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

Oortwijn, M.B.

CHAPTER 4

The effect of teacher stimulation on math-related talk of elementary age pupils during cooperative learning

Abstract

We investigated whether teachers who encourage pupils to use high-quality helping behavior (experimental group) during cooperative learning (CL) stimulate pupils' math-related linguistic proficiency more than teachers who do not stimulate pupils' use of high-quality helping behavior (control group). We were specifically interested in differences between national and immigrant pupils. Additionally, we studied whether math-related linguistic proficiency boosts math post-test scores. Subjects were a subsample of study 1, namely 59 elementary age pupils who were videotaped while cooperatively working on math assignments.

Math-related linguistic proficiency in general was higher in the experimental group. Furthermore, we found that in the experimental condition immigrant pupils' use of high-quality question markers was raised. Interestingly, low-quality question markers were negatively related to math post-test scores, but only for immigrant pupils in the control group. The findings support the sociocultural assumption that language proficiency and learning gains are intertwined and need to be structured by the teacher.

Key words: Cooperative learning, teacher stimulation, linguistic proficiency, immigrant pupils

1. Introduction

Increasingly, researchers recognize that peer interactions are an essential force that drives students' cognitive development in cooperative learning (CL) settings (e.g. Gillies & Ashman, 2000; 2004; Keefer, Zeitz & Resnick, 2000; Webb, Farivar & Mastergeorge, 2002; Wegerif, Mercer & Dawes, 1999). Most of the earlier mentioned studies have focused on different verbal aspects of peer interactions to assess its relationship with cognitive development. For instance, Webb et al. (2002) found a positive relationship between helping behavior and math post-test scores. On the other hand, Wegerif et al. (1999) found evidence for a positive association between the occurrence of group discussions during peer interactions and subsequent cognitive development. The earlier mentioned researchers all demonstrated that there is empirical support for the idea that peer interactions are related to cognitive growth. However, there are differences in opinion about which part of the peer interaction process positively influences cognitive development.

From a socio-cultural point of view, linguistic proficiency is seen as a prerequisite for the acquisition of cognitive skills in a social learning context. Moreover, some have argued that cognitive development is mostly shaped by context specific factors (e.g. Kumpulainen & Mutanen, 1999; Rogoff, 1995). One of the most important context specific factors in the classroom is the teacher (Cohen, 1994; De la Mata Benitez, 2003; Hoek, Van den Eeden & Terwel, 1999; Webb & Farivar, 1994). In this study, we investigated how the teacher influences pupils' linguistic command of math-related concepts in a CL setting, and whether this is related to pupils' math post-test scores.

1.1. Theoretical underpinning of social learning

The recent surge in studies about CL methods is driven by an increasing interest of investigators in the context in which individuals learn (Kumpulainen & Mutanen, 1999; Lave & Wenger, 1991; Salomon & Perkins, 1998). Theoretically speaking, several research approaches to contextual learning might be distinguished, of which the traditional CL research and the socio-cognitive research are two of the most influential branches. The main stronghold of the traditional CL research concerns the assertion that social learning is not rewarding in itself but has to be reinforced by incentives (Slavin, 1996). Motivational specialists have debunked the assumption that the application of incentives will guarantee successful CL (for a meta-analysis see Deci, Koestner & Ryan, 2001). Moreover, these researchers found that, especially for pupils in elementary schools, the provision of verbal as well as tangible rewards might be detrimental to students' performance. The emphasis on external incentives during CL also seems undesirable from a theoretical perspective because it is argued that it only stimulates motivation for superficial learning (i.e. earning the grade/passing the exam), rather than encouraging a drive for deep-level understanding of the problem-solving process and the learning process itself (Covington, 2000).

On the other hand, proponents of the socio-cultural approach argue that CL is successful by itself, provided that the students are working on a challenging task and are of roughly the same mental age. That is, students should be in the same developmental zone, or *zone of proximal development* (ZPD) (Valsiner & Van Der Veer, 2000). When this happens, students can benefit from each other by interacting. As such, the socio-culturists contend that it is linguistic interaction that propels cognitive development (Kumpulainen & Mutanen, 1999; Valsiner & Van der Veer, 1993; 2000).

1.2. The role of culture in the sociocultural learning approach

According to the founder of the socio-cultural perspective, Lev Vygotsky, language and learning are to a large extent intertwined. He distinguished two kinds of learning, 'lower-order thinking' and 'higher-order thinking' (Valsiner & Van Der Veer, 2000). Lower-order, or natural thinking, entails the skills that are learned naturally, like associative memory and reactive attention. Higher-order thinking concerns the cultural construction of new skills (called signs) on the basis of the lower-order skills (for instance learning to use fractions). The acquisition of higher-order knowledge must be supervised by an adult expert. This is deemed crucial, since this supervisor 'scaffolds', or supports, the learner's transition from lower-order to higher-order knowledge. Therefore, Vygotsky defined development as the transformation of lower-order thinking into higher-order thinking, which in his perception is mediated by the use of signs, most notably language. Vygotsky argued that verbal interaction is essential for higher-order thinking, but that it is at the same time limited because of cultural differences in the content and form of verbal interaction. This notion is also apparent in the Sapir-Whorf hypothesis, known as 'linguistic relativity hypothesis' (Lee, 1997). Linguistic relativity refers to the assertion that in the ideal learning setting, cultural boundaries might be overcome by paying attention to the different cultural perspectives. The Sapir-Whorf hypothesis is mentioned here to illustrate that there is a broad theoretical support for the assertion that culture can both aid in and hinder the learner's development. Empirical studies have corroborated the assertion that linguistic proficiency is related to academic development (e.g., Abedi & Lord, 2001; Cardelle-Elewar, 1992).

1.3. Factors that influence CL: empirical evidence

The abovementioned theoretical importance of interaction as a facilitator of cognitive development stands in sharp contrast to the instructional methods that are actually used in everyday classrooms in the Netherlands. That is, the average classroom is still characterized by predominantly direct teaching (Veenman & Krol, 1999). Although the benefits of CL may be abundant (e.g. for reviews see Qin, Johnson & Johnson, 1996; Rohrbeck, Ginsburg-Block, Fantuzzo & Miller, 2003), it is less obvious *how* CL should be implemented to maximize learning gains. Important factors that influence the effectiveness of CL include teacher stimulation (Gillies & Ashman, 2000), group composition (Oetzel, 2001), prior knowledge of CL skills (Butler & Kedar, 1990; Gillies & Ashman, 1997), task structure (Cohen, 1994; Cordova & Lepper, 1996), and reward structures (Covington, 2000; Slavin, 1996). With respect to teacher stimulation, it has been found that teachers who stimulate peer interactions boost pupils' mathematical performance (e.g. Gillies, 2004; De la Mata Benitez, 2003; Webb & Farivar, 1994). Since the teacher is the most dominant influence in the classroom, we studied his or her effect on pupils' linguistic development and math post-test scores.

1.4. Measuring linguistic development

In the research literature, several methods have been advanced to analyze the linguistic quality of students' verbal interactions. We distinguish two approaches. The first approach is aimed at the frequency with which students use specific signaling words, like question markers, conjunctions, and words that are associated with meta-linguistic and mental activities, such as 'but' and 'because' (Vedder, Kook & Muysken, 1996). In the second approach a set of semantically linked words within a specific context are investigated. That is, the focus lies on the comprehension of figurative speech, also known as idioms (e.g. Cain, Oakhill & Lemmon, 2005). In this approach, the degree to which students take idioms literally is researched. For instance, a study might focus on whether pupils understand that when someone 'hits the road', this does not mean that this person is going to punch the road, but rather that he or she is leaving. Research has demonstrated that especially linguistically low achievers are having difficulties with the correct interpretation of figurative speech (Levorato & Cacciari, 1995).

In this study we combined elements from both approaches in the context of a math curriculum. On the one hand, we investigated how the math curriculum influences pupils' use of questions markers and conjunctions. On the other hand, we explored how the math curriculum influenced pupils' use of math-related idioms. We argue here that pupils' understanding of the mathematical meaning of these particular concepts, raises their math achievement. Following

Niemi (1996), *understanding* was defined here in two different ways. Firstly, we distinguish *semantic* understanding. We interpret this as the pupils' explicit knowledge of the specified math concepts. That is, correctly verbalizing the definition. Secondly, we distinguish *general* understanding, which is a more implicit grasp of the conceptual framework in which the specified math concepts are embedded. More specifically, general understanding was interpreted here as the 'operationalization' of the concept in a specific context.

1.5. Performance of pupils in Dutch multi-ethnic elementary schools

In the Netherlands many elementary schools have a multicultural make up. These schools are characterized by a high percentage of immigrant pupils. On average, throughout the primary school period, the linguistic performance of immigrant children is below the national mean. Not only do teachers have insufficient resources to cope with the highly divergent language abilities of the pupils, but due to the high communicative requirements of most math curricula, they also lack the skills to prevent pupils' math performance from falling behind when compared to the national mean. This is translated into a significant disadvantage in math and language performance of the pupils at the end of elementary school (Tesser & Iedema, 2001).

Both international and Dutch studies have demonstrated that the social, linguistic, and mathematical performance of immigrant pupils may profit from CL (Cooper & Slavin, 1999; De Haan en Elbers, 2003; Webb & Farivar, 1994). In this process, the teachers play an important role. Only when they actively stimulate pupils' peer interactions, pupils' school performance might be raised (Webb & Farivar, 1994). Otherwise, CL can even be detrimental to pupils' school performance (Cohen, 1994). With respect to reading skills, Calderón, Hertz-Lazarowitz, and Slavin (1998) found that teachers who used a structured CL educational program boosted students' reading performance more as compared to teachers who led the students fend for themselves. In addition, a study by Webb and Farivar (1994) revealed that teachers who stimulated pupils' high-quality helping behavior (characterized by asking, giving, and applying explanations) increased the high-quality helping behavior of immigrant pupils as compared to teachers who did not stimulate their helping behavior. However, they did not evaluate whether the linguistic proficiency of the immigrant pupils was augmented as a consequence of the more intensive peer interactions in the experimental condition (characterized by the teachers' stimulation of pupils' use of high-quality helping behavior). In this study we attempted to replicate the findings from the study by Calderón et al. (1998). In addition, we examined whether teacher stimulation of the pupils' use of high-quality helping behavior augments the linguistic proficiency of immigrant pupils.

2. Aims - Hypotheses

In this study we created two conditions to investigate teacher stimulation on the math-related talk of pupils in multi-ethnic elementary schools, namely; an experimental condition (requiring the teacher to stimulate pupils' use of high-quality helping behavior) and a control condition (in which the teacher was required to refrain from stimulating pupils' use of high-quality helping behavior). The concrete hypotheses for this study were:

1. Teachers who stimulate the pupils' use of high-quality helping behavior (henceforth experimental condition) raise their math-related talk more –i.e., the frequency of question markers, conjunctions and the quality of mathematical idioms- as compared to a control condition in which teachers do not stimulate pupils' use of high-quality helping behavior.
2. The math-related talk of immigrant pupils is raised more than that of national pupils if teachers stimulate pupils' use of high-quality helping behavior.
3. Math-related talk is positively related to math post-test scores.

3. Method

3.1. Sample

The study was carried out in ten fifth grade classes and draws on the same pupils as reported in Chapter 2 and 3 -see also paragraph 5 of Chapter 1 of this thesis, entitled: *Overview of the thesis and hypotheses*. During two lessons video recordings were made of the peer interactions of 29 teams. From these teams, we selected those of which: a) there were two video recording episodes, b) assignments of both recordings dealt with comparable math topics. In addition, video recordings were made of teacher-pupil interactions during three lessons. Due to technical failure, recordings were available of eight teachers only. In all, 18 recordings of teacher-pupil interactions were codeable.

As illustrated in Table 1, fifteen groups (mean age 134.3 months, *SD* 6.3 months) met the earlier mentioned selection criteria, totaling 59 pupils. In the control condition there were nine groups, comprising 35 pupils (mean age = 133.4 months, *SD* = 5.9; 16 male, 19 female, 8 Dutch, 27 immigrant). There were six groups in the experimental condition, comprising 24 pupils (mean age = 135.5 months, *SD* = 6.9; 12 male, 12 female, 11 Dutch, 13 immigrant). With respect to ethnicity, pupils were defined as national when both parents were Dutch and defined as immigrant when one or both parents had a non Dutch nationality.

Starting point of all recordings was the instance that pupils commenced with a new math assignment. The average length of the videorecordings was 941.1 seconds (*SD* = 229.0) and did not differ between the experimental and the control condition. We also made video recordings of teacher-pupil interactions during two randomly selected lessons to check the treatment integrity.

Table 1

Sample characteristics

Group	Number of groups	Number of pupils	Mean age (*SD*)	Ethnicity	Gender
1	9	35	133.4 (5.9)	8 Dutch	16 Male
				27 Immigrant	19 Female
2	6	24	135.5 (6.9)	11 Dutch	12 Male
				13 Immigrant	12 Female
Total	15	59			

3.2. Procedure

The CL curriculum consisted of 11 CL lessons. In the first two CL lessons the teacher instructed the pupils how to use particular CL rules. These were ‘everyone cooperates’, ‘everyone listens to each other’, ‘everyone shares their knowledge and opinions’, and ‘checks whether everyone agrees’, ‘ask precise questions’, ‘continue asking in case of ambiguities’, ‘think before asking a question’, ‘ask for help on time’, ‘fine-tuning of the level of guidance to the need for help that is requested’, ‘giving a clear and precise answer, ‘ giving the help receiver a chance to apply the help given’, ‘continuing to ask if the question for help is unclear’, and ‘giving help when needed’. The CL rules were practiced and then posted in front of the class. After these two CL training lessons (similar for both the experimental and the control condition), all pupils received the math curriculum of nine one-hour CL lessons. Each lesson consisted of two open-ended authentic math assignments with a narrative structure. All assignments started with an individual component (to enhance individual accountability) followed by a group component. The curriculum was designed to specifically target the concepts of circumference, surface, scale, fractions, and estimation. After the math lessons, pupils completed a curriculum dependent math exam. In all, the length of the CL curriculum amassed six weeks.

3.3. Manipulation

All teachers received a lesson-to-lesson protocol. The teachers who were randomly assigned to the control condition were instructed to guide the group work only when pupils: 1) talked too loudly (disturbed other groups), 2) did not listen to each other, 3) made fun of each other. Teachers in the experimental condition were instructed to structure group work by stimulating the rules that were posted in front of the class. Additionally, the teachers discussed the group work with the whole class at the end of each lesson or at the beginning of the following lesson. Finally, pupils in the experimental condition were required to fill in checklists on their use of CL rules during group work.

3.4. Instruments used to test the hypotheses

3.4.1. Coding math-related talk

On the basis of Levorato and Cacciari (1995), Niemi (1996) and Vedder et al. (1996), a coding scheme was developed. In this coding scheme, three dimensions were distinguished. Dimension one consisted of: A) question markers, 'what' question markers (value 1) and 'why' question markers (value 2). B) The application of conjunctions, like 'because' and 'unless', which was also scored as a dichotomous variable. Dimension two concerned the frequency and quality of understanding of the mathematical concepts 'scale', 'surface', 'circumference', and 'estimation'. For each concept, two levels were distinguished: 1) low-level application of the mathematical concept (inappropriate use of a definition, verbalization of a math concept only, use of numbers only), 2) high-level application of a mathematical concept (use of a context definition, sharing a definition by two pupils, or use of an abstract definition). Dimension three comprised a dichotomous variable regarding the occurrence of a number of mathematical words, which could enrich the specific problem-solving context (e.g. 'fraction', 'divide'). Two coders were trained in the scoring procedure. Subsequently, both coders independently coded approximately 20 percent of the data (six video recordings) to establish the inter-coder reliability. Three measures of concurrence were used, namely the inter-coder agreement, Cohen's kappa and Krippendorff's alpha (Krippendorff, 2004). Krippendorff's alpha was used for dimension 1a, 1c and dimension 3, because Cohen's kappa might give problems when applied to dichotomous data (Weinberger & Fischer, 2006). Using Krippendorff's alpha, not only attention is paid to non-occurrence (i.e. one of the coders has not given a score) but also to the co-occurrence of non-occurring utterances (i.e. agreement between the coders that a specific utterance does not occur). For dimension 1a, kappa was .79, Krippendorff's alpha was .80, for 1b the inter-coder agreement was 84%, and Krippendorff's alpha was .62. For dimension 2, inter-coder agreement was 93%, Krippendorff's alpha was .62. For dimension 3, inter-coder agreement was 94%, Krippendorff's alpha was .93. The second coder, who was blind to the experimental manipulation, then coded the remaining part of the recordings.

To code the video recordings, the coders had at their disposal comprehensive coding instructions, the math assignments that the pupils worked on, and a list with all the correct problem-solving steps and the right solutions for the assignments. The video recordings were coded with the software program Observer 5.0 (Noldus, 2003). With this program it is possible to mark specific behavioral events or states on a timeline.

3.4.2. Prior linguistic proficiency

This test was taken from the national testing institute, used to assess pupil learning progress at the elementary school level (Janssen, Kraemer & Noteboom, 1996). The scores of the two dimensions, 'vocabulary' and 'reading comprehension' were averaged into a new variable labeled linguistic proficiency. The test was taken before the onset of the CL curriculum.

3.4.3. Prior math knowledge

We used scores from a curriculum independent test (CITO; Janssen et al., 1996) to assess whether there were initial differences between the two conditions. It is well validated and reliable, α = .94. CITO is widely used at Dutch elementary schools to monitor children's mathematical progress. Normally, raw scores are transformed to a standardized 5-points rating scale ranging from one to five, five being the highest. Because some schools only provided standardized scores, all CITO scores used here were transformed into this 5-points rating scale.

3.4.4. Math post-test

This is an exam (with possible scores ranging from 1 to 10) that covered the math domains that the pupils learned during the CL curriculum. A previous study demonstrated that the curriculum independent math test significantly correlated with the math post-test, $r = 0.77$, $p < 0{,}001$ (Oortwijn, Boekaerts & Vedder, 2005).

3.4.5. Teacher checklist on CL implementation

The teachers completed a checklist at the end of every other lesson, on which they indicated on a 4-points Likert-scale (1 = 'very often' and 4 = 'very little'); a) to what extent they had implemented the CL rules, and b) their teaching activities during the last CL lesson. The items of the checklist corresponded to the CL instructions in the lesson-to-lesson protocol for the experimental condition.

3.4.6. Videotaped teacher-pupil interactions

All teachers were videotaped during two, randomly selected, lessons. The teachers were not told in advance which CL lessons we would videotape. All recordings were rated by two independent scorers, one of whom was double blind to the experimental manipulation. The coders filled in a coding scheme of 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 (6 items, Cronbach's $\alpha = .71$) concerned whole-class reflection on the group work (e.g., ‘‘Did the teacher reflect on group performance in the prior lesson?’’). The second factor (8 items, Cronbach's $\alpha = .86$) regarded the teacher's activities during the group work (e.g., “Did the teacher encourage group members to ask each other questions?”). The items were rated on 3-point Likert-scale (1 = ‘little’ and 3 = ‘often’). The inter-coder reliability (calculated over two recordings, approximately ten percent of the total number) was satisfactory: for Factor 1 kappa = .73 and for Factor 2 kappa = .62.

4. Results

4.1 Preliminary analyses

4.1.1. Differences between conditions in math pretest scores and linguistic proficiency

An independent sample T-test taken prior to the CL curriculum revealed that the average scores of the pupils on the math pretest did not differ between the two conditions. Pupils in the experimental condition had a higher score (mean 2.80, $SD = .69$) on the test of linguistic proficiency than pupils in the control condition (mean 2.40, $SD = .83$), $t(57) = -2.24$, $p < .02$, Cohen's $d = .60$.

4.1.2. Teacher checklist on CL implementation

An independent samples T-test showed that teachers in the experimental condition reported instructing pupils significantly more in the use of helping skills, $t(21) = -3.37$, $p < .005$, Cohen's $d = 1.48$, than the teachers in the control condition. No differences were found on the dimensions ‘general social rules’ and ‘extent of feedback on the group process’ between the two conditions.

4.1.3. Videotaped teacher-pupils interactions

An independent samples T-test on the dimensions ‘whole-class reflection on group work’ and ‘feedback on group work during CL’ was performed. Although the samples were small, homogeneity of variance did not differ between the two conditions. Also, the data were not significantly skewed and had no significant kurtosis. Analysis of the coded lessons showed that teachers in the experimental condition reflected more on the group work at the start or end of the CL lessons than teachers in the control condition, $t(16) = -1.78$, $p < .05$, Cohen's $d = .58$. For the dimension ‘feedback on group work during CL’, no differences between the two conditions were found.

Table 2

Relationship of group (experimental or control condition) with linguistic performance

Dimensions of the linguistic coding scheme	Mean experimental condition (*SD*)	Mean control condition (*SD*)	*SS*	*Df*	*MS*	*F*	η^2
High-quality question markers	8.20 (2.87)	1.43 (.30)	.69	44	.69	9.33***	.18
Low-quality question markers	17.80 (2.65)	14.43 (4.66)	.65	54	.65	10.34***	.16
Conjunctions	6.20 (1.28)	1.71 (.29)	.53	34	.53	4.50*	.12
Low-level math understanding	3.40 (1.44)	2.29 (.47)	.10	16	.10	1.78	.10
Mathematical words	5.40 (1.44)	2.86 (.59)	1.76	42	1.76	15.56***	.27

* $p < .05$, ** $p < .01$, *** $p < .005$.

4.2. Main analyses

4.2.1. Hypothesis 1: Relationship of teacher stimulation with math-related talk

Since the data were substantially skewed, and there was a significant heterogeneity of variance between the conditions, the data were rotated using the formula LG10(X) (See also Tabachnick & Fidell, 2001). With respect to dimension two, only attention was paid to the use of low-level mathematical concepts (66 of the 2041 coded utterances, 3.2 %), since only a negligible number of all utterances were related to high-level mathematical concepts (7 of the 2041 coded utterances). After this, univariate analyses of covariance were executed for the relationship of the independent variables, ‘condition’ and ‘ethnicity’ with each of the dependent variables ‘high-quality question markers’, ‘low-quality question markers’, ‘conjunctions’, ‘low-level understanding of math concepts’ and ‘use of mathematical words’. ‘Linguistic proficiency’ was the covariate.

The analyses yielded the following results (see Table 2). There were significant main effects for ‘condition’ with ‘high-quality question markers’, $F(1,44) = 9.33$, $p < .005$ [$\eta^2 = .18$], which explained 18% of the variance, ‘low-quality question markers’, $F(1,54) = 10.34$, $p < .003$ [$\eta^2 = .16$], explaining 16% of the variance, ‘conjunctions’, $F(1,34) = 4.50$, $p < .05$ [$\eta^2 = .12$], which explained 12% of the variance, and ‘use of mathematical words’ $F(1,42) = 15.56$, $p < .001$ [$\eta^2 = .27$], explaining 27% of the variance.

4.2.2. Hypothesis 2: Interaction of condition and ethnicity with math-related talk

The analyses revealed a significant two-way interaction effect for ‘high-quality question markers’, $F(1,44) = 5.11$, $p < .03$ [$\eta^2 = .10$], explaining 10% of the found variance (see Figure 1).

Because of the small sample size, a Mann-Whitney U test was carried out to cross-validate the finding. This test also showed a significant effect, $t(49) = -2.41$, $p < .02$. Figure 1 illustrates that immigrant pupils displayed more high-quality question markers in the experimental condition (mean 9.64, *SD* = 6.00) than immigrant pupils in the control condition (mean 3.67, *SD* = 2.15), $F(1,29) = 18.80$, $p < .001$ [$\eta^2 = .39$], explaining 39% of the variance. On the other hand, national pupils did not display a higher frequency of high-quality question markers in the experimental condition than in the control condition. Furthermore, immigrant pupils in the experimental condition uttered significantly more high-quality question markers (mean 9.64, *SD* = 6.00) than national pupils in the experimental condition (mean 3.30, *SD* = 2.26), $F(1,18) = 14.35$, $p < 0.001$ [$\eta^2 = .44$], explaining 44% of the variance.

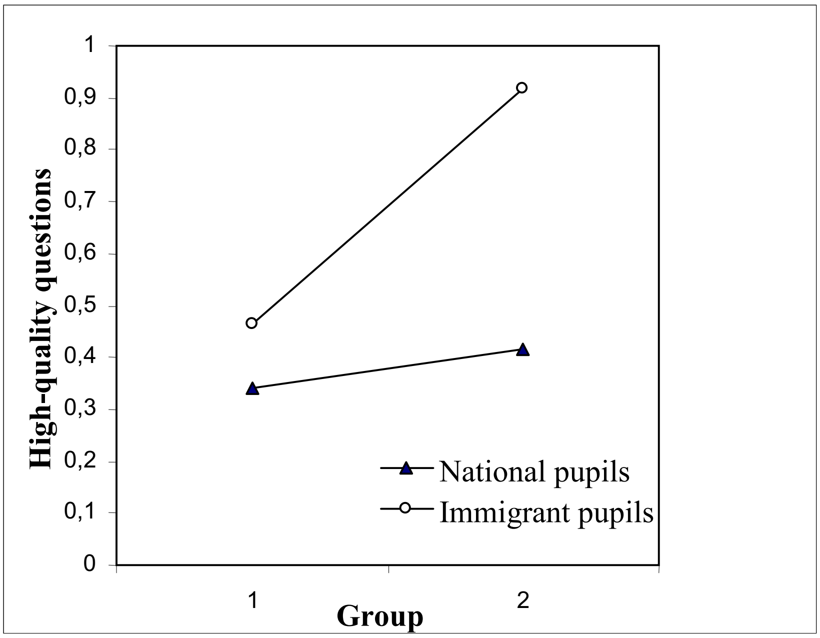

Figure 1. *Interaction of group (experimental or control condition) and ethnicity with the number of high-quality question markers*

4.2.3 Math-related talk and math post-test scores

Partial correlations were calculated to assess the relationship between ‘math post-test scores’ with ‘high-quality question markers’, ‘low-quality question markers’, ‘conjunctions’, ‘low-level understanding of math concepts’ and ‘use of mathematical words’. ‘Linguistic proficiency’ served as control variable. The analysis showed that for immigrant pupils in general the frequency with which they made use of low-quality question markers was negatively related to math post-test scores, $r = -.40$, $p < .02$. However, a further exploration revealed that actually the relationship was only significant in the control condition, $r = -.45$, $p < .03$, not in the experimental condition, $r = .12$, $p = .71$. This provides indirect support for our hypothesis, which asserted that the quality of question markers in the experimental condition positively influenced math post-test scores.

4.3 Summary of findings

In this study an answer was sought on the question whether teacher stimulation of pupils' high-quality helping behavior during a CL curriculum influences pupils' math-related talk. More specifically, we were interested to find out to what extent teacher stimulation of pupils' high-quality helping behavior is related to the use of specific question markers, conjunctions, and the use of mathematical concepts and words. Also, we wanted to know whether this relationship is moderated by ethnicity. Finally, we examined whether the use of math-related talk is related to math post-test scores.

The results show that the use of both high and low-quality question markers, of conjunctions, and of mathematical words was stimulated more in the experimental condition than in the control condition. Furthermore, there was an interaction between the condition pupils were in, their ethnicity, and their use of high-quality question markers. That is, immigrant pupils used more high-quality question markers in the experimental condition than in the control condition. Finally, we found evidence that the frequency with which immigrant pupils in the control condition verbalized low-quality question markers was inversely related to their math post-test scores.

5. Discussion

In this study we investigated to what extent the teacher's stimulation of pupils' high-quality helping behavior is related to the use of specific question markers, conjunctions, and use of mathematical concepts and words. Also, we studied whether this relationship is moderated by ethnicity. Finally, we examined whether the use of math-related talk of immigrant pupils is related to their subsequent math performance. The results show that pupils used more high-quality and low-quality question markers, conjunctions, and mathematical words in the experimental condition than pupils in the control condition. Furthermore, there was an interaction between the condition pupils were in, their ethnicity, and their use of high-quality question markers. That is, immigrant pupils used more high-quality question markers in the experimental condition than pupils in the control condition. No such effect occurred for national pupils. Finally, we found evidence that the frequency with which pupils in the control condition verbalized low-quality question markers was inversely related to their subsequent math performance, but only for the immigrant pupils.

The teachers' efforts to stimulate pupils to give and receive help positively affected pupils' use of math-related words, questions -both high-quality and low-quality-, and conjunctions. These results extend earlier finding from other studies (Webb & Farivar, 1994; Webb et al., 2002). Additionally, we found that pupils whose high-quality helping behavior was not stimulated tended to stick to a basic level of peer interactions, characterized by low-quality question markers. The use of low-quality question markers was most detrimental for immigrant pupils, whose posttest math performance was negatively influenced by it.

No teacher effect on pupils' use of mathematical idioms was demonstrated. Possibly this was due to the overall low use of mathematical idioms by the pupils. Two reasons are put forward here for the possible low use of mathematical idioms by the pupils. The first is methodological: the inclusion criteria for mathematical idiom were too rigid. Although pupils frequently verbalized a numerical rule without reference to a mathematical concept, these verbalizations usually were too ambiguous to make accurate coding possible. For instance, pupils frequently multiplied two numbers while working on area and scale problems. Nevertheless, it was sometimes unclear whether they were (incorrectly) calculating the scale, or whether they were calculating an area.

The second reason put forward here is that pupils are just not accustomed to explicitly labelling the mathematical operation when referring to a specific math concept in their peer interactions. This can not be caused by inexperience with the use of such concepts, since the pupils did make frequent use of other math-related talk, like mathematical words. Another explanation is suggested by Cain et al. (2005) who argued that there are two approaches to understanding idioms: through semantic analysis or by making use of the context. Perhaps the pupils in our study, although they did make use of the context in order to work with the mathematical idioms, had difficulty verbalizing implicit knowledge. This was noticeable on a number of recordings. In some groups, different pupils simultaneously used definitions of both area and circumference when interacting about how to calculate the area of a classroom, without correcting each other. This suggests that, in spite of a shared understanding of the mathematical concept, evidenced by correct mathematical solutions, pupils still seemed to have problems to relate this implicit knowledge to the right mathematical idiom.

As mentioned before, pupils in the experimental condition used both more low and high-quality question markers than pupils in the control condition. How might this apparent contradiction be explained? Two possible explanations are discussed here. The first explanation is that the higher frequency of high-quality question markers neutralized the adverse effect of the low-quality question markers. Although not directly related to post test math performance, the high-quality question markers could have influenced the relationship of low-quality question markers with posttest math performance. There is a large body of literature which suggests that high-quality questioning is positively related to learning gains (Fantuzzo, Riggio, Connelly, & Dimeff, 1989; King, 2002). A larger sample might be required to replicate the positive effect of high-quality question markers with subsequent performance.

A second possible explanation is that a functional differentiation occurred: it might be that in the experimental condition, the low-quality question markers were used more for the management of the group's CL process and the high-quality question markers were used more for the math-related problem solving process. There is no direct evidence for this, but the results did show that pupils used more mathematical words in the experimental condition as compared to the control condition. This suggests that the pupils in the experimental condition talked more about math.

5.1. Limitations

Two limitations are mentioned here. First of all, the sample size is relatively small. A larger sample is needed to corroborate the findings reported in this paper. Secondly, there were more immigrant pupils in the control condition than in the experimental condition. This may have distorted some of our findings.

5.2. Conclusion

There is a growing understanding that it is impossible to separate linguistic development from other types of development (e.g. Kumpulainen & Mutanen, 1999; Van Der Veer & Valsiner, 2000). This is reminiscent of the sociocultural approach, which states that language and learning are intertwined. Moreover, high order knowledge will only occur in a social learning context, as a result of interaction with other learners, under the supervision of an adult expert (Van Der Veer & Valsiner, 2000). In line with this, our study suggests that the CL context requires the presence of a supervisor or teacher to support the pupils' development of math-related talk. Furthermore, the study implies that for this supervision to be most effective, it is advisable that teachers pay attention not only to pupils' cognitive development, but also to their development of high-quality helping behavior. Pupils whose high-quality peer interactions are not stimulated resort to low-quality interaction patterns and their subsequent performance might be hindered.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*, 219-234.

Butler, R., & Kedar, A. (1990). Effects of intergroup competition and school philosophy on student perceptions, group processes, and performance. *Contemporary Educational Psychology*, *15*, 301-318.

Cain, K., Oakhill, J., & Lemmon, K. (2005). The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology*, *90*, 65-87.

Cardelle-Elawar, M. (1992). Effects of teaching metacognitive skills to students with low mathematics ability. *Teaching and Teacher Education*, 8, 109-121.

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

Cooper, R., & Slavin, R. E. (1999). Reducing racial prejudice, discrimination, and stereotyping: Translating research into programs - improving intergroup relations: lessons learned from cooperative learning programs. *The Journal of Social Issues*, *55*, 647-664.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*, 715-730.

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, *51*, 171-200.

Deci, E.L., Koestner, R., & Ryan, R.M. (2001). Extrinsic rewards and intrinsic motivation in education: reconsidered once again. *Review of Educational Research*, *71*, 1-27.

De Haan, M., & Elbers, E. (2003). Samenwerkend leren [Cooperative learning]. *Tijdschrift voor Ontwikkelingsgericht Onderwijs*, *2*, 4-6.

De la Mata Benitez, M. L. (2003). Text remembering as a social process: the role of teacher-student interaction in the acquisition of structure strategy. *Learning and Instruction*, *13*, 93-115.

Fantuzzo, J. W., Riggio, R. E., Connelly, S., & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology*, *81*, 173-177.

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. (1997). Children's cooperative behavior and interactions in trained and untrained work groups in regular classrooms. *Journal of School Psychology*, *35*, 261-279.

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

Hoek, D., Van den Eeden, P., & Terwel, J. (1999). The effects of integrated social and cognitive strategy instruction on the mathematics achievement in secondary education. *Learning and Instruction*, *9*, 427-448.

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

Keefer, M. W., Zeitz, C. M., & Resnick, L. B. (2000). Judging the quality of peer-led student dialogues. *Cognition and Instruction*, *18*, 53-81.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal*, *29*, 303-323.

King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice*, *41*, 33-39.

Krippendorff, K. (2004). Measuring the Reliability of Qualitative Text Analysis Data. *Quality and Quantity: International Journal of Methodology*, *38*, 787-800.

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

Lave, J., & Wenger, E. (Eds.) (1991). *Situated learning: Legitimate Peripheral Participation*. New York, NY, US: Cambridge University Press.

Lee, P. (1997). Language in thinking and learning: Pedagogy and the new Whorfian framework. *Harvard Educational Review*, *67*, 430-471.

Levorato, M. C., & Cacciari, C. (1995). The effects of different tasks on the comprehension and production of idioms in children. *Journal of Experimental Child Psychology*, *60*, 261-283.

Niemi, D. (1996). Assessing conceptual understanding in mathematics: Representations, problem solutions, justifications and explanations. *The Journal of Educational Research*, *89*, 351-363.

Noldus L. P. J. J. (2003). *The Observer (Version 5.0)* [Computer software]. Wageningen, The Netherlands: Noldus Information Technology Inc.

Oetzel, J. G. (2001). Self-construals, communication processes, and group outcomes in homogenous and heterogeneous groups. *Small Group Research*, *32*, 19-54.

Oortwijn, M. B., Boekaerts, M., & Vedder, P. (2005). *Is coöperatief rekenen op multiculturele basisscholen effectiever dan klassikaal leren?* [Is collaborative learning during math at multi-ethnic elementary schools more effective than frontal learning?]. *Panama-Post*, *24*, 3-11.

Qin, Z., Johnson, D. W., & Johnson, R. T. (1996). Cooperative versus competitive efforts and problem solving. *Review of Educational Research*, *65*, 129-143.

Rogoff, B. (1995). Observing sociocultural activity on three planes: participatory appropriation, guided participation, and apprenticeship. In J. V. Wertsch, P. Del Rio, & A. Avarez (Eds.), *Sociocultural Studies of Mind* (pp. 139-164). New York, NY: Cambridge University Press.

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.

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* (pp. 1-24). Washington DC: American Educational Research Association.

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

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

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate statistics* (4th ed.). Boston: Allyn and Bacon.

Valsiner, J., & Van der Veer, R., (1993). The encoding of distance: The concept of the zone of proximal development and its interpretations. In Cocking, R. R., & Renninger, K. A. (Eds.), *The Development and Meaning of Psychological Distance*. (pp. 35-62). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

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

Vedder, P., Kook, H., & Muysken, P. (1996). Language choice and functional differentiation of languages in bilingual parent-child reading. *Applied Psychololinguistics*, *17*, 461-484.

Veenman, S., & Krol, K. (1999). Implementatie van coöperatief leren in het basisonderwijs [Implementation of cooperative learning in elementary education]. *Pedagogische Studiën*, *76*, 139-156.

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., Farivar, S. H., & Mastergeorge, A. M. (2002). Productive helping in cooperative groups. *Theory into Practice*, *41*, 13-20.

Wegerif, R., Mercer, N., & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal*, *25*, 95-111.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers and Education*, *46*, 71-95.