch studies (e.g. Tesser & Iedema, 2001) have shown that the school performance of mixed pupils resembles that of national pupils more than that of immigrant pupils.

*3.2. Design of the study*

The CL curriculum was composed of three phases. The first phase consisted of a mini-workshop of about two hours, in which the researcher explained to the teachers, irrespective of condition, exactly what the CL curriculum was about. Next, the teachers were given the most essential condition specific instructions. Subsequently, the teachers were asked to carefully read the lesson-to-lesson CL protocol and they were encouraged to ask clarifying questions prior to the first CL training lesson.

In the second phase, all teachers placed their pupils in teams that were narrow-heterogeneous in math ability (high-middle, or low-middle). Then the teacher trained the children to cooperate effectively in two one-hour lessons. In lesson 1, general social CL rules were taught and practiced. These rules were: ‘‘everyone cooperates’’, ‘‘everyone listens to each other’’, ‘‘everyone shares their knowledge and opinions’’, and ‘‘checks whether everyone agrees’’. In lesson 2, more specific CL rules were mentioned and practiced. Adapted from Webb and Farivar (1994), these rules all dealt with giving and receiving help. With respect to asking for help, we distinguished; a) ask precise questions, b) continue asking in case of ambiguities, c) think before asking a question, d) ask for help on time. With respect to giving help, we distinguished; a) fine-tuning of the level of guidance to the need for help that is requested, b) giving a clear and precise answer, c) giving the help receiver a chance to apply the help given, d) continuing to ask if the question for help is unclear, and e) giving help when needed. All CL rules (both the general and the more specific CL rules) were written down on a poster, which was displayed in the classroom and was clearly visible to all children of all classes of both conditions. This poster remained there throughout the whole CL curriculum as a memory aid for the pupils.

In addition to the poster, another memory aid for the pupils was a short checklist which they were required to fill in during each lesson. It also served as a check for pupils to see for themselves which CL rules they used inadequately. These checklists asked for the level of application of the general social CL rules that were taught in lesson 1 of the CL training and the amount of help given and help received (lesson 2 of the CL training). These checklists were not used for analysis.

Phase three consisted of a CL math curriculum of nine one-hour lessons, covering five weeks. The teacher carried out two lessons per week.

### *3.3. Experimental condition*

Following Webb and Farivar (1994), the impact of teacher stimulation of pupils' high-quality helping behavior on their math post-test scores was investigated (see Table 2). Two conditions were created: an experimental and a control condition. Teachers in the control condition were trained to do nothing to stimulate pupils' high-quality helping behavior. The teachers only managed the group work in the instances that team members; a) talked too loudly (disturbed other teams), b) did not listen to each other, c) made fun of each other. In addition, the teachers in the experimental condition stimulated pupils' high-quality helping behavior as much as possible. The CL protocol was condition specific: teachers in the experimental condition had a different protocol than the teachers in the control condition. However, in both protocols detailed descriptions of all the math assignments were provided. Additionally, the protocol contained detailed lesson-to-lesson instructions about how the teachers should apply the CL rules in their condition. In this way, differences between the experimental and the control condition were optimized, enabling a better assessment of the effect of teacher stimulation of pupils' high-quality helping behavior on math post-test scores.

Table 2

*Summary of the CL curriculum*

| *A. Teacher activity* | | Duration | |
|---|---|---|---|
| | | *Control condition* | *Experimental condition* |
| 1. Appointment of chairmen by the teacher. | | No | Yes |
| 2. Direct teaching episodes. | During CL to restore order. | Yes | Yes |
| | Evaluation of the condition work. | No | Yes |
| 3. Providing CL feedback (circulating among the teams). | Repeat + explain general basic CL rules. | Yes | Yes |
| | Verbally rewarding use of rules for help giving and receiving. | No | Yes |
| | | No | Yes |
| *B. Task structure* | | | |
| 4. Group assignment checks? | | Yes | Yes |
| 5. Explicitly mention in the assignments: | | | |
| | a. responsibility of the chairman as role divider. | No | Yes |
| | b. The need to share and discuss the solutions. | No | Yes |

### *3.4. Task structure*

The assignments dealt with surface, percentage, scale, estimation, and fractions. All assignments were adapted for CL purposes from a regular math curriculum that employs realistic math problems with a narrative composition. These are math problems characterized by an emphasis on contexts that are familiar to the children, like the zoo and the school yard. After the adaptation, the assignments (all open-ended) and the math exam were first scrutinized by five teachers and then tested in a CL trial study in five classes. This pilot study had exactly the same form as the main study reported here. On the basis of the pilot findings, further adaptations were made. The adaptations mainly dealt with correcting textual errors, simplifying certain phrases and words, and making the lay-out and the text of the assignments more structured by numbering the tasks and simplifying the lay-out. The assignments that were eventually used in the CL curriculum were moderately structured, open-ended, narrative math assignments, all of which consisted of three parts. Firstly, team members had to work individually on a part of the math task. Secondly, they had to discuss their findings. Subsequently, all team members were required to cooperate to solve the last part of the math task. In the protocol, the teachers were asked to emphasize in their instructions to the pupils that the focus in the CL curriculum was on understanding the math tasks rather than completing them.

### *3.5. Reward structure*

The pupils completed an individual exam at the end of the CL curriculum. During the curriculum the teachers in both conditions took in the worksheet of a random chosen team member of every team at the end of each lesson. The worksheets were only discussed in the classroom: they were not taken into account when pupils' scores on the math exam were calculated.

### *3.6. Instruments*

To check the integrity of the manipulation we used a teacher checklist of helping behavior and videotaped teacher-pupil interactions. The pupils' math ability was tested with a math pre and post-test and a pupil questionnaire on the quality of CL.

#### *3.6.1. Teacher checklist on CL implementation*

Teachers rated on a 4-point Likert-scale (1 = 'very often' and 4 = 'very little') the extent to which they had implemented a number of CL rules. A principal component analysis with varimax rotation revealed a three-factor solution. The solution explained 71 % of the variance. All factor loadings were higher than .50. The first factor (18 items, Cronbach's $\alpha$ = .97) comprised statements about general CL rules (e.g., ''I teach the children not to interrupt each other''). The second factor (5 items, Cronbach's $\alpha$ = .81) referred to the rules for giving help and receiving help (e.g., ''I teach the children to keep asking when someone poses an unclear question''). The third factor (4 items, Cronbach's $\alpha$ = .84) regarded the feedback on the CL process (e.g., ''At the end of each lesson I discuss with each group what is going well and what should be improved''). Each teacher completed the checklist at the end of every other mathematical lesson, starting at the first lesson, amassing five checklists in total. Next to this questionnaire, teachers were required to indicate whether they implemented CL during regular lessons and whether they had made more use of CL for the regular program during the CL curriculum than before the CL curriculum started.

*3.6.2. Videotaped teacher-pupil interactions*

All teachers were videotaped during two or more lessons to know whether the two conditions differed regarding the implementation of CL rules. All recordings were rated by two independent scorers, one of whom was double blind to the experimental manipulation. The coding scheme comprised 14 items. A principal component analysis with varimax rotation was applied; 62 % of the variance was explained. All factor loadings were .50 or higher. The first factor (six items, $\alpha = .71$) was about the presence of whole-class reflection on the group work (eigenvalue 3.4, explained variance 56 %). A sample item is ''Does the teacher reflect on group performance in the previous lesson?'' The second factor (eight items, $\alpha = .86$) covered items that were about the teacher's whereabouts and activities during group work (eigenvalue 4.3, explained variance 53 %). A sample item is ''Does the teacher encourage team members to ask each other clear questions?'' The items were rated on a 3-points scale. The higher the score, the more the teacher was perceived to encourage pupils' high-quality helping behavior. Due to technical failure, recordings were available of eight teachers only. In all, 18 recordings of teacher-pupil interactions could be coded. The overall inter-coder reliability was assessed on the basis of Cohen's kappa, calculated on two recordings (approximately ten percent) and was found to be satisfactory: kappa = .68.

*3.6.3. Prior math knowledge (math pre-test)*

Scores from a curriculum independent math test (CITO; Janssen, Kraemer, & Noteboom, 1996) were used to assess the baseline math performance of all pupils. Since the teachers did not provide us with the data needed for the calculation of the internal validity, we refer to earlier research which has demonstrated that the curriculum independent math test has a good reliability, $\alpha = .94$ (Evers, Van Vliet-Mulder, & Groot, 2000). Because some schools only provided the standardized 5-point scores, all CITO scores used in this study were transformed into this 5-point rating scale. A Pearson's correlation test showed that the pre-test significantly correlated with the post-test, $r = 0.86$, $p < .001$.

*3.6.4. Math post-test*

This is an exam (with possible scores ranging from 1 to 10) that consists of seven items covering the math domains that the children learned during the math curriculum. All teachers applied the same curriculum-dependent math exam after the CL curriculum. A reliability analysis of the data obtained in this study revealed that the internal consistency was satisfactory, $\alpha = .75$.

*3.6.5. Pupils' questionnaire on the quality of cooperative learning (QCL)*

This questionnaire is adapted from Hijzen, Boekaerts, and Vedder (2006). Items of the original questionnaire, which was intended for pupils from secondary vocational education, were reformulated for elementary school pupils. The questionnaire consists of two dimensions: *CL instruction* and *CL motivation*. The pupils filled in the dimension CL instruction only once: before the start of the CL curriculum, to check for differences in CL experiences between conditions. The dimension CL motivation was filled in twice, namely a first time prior to the CL math curriculum, but after the CL training (T1), and a second time after the math exam (T2).

In total, the QCL consists of 30 items. A 4-point Likert-scale (1 = 'very often' and 4 = 'very little') was chosen instead of five points to avoid pupils opting for the middle, neutral category. The dimension *CL instruction* is made up of three scales. The scale 'learned CL rules' (six items, $\alpha = .72$) is about pupils' perception of the CL rules that they learned from their teacher (eigenvalue = 2.6, explained variance = 37.7%). A sample item is ''The teacher has taught us to listen to the other team members during group work''. The scale 'planning of CL' (nine items, $\alpha = .81$) covers pupils' opinion about the teacher's preparation for group work (eigenvalue = 3.56, explained variance = 39.5%). A sample item is ''Before beginning, the teacher tells us what we have to learn from the task''. The scale 'activity of the teacher during CL' (seven items, $\alpha = .75$) is about how pupils perceive teacher stimulation during group work (eigenvalue = 2.8, explained variance = 40.5%). A sample item is ''During group work, the teacher frequently asks how we are getting along with the task''. The dimension *CL motivation* (eight items, $\alpha = .83$) is about pupils' motivation to cooperate (eigenvalue = 3.7, explained variance = 46.1%). A sample item is ''I think it's more fun to work together than to work alone''.

An overview of all experimental measures is presented in Table 3.

## 4. Results

We started the analyses by checking differences between conditions in teachers' and pupils' experiences with CL, differences in prior math knowledge, and assessing the manipulation integrity. Then we proceeded with analyses of the relations between the independent variables 'condition' (experimental condition or control condition), 'ethnicity', and 'prior math knowledge' with the dependent variable 'math post-test scores', both at the individual and the group level. Next, we examined whether there were any differences in pupils' CL motivation that were associated with the condition they were in, their ethnicity, and their prior math knowledge, again both at the individual and the group level.

Table 3

*Overview of the instruments used in this study*

| Implemented instruments | Assessment of: | Number and times of measurement |
|---|---|---|
| Teacher checklist of helping behavior | Treatment integrity | Five measurements. At the end of every other lesson |
| Videotaped teacher-pupils interactions | Treatment integrity | Two video recordings during two CL math lessons |
| Questionnaire on CL (QCL) | Pupils' perceived quality of CL. | Part A (learned CL rules, planning of CL, activity of the teacher during CL): one measurement - before the CL curriculum<br>Part B (CL motivation): Two measurements, one before the CL curriculum, one after |
| Pre-test math scores | Math ability of pupils prior to the CL curriculum. | One measurement - before the CL curriculum |
| Post-test math scores | Pupils' knowledge of the CL math curriculum | One measurement - after the CL curriculum |

*4.1. Preliminary analyses*

*4.1.1. Teachers' and pupils' experiences with CL*

The two conditions did not differ with respect to teachers' self-reported implementation of CL during regular lessons. Although teachers in both the experimental and the control condition indicated using CL more often for the regular program during the CL curriculum than prior to the CL curriculum, the two conditions did not differ from each other. There were also no differences at the start of the CL curriculum regarding pupils' perception of CL instruction between the two conditions.

*4.1.2. Manipulation integrity*

To assess the manipulation integrity we used the teachers' checklist of helping behavior and the video recordings of teacher-pupil interactions. The teachers in the experimental condition reported instructing pupils significantly more in the use of high-quality helping behavior, $t(21) = -3.37$, $p < .005$, Cohen's $d = 1.48$, than did the teachers in the control condition. We found no differences for the dimensions 'general social rules' and 'extent of feedback on the group processes'. Analysis of the coded lessons showed that teachers in the experimental condition did provide more whole-class reflections on the group work than teachers in the control condition, $t(16) = -1.78$, $p < .05$, Cohen's $d = .58$. No differences were found for the dimension 'CL activities during group work'.

Table 4

*Mean scores of the pupils on the math post-test*

| Condition | Ethnicity | Prior math knowledge | Mean individual scores (SD) | Mean team scores (SD) |
|---|---|---|---|---|
| Experimental | National | Low | 3.74 (1.16) | 4.35 (.40) |
| | | Moderate | 5.32 (1.67) | 5.42 (.79) |
| | | High | 6.82 (1.51) | 5.92 (1.18) |
| | | All | 5.59 (1.94) | 5.46 (1.05) |
| | Immigrant | Low | 3.86 (1.69) | 5.81 (.05) |
| | | Moderate | 5.03 (1.24) | 4.64 (1.19) |
| | | High | 6.01 (1.63) | 4.7* |
| | | All | 4.78 (1.72) | 4.94 (1.05) |
| Control | National | Low | 2.64 (1.23) | 3.43 (.83) |
| | | Moderate | 3.53 (.63) | 2.6* |
| | | High | 6.09 (1.62) | 5.56 (1.03) |
| | | All | 4.37 (2.10) | 4.52 (1.51) |
| | Immigrant | Low | 3.53 (1.46) | 3.62 (1.56) |
| | | Moderate | 5.01 (2.27) | 4.36 (1.07) |
| | | High | 7.02 (1.52) | 7.66 (.75) |
| | | All | 5.31 (2.27) | 5.21 (2.12) |

* These cells consisted of only one group. Therefore, the *SD* could not be calculated. Removal of these single measurements from analysis did not alter the significant finding.

### *4.2. Main analyses*

#### *4.2.1. Hypothesis 1*

A repeated measures test was performed. The independent variables were ethnicity, condition (experimental or control condition), and prior math knowledge. Checks of the assumptions showed that normality, linearity, and homogeneity of variance were satisfactory. No univariate or multivariate outliers were found. No main effects were found (see Table 4). We did find a significant 2-way interaction effect for ‘condition’ x ethnicity, $F(1,161) = 4.51$, $p < .04$, explaining 3% of the variance [$\eta^2 = .03$] (see Figure 1). This means that the post-test math scores of immigrant pupils in the control condition were significantly better than that of national pupils in the control condition, $F(1,68) = 5.9$, $p < .02$, $\eta^2 = .08$. Also, the post-test math post-test scores of national pupils in the control condition were significantly lower than that of the national pupils in the experimental condition, $F(1,82) = 8.02$, $p < .007$, explaining 9% of the variance [$\eta^2 = .09$]. Thus, national pupils did perform as we hypothesized, showing higher learning gains in the experimental condition. In contrast to our hypothesis, immigrant pupils performed better in the control condition. Furthermore, we could not demonstrate a positive effect of the stimulation of high-quality helping behavior on the performance of pupils with low prior math knowledge.

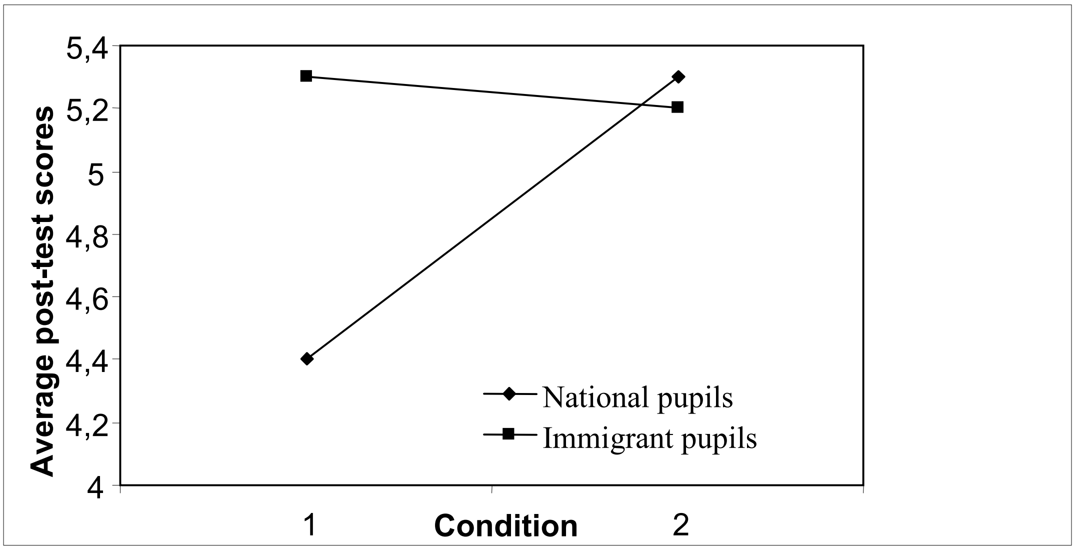


Figure 1. *Interaction ‘condition’ X ‘ethnicity’ at the individual level*

*Analyses at the group level*. Since pupils completed the math exam individually, the pupils’ individual math scores were used as the unit of analysis. Nevertheless, as group work was the focus of research in this study, an explorative analysis of the math post-test scores at the group level was also incorporated. Due to the small sample size, teacher stimulation could not be evaluated using a multilevel approach. Inspired by the studies conducted by Gillies and Ashman (2000), Webb and Farivar (1994), and Saleh, Lazonder, and de Jong (2005), we used analyses at the group level, aggregating pupils’ math scores from the pre-test as well as the post-test and dividing these by the number of pupils in the teams. Regarding ‘ethnicity’, a new variable was created (1 = majority of children have at least one Dutch parent, 2 = majority of pupils have immigrant parents). Also a new variable was created for ‘prior math knowledge’ (1 = mean group pre-test math score is below average, 2 = mean group pre-test math score is on average, 3 = mean group pre-test math score is above average).

The group data ($n$ = 48) were analyzed with a repeated measure design, in which 'condition' (experimental or control condition), 'ethnicity', and 'average prior math knowledge' were the independent variables. No main effects were found. However, a significant 2-way interaction effect was found for ethnicity with 'condition', $F(1,36) = 5.04$, $p < .04$ [$\eta^2 = .12$], indicating that teams with national pupils whose use of high-quality helping behavior was stimulated by the teacher performed better than teams with national pupils whose helping behavior was not stimulated. Furthermore, a 2-way interaction effect was found for 'condition' x 'average prior math knowledge', $F(2,36) = 4.55$, $p < .02$, explaining 20% of the variance [$\eta^2 = .20$] (see Figure 2).

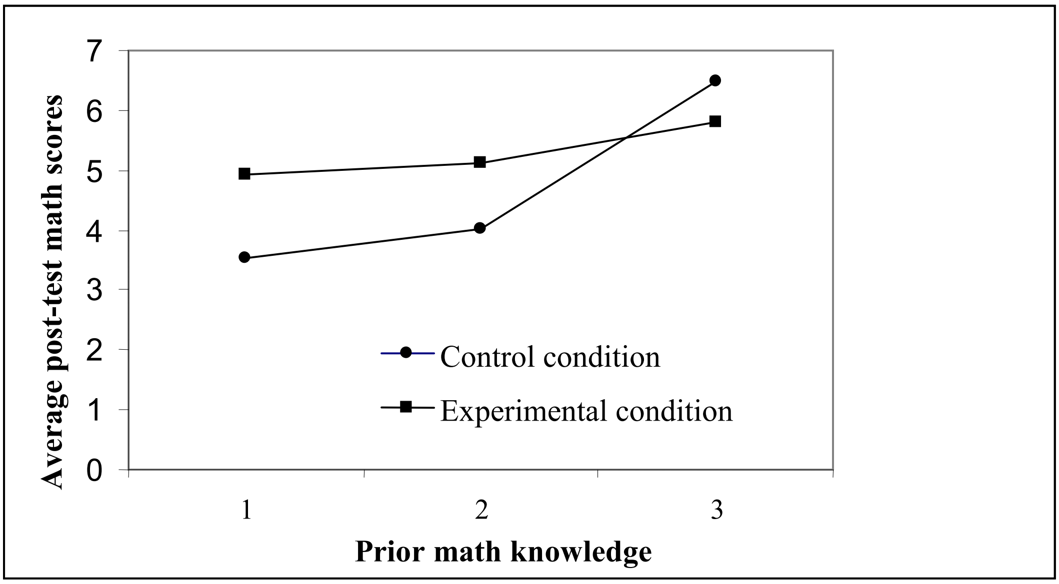


Figure 2. *Interaction 'condition' X 'prior math knowledge' at the group level*

Teams with high prior math knowledge performed better than teams with low prior math knowledge in the control condition, $F(2,18) = 11.8$, $p < .005$. Because of the small sample, a Kruskal Wallis test was carried out to cross validate this finding: it yielded a similar result, $\chi^2 = 11.03$, $df$ 2, $p < .005$. No differences between teams with high and low prior math knowledge were present in the experimental condition. Furthermore, a Mann Whitney test revealed that the stimulation of high-quality helping behavior was only positively related to math post-test scores of teams with low prior math knowledge, $Z = -2.30$, $p < .05$. For teams with medium prior math knowledge a trend was found, $Z = -78$, $p < .08$. Finally, a 3-way interaction effect was found for 'condition' x 'average prior math knowledge' x 'ethnicity', $F(2,36) = 3.26$, $p = .05$, which explained 15% of the variance [$\eta^2 = .15$]. A Kruskal-Wallis revealed that teams consisting of national pupils with high prior math knowledge performed better in the control condition than teams with national pupils with low prior math knowledge, $\chi^2 = 6.04$, *df* 2, $p < .05$. The same pattern emerged in the experimental condition, $\chi^2 = 6.05$, *df* 2, $p < .05$. For teams with immigrant pupils, the picture was slightly different: there was a difference between immigrant teams with high and low prior math knowledge in the control condition, $\chi^2 = 7.73$, *df* 2, $p < .03$, but not in the experimental condition.

It seems that teams with low prior math knowledge are generally better off in the experimental condition, which is in line with our expectations. Unexpectedly, the immigrant teams with high prior math knowledge did not outperform the immigrant teams with low prior math knowledge in the experimental condition.

*4.2.2. Hypothesis 2*

Pupils in the experimental condition are more motivated to cooperate than are their counterparts in the control condition, especially pupils with low prior math knowledge and immigrant pupils. In addition to 6 pupils who filled in only one questionnaire, one class unwittingly filled in the second questionnaire only. Therefore, the sample amassed 149 pupils: 67 in the control condition (21 national, 46 immigrant) and 82 in the experimental condition (58 national, 24 immigrant). To check for initial differences between the two conditions, scores on the scales ‘learned CL rules’, ‘planning of CL rules’ and ‘activity of the teacher during CL’ of the dimension ‘CL instruction’ were compared between the two conditions prior to the CL curriculum (T1) using a MANOVA (see Table 5). No differences on these scales between the two conditions were found prior to the CL curriculum.

The effect of the CL curriculum on pupils’ self-reported CL motivation was analyzed at T1 and after the curriculum (T2) with a repeated measures test in which the independent variables were ‘condition’, ‘ethnicity’ and ‘prior math knowledge’. No significant main effects were found (see Table 5). However, the analysis did reveal a significant 2-way interaction effect, Wilks’ $F(2,130) = 3.20$, $p < .05$, which explained 5% of the variance [$\eta^2 = .05$]. That is, pupils with low prior math knowledge in the experimental condition were more motivated to cooperate at the end of the CL curriculum than pupils with high prior math knowledge in the experimental condition. Further analysis showed that the effect was related to ethnicity, $F(2,59) = 3.78$, $p < .05$ [$\eta^2 = .11$]. That is, whereas immigrant pupils with high prior math knowledge were more motivated to cooperate in the control condition, immigrant pupils with low prior math knowledge were more motivated to cooperate in the experimental condition.

The fact that immigrant pupils with high prior math knowledge were more motivated to cooperate provided they did not receive feedback regarding their high-quality helping behavior resembles the analyses of pupils' math post-test scores, which showed that immigrant pupils had higher math scores in the control condition as compared to the experimental condition. These findings contradict our expectations. In order to understand why, we explored these findings further.

Table 5

*Comparison of the control and experimental condition with respect to pupils' scores on the QCL at T1 and T2*

| Dimension | Scale | T1 Control (*SD*) | T1 Experimental (*SD*) | T2 Control (*SD*) | T2 Experimental (*SD*) |
|---|---|---|---|---|---|
| 1 (CL instruction) | Learned CL rules[1] | 3.54 (.36) | 3.53 (.36) | | |
| | Planning of CL[1] | 3.31 (.47) | 3.32 (.61) | | |
| | Activity of the teacher during CL[1] | 3.34 (.51) | 3.43 (.50) | | |
| 2 (Use of CL skills) | Application of general CL rules | 3.66 (.44) | 3.60 (.38) | 3.69 (.42) | 3.56 (.51) |
| | Application of helping rules | 3.59 (.35) | 3.51 (.45) | 3.64 (.40) | 3.48 (.48) |
| 3 (CL motivation) | CL motivation | 3.50 (.52) | 3.34 (.59) | 3.46 (.56) | 3.19 (.66) |

Higher mean scores indicate a higher perceived quality of CL.

[1] Only filled in by pupils prior to the CL curriculum to check for initial differences.

In the present study, the main difference between the two conditions was teacher stimulation of high-quality helping behavior. Intensive peer interaction presupposes a reasonable command of the language. Since immigrant pupils are known to perform less well on tests of linguistic ability (Tesser & Iedema, 2001), it could be that immigrant pupils with high prior math knowledge had lower linguistic proficiency and therefore were more motivated to cooperate in the control condition. To test this hypothesis, a repeated measures analysis was performed again but now with 'linguistic proficiency' added as a covariate. 'Linguistic proficiency' was measured prior to the CL curriculum with the CITO's 'vocabulary' scale (CITO is a national testing service in the Netherlands: Janssen et al., 1996). The analysis showed that with the addition of 'linguistic proficiency', the significant effect disappeared. This suggests that immigrant pupils with high prior math knowledge may have shown lower CL motivation in the experimental condition due to more limited linguistic proficiency.

We also added linguistic proficiency as a covariate to the analyses of the relationship of condition (experimental or control) with post-test math scores to explore whether this might explain the unexpected findings. The addition of the covariate weakened the relation of the interaction of condition and ethnicity with math post-test scores. Interestingly, the linguistic proficiency of immigrant pupils with high prior math knowledge was lower than that of national pupils with high prior math knowledge in both the control and the experimental condition, respectively, $Z(29) = -2.87$, $p < .005$, and $Z(20) = -2.38$, $p < .02$.

*Analyses at the group level*. Analyses to examine pupils' perception of CL at the group level revealed no significant effects. Therefore our prediction that teams in the experimental condition would be more motivated to cooperate could not be confirmed.

### *4.3. Summary of findings*

The results partly supported our hypotheses. It was found that national pupils achieved a higher math score in the experimental condition than national pupils in the control condition.

This effect was corroborated at the group level. In the experimental condition, post-test math scores of the teams with low and medium prior math knowledge were higher than the post-test scores of teams with low and medium prior math knowledge in the control condition. Split for ethnicity the analyses revealed that, in contrast to national teams with high prior math knowledge, the immigrant teams with high prior math knowledge did not score higher in the experimental condition than did immigrant teams with low prior math knowledge.

With respect to the pupils' motivation to cooperate, we showed that immigrant pupils with high prior math knowledge were more motivated to cooperate in the control condition as compared to the experimental condition. For immigrant pupils with low prior math knowledge the opposite pattern emerged.

## **5. Discussion**

Even though the CL curriculum was of short duration, the teachers did influence the development of pupils' math post-test scores. The finding that the stimulation of pupils' use of high-quality helping behavior by the teachers resulted in better math post-test scores is in line with earlier findings (e.g. Gillies, 2004; Gillies & Ashman, 2000), but is in conflict with other studies that demonstrated that immigrant pupils' performance is best served by stimulating their use of high-quality helping behavior (e.g. Webb & Farivar, 1994). A study by Kirchmeyer (1993) showed that immigrant students who worked in ethnically heterogeneous teams were less active than national students. In the present study there was an even distribution of ethnically heterogeneous teams and teams with only immigrant pupils in the control condition. In contrast, in the experimental condition the majority of teams were ethnically heterogeneous. Thus it could be that the presence of national pupils in most teams in the experimental condition lowered the activity of the immigrant pupils.

In keeping with the expectations, the performance of teams with low prior math knowledge benefited by the teachers' stimulation of their use of high-quality helping behavior. Moreover, these pupils also were more motivated to cooperate as compared to pupils whose helping behavior was not stimulated. This latter result is in line with earlier findings (Gillies & Ashman, 2000; Johnson & Johnson, 2003).

For the teams with high prior math knowledge, the picture was different: their performance was not influenced by whether or not the teacher encouraged them to use high-quality helping behavior. Several studies have suggested that pupils who are able to effectively monitor their own learning process need less feedback from the teacher about how they cooperate (e.g. Cohen, 1994; Puustinen, 1998). Moreover, the motivation of these pupils, typically pupils with high prior math knowledge (Puustinen, 1998; Stevens, Slavin & Farnish, 1991), to cooperate effectively might be undermined when their level of autonomy is restricted (Cohen, 1994). We found partial support for this assertion: we did find that pupils with high prior math knowledge were more motivated to cooperate in the control condition, but this held true for immigrant pupils only.

The sample was too small to use a statistical multilevel approach. In an attempt to overcome this shortcoming, we conducted analyses at the individual as well as at the group level. Both levels of analyses yielded a different outcome regarding the role of ethnicity. Whereas at the group level ethnicity added explanatory value to the relationship between teacher stimulation and prior math knowledge with math post-test scores, no effect of ethnicity was found at the individual level. This seems a puzzling finding. It has been suggested that individual characteristics such as prior math knowledge or the quantity of talk during CL cannot properly account for the learning process at the group level (Barron, 2003). In our analyses we defined prior math knowledge as student prior math ability. Barron's study suggests that team success is best predicted by joint attention to the task at hand and a supportive climate for different ideas. Barron argued that more attention should be paid to interrelational and situated factors, such as the opportunity for positive relational talk, the discussion of ideas, and whether team members feel comfortable with each other. In this study, attention was paid only to prior math knowledge and learning outcomes: no specific attention was paid to process factors like the discussion of ideas. Research has demonstrated this can provide valuable insights into the mechanisms that drive learning gains (e.g. Kumpulainen & Mutanen, 1999; Webb et al., 1995; Wegerif, Mercer & Dawes, 1999).

Some mention must be made of the mixed findings as regards the manipulation check. There was a discrepancy between the teachers' own views and that of the coders. The teachers in the experimental condition indicated that they were more actively teaching high-quality helping behavior *during* group work than teachers in the control condition. On the other hand, the coders only detected more discussion of high-quality helping behavior in the experimental condition *prior* or *after* to the group work. In accordance with other studies this study also suggests that there is a discrepancy between what the teachers think they are capable of with respect to group work and what they are actually doing (Sharan, 1990; Vedder & Veendrick, 2003). This study demonstrates that, even with a limited amount of time and resources, teachers are able to master at least some of the skills that are needed to successfully carry out group work in multi-ethnic classes. With more training, teachers may not only become more experienced in the implementation of specific CL skills (like helping behavior), but also become more aware of their own teaching behavior during CL. In such a training explicit attention should also be devoted to teachers' unique teaching style. In our study we did not incorporate teacher background variables. Recent research has shown that the teacher's educational style can be influential in the classroom (e.g., Webb, Nemer & Ing, 2006). In future studies attention to the teacher's educational style and teacher background variables is warranted to extend the findings we reported regarding the effectiveness of the teacher during CL.

## References

Barron, B. When smart groups fail. *The Journal of the Learning Sciences*, *12*, 307-359.

Calderón, M., Hertz-Lazarowitz, R., & Slavin, R. (1998). Effects of bilingual cooperative integrated reading and composition on students making the transition from Spanish to English reading. *The Elementary School Journal*, *99*, 153-165.

Chinn, C. A., O'Donnell, A. M., & Jinks, T. S. (2000). The structure of discourse in collaborative learning. *The Journal of Experimental Education*, *69*, 77-97.

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

DeVries, D. L., & Slavin, R. E. (1978). Teams-games-tournaments (TTG): Review of ten classroom experiments. *Journal of Research and Development in Education*, *12*, 28-38.

Evers, A., van Vliet-Mulder, J.C., & Groot, C.J. (2000). *COTAN Documentatie van tests en testresearch in Nederland* [Documentation on tests and test research in the Netherlands]. Amsterdam: Boom test uitgevers.

Gillies, R. M. (2004). The effects of cooperative learning on junior high school students during small group learning. *Learning and instruction*, *14*, 197-213.

Gillies, R. M., & Ashman, A. F. (2000). The effects of cooperative learning on students with learning difficulties in the lower elementary school. *The Journal of Special Education*, *34*, 19-27.

Hmelo, C. E, Nagarajan, A, & Day, R. S. (2000). Effects of high and low prior knowledge on construction of a joint problem space. *Journal of Experimental Education*, *69*, 36-56.

Hijzen, D., Boekaerts, M., & Vedder, P. (2006). The relationship between the quality of cooperative learning, students' goal preferences and perceptions of contextual factors in the classroom. *Scandinavian Journal of Psychology*, *47*, 9–21.

Janssen, J., Kraemer, J. M., & Noteboom, A. (1996). *Rekenen-Wiskunde 3. Handleiding* [Arithmetic-Math 3. Manual]. Arnhem: Citogroep.

Johnson, D. W., & Johnson, F. P. (2003). *Joining together, group theory and group skills* (8th Ed.). Boston: Pearson Education, Inc.

Kagan, S., & Knight, G. P. (1981). Social motives among Anglo American and Mexican American children: Experimental and projective measures. *Journal of Research in Personality*, *15*, 93-106.

Kirchmeyer, C. (1993). Multicultural task groups: An account of the low contribution level of minorities. *Small Group Research*, *24*, 127-148.

Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. *The Elementary School Journal*, *99*, 3-22.

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

O'Donnell, A. M, & Dansereau, D. F. (2000). Interactive effects of prior knowledge and material format on cooperative teaching. *Journal of Experimental Education*, *68*, 101-18.

Pollock, E., Chandler, P, & Sweller, J. (2002). Assimilating complex information, *Learning and Instruction*, *12*, 61-86.

Puustinen, M. (1998). Help-seeking behavior in a problem-solving situation: Development of self-regulation. *European Journal of Psychology of Education*, *13*, 271-82.

Qin, Z., Johnson, D. W., & Johnson, R. T. (1995). Cooperative versus competitive efforts and problem solving. *Review of Educational Research*, *65*, 129-143.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, *95*, 240-257.

Saleh, M., Lazonder, A. W., & de Jong, T. (2005). Effects of within-class ability grouping on social interaction, achievement, and motivation. *Instructional Science*, *33*, 105-119.

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

Sharan, S. (1990). Cooperative learning and helping behavior in the multi-ethnic classroom. In H. C. Foot, M. J. Morgan, & R. H. Shute (Eds.), *Children Helping Children* (pp. 151-176). Chichester: John Wiley.

Slavin, R. E., & Cooper, R. (1999). Improving intergroup relations: Lessons learned from cooperative learning programs. *Journal of Social Issues*, *55*, 647-663.

Stevens, R. J., Slavin, R. E., & Farnish, A. M. (1991). The effects of cooperative learning and direct instruction in reading comprehension strategies on main idea identification. *Journal of Educational Psychology*, *83*, 8-16.

Tesser, P. T. M., & Iedema, J. (2001). *Rapportage Minderheden 2001* [Minorities Report 2001]. SCP: Den Haag.

Valsiner, J., & Van der Veer, R. (2000). *The Social Mind: Construction of the Idea*. New York, NY: Cambridge University Press.

Vedder, P., & Veendrick, A. (2003). The role of the task and reward structure in cooperative learning. *Scandinavian Journal of Educational Research*, *47*, 529-542.

Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and individual differences*, *15*, 159-176.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press. (Original work published in 1930).

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

Webb, N. M., Nemer, K. M, & Ing, M. (2006). Small group reflections: parallels between teacher discourse and student behavior in peer-directed groups. *The Journal of the Learning Sciences*, *15*, 63-119.

Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner, & R. C. Cafree (Eds.), *Handbook of Educational Psychology* (pp. 841-873). New York: Simon & Shuster MacMillan.

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

Wegerif, R., Mercer, N., & Dawes, L. (1999). From social interaction to individual reasoning: an empirical investigation of a possible socio-cultural model of cognitive development. *Learning and Instruction*, *9*, 493-516.