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

Oortwijn, M.B.

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

Michiel B. Oortwijn

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

Proefschrift
ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof. mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 14 juni 2007
klokke 15.00 uur

door
Michiel Oortwijn
geboren te Heemskerk
in 1977

Promotoren
Prof. Dr. P. Vedder;
Prof. Dr. M. Boekaerts

Beoordelingscommissie
Prof. Dr. M. Segers;
Prof. Dr. W.C.M. Resing;
Prof. Dr. R. Martens (Open Universiteit);
Prof. Dr. G. Kanselaar (Universiteit Utrecht);
Prof. Dr. J. Terwel (Vrije Universiteit)

Voorwoord

November 2002 studeerde ik af in de experimentele psychologie (functieleer) aan de UvA. Dit is een heel andere richting dan een onderwijskundig onderzoek naar leerprocessen op multiculturele basisscholen waar ik me als aio mee bezig heb gehouden. Het werken in een pedagogische setting heeft dan ook behoorlijk wat aanpassing gevergd, niet in het minst omdat het onderzoek zich voltrok op basisscholen overal in de Randstad (en zelfs daarbuiten), in plaats van op de comfortabele universiteit. Toch heb ik geen enkel moment spijt van deze stap gehad.

Mijn dank gaat in de eerste plaats uit naar alle leerkrachten die hebben geparticipeerd in dit onderzoek; Ineke Brinkman, Loes Ostendorf, Debbie, Janneke Broerse, Frank Aafjes, Karin Eichhorn, Truus, Marije, Nicole Hagenaars-Wesseling, Marion van der Heijden, Rob Tops, Gretl Walraven, Mireille Rozeboom en Monique Reijnders. Zonder hen was dit proefschrift niet mogelijk geweest. Ten tweede wil ik mijn ‘collega’s van het eerste uur’ Daphne en Marie bedanken; zij waren eerder dan ik begonnen en hebben in veel opzichten het promotietraject voor mij meer begaanbaar en overzichtelijk gemaakt. Als we aan het eind van de dag niet verder meer kwamen en er geen gat meer inzagen, begaven we ons naar café de Bruine Boon (BB) waar we op de één of andere manier de dingen altijd veel scherper konden zien. Helaas was deze scherpe blik de volgende dag vaak weer verdwenen... Andere collega’s die mij veel hebben ondersteund in zware tijden zijn Nanine (altijd goed voor kinderen voor kinderen liedjes om de moraal hoog te houden), Janna (altijd in voor gesprekken over sociale netwerken), Jan-Willem (altijd in voor een gesprek over werk) en Jeroen (altijd gehaast, maar altijd vriendelijk).

Verder wil ik Dorothée Venrooij bedanken voor haar inzet en hulp. Niet alleen bij het verzamelen van de data, waarbij we vele uren op de scholen hebben gespendeerd aan het maken van video-opnames van de leerkrachten en de leerlingen, maar ook bij het coderen van de video-opnames. In dit verband ben ik ook dank verschuldigd aan Coen Kouwenhoven. Met schijnbaar eindeloos geduld heeft hij vele uren aan video-opnames met leerkrachtgedrag en samenwerkende leerlingen geobserveerd en beoordeeld.

Tijdens het lange, eenzame ploeteren kwam ik Katka tegen. Zij heeft het begrip ‘een leven lang leren’ niet alleen een nieuwe dimensie gegeven (namelijk; Tsjechisch leren), maar ook laten zien dat een leven lang leren weinig voorstelt zonder een levenlang liefhebben.

Michiel Oortwijn
Leiden, Maart 2007

Contents

CHAPTER 1

Learning in teams. An introduction

1. Introduction

An increasing number of elementary schools in the Netherlands have a multi-ethnic make-up, and the number of immigrant pupils is still on the rise (Gijsberts, 2004). The study by Gijsberts shows that the increasing number of immigrants is accompanied by an escalating segregation, especially in the large cities. The rising segregation in urban areas occurs in many countries, including the USA (see for a recent study, Shelton & Richeson, 2005) and Great Britain (Dixon, Durrheim & Tredoux, 2005), resulting in so-called black schools - i.e. schools where the majority of students are of foreign origin. These multi-ethnic schools, most of which are located in the densely populated western part of the country, generally have an educational disadvantage as compared to the national mean (Tesser & Iedema, 2001). These schools face social as well as educational problems: interethnic bias is daily practice. Thus, one of the major challenges these schools face is how to let students with different ethnic backgrounds get along with each other. In addition, these schools face the challenge to reduce their educational disadvantage. Several educational researchers have suggested that cooperative learning (CL) is an effective educational tool to stimulate both interethnic relations (Cohen, 1994; Warring, Johnson, Maruyama & Johnson, 1985) and learning gains (for meta analyses, see e.g., Qin, Johnson & Johnson, 1995; Rohrbeck, Ginsburg-Block, Fantuzzo & Miller, 2003). Not surprisingly, many elementary schools have taken the step to introduce CL in the classrooms (School Inspectorate, 2005; Gillies, 2004). This may be a small step for the schools, but it is a giant leap for both pupils and teachers. Research reveals that introducing CL is not particularly difficult, but implementing it successfully is (Gillies, 2003; Hoy & Tschannen-Moran, 1999; Meloth & Deering, 1999). Most teachers are not equipped to implement CL effectively (Gillies, 2003). In the case of the pupils the same applies: learning to effectively work together takes considerable time and effort (Cohen, 1994).

In this study we focused on the teacher's behavior and pupil background characteristics as important determinants of the effectiveness of CL (in terms of pupils' academic as well as social performance). In the remainder of this introduction we will outline the major theoretical views on CL and sketch what distinguishes it from other educational methods. Next, we will discuss some earlier studies in which the teacher's behavior and pupil background characteristics were investigated in the context of CL. We will now discuss our perspective on CL, our hypotheses and provide an outline of the content of this thesis.

1.1. Deutsch: three perspectives on learning

According to Deutsch (1949), three types of learning can be distinguished; 1), competitive, 2) individualistic, and 3) cooperative learning. Competitive learning is characterized by a negative relation in goal attainment between peers, which means that students can only reach their goals when their classroom peers fail to do so. This may result in discouraging peers' learning efforts and withholding information or even sharing false information. Competitive learning can be effective with regard to well-learned, simple tasks (Johnson & Johnson, 1994), but appears ineffective with new or complex tasks - defined as cognitively challenging tasks requiring expertise in multiple-knowledge domains- (Qin et al., 1995). Kohn (1986) argued that gaining success by making others fail is an unproductive way of learning, because, inevitably, only a few winners benefit. The losers, who make up the majority, learn to perceive learning as boring and unfair and evaluate themselves more negatively (Crockenberg, Bryant & Wilce, 1976; Kohn, 1986). The second type of learning is individualistic learning, characterized by the absence of a relation in goal attainment between peers. Students are expected to motivate themselves to perform learning tasks and to refrain from interactions with peers. Again, this type of education can be beneficial for gifted students or with simple, unchallenging tasks. However, with tasks that are new or complex, individualistic learning is less effective since pupils are unlikely to grasp the essence of all knowledge domains that make up the tasks (Cohen, 1994; Qin et al., 1995).

The third type of learning is cooperative learning, or CL, which is characterized by a positive relation between peers in goal attainment: Students need each other to successfully solve a given task. CL is the subject of this study. A large body of literature shows that CL can yield higher learning gains than individualistic and competitive learning methods (e.g., Qin et al., 1995; Rohrbeck et al., 2003) and that it impacts on a wide range of academic and social skills. In the context of this thesis, the most important skills that can be boosted by CL are math performance (e.g., De la Mata Benitez, 2003; Webb & Farivar, 1994), reading performance (e.g., Calderón, Hertz-Lazarowitz & Slavin, 1998; Morrow & Smith, 1990), social acceptation (e.g., Prater, Bruhl, and Serna, 1998; Slavin & Cooper, 1999), and peer communication (e.g., Keefer, Zeitz & Resnick, 2000; Mercer, 1996; Webb & Mastergeorge, 2003). Following Cohen (1994) we define CL in this thesis as an educational setting in which pupils work together in a group small enough for everyone to be able to participate in a clearly assigned collective task, without direct supervision from the teacher. Now that we have defined CL we will take a closer look at its theoretical underpinnings.

1.2. The theoretical roots of CL

1.2.1. Socio-cultural theory

Most modern research paradigms investigating the effect of social learning on academic performance build on the sociocultural approach, which originates from Vygotsky. In Vygotsky's view, learning is defined as a social process in which individual learners internalize knowledge that has been shared between learners by mediational means (Vygotsky, 1978, first published in 1930). Mediational means are semiotic tools by means of which a social action is communicated (i.e. all aspects of the context of social learning by which information can be conveyed). In his view, the process of interaction (speech) is essential for cognitive development. Vygotsky asserted that there are two types of knowledge: lower-order, or natural, knowledge and higher-order, or sociocultural, knowledge. Higher-order knowledge (e.g., memory functions), other than lower-order knowledge -which develops naturally (e.g., reflexive behavior)- is a culturally intersubjective process of internalizing (i.e. social) stimuli that was initially external, expressed in language (Moll, 1994). Intersubjectivity is defined as shared knowledge of what is being interacted or worked on (Levine, Resnick & Higgins, 1993). Development occurs as the lower-order knowledge is transformed into higher-order knowledge, which is mediated by the use of cultural tools, for instance language. As such, development can be thought of as an internalization of external, social stimuli. Intersubjectivity between young learners only occurs if the interactors have an adult supervisor to guide them or if they differ in their developmental level of cognitive capacities. However, learning will be hindered if the distance between the least and the most cognitively able interactors becomes too large, (e.g., there is no sense in explaining children how to use the internet when they do not know yet how to read and write). The distance between the most and the least capable learners should be just large enough for the least capable learner to benefit from the most capable learner. Vygotsky called this the zone of proximal development (ZPD). He also argued that cultural tools (e.g., language, culturally based customs) are essential for sharing knowledge between interactors. He emphasized that cultural diversity could enrich the process of social learning, because it offers multiple perspectives on the knowledge being shared.

When there is an adult supervisor (e.g., the teacher) to guide the pupils, the transition of the students from assisted (other-regulated) performance to independent (self-regulated) performance is facilitated. This process is called *scaffolding*. Following Webb and Farivar (1994), we view the teachers' behavior during CL as an educational tool that they use to support the pupils' development of high-quality helping behavior. In this thesis, we focus on how teachers stimulate high-quality helping behavior and refer to this type of teacher behavior as *teacher stimulation*.

1.2.2. Observational learning theory

The observational learning theory by Bandura (1997) resembles Vygotsky's socio-cultural theory in some respects. Bandura argues that learning is facilitated when learners are of equal ability and/or age level, because this heightens perceptions of similarity and self-efficacy. The process by which peers explicitly compare each other's performances is called shared social appraisal. It strongly affects students' motivation to learn and the degree to which they consider themselves successful learners (Bandura, 1997). Bandura asserted that observing the successful behavior of better learners stimulates some learners to imitate their learning behavior. He argued that in a typical classroom setting students with a high-level of prior knowledge are perceived as role models for the other learners, not only by themselves but also by their peers and the teacher. However, if the gap between the learning behavior of the role model and that of the other learners becomes too big, learners with low prior knowledge will more readily attribute good marks to external sources (e.g. luck or chance), undermining their motivation to learn. This downward process is reinforced by their peers and the teacher.

1.3. Observational learning and ethnicity

In observational learning, the label 'high prior knowledge' is not restrictively limited to students with high academic prior knowledge, but is also used for students with high social knowledge (i.e. popular students). Bandura asserts that students who are part from an ethnic minority are less likely to be perceived as models. Moreover, minority students are less likely to be perceived as popular than majority students (cf. Coie, Dodge & Copotelli, 1982). Allport (1954) proposed the *intergroup contact hypothesis* to reduce this behavior. This hypothesis states that forming multi-ethnic groups is not enough to combat interethnic bias. Interethnic bias can/will only be countered if four criteria are met. These are: cooperation instead of competition, equal status, common goals, and support from authorities and institutions (Allport, 1954; Van Dick et al., 2004). A meta-analysis carried out by Pettigrew and Tropp (2006) provide support for the importance of the four criteria as specified by Allport. However, Pettigrew and Tropp also demonstrated that these four criteria are not essential to a reduction in interethnic bias. Rather, their presence facilitates positive interethnic relations. Pettigrew and Tropp asserted that it is not the presence of the four conditions, but the exposure time to ethnically distinct groups that is essential for a decrease in bias. That is, the more people from different ethnic groups get to know each other, the more they are inclined to like each other. As such, the intergroup contact hypothesis is interpreted as a longitudinal model in which a fifth criterion, the opportunity to let people become friends, is the core feature (see also Pettigrew, 1998). This notion has received support from other studies (e.g., Eller & Abrams, 2004).

Based on these insights we identified important concepts in CL and set up a study to investigate how it functions in the multi-ethnic classroom. More specifically, we studied the impact of the teacher and two pupil background characteristics (ethnicity and prior knowledge) on CL effectiveness. What makes CL effective? Webb and Palincsar (1996) show in their comprehensive review study that there are different views as to what constitutes effective CL. The studies reviewed revealed that investigating the effectiveness of CL necessitates not only looking at pupils' academic performance, but also investigating their verbal behavior during team work and, not in the last instance, their socio-emotional behavior. Accordingly, in this thesis we focused not only on academic learning gains (i.e., math performance), but also on peer interactions (in terms of helping behavior) and on social skills. The latter is operationalized as pupils' popularity, their motivation to work in teams, and their perceived non-cooperativeness. Other research has corroborated the importance of these socio-emotional behaviors for the pupils' academic performance (Gillies & Ashman, 1997; Hijzen, Boekaerts & Vedder, 2006; Slavin & Cooper, 1999; Wentzel, 1993). In the next section we will discuss some of the results found by educational researchers regarding teacher stimulation and pupil background characteristics in the context of CL.

1.4. Research on CL

1.4.1. Teacher stimulation

Teachers play an important part in CL; what they are doing and not doing affects the quality of their pupils' problem-solving processes considerably. Teachers who encourage pupils to use high-quality helping behavior -characterized by asking for explanations, giving explanations, and applying explanations- boost pupils' quality of peer interactions (Fuchs, Fuchs, Kazdan & Allen, 1999; Gillies & Ashman, 1997; 2000). Gillies and Ashman (1997) demonstrated that teacher stimulation of pupils' high-quality helping behavior increased their cooperativeness, helpfulness, and the provision of explanations -both solicited and unsolicited. In the same vein, stimulating pupils' high-quality helping behavior has been demonstrated to enhance peer tutoring (Nixon & Topping, 2001; Topping, 2005). In an extension of their 1997 study, Gillies and Ashman (2000) found that this also held for pupils with low prior knowledge: these pupils displayed more group involvement and were more helpful towards each other if the teacher stimulated their use of high-quality helping behavior. Moreover, high-quality helping behavior has been found to augment learning gains. For instance, Webb and her colleagues (Webb & Farivar, 1994; Webb & Mastergeorge, 2003; Webb, Troper & Fall, 1995) showed that high-quality helping behavior is related to higher learning gains (see also Topping, 2005). The studies by Webb and her colleagues have revealed that not all types of helping behavior stimulate

pupils' learning gains (see also Vedder, 1985). For instance, providing providing only the right answer without explanation even obstructs learning gains (Webb & Mastergeorge, 2003). Instead, the help providers have to explain *how* they arrived at a given outcome and provide the receiver with the opportunity to apply the help. Webb, Nemer, and Ing (2006) investigated whether and how pupils replicated teacher feedback during a CL program. Their study showed that if a teacher provides feedback that is characterized by low-quality helping behavior, pupils are more inclined to resort to low-quality helping behavior in their team.

Exactly how much the teacher should encourage pupils' helping behavior is still open for debate (see also Cohen, 1994; Webb et al., 2006). Various studies have revealed that adjustments in the stimulation of pupils' high-quality helping behavior should be based on background characteristics of the individual team members (Calderón et al., 1998; Cohen, 1994; Slavin & Cooper, 1999) as well as the team as a whole (Oetzel, 2001). In addition to teacher stimulation we also studied the role of ethnicity and prior knowledge.

1.4.2. Student characteristics: ethnicity and prior knowledge

Several studies have evidenced that pupils' ethnicity and prior knowledge are important characteristics that influence high-quality helping behavior and learning gains in a CL setting.

Ethnicity. A number of studies have shown that the teachers' stimulation of high-quality helping behavior in pupils increases their academic achievement. For instance, Calderón, Hertz-Lazarowitz, and Slavin (1998) showed that the reading performance of immigrant pupils who were stimulated by the teacher to use high-quality helping behavior (structured CL) was enhanced as compared to the performance of pupils in a control condition receiving direct instruction. Webb and Farivar (1994) found that teacher efficacy, related to the active promotion of high-quality peer interactions (provision and reception of help), was associated with students' ethnicity. Active stimulation of pupils' help giving and help receiving behavior increased immigrant pupils' high-quality helping behavior, and reduced their math disadvantage compared to Dutch pupils. As mentioned previously, immigrant pupils in the Netherlands on average have an educational barrier as compared to the national mean (Bosker & Guldemond, 2004; Tesser & Iedema, 2001). Structured CL appears to have the potential to address this problem (Slavin & Cooper, 1999; Webb & Farivar, 1994).

A number of researchers have argued that, in addition to augmenting learning gains, CL is an effective way to reduce interethnic bias (e.g., Eller & Abrams, 2004; McGlothlin & Killen, 2005). Pettigrew (1998) identify pupils' inclination to form interethnic friendships as a reason why CL can reduce inter-ethnic bias (see also Webb & Palincsar, 1996). Pettigrew asserts that the formation of interethnic friendships is the number one reason why inter-ethnic bias decreases. In this thesis we also investigated the effect of CL on interethnic bias. However, our approach

differed from the earlier mentioned research on two grounds. Firstly, we carried out the study in a classroom CL setting instead of a direct teaching setting. Secondly, we investigated whether prolonged interethnic contact during CL stimulated interethnic friendships in addition to strengthening popularity, and decreased non-cooperative behavior in multi-ethnic teams.

Prior knowledge. As mentioned before, ethnicity often interacts with prior knowledge. That is, immigrant pupils often have an educational disadvantage, resulting in lower task relevant knowledge, which we refer to as poor *academic prior knowledge*. We distinguish this type of prior knowledge from a second form, namely *prior knowledge of CL skills*. With respect to academic prior knowledge, Cohen (1994) suggests that teams homogeneous as regards academic prior knowledge perform less well than heterogeneous teams. She argued that pupils with low academic prior knowledge benefit from the help provided by pupils with high academic prior knowledge. Cohen qualifies this notion by suggesting that the cognitive gap between pupils with high and low academic prior knowledge should not be too large. Other studies have revealed that students with high academic prior knowledge are more able to maintain focused on the group tasks and to plan and evaluate their actions (Hmelo, Nagarajan & Day, 2000; O'Donnell & Dansereau, 2000). Puustinen (1998) argues that teacher stimulation is important to compensate for the differences in academic prior knowledge between pupils. More precisely, she asserted that pupils with low academic prior knowledge are less able to self-regulate their learning, which heightens their need for external regulation in a structured CL context. In this thesis we defined academic prior knowledge as prior math knowledge. Regarding the effect of social prior knowledge on CL skills, research has revealed that introducing pupils to CL boosts their social skills (Gillies & Ashman, 1997; Johnson & Johnson, 1994). On the other hand, there is also evidence that the academic learning gains of students without prior knowledge of CL skills are obstructed if they are required to work in a CL setting (Hijzen, Boekaerts & Vedder, 2006). Hijzen et al. argue that this may spring from teachers' unfamiliarity with CL and uncertainty as to how to implement it effectively.

The earlier mentioned studies into prior knowledge and ethnic background support Vygotsky's assertion that cultural diversity can enrich the social learning process. At the same time they qualify Vygotsky's assertion by highlighting that learners must have prior knowledge of CL skills to be able to profit from cultural diversity.

2. Design of the study

Webb and Palincsar (1996) conceive of CL as an elaborate interplay between input, process and outcome processes. They emphasized that the effectiveness of CL had best be assessed by investigating both progress in performance and conceptual development, and socio-emotional

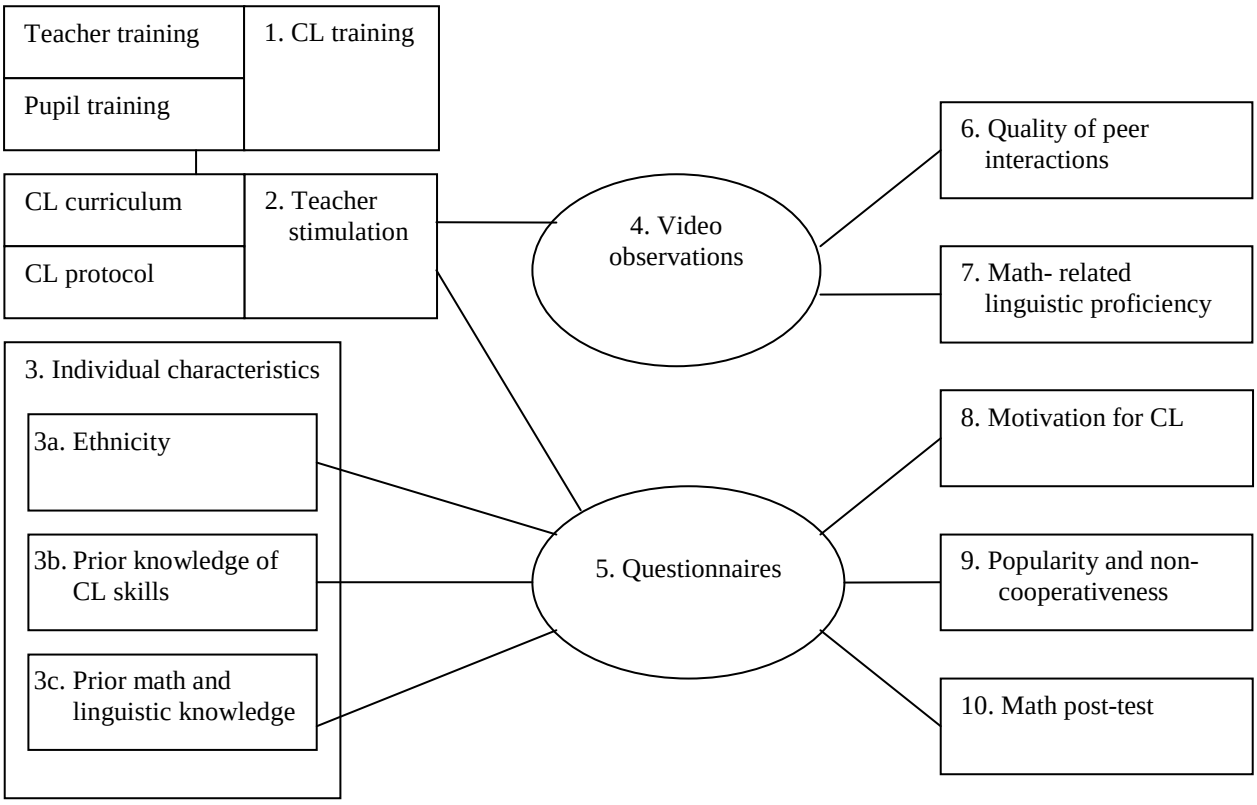

Figure 1. Design of the study

progress. The former type of assessment refers to testing student progress by using formal and informal tests that vary in depth (from automatized skills to complex concepts and reasoning) and range (from a test on a specific concept to a complete curriculum unit). The latter type of assessment refers to tools that register the quality of intergroup relations (quality of peer interactions, perceived peer support, liking classmates and feeling liked by classmates, and non-cooperativeness). In keeping with this, we focused in this thesis on pupils' academic learning gains as well as their gains in the quality of peer interactions and social skills.

2.1. CL training

In the CL study that we carried out to study the effectiveness of CL both the teachers and the pupils were trained how to cooperate effectively (see Figure 1, component 1).

2.1.1. Teacher training

Teachers were randomly assigned to one of two conditions: the experimental or the control condition. After this, the researcher gave a mini-workshop of about two hours to each

teacher individually. Eight of the ten teachers who participated in this study did not have any prior knowledge of CL skills. Therefore, the researcher explained to the teacher what CL was about and how the CL curriculum was made up. Subsequently, the teacher was instructed how to implement CL in their classroom. Following this instruction, the teachers received a written protocol of the two CL training lessons. They were asked to carefully read this protocol and were encouraged to ask clarifying questions prior to the first CL training lesson. The CL protocol for the two CL training lessons was identical for the teachers in the experimental and the control condition.

2.1.2. Pupil training

The teachers trained the children to cooperate effectively in two one-hour lessons. The researcher was present in both lessons to provide the teachers with feedback regarding their implementation of the CL rules. 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’. The rules were practiced in an assignment in which pupils were required to build a bridge between their tables that could bear a small weight -such as a ruler. 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. To ensure that pupils would more readily apply these CL rules, a series of video fragments was shown in which each rule was demonstrated both correctly and incorrectly by two actors. The pupils were asked to indicate in which fragments the CL rules were used correctly and to explain why they thought so. After the CL rules had been demonstrated and discussed, pupils practiced the rules while engaged in a CL math assignment. 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 in all classes, indifferent of the condition. This poster remained there throughout the whole CL curriculum as a memory aid for the pupils. In addition to the poster, all pupils were required to fill in a short checklist halfway each lesson. It served as a check for pupils to see for themselves which CL rules they used (in)adequately. 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; they solely served as a memory booster for the pupils.

2.2. Teacher stimulation

The difference between the experimental condition and the control condition was that the teachers in the experimental condition were instructed to stimulate pupils' high-quality helping behavior (Figure 1, component 2). In the control condition teachers were instructed not to intervene with pupils' use of high-quality helping behavior. That is, pupils were not encouraged to use high-quality helping behavior, but when pupils did make use of high-quality helping behavior they were not discouraged in doing so either. Teachers in the control condition did not object to the fact that they were required to let pupils fend for themselves, because they were not experienced in stimulating the pupils' high-quality helping behavior in any case and they did not have firm believes as to whether stimulating pupils to help other pupils was beneficial for learning. After the CL training, but preceding the implementation of the CL curriculum, the teachers received a protocol with detailed instructions regarding the nine CL math lessons. The researcher observed the first four lessons of the teachers in both conditions and provided feedback as to whether the teachers behaved in accordance with the conditional requirements.

The CL curriculum that the teachers carried out consisted of nine one-hour math lessons, covering five weeks. Each lesson was made up of two CL math assignments, totaling 18 assignments. The assignments dealt with surface, percentage, scale, estimation, and fractions. All assignments were adapted for CL purposes from the *Pluspunt* math curriculum (Bergervoet, Roijackers & Rouvroye, 2001) that employs realistic math problems with a narrative composition. Realistic math problems are characterized by an emphasis on situating math problems in contexts which are familiar to the children, like the zoo and the school yard. The assignments that were 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 individually work on a part of the math task. Secondly, they had to discuss their findings. Thirdly, all team members were required to cooperate to solve the last part of the math task. Each teacher was instructed to convey to the pupils that their job was to understand the CL math assignments, rather than completing them.

2.3. Individual background characteristics

'Ethnicity' and 'prior math knowledge' were assessed prior to the CL curriculum (Figure 1, component 3). Prior math knowledge and linguistic proficiency were assessed with curriculum

independent tests from the national testing service (CITO; Janssen, Kraemer, & Noteboom, 1996). On the basis of pupils' prior math knowledge they were placed in teams that were narrow-heterogeneous in prior math ability (high-middle, or low-middle). The ethnic composition of the teams was not manipulated. 'Prior knowledge of CL skills' was assessed by a teacher questionnaire, on which teachers were required to indicate how much experience they themselves and their pupils had with CL.

2.4. Video observations

We made video recordings of both the peer interactions and the teacher-student interactions (Figure 1, component 4). The teacher-student interactions were recorded to assess the treatment integrity. The recordings of the teachers in the control and the experimental condition were used to assess their stimulation of helping behavior preceding, during, and after the group work for two CL lessons. The peer interactions were videotaped twice to investigate pupils' use of helping behavior (Figure 1, component 6) and their math-related linguistic proficiency (Figure 1, component 7).

2.5. Questionnaires

A number of pupil questionnaires was used to assess background characteristics, math performance, and social learning gains (Figure 1, component 5). All questionnaires were filled in individually. Regarding pupils' social learning gains, they filled in a questionnaire about their motivation for CL (Figure 1, component 8), a questionnaire that assessed their perception of the non-cooperativeness of the other team members, and a questionnaire that required them to rate the popularity of their team members as perceived by the whole class (Figure 1, component 9). These questionnaires were filled in both at the onset and the end of the CL curriculum. The 'math post-test' is a curriculum dependent math test that was filled in by the pupils at the end of the CL curriculum (Figure 1, component 10). It was incorporated to measure pupils' math knowledge of the 19 CL math assignments.

In addition to the pupil questionnaires, we administered a questionnaire to the teachers regarding how much they perceived themselves to stimulate pupils' high-quality helping behavior (related to Component 2, Figure 1). Teachers were required to complete these questionnaires at the end of every week.

3. Sample

A letter with a request to participate in this CL study was sent to 200 schools. The schools were telephoned after approximately a week to inquire whether the teacher(s) from the 5th grade wanted to participate in this study. Only those teachers who met the following three criteria were enrolled in the study: 1) willing to spend time outside the regular curriculum on the preparation of the math lessons, 2) at least 25% of the pupils in their class was immigrant, and 3) they had to employ an authentic math curriculum (Pluspunt).

Ten teachers from ten schools met the earlier mentioned criteria. Each teacher and his/her classroom were then randomly assigned to the experimental and the control conditions. In eight of the ten classes both pupils and teachers had no prior knowledge of CL skills. In the other two classes the teachers had implemented it regularly (one in the control condition and one in the experimental condition). The total sample consisted of 48 teams (N = 166).

4. Analytical perspective

In the past, traditional cognitive psychology viewed mental processes as individual, domain-independent skills (cf. Mayer, 1987). This perspective resonated in early CL studies, where only the effect of CL on the performance of *individual* learners was investigated (e.g., Okebukola, 1986; Peterson & Swing, 1985; Slavin, 1980; Warring, Johnson, Muruyama & Johnson, 1985). Nowadays, most scholars agree that 'experts' utilize general skills as a stepping stone for swifter acquisition of domain-specific, or situational skills (Hatano & Wertsch, 2001). Their emphasis on the importance of situational factors has brought about that CL effectiveness is increasingly being investigated with a multilevel approach. For instance, Rogoff (1995) argued that to assess the full impact of CL on a learner, one has to look at the individual (individual plane), the group of which the individual is part (interpersonal plane), and the social learning setting wherein the group is located (community plane). Theorists like Rogoff (see also Kumpulainen & Mutanen, 1999; Levine et al., 1993; Wertsch, Del Río & Alvarez, 1995) are part of a research tradition called situated cognition. They argue that rather than mediating cognition, social factors *constitute* cognition. That is, the individual is not assumed to be able to take a cognitive skill acquired in one context and apply it in another context, unless the new context facilitates this transfer (see for a more detailed discussion Anderson, Reder & Simon, 1996). In this thesis we acknowledge the interdependence between learners working in teams. Although our samples were too small to adopt a multilevel approach, we have attempted to take the within-group dependency into consideration by analyzing the data at both the individual and the team level.

5. Overview of the thesis and hypotheses

In this thesis we attempted to shed light on teacher stimulation and two pupil background characteristics (i.e. ethnicity and prior knowledge) on the math performance, quality of peer interactions, popularity, motivation for CL, and perceived non-cooperativeness of elementary school pupils aged 10 to 12 year old in a CL math setting. Figure 2 depicts the relations that were investigated. We will present our hypotheses in the order of the output variables.

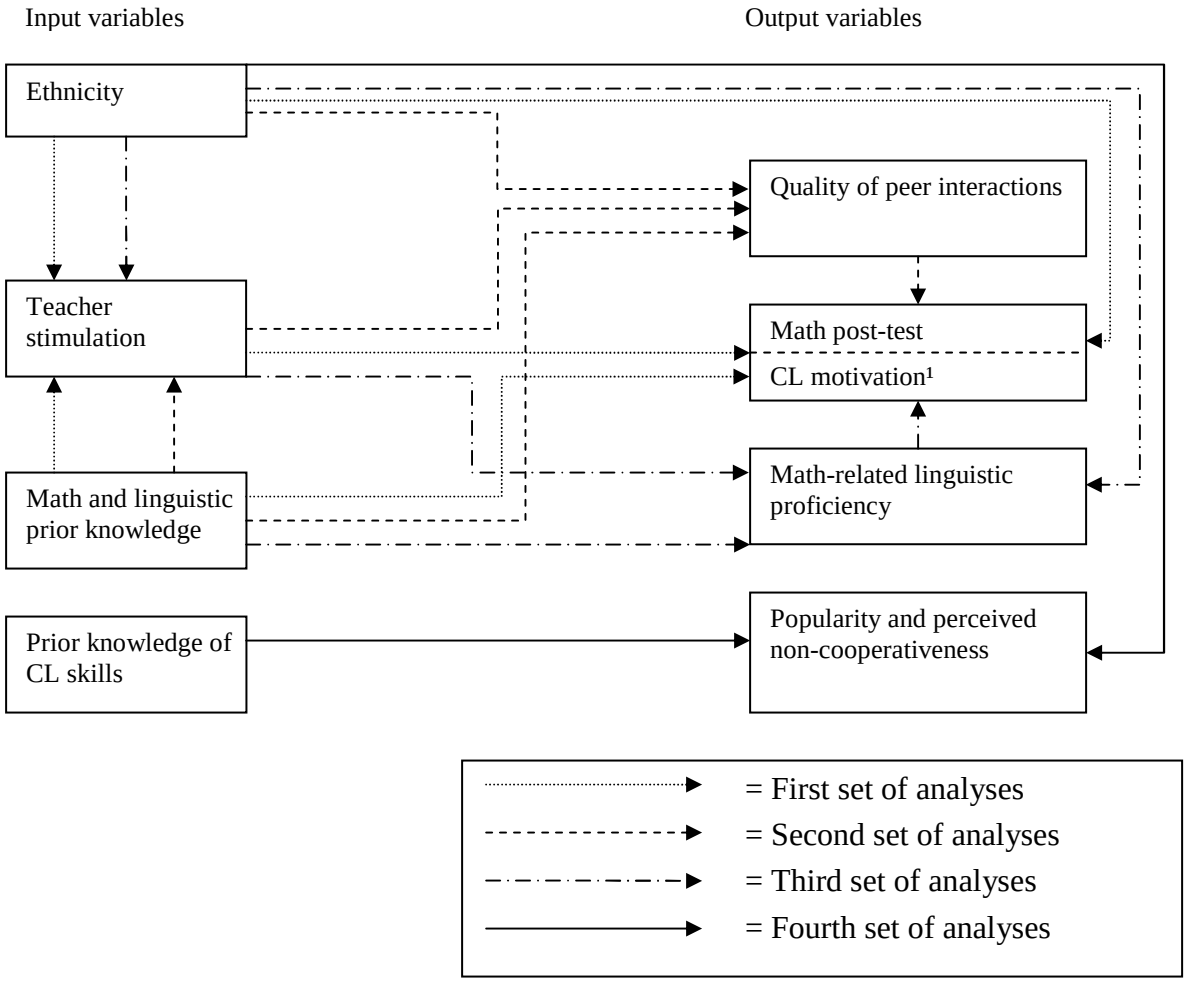

Figure 2. The relations that are investigated in this thesis

¹ This output variable was only measured in study 1.

The empirical body of this thesis covers the four sets of analyses that are visualized in Figure 1. These sets of analyses draw on the same sample (N = 166), but each set highlights a different process or output variable. The sample sizes differ for each set of analyses, due to the

time-consuming nature of the measurement of some variables -such as the videotaped peer interactions- and missing data. The first set of relations we have investigated -depicted by a dotted line in Figure 2- regards the impact of teacher stimulation on pupils' math post-test scores and motivation for CL and how these relations are affected by the pupils' ethnicity and prior math knowledge. In accordance with the studies mentioned previously (e.g., Gillies & Ashman, 1997) we hypothesized that teacher stimulation of the pupils' use of high-quality helping behavior (the experimental condition) boosts their math post-test scores as compared to the control condition, where the teacher does not encourage pupils to use high-quality helping behavior. Furthermore, we investigated how the stimulation of pupils' high-quality helping behavior interacts with the pupil background characteristics ethnicity and prior math knowledge. Whereas Webb and Farivar (1994) focused on ethnicity, and Gillies and Ashman (2000) focused on prior math knowledge, we examined both. We hypothesized that the math post-test scores of the immigrant pupils and pupils with low prior math knowledge are higher if their high-quality helping behavior is stimulated. In addition, we hypothesized that teacher stimulation of the pupils' use of high-quality helping behavior boosts the pupils' motivation to cooperate as compared to the control condition, where the teacher does not stimulate pupils' use of high-quality helping behavior. Extending the findings of Gillies and Ashman (2000), we hypothesized that the motivation to cooperate is higher for the immigrant pupils and pupils with low prior math knowledge in the experimental condition than for immigrant pupils and pupils with prior math knowledge in the control condition. Lastly, we explored whether including linguistic ability in the statistical analyses model results in a better prediction of math post-test scores.

The second and third set of hypotheses that were examined in this thesis concentrated on the pupils' verbal behavior during CL. The second set -depicted by a broken line in Figure 2- targeted the quality of the verbal helping behavior of the pupils (Chapter 3) and the third set -depicted by a broken / dotted line in Figure 2- focused on pupils' math-related linguistic proficiency (Chapter 4). Regarding the quality of verbal helping behavior, we videotaped the pupils' helping behavior and analyzed this with the use of a coding scheme adapted from Webb and Mastergeorge (2003). We aimed to generalize the findings of Webb and Mastergeorge (2003) by demonstrating that not only high-quality helping behavior is positively related to math post-test scores, but tutoring behavior as well. Additionally, we hypothesized that the quality of helping behavior is higher in the experimental condition than in the control condition. In keeping with Gillies and Ashman (2000) we hypothesized that teacher stimulation increases the use of high-quality helping behavior by pupils with low prior math knowledge. In line with the findings by Webb and Farivar (1994) we hypothesized that teacher stimulation augments immigrant pupils' high-quality helping behavior more than that of national pupils. Finally, we investigated the interaction of ethnicity with prior math knowledge: national pupils with low prior math

knowledge were hypothesized to display more high-quality helping behavior in the experimental condition as compared to immigrant pupils with low prior math knowledge.

With respect to pupils' math-related linguistic proficiency we also videotaped the peer interactions and analyzed this with a coding scheme based on studies by Vedder, Kook and Muysken (1996), Levorato and Cacciari (1995), and Niemi (1996). We hypothesized that pupils' math-related linguistic proficiency is positively related to their math post-test scores. In keeping with Calderón et al. (1998), we hypothesized that teachers who stimulate the pupils' use of high-quality helping behavior boost their math-related linguistic proficiency as compared pupils who do are not stimulated by the teachers in their use of high-quality helping behavior. With regard to the experimental condition, we proposed that the math-related linguistic proficiency of immigrant pupils increases more than that of national pupils.

Finally, we investigated the effect of prior knowledge of CL skills on pupils' popularity and perceived non-cooperativeness (Chapter 5). We proposed that team members' perception of intragroup popularity increases and their perceived non-cooperativeness decreases in function of the time they spend working in their team. In keeping with Slavin and Cooper (1999) we hypothesized that both the popularity of immigrant pupils increases and their perceived non-cooperativeness decreases with more CL experience. Lastly, we aimed to extend the finding that the positive effect of CL time on reducing interethnic bias also holds for popularity and perceived non-cooperativeness: we expected that the popularity within ethnically heterogeneous teams is augmented and their perceived non-cooperativeness reduced with increasing CL experience.

These four empirical chapters are followed by a concluding chapter discussing the extent to which the findings reported in the four studies confirmed our hypotheses, what the implications are for future studies, and how our findings translate to the educational setting.

References

Allport, G. W. (1954). *The nature of prejudice*. Oxford, England: Addison-Wesley.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*, 5-11.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W H Freeman / Times Books / Henry Holt & Co.

Bergervoet, S., Roijackers, P., & Rouvroye, R. (2001). *Pluspunt: reken-wiskundemethode voor de basisschool* [Pluspunt: arithmetic-math method for the elementary school]. Den Bosch, the Netherlands: Malmberg.

Bosker, R. J., & Guldemond, H. (2004). *Een herijking van de gewichtenregeling* [A recalibration of the financial arrangements for students with an educational disadvantage]. Groningen, The Netherlands: GION.

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.

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

Coie, J. D., Dodge, K. A., & Copotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology, 18*, 557-570.

Crockenberg, S. B., Bryant, B. K., & Wilce, L. S. (1976). The effects of cooperatively and competitively structured learning environments on inter- and intrapersonal behavior. *Child Development*, *47*, 386-396.

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.

Deutsch, M. (1949). A theory of cooperation and competition. *Human Relations*, *2*, 129-152.

Dixon, J., Durrheim, K., & Tredoux, C. (2005). Beyond the optimal contact strategy: A reality check for the contact hypothesis. *American Psychologist*, *60*, 697-711.

Eller, A., & Abrams, D. (2004). Come together: longitudinal comparisons of Pettigrew's reformulated intergroup contact model and the common ingroup identity model in Anglo-French and Mexican-American contexts. *European Journal of Social Psychology, 34*, 229-256.

Fuchs, L. S., Fuchs, D., Kazdan, S., & Allen, S. (1999). Effects of peer-assisted learning strategies in reading with and without training in elaborated help giving. *Elementary School Journal, 99*, 201-219.

Gillies, R. M. (2003). Structuring co-operative learning experiences in primary school. In R. M. Gillies & A. F. Ashman (Eds.), *Co-operative learning, the social and intellectual outcomes of learning in groups* (pp. 36-53). London: Routledge Falmer.

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 CL on students with learning difficulties in the lower elementary school. *The Journal of Special Education, 34*, 19-27.

Gijsberts, M. (2004). *Ethnic minorities and integration: Outlook for the future*. The Hague, Netherlands: Social and Cultural Planning Office (SCP).

Hatano, G., Wertsch, J.V. (2001). Sociocultural approaches to cognitive development: The constitutions of culture of mind. *Human Development, 44*, 77-83.

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.

Hoy, A. W., & Tschannen-Moran, M. (1999). Implications of cognitive approaches to peer learning for teacher education. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp.257-283). Mahwah, NJ: Lawrence Erlbaum Associates.

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, R. T. (1994). *Learning together and alone: Cooperative, competitive, and individualistic learning*. Englewood Cliffs, NJ: Prentice-Hall.

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

Kohn, A. (1986). *No contest: The case against competition* (8th edition). Boston, MA: Houghton, Mifflin & Co.

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

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.

Levine, J. M., Resnick, L. B., & Higgins, E. T. (1993). Social foundations of cognition. *Annual Review of Psychology, 44*, 585-612.

Mayer, R. E. (1987). *Educational Psychology: A cognitive approach*. Boston: Little, Brown and company.

McGlothlin, H., & Killen, M. (2005). Children's perceptions of intergroup and intragroup similarity and the role of social experience. *Applied Developmental Psychology, 26*, 680-698.

Meloth, M. S., & Deering, P. D. (1999). The role of the teacher in promoting cognitive processing during collaborative learning. In A. M. O'Donnell, & A. King (Eds.), *Cognitive Perspectives on Peer Learning* (pp. 235-255). Mahwah, NJ: Lawrence Erlbaum Associates.

Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction, 6*, 359-377.

Moll, I. (1994). Reclaiming the natural line in Vygotsky's theory of cognitive development. *Human Development, 37*, 333-342.

Morrow, L.M., Smith, J.K. (1990). The effects of group size on interactive storybook reading. *Reading Research Quarterly, 25*, 213-231.

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

Nixon, J. G., & Topping, K. J. (2001). Emergent writing: The impact of structured peer interaction. *Educational Psychology, 21*, 41-58.

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.

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

Okebukola, P. A. (1986). Impact of extended cooperative and competitive relationships on the performance of students in science. *Human Relations, 39*, 673-682.

Peterson, P. L., & Swing, S. R. (1985). Students' cognitions as mediators of the effectiveness of small-group learning. *Journal of Educational Psychology, 77*, 299-312.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*, 65-85.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*, 751-783.

Prater, M. A., Bruhl, S., & Serna, L. A. (1998). Acquiring social skills through cooperative learning and teacher-directed instruction. *Remedial and Special Education, 19*, 160-72.

Puustinen, M. (1998). Help-seeking behavior in a problem-solving situation: Development of self-regulation. *European Journal of Psychology of Education, 13*, 271-282.

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

Rogoff, B. (1995). Sociocultural activity on three planes. In Wertsch, J. V., Del Río, P., & Alvarez, A. (Eds.) *Sociocultural Studies of Mind*. 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.

School Inspectorate (2005). *De staat van het onderwijs: Onderwijsverslag 2004/2005* [State of the educational system: Educational report 2004/2005]. Utrecht: Inspectie van het Onderwijs.

Shelton, J. N., & Richeson, J. A. (2005). Intergroup contact and pluralistic ignorance. *Journal of Personality and Social Psychology, 88,* 91-107.

Slavin, R. E. (1980). Effects of student teams and peer tutoring on academic achievement and time on-task. *Journal of Experimental Education, 48,* 252-257.

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

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

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology, 25,* 631-645.

Van Dick, R., Wagner, U., Pettigrew, T. F., Christ, O., Wolf, C., Petzel, T., Castro, V. S., & Jackson, J. S. (2004). Role of perceived importance in intergroup contact. *Journal of Personality and Social Psychology, 87,* 211-227.

Vedder, P. (1985). *Cooperative learning: A study on processes and effects of cooperation between primary school children*. Dissertation: University of Groningen.

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

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

Warring, D., Johnson, D. W., Maruyama, G., & Johnson, R. (1985). Impact of different types of cooperative learning on cross-ethnic and cross-sex relationships. *Journal of Educational Psychology, 77,* 53-59.

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., & Mastergeorge, A. M. (2003). The development of students' helping behavior and learning in peer-directed small groups. *Cognition and Instruction, 21,* 361-428.

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

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology, 85*, 357-364.

Wertsch, J. V., Del Río, P., & Alvarez, A. (1995). *Sociocultural Studies of Mind*. Cambridge: University Press.

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	95	52 female	40 national
condition			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
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 α = .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 α = .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 α = .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, α = .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, α = .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, α = .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, α = .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, α = .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, α = .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, α = .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, α = .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 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.

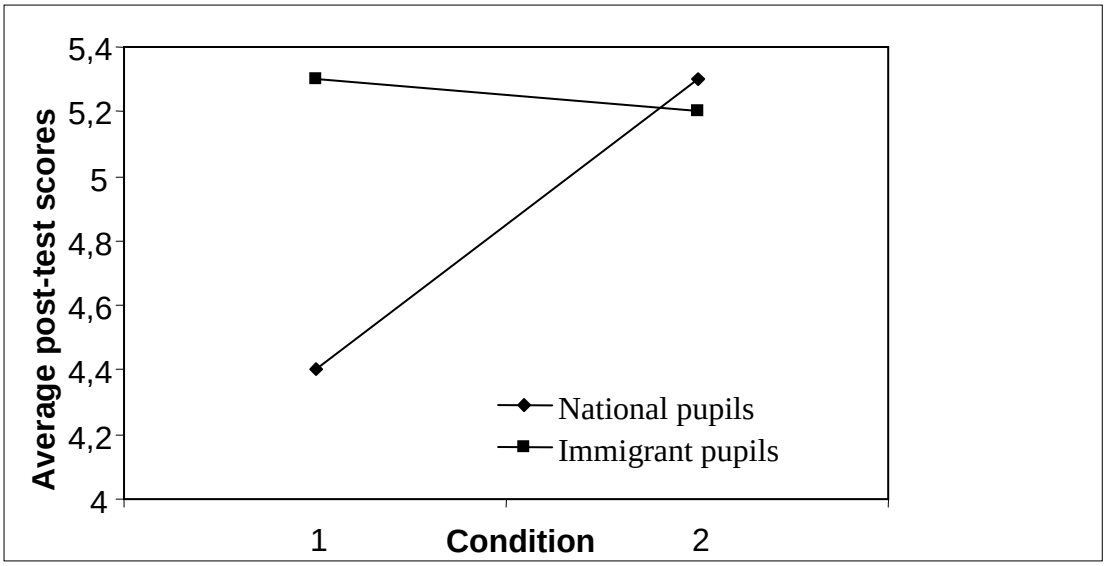

Figure 1. *Interaction 'condition' X 'ethnicity' at the individual level*

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).

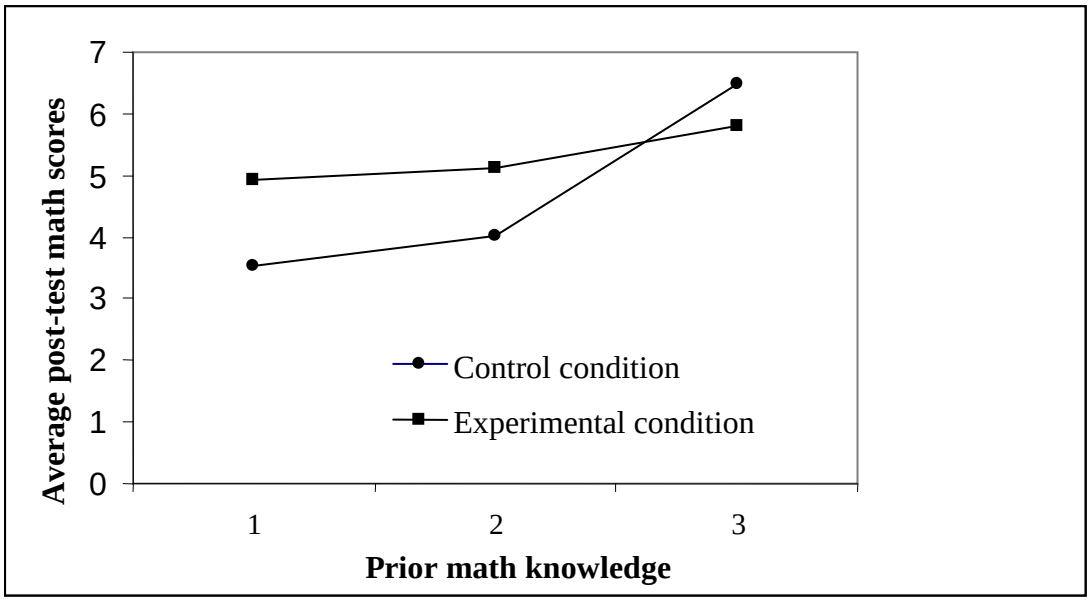

Figure 2. *Interaction 'condition' X 'prior math knowledge' at the group level*

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 [η^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, χ^2 = 6.04, *df* 2, $p < .05$. The same pattern emerged in the experimental condition, χ^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, χ^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 [η^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, $F(2,59) = 3.78$, $p < .05$ [η^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		T2	
		Control (*SD*)	Experimental (*SD*)	Control (*SD*)	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.

CHAPTER 3

Helping behavior during cooperative learning and learning gains: The role of the teacher and of pupils' prior knowledge and ethnic background[1]

Abstract

Is helping behavior (i.e., solicited help and peer tutoring) during cooperative learning (CL) related to subsequent learning gains? And can teachers influence pupils' helping behavior? Subjects were a subsample of study 1. One hundred one 5th grade pupils from multiethnic schools, 10-12 years old, participated in the study. Forty two pupils (31 immigrant) worked in an experimental condition, characterized by the stimulation of solicited high quality help and 59 (24 immigrant) worked in a control condition. It was found that learning gains were predicted positively by pupils' unsolicited helping behavior (i.e., peer tutoring) and negatively by solicited help. Furthermore, teachers were able to affect pupils' low quality solicited help only. Lastly, immigrant pupils used less helping behavior than local pupils, irrespective of CL setting.

Key words: Helping behavior; Mathematical ability; Ethnicity; Cooperative learning.

[1] This Chapter is based on: Oortwijn, Boekaerts, Vedder & Strijbos (in press). Helping behaviour during cooperative learning and learning gains: The role of the teacher and of pupils' prior knowledge and ethnic background. To be published in *Learning and Instruction*.

1. Introduction

Often, during direct teaching in elementary schools there is little room for pupils to ask their classmates for help (Johnson & Johnson, 1994). Pupils often are inclined to ask the teacher for help, since they view the teachers as more able to facilitate learning as compared to their peers (Newman & Schwager, 1993). This is unfortunate since educational research has demonstrated that interactions between peers can augment their learning gains (e.g., Chinn, O'Donnell, & Jinks, 2000; Gillies & Ashman, 2000; Webb & Mastergeorge, 2003). Thus, there is increased interest in the mechanisms that bring about effective peer interactions, that is, the constituents of peer interactions that are related to higher learning gains, the context factors that affect peer interactions, and the relation of peer interactions with learning gains (e.g., Fuchs, Fuchs, Kazdan, & Allen, 1999; Gillies, 2004).

One of the most consistent findings in the literature is the positive effect of high quality verbal helping behavior on learning gains (Gillies & Ashman, 1997; King, 2002; Topping, 2005; Webb & Farivar, 1994; Webb & Mastergeorge, 2003). Following Webb and Mastergeorge (2003), high quality verbal helping behavior is defined here as those utterances of peers that ask for explanations, give explanations, or apply them on the task at hand. In the present study we investigated how high quality helping behavior during cooperative learning (CL) affects pupils' subsequent learning gains. Our study differs from that of Webb and Mastergeorge (2003) in that we investigated both solicited and unsolicited helping behavior and additionally considered three factors that might mediate the relationship between high quality verbal helping behavior and learning gains, namely the role of the teacher and the pupil background characteristics, such as ethnicity and prior knowledge.

1.1. High quality helping behavior

Researchers of peer interactions and learning tend to use different concepts and distinctions when exploring peer interactions. For instance, Nelson-Le Gall and Clor-Scheib (1985) distinguished executive help seeking (i.e., asking for an answer) from instrumental help seeking (i.e., asking for an explanation). Vedder (1985) proposed that for instrumental help to be effective, the help receiver must understand the help given, have an opportunity to apply it, and actually apply it. Webb and her colleagues (Webb, Troper, & Fall, 1995) integrated the aforementioned elements of solicited helping behavior into an elaborate coding scheme that distinguished between high and low quality verbal helping behavior in asking, giving, and applying help.

The effects of both solicited help giving and help receiving have been extensively studied. Studies have consistently reported that the help giver benefits from providing high quality solicited help (e.g., King, 2002; Webb & Mastergeorge, 2003). King (2002) asserts that this is because it stimulates the help giver to elucidate and reorganize knowledge and to recognize mistakes. In a group setting, a learner who is explaining a concept to a peer has to 'tune' the help to the cognitive level of that peer. By reorganizing the knowledge needed for the attuning, the learner who gives the explanation comes to understand the concept more thoroughly. Clearly, solicited high quality helping behavior may also benefit the help receiver. Webb and Mastergeorge (2003) emphasize that high quality help is only useful to the receiver when it is sufficiently elaborated, correct, on time, and links up to the need for help. However, the most accurate predictor of learning gains is whether the help receiver applies the help that is given.

Not all help is asked for. Sometimes a pupil assumes the role of tutor, guiding the problem-solving process of another pupil, the tutee, by asking problem-solving questions or giving assignments that are aimed at solving the problem. We refer to this type of unsolicited helping behavior here as peer tutoring. Topping (2005) showed in a review study that peer tutoring can increase the learning gains of both the tutor and the tutee. Most studies on peer tutoring have focused on cross-age peer tutoring and generally have provided support for a positive relation between peer tutoring and learning gains (Topping, Peter, Stephen, & Whale, 2004). Van Keer and Verhaeghe (2005) suggested that the positive effect on reading performance was larger for cross-age peer tutoring than for same-age peer tutoring. Topping et al. (2004) found that tutoring not only boosted the learning gains of the tutees, but also of the tutors. In their study, the cognitive ability of the tutors was roughly of the same level as that of the tutees. They concluded that same-age peer tutoring might be just as effective as cross-age peer tutoring for students' learning gains. However, they also suggested that peer tutoring is most effective if the participating students are free to be a tutor or a tutee, depending on the nature of the problem. This is in keeping with a study by Robinson, Schofield, and Steers-Wentzell (2003) who argued that cross-age tutoring makes it almost impossible to establish reciprocal tutoring and therefore is less effective than reciprocal same-age peer tutoring.

1.2. Peer interactions during CL and learning gains

1.2.1. The teacher's role

Teachers play an important role in CL. What they are doing and not doing affects the quality of pupils' problem-solving process considerably. Teachers who promote complex cognitive communication between pupils boost the quality of peer interactions and performance (Chinn et al., 2000; Gillies & Ashman, 2000). However, teachers seem to have difficulties to

teach pupils to use high quality helping behavior (Gillies, 2003; Meloth & Deering, 1999). A possible reason is a high task load for the teacher resulting from the requirements for CL (Turner, Meyer, Cox, Logan, DiCintio, & Thomas, 1998). Specifically, teachers need to plan learning activities geared at the acquisition of content knowledge and new domain skills as well as learning activities that help students to cooperate effectively. Additionally, teachers need to guide the CL skills of both individual students and groups as a whole. Effectively and efficiently satisfying these combined requirements takes effort and time on the part of the teacher, which means that promoting effective CL is a long-term project (Webb, Nemer, & Ing, 2006).

Gillies and Ashman (1997, 2000) demonstrated that when teachers successfully stimulate high quality helping behavior, pupils' communicative skills and performance are boosted: they found that pupils had higher learning gains, and their interactions were characterized by more high quality solicited helping behavior and peer tutoring. Other studies have confirmed the positive effect of the stimulation of pupils' elaborated helping behavior on their use of high quality helping behavior (Fuchs et al., 1999) and peer tutoring (Nixon & Topping, 2001).

1.2.2. Student characteristics

Several studies have shown that students' characteristics, such as ethnicity and prior knowledge, influence their helping behavior and learning gains in a CL setting.

Ethnicity. In the Netherlands, there are three major discernable ethnic groups with respect to their performance at school: (a) Moroccan, Turkish, and Antillean youth, (b) Surinamese and other ethnic youth groups (e.g., Asian, former Yugoslavia), and (c) Dutch youth. Tesser and Iedema (2001) have shown that especially the performance of the Moroccan, Turkish, and Antillean groups falls behind. Research has revealed that the linguistic disadvantage of these groups is a possible explanation of their low academic performance (Tesser & Iedema, 2001; Vedder & Horenczyk, 2006). Since these ethnic groups together form the majority of immigrant youth in the Netherlands, it may come as no surprise that the academic performance of pupils in multicultural schools falls below the national mean (Bosker & Guldemond, 2004). However, a study by Webb and Farivar (1994) revealed that the implementation of CL can decrease the educational disadvantage of immigrant pupils. They found that teachers who promoted pupils' use of solicited high quality helping behavior boosted the solicited high quality helping behavior of immigrant pupils and their mathematical disadvantage was reduced, as compared to national pupils (see also Calderón, Hertz-Lazarowitz, & Slavin, 1998).

Prior knowledge. Several studies have shown that students with high prior knowledge are more able to stay focused on the group task and to plan and evaluate their actions (Hmelo, Nagarajan, & Day, 2000; O'Donnell & Dansereau, 2000). Puustinen (1998) argued that pupils with low prior knowledge are less able to self-regulate their learning, which heightens their need

for structured group work and guidance by the teacher. Evidently, immigrant students in the Netherlands fall into this category of pupils.

2. Aims - Hypotheses

In this study we investigated how high quality verbal helping behavior is related to learning gains in a CL setting. We were interested in how the interaction between ethnicity and prior knowledge with the teacher's role affects pupils' helping behavior (both solicited and tutoring) and their subsequent learning. To assess the effect of solicited high quality helping behavior on learning gains we constructed a CL curriculum that borrowed rules for solicited high quality helping behavior from Webb and Farivar (1994) and Webb, Troper, and Fall (1995). We manipulated the role of the teacher: one group of teachers was required to stimulate pupils' solicited high quality helping behavior (experimental condition), while the other group was required not to do this (control condition). Pupils' prior knowledge in mathematics was defined as "mathematical ability". Pupils were enrolled in ethnically diverse classrooms.

Our hypotheses were the following:

Both tutoring behavior and high quality solicited verbal helping behavior will be positively related to posttest mathematical performance (Hypothesis 1). National pupils will have an advantage over immigrant ones as regards the frequency with which they display tutoring behavior and high quality solicited verbal helping behavior because of the more limited linguistic proficiency of immigrant pupils (Hypothesis 2). Pupils in the experimental condition will use more high quality solicited verbal helping behavior than in the control condition (Hypothesis 3). Following Webb and Farivar (1994) and Gillies and Ashman (2000), however, we expected that there will also be interactions of condition, ethnicity, and mathematical ability. Specifically, we predicted that immigrant pupils and pupils with low mathematical ability will display higher learning gains and more solicited high quality helping behavior in the experimental condition. Also, national pupils with low mathematical ability will display more solicited high quality helping behavior in the experimental condition as compared to immigrant pupils with low mathematical ability (Hypothesis 4).

3. Method

3.1. Sample

The total sample comprised 48 groups of 3 to 4 pupils each (N = 166), from 10 classes of 5th grade - see also paragraph 5 of Chapter 1 of this thesis, entitled: *Overview of the thesis and*

hypotheses. Each teacher and his/her classroom were randomly assigned to the experimental or control condition. Ten teachers (nine Dutch, one immigrant; one male, nine female) participated in this study (M = 41 years, SD = 8.6). Four of them were experienced teachers (i.e., teaching more than 10 years, two in the experimental and two in the control condition). Two CL lessons were video-recorded in order to gain a representative picture of the interactions of the pupils. Since we were only interested in the peer interactions, not in the teacher-pupil interactions, we selected only those recordings in which the peer interactions were not interrupted by the teacher. A sub-sample of 27 groups (n = 101) qualified for further analysis, consisting of 53 boys and 48 girls (M = 135.2 months, SD = 6.4). The mean length of the two video recording episodes was 1372.4 seconds (SD = 142.4) and did not differ between conditions.

Table 1

Sample characteristics of the recorded groups

Condition	Number of groups	Ethnicity (number of pupils)	Mean mathematical ability[1] (*SD*)	Mean linguistic proficiency (*SD*)
Control	12	National (11)	3.27 (1.42)	2.35 (.77)
		Immigrant (31)	3.16 (1.10)	2.32 (.81)
Experimental	15	National (35)	3.35 (1.43)	3.15 (.95)
		Immigrant (24)	2.75 (1.19)	2.56 (.81)
Total	27	101		

There were 12 groups in the control condition (n = 42) and 15 groups in the experimental condition (n = 59). The groups were narrow-heterogeneous in terms of mathematical ability (high-middle or low-middle) and were formed by the teacher and the researcher. As regards ethnicity, pupils were considered national when at least one parent was of Dutch origin and immigrant when both parents were of foreign origin. There were 11 national and 31 immigrant pupils in the control condition. In the experimental condition there were 35 national pupils and 24 were immigrant pupils (see Table 1). The majority of the immigrant pupils were of Moroccan or Turkish descent: 67% in the experimental condition and 87% in the control condition. Mathematical ability and linguistic proficiency were used as covariates in the analyses.

3.2. Instruments

As a manipulation check a teacher checklist on CL implementation and videotaped teacher-pupil interactions was used. Pupil performance was assessed with a mathematical pre-

and posttest, a linguistic proficiency test. Helping behavior was assessed with a coding scheme for verbal peer interactions.

3.2.1. 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 α = .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 α = .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 α = .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.

3.2.2. 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 α = .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 α = .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.

3.2.3. Mathematical ability

Scores from a curriculum-independent mathematical test by the Central Institute for Test Research (CITO; Janssen, Kraemer, & Noteboom, 1996) were used to assess the baseline mathematical performance of all pupils. The teachers scored the test for all pupils. Previous research has shown that CITO has a good internal consistency, Cronbach's α = .94 (Evers, Van Vliet-Mulder, & Groot, 2000).

3.2.4. Mathematical posttest

The mathematical posttest consisted of multiple choice items that assessed general knowledge of area, scale, fractions, percentage, and circle diagrams. A reliability analysis of the data obtained in this study revealed that the internal consistency was satisfactory, Cronbach's α = .75. A previous study demonstrated that the mathematical posttest significantly correlated with CITO, r = .77, p < .001 (Oortwijn, Boekaerts, & Vedder, 2005).

3.2.5. Linguistic proficiency

This test was taken from the National Testing Service, used to assess pupils' learning progress in elementary schools (Janssen et al., 1996). The scores of the two dimensions of the test, namely vocabulary and reading comprehension, were averaged in our study into the new variable *linguistic proficiency*. This variable was used to determine whether pupils' language proficiency affected their helping behavior. Thus, it was used as relevant background characteristic of the pupils.

Table 2

The coding scheme of Webb and Mastergeorge (2003) and the coding scheme used in this study

Webb & Mastergeorge (2003)	Present study	Description	Examples
I. Need for help	1. Need for help		
Ia. Request for information	1a. Asking for an answer	No intention to ask for an explanation, typically a yes/no question	L1: This is the area, right?
Ib. Request for general information	1b. Request for an explanation	Typically an open ended question, that asks for a process rather an answer	L1: How do you know that's 9 liters?
Ic. Request for specific explanation	------		
II. Level of verbally received help	2. Level of verbally provided help*		
IIa. Low quality help	2a. Low quality help	Help that only includes an answer / answers	L1: You want to measure how much is… uh.. is in here. For instance, in this one there is 9.

(table continues)

Table 2 (continued)

The coding scheme of Webb and Mastergeorge (2003) and the coding scheme used in this study

Webb & Mastergeorge (2003)	Present study	Description	Examples
IIb. High quality help	2b. High quality help	Help that includes an explanation (with or without answer(s))	L1: You have to measure the length and width and then …uh … you multiply them.
III. Constructive activity on current problem	3. Constructive activity on current problem		
IIIa. Low quality activity	3a. Low quality activity	Help application that does not contain new information (copying / finishing another's calculation)	L1: So it has to do with 6.
IIIb. High quality activity	3b. High quality activity	Help application that includes new information (explanation with or without answer(s))	L1: Ah, I get it. You multiply 3 with 2 to get the area. That's 6.
------	4. Tutor actions (unsolicited help)	Utterance targeted at provoking a problem-solving response from a peer	L1: Area is times. So, the length times the width. That's the area. So, 3 times 2 is? L2: 6.
IV. Constructive activity on next problem	------		

* Category 2 targeted the level of helping behavior of the help provider.

3.2.6. Coding of verbal peer interactions

The coding scheme of Webb and Mastergeorge (2003) on solicited verbal help was adapted to suit the needs of this study. The resulting coding scheme was made up of four categories (see Table 2). Webb and Mastergeorge (2003) discussed the relationship of both help giving and help receiving with learning gains and investigated the relationship of help receiving with learning gains. We investigated both relationships.

Category 1, *need for help*, was composed of two subcategories: (a) *request for an answer* (low quality questions); it comprised *request for information*, e.g., ''What is the answer to this one?'', and *general request for help*, e.g., ''I don't get it'' (see also Webb, Ing, Kersting, and

Nemer (2006) for a discussion of help that is not indicated by a question). (b) *request for an explanation* (high quality questions).

Category 2 targeted the *level of verbally provided help* by the help giver. It was composed of two subcategories: (a) *low quality help*, comprising unclear help, undesired help, and (numerical) outcome only, and (b) *high quality help* (explanation with a (part of a) problem-solving step).

Category 3, *constructive activity on the current problem* by the help receiver, was composed of two subcategories: (a) *low quality constructive activity*, that is, no reaction and acknowledgement of the help received and copying the provided (numerical) outcome, and (b) *high quality constructive activity*, that is, working out part of a problem-solving step and working out one problem-solving step.

We excluded the category *constructive activity on the next problem* from our study. Since authentic, open-ended mathematical tasks were used, none of them were truly independent from each other. Therefore, no satisfactory distinction could be made between the current and the next problem. However, we added another category (Category 4) that assessed the frequency of peer tutoring by counting the *number of tutor actions*. Tutor actions were defined as unsolicited utterances aimed at stimulating a group member to give a problem-solving oriented response by asking a question, giving an assignment, or giving help. In the experimental condition, we only manipulated solicited help, not peer tutoring, and had no pre-set idea of which pupils should be tutors frequently and which pupils would be frequent tutees. A pupil action was scored as tutoring when (a) a tutor asked a problem-solving oriented question, or gave a problem-solving oriented assignment, or explanation, and (b) the tutee gave a problem-solving oriented response – see also the example in Table 2. We marked for each tutor action which pupil took on the role of tutor and which pupil(s) assumed the role of tutee (see Appendix for an example).

The inter-coder reliability was calculated on six recordings (approximately 10% of the total sample) between two observers (the first author and a second coder, unfamiliar with the study). For Category 1 the agreement between the two coders was 83%, and kappa was .73. For Category 2, the agreement was 76%, and kappa was .60. The agreement was 72% for Category 3, and kappa was also .60. For Category 4 (unsolicited help), the agreement was 88% and kappa was .76. In a number of cases (5% of all utterances) one or both of the coders found an utterance to be impossible to be coded. These utterances were removed from the dataset for both coders (pair wise deletion). The second coder, who was blind to the manipulation, individually scored all the videotaped peer interactions.

3.3. Procedure

3.3.1. CL training

During a two-hour workshop the first author explained the essentials of effective CL to the teachers and instructed them how to implement it in the classroom. Subsequently, the teachers trained their pupils in two lessons how to effectively work in groups. In the first lesson general social CL rules were discussed and practiced. These rules required pupils to check whether: ‘‘everyone cooperates’’, ‘‘everyone listens to each other’’, ‘‘everyone shares their knowledge and opinions’’, and ‘‘everyone agrees’’. In the second lesson more specific CL rules were discussed and practiced. Adapted from Webb and Farivar (1994), these rules were about high quality helping behavior. Regarding receiving help, pupils were instructed to (a) ask precise questions, (b) continue asking in case of ambiguities, (c) think before asking a question, and (d) ask for help on time. With respect to giving help, pupils were instructed to (a) fine-tune the level of help to the need for help that is being requested, (b) give a clear and precise answer, (c) let the help receiver apply the help that is given, (d) continue to ask if the question for help is unclear, and (e) give help when needed.

3.3.2. CL mathematical curriculum

The CL mathematical curriculum consisted of nine one-hour lessons followed by a mathematical exam. The teachers in the control condition were required not to intervene in pupils’ interactions. They were instructed only to interact with the pupils to tell them to talk less loudly (and not disturbing other groups), to listen to each other, or to stop making fun of each other. Teachers in the experimental condition were instructed to stimulate pupils’ solicited high quality helping behavior, as specified in the second lesson of the CL training, and to promote their use of general CL rules of the first lesson as much as possible. Lesson-to-lesson protocols were used to help the teachers implement CL in their condition (experimental and control condition).

The mathematical assignments used in this CL curriculum were authentic mathematical assignments. These are mathematical tasks with a strong narrative structure and which are embedded in contexts familiar to the children, such as calculating the area of classrooms in their school. They dealt with area, scale, fractions, percentage, and circle diagrams. All mathematical assignments were adjusted for CL purposes using authentic mathematical assignments from the regular mathematical curriculum. Pupils worked on two assignments per lesson.

During two randomly selected lessons – one somewhere at the beginning (Lesson 1-5) and one near the end of the CL curriculum (Lesson 7-9) – video recordings were made of the peer

interactions in both the experimental and the control condition. After the CL mathematical curriculum all pupils individually completed a mathematical exam.

4. Results

We start with the manipulation check. Regarding the teacher checklist, we found that teachers in the experimental condition reported instructing pupils more in the use of helping behavior throughout the CL curriculum, $t(21) = -3.37$, $p < .005$, than the teachers in the control condition, with a large effect size, Cohen's $d = 1.48$. Remarkably, no differences were found on general CL rules and on extent of feedback on the CL process.

Regarding the videotaped teacher-pupils interactions we found that teachers in the experimental condition elaborated more on the group work at the start and the end of the lesson than teachers in the control condition, $t(16) = -1.78$, $p < .05$, which equates to a moderate effect size, Cohen's $d = .58$. No differences were found for the factor CL activities during group work.

Table 3

Hierarchical regression analysis of pretest mathematical score, request for an explanation, condition, and ethnicity on posttest mathematical performance

Variables	*B*	*SE*	ß	*T*	*F*	*df*	Δ*F*
Step 1						1, 56	
Mathematical pretest	.83	.13	.65***	6.33	40.05		
Step 2						2, 55	5.25
Mathematical pretest	.79	.13	.62***	6.22	24.17		
Request for explanation	-.18	.08	-.23*	-2.29			
Ethnicity	.20	.36	.06	.54			
Condition	.17	.35	.05	.49			

$R^2 = .42$ for Step 1; $R^2 = .47$ for Step 2; $\Delta R^2 = .05$ ($p < .03$).

* $p < .05$, ** $p < .01$, *** $p < .001$.

4.1. Hypothesis 1

4.1.1. Analyses at the individual level

Hierarchical regression analyses of the performance of the individual pupils were employed to investigate whether peer tutoring and high quality solicited helping behavior predicted subsequent mathematical performance. The predictor variables were ethnicity,

condition, mathematical ability, requests for explanations, number of tutor actions provided by the tutor, number of tutor actions received by the tutee, high quality help, and high quality constructive activity. Posttest mathematical performance was the criterion variable (see Table 3). The analysis showed that mathematical ability was the main predictor of posttest mathematical performance. Of the other variables entered in Step 2 mathematical ability and requests for explanations were significant predictors. Requests for explanations were negatively associated with posttest mathematical scores. Remarkably, inclusion of the variables Condition and Ethnicity in the equation did not significantly changed the explained variance.

In a similar hierarchical regression analysis, in which the Number of Tutor Actions Provided by the Tutor and Number of Tutor Actions Received by the Tutee instead of Requests for Explanations were included, only a positive relation between Number of Tutor Actions Provided by the Tutor and posttest mathematical performance was found. After Step 2, mathematical ability and the number of tutor actions provided by the tutor were positively predicting posttest mathematical performance (see Table 4).

Table 4

Hierarchical regression analysis of pretest mathematical score, tutor actions and ethnicity on posttest mathematical performance

Variables	*B*	*SE*	ß	*T*	*F*	*df*	Δ*F*
Step 1					12.58	1, 25	
Mathematical pretest	.76	.22	.58**	3.55			
Step 2					10.43	2, 24	5.84
Mathematical pretest	.68	.20	.51**	3.39			
Tutor actions provided by the tutor	.05	.02	.37*	2.42			
Tutor actions received by the tutee	.05	.05	.18	1.07			
Ethnicity	-1.03	.66	-.29	-.95			
Condition	-.62	.65	-.17	-1.55			

R^2 = .34 for Step 1; R^2 =.47 for Step 2; ΔR^2 = .13 (p < .03).

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

4.1.2. Analyses at the group level

We performed analyses at the group level in an attempt to corroborate the findings we found at the individual level. A number of researchers have suggested that more valid

conclusions on learning in a social setting can be drawn when one takes into account multiple analytical perspectives (e.g., Rogoff, 1995). Due to the small sample size, the relationship of helping behavior with posttest mathematical performance could not be evaluated with a multi-level approach. Inspired by earlier studies using a similar approach (Gillies & Ashman, 2000; Webb & Farivar, 1994), we conducted analyses at the group level by aggregating individual scores for each team.

The variable Group Level Tutoring Behavior was created as the mean number of tutor actions provided by the tutors in each group. The variable had a severe skewness and kurtosis. A logarithmic transformation reduced the kurtosis and skewness to a value < 1, which is acceptable (DeCarlo, 1997). Partial correlations were calculated for requests for explanations and group level tutoring behavior, corrected for mathematical ability. We found a negative trend for requests for explanations (M = 1.65, SD = 1.34), r = -.36, p < .08 (two-tailed). No relationship was found between group level tutoring behavior and posttest mathematical performance. However, when we selected only those groups with a high number of tutor actions (i.e., at least one tutor action per group member, n = 10), we did find a positive trend, provided we corrected for mathematical ability (M = .57, SD = .44), r = .63, p < .07 (two-tailed).

4.2. Hypothesis 2

The linguistic proficiency of the immigrant pupils was lower than that of the national pupils, $t(88) = 3.46$, $p = .001$. We explored whether the lower linguistic proficiency of the immigrant pupils was related to the use of tutor actions. We expected that if national pupils provided more tutor actions than immigrant pupils, this would be related to a more limited linguistic proficiency of immigrant pupils. An independent samples t-test revealed a significant effect, $t(21) = 2.89$, $p < .01$. A Mann-Whitney test corroborated this finding, $Z(21) = -2.42$, $p < .02$. National pupils who provided tutor actions had a higher linguistic proficiency than immigrant tutors.

4.3. Hypothesis 3 and 4

We carried out a 2(condition) x 2(ethnicity) MANCOVA to analyze the relationship of condition with high quality solicited helping behavior. Mathematical ability was the covariate. Additionally, we carried out a regression analysis to investigate the effect of the interaction Condition x Mathematical ability and the interaction Ethnicity x Mathematical ability on high quality solicited helping behavior.

4.3.1. Analyses at the individual level

In the above MANCOVA the dependent variables were the seven categories of the coding scheme (i.e., request for an answer, request for an explanation, low and high quality of provided help, low and high quality constructive activity, and tutor actions).

No main effects were found for condition and ethnicity, although there was a trend for condition. We found a 2-way interaction effect for the relation of condition and ethnicity with the level of verbally provided help, Wilks' Lambda = .89, $F(2, 95) = 5.76$, $p < .005$, $\eta^2 = .11$. National pupils in the control condition provided more low quality help than immigrant pupils, $F(1, 38) = 4.94$, $p < .04$, $\eta^2 = .12$ (see Figure 1). In the experimental condition no such difference was found. Regarding low and high quality constructive activity and tutoring no effects were found.

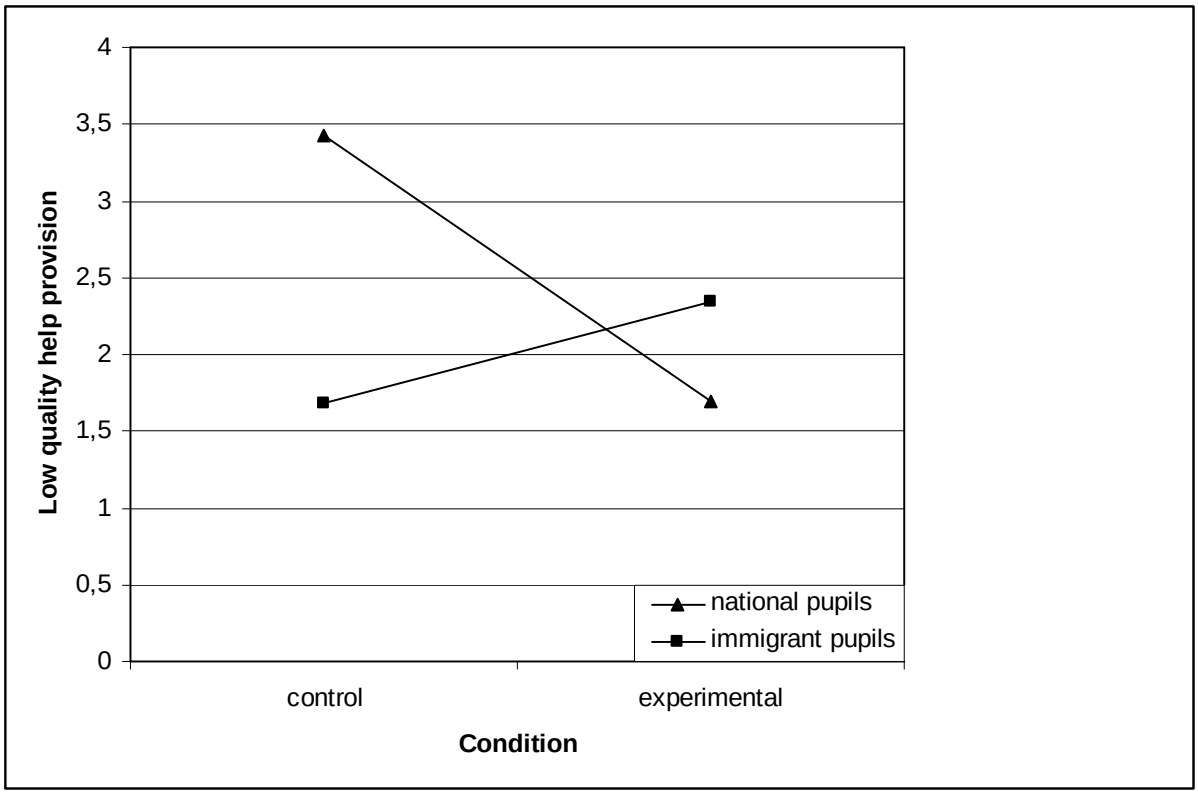

Figure 1. *Mean individual low quality help provided by national pupils (n = 45) and immigrant pupils (n = 56), corrected for mathematical ability in the control and the experimental condition*

We conducted a regression analysis to test the impact of the Mathematical ability x Condition and Mathematical ability x Ethnicity interactions (predictor variables) on the categories of the coding scheme (criterion variables). To compensate for effects of multiple testing, we used a stricter alpha ($\alpha = .01$). No effects were found on the subcategories of Need for

Help. In the subcategories of Level of Verbally Provided Help, the interaction Ethnicity x Mathematical ability explained 17% of the variance of high quality help, β = -.42, $t(43)$ = -2.79, p < .01 (see Table 5). Immigrant pupils with low mathematical ability provided more high quality help than national pupils with a low one. Immigrant pupils with medium to high mathematical ability provided less high quality help than national pupils with a medium to high one. Regarding the subcategories of Constructive Activity on the Current Problem the interaction of ethnicity with mathematical ability explained 19% of the variance in low quality constructive activity, β = -.36, $t(67)$ = -3.08, p < .004. Immigrant pupils with low mathematical ability showed more low quality constructive activity than the respective national pupils. In contrast, immigrant pupils with medium to high mathematical ability showed more low quality constructive activity than the respective national pupils. We found no relations between the aforementioned criterion variables and tutoring (Category 4).

Table 5

Regression analyses of the interaction of mathematical ability and condition, and of ethnicity and mathematical ability at the individual level on the categories of the coding scheme

					Condition x Mathematical ability		Ethnicity x Mathematical ability	
Category	*F*	*df*	R^2	*n*	*T*	β	*T*	β
1. Need for help								
1a. Request for answer	2.33	2, 90	.12	92	-2.27	-.25	-.83	-.09
1b. Request for explanation	.49	2, 56	.04	58	-.15	-.02	-1.09	-.15
2. Verbally provided help								
2a. Low quality help	2.07	2, 94	.04	96	-1.94	-.21	-1.19	-.13
2b. High quality help	4.17	2, 41	.17	43	-1.67	-.25	-2.79*	-.42
3. Constructive activity on current problem								
3a. Low quality activity	7.81	2, 65	.19	67	-2.39	-.29	-3.08*	-.36
3b. High quality activity	3.35	2, 38	.33	40	-2.29	-.37	-2.03	-.33
4. Tutor actions	.90	2, 24	.18	26	-.21	-.05	-1.01	-.26

* p < .01

4.3.2. Analyses at the group level

We investigated whether the interaction effect of ethnicity and condition with low quality help that we found at the individual level was maintained at the group level. We recoded ethnicity into a group-level variable: we considered a group national if there was one or no immigrant

pupil present and we regarded a group as immigrant when there was one or no national pupil in the group. The groups with an even distribution of immigrant and national pupils (n = 4) were left out of the group level analyses. We conducted an ANCOVA analysis in order to examine whether we could support the two-way interaction effect of condition and ethnicity with low quality help, corrected for mathematical ability, which we found at the individual level. The group level analysis corroborated the 2-way interaction effect we found at the individual level, $F(1, 18) = 6.63$, $p < .02$, $\eta^2 = .27$. National groups provided more low quality help in the control condition than immigrant groups. We could not confirm the relation between mathematical ability and ethnicity with high quality help and low quality constructive activity which we found in the analyses at the individual level.

4.4. Summary of the findings

The frequency of requests for explanations was negatively related to posttest mathematical performance and the number of tutor actions provided by the tutor was positively related to posttest mathematical performance, both at the individual and the group level. These findings partly support Hypothesis 1. In addition, we found that national pupils provided more tutor actions than immigrant pupils. This was associated most notably with a lower linguistic proficiency of immigrant tutors, as Hypothesis 2 predicted. Hypothesis 3, however, that predicted a condition effect on high quality helping behavior was not verified. Instead there was an interaction of condition with ethnicity and mathematical ability as Hypothesis 4 predicted. National pupils in the control condition provided more low quality help than immigrant pupils. This finding was corroborated at the group level. Analyses at the individual level further showed that, regardless of condition, immigrant pupils with low mathematical ability provided more high quality help and used more low quality constructive activity than their national counterparts. In contrast, immigrant pupils with medium to high mathematical ability provided less high quality help and used less low quality constructive activity than their national counterparts.

5. Discussion

The finding that the number of tutor actions provided by the tutor was positively related to subsequent mathematical performance of the tutor is in line with other studies that also pointed out the significance of peer tutoring in CL (Duran & Monereo, 2005; Gillies & Ashman, 1997, 2000; Topping, 2005). Topping (2005) suggested that for peer tutoring to be effective for both the tutor and the tutee, it has to be stimulated by the teacher. However, our instructions to the teachers specifically targeted the solicited helping behaviors and not peer tutoring. On the other

hand, the study of Duran and Monereo (2005) suggests that peer tutoring is most successful when the tutor and tutee interact on an equal or reciprocal basis. A stronger relation between peer tutoring and posttest mathematical performance might have been found if the teachers had been instructed to stimulate reciprocal tutoring.

This study also demonstrated that pupils do not have to be the best in mathematics to be successful tutors. There is mounting evidence that peer tutoring is most successful when the tutor and tutee cognitively challenge each other, meaning that their cognitive abilities are roughly the same (Topping, 2005; Topping et al., 2004). The fact that national pupils assumed the role of tutor more often might have to do with their higher linguistic proficiency. It might also be related to their willingness to assume the tutor role.

We found no relationship of provided high quality help and high quality constructive activity with subsequent mathematical performance. A reason could be a transfer problem. In the present study, and different from the Webb and Mastergeorge (2003) study, the mathematical posttest contained problems that were meant to be different from the type of assignments that the pupils completed during the CL lessons. Webb and Mastergeorge (2003) used a program-dependent mathematical test, whereas we used a mathematical test that was more general and program-independent. A second reason might be pupils' lack of CL experience. Limón (2001) suggested that minimal prior CL knowledge hampers productive participation in CL. Our study showed that pupils had minimal experience with and knowledge of CL. It is possible that the pupils adopted the solicited low quality helping behavior when solving a mathematical assignment because they were more comfortable with it. Other studies also reported that teachers are less successful in increasing high quality helping behavior if pupils lack the skills for high quality peer interactions (Prichard, Stratford, & Bizo, 2006). In addition to a lack of CL experience, the mathematical tasks may have been too complex (both linguistically and mathematically) for the pupils – of whom a significant part had both a linguistic and a mathematical disadvantage – resulting in a cognitive overload. Research by Pollock, Chandler, and Sweller (2002) has suggested that pupils who lack the necessary skills to solve complex tasks (i.e., linguistic and mathematical low-achievers) are prone to a cognitive overload. These pupils might be helped more with a highly structured direct teaching setting (see also Tesser & Iedema, 2001).

Regarding the teacher's role during CL we found that teachers were unable to stimulate pupils' high quality helping behavior, although they were successful in reducing low quality helping behavior. Three reasons are put forward here for this finding. First, the CL curriculum was not long enough. Researchers have demonstrated that teaching pupils to use high quality helping behavior is a long-term, time consuming process (Cohen, 1994; Webb et al., 2006). Nevertheless, an earlier study (Webb & Farivar, 1994) did reveal that the teachers were able to

stimulate pupils' solicited high quality helping behavior on a relatively short term. These researchers implemented a CL mathematical curriculum that covered ten weeks and audiotaped the peer interactions in the eighth week. This differs from our study, in which we videotaped the peer interactions of each group twice, the first near the beginning and the second near the end of the CL curriculum. We did this in order to gain a more representative picture of the frequency of the occurrence of the (sub)categories of the coding scheme. It is plausible that we could not corroborate Webb and Farivar's (1994) results because the children were not yet familiar with the use of high quality helping behavior during the first video recording.

Second, teachers in the experimental condition reported instructing their pupils more in the use of solicited high quality helping behavior than in the control condition. Nevertheless, observation of videotaped teacher-pupil interactions revealed that the teachers only instructed their pupils in the use of solicited high quality helping behavior at the beginning and end of the group work; not during the group work. Why did the teachers not give CL feedback when the pupils required it? It might be that the teachers did not accurately perceive the effectiveness of their own CL activities. Earlier studies have made similar suggestions (e.g., Vedder & Veendrick, 2003). Indeed, researchers have suggested that teachers are not well equipped to implement effective CL (e.g., Gillies, 2003). It could also be that, although the teachers in this study were given detailed instructions (both orally and in written form), individual differences in teaching style blurred differences between the two conditions. Previous studies have suggested that teaching style is hard to change (see Gill, Ashton, & Algina, 2004 for a more detailed discussion), which might well mean that changing the teaching style in order to let teachers implement more effective CL takes longer than the 11 lessons this CL curriculum consisted of.

The third reason, which is related to the second, regards the lack of experience of both the pupils and the teachers with CL. Webb et al. (2006) pointed out that pupils tend to copy teacher-pupil interactions in their own interactions with fellow group members. When a lack of CL experience results in the teachers not instructing pupils properly in the use of helping skills and in giving no example of good practice, pupils are likely to model behavior that is poor on examples of help giving. This might explain why we found no difference between the two conditions both with respect to the instruction of solicited high quality help by the teacher and the use of solicited high quality help by the pupils. The aforementioned finding suggests that further research is warranted to investigate how teacher background characteristics (e.g., experience with CL) affect the effectiveness of their teaching behavior during CL.

Finally, we found mixed results for the role of ethnicity in helping behavior. We could not find support for our hypothesis that immigrant pupils with low mathematical ability would use less high quality helping behavior than the respective national pupils. However, we did find that the immigrant pupils in general incorporated less verbal helping behavior in their

communication. These findings are in line with other studies suggesting that immigrant pupils are less actively involved in group work (e.g., Kirchmeyer, 1993). In our study we found that this was due most notably to a lower linguistic proficiency of immigrant pupils as compared to national pupils.

5.1. Limitations

It is conceivable that the frequency of high quality help is higher when the pupils are accustomed to CL. Thus, if we had videotaped two lessons at the end of the CL curriculum we might have found more pronounced differences in high quality helping behavior between the experimental and the control condition. Also, the immigrant pupils were overrepresented in the control condition where in fact they formed the majority. This might have affected the results, although it is important to point out here that the immigrant pupils in the control condition did not differ from the immigrant pupils in the experimental condition with regard to their use of helping behavior.

5.2. Conclusion

Earlier studies have shown that the stimulation of high quality helping behavior during CL is associated with higher learning gains. Our study suggests, however, that incorporating high quality helping behavior in CL not only strains pupils' cognitive capacities, but also places high demands on teacher behavior. It is recommended that future studies implement training programs that take into account relevant background characteristics of both pupils and teachers (see also Webb et al., 2006). Additionally, the data revealed that, although pupils were not trained in the use of peer tutoring, there was a positive relation between peer tutoring and posttest mathematical performance. This suggests that peer tutoring requires less training than solicited high quality help and is better suited to augment mathematical performance. Moreover, peer tutoring may be more effective to reduce pupils' educational disadvantage.

Appendix. Example of a coded interaction fragment

Utterance		Coded category
Pupil V	O!, can I use your ruler, I don't have one.	Organizational utterance (not used in the analyses)
Pupil B	Sure. Do you know what to do now? This is 8.5 and this is 19, ok? So 8 and a half times 2 is 19. So you have to write 8.5 centimeters here and 19 here, ok? (shows on work sheet of V)	Tutoring behavior
Pupil V	Yes, so I have to write here 8.5 centimeters first.	Low quality constructive activity
Pupil B	No! No, don't write. You have to do 8.5 centimeters in length and 19 centimeters in width.	Tutoring behavior
Pupil V	Yes… (starts calculating)	Low quality constructive activity
Pupil B	But not like that! Look like this (writes on the worksheet of V)…	Tutoring behavior (part 1)
Pupil V	(interrupts B) No but…(unintelligible)	Unclear utterance (not coded)
Pupil B	…look 8.5 centimeters. You have to do it like this, like this (shows by writing on V's worksheet)	Tutoring behavior (part 2)*
Pupil V	Is 8…	Low quality constructive activity
Pupil B	No, 8.5.	Low quality help
Pupil V	(writes answer down) And this is 10, right?	Need for help
Pupil B	No, we don't have to do that one yet. Just finish this one.	- Low quality help - Organizational utterance (not used in the analyses)
Pupil A	Has everyone finished?	Organizational utterance (not used in the analyses)
Pupil V	Almost, just filling in the numbers…	Organizational utterance (not used in the analyses)

* Coded as a single tutor action. Although the first tutor action is interrupted by the tutee, the utterance of the tutee is unintelligible and the tutor is not distracted by the interruption, continuing the tutor action.

References

Bosker, R. J., & Guldemond, H. (2004). *Een herijking van de gewichtenregeling* [A recalibration of the financial arrangements for students with an educational disadvantage]. Groningen, The Netherlands: GION.

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.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 3,* 292-307.

Duran, D., & Monereo, C. (2005). Styles and sequences of cooperative interaction in fixed and reciprocal peer tutoring. *Learning and Instruction, 15,* 179-199.

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

Fuchs, L. S., Fuchs, D., Kazdan, S., & Allen, S. (1999). Effects of peer-assisted learning strategies in reading with and without training in elaborated help giving. *Elementary School Journal, 99,* 201-219.

Gill, M. G., Ashton, P. T., & Algina, J. (2004). Changing preservice teachers' epistemological beliefs about teaching and learning in mathematics: An intervention study. *Contemporary Educational Psychology, 29,* 164-185.

Gillies, R. M. (2003). Structuring co-operative learning experiences in primary school. In R. M. Gillies & A. F. Ashman (Eds.), *Co-operative learning, the social and intellectual outcomes of learning in groups* (pp. 36-53). London: Routledge Falmer.

Gillies, R. M. (2004). The effects of communication training on teachers' and students' verbal behaviours during cooperative learning. *International Journal of Educational Research, 41,* 257-279.

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.

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.

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

Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperative, competitive, and individualistic learning*. Englewood Cliffs, NJ: Prentice Hall.

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

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

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction, 11*, 357-380.

Meloth, M. S., & Deering, P. D. (1999). The role of the teacher in promoting cognitive processing during collaborative learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 235-255). Mahwah, NJ: Erlbaum.

Nelson-Le Gall, S., & Glor-Scheib, S. (1985). Help seeking in elementary classrooms: An observational study. *Contemporary Educational Psychology, 10*, 58-71.

Newman, R. S., & Schwager, M. T. (1993). Students' perceptions of the teacher and classmates in relation to reported help seeking in math class. *The Elementary School Journal, 94*, 3-17.

Nixon, J. G., & Topping, K. J. (2001). Emergent writing: The impact of structured peer interaction. *Educational Psychology, 21*, 41-58.

O'Donnell, A. M., & Dansereau, D. F. (2000). Interactive effects of prior knowledge and material format on cooperative teaching. *Journal of Experimental Education, 68*, 101-118.

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

Prichard, J. S., Stratford, R. J., & Bizo, L. A. (2006). Team-skills training enhances collaborative learning. *Learning and Instruction, 16*, 256-265.

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*, 271-282.

Robinson, D. R., Schofield, J. W., & Steers-Wentzell, K. L. (2003). Peer and cross-age tutoring in math: Outcomes and their design implications. *Educational Psychology Review, 17,* 327-362.

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: Cambridge University Press.

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

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology, 25,* 631-645.

Topping, K. J, Peter, C., Stephen, P., & Whale, M. (2004). Cross-age peer tutoring of science in the primary school: Influence on scientific language and thinking. *Educational Psychology, 24,* 57-75.

Turner, J. C., Meyer, D. K., Cox, K. E., Logan, C., DiCintio, M., & Thomas, C. T. (1998). Creating contexts for involvement in mathematics. *Journal of Educational Psychology, 90,* 730-745.

Van Keer, H., & Verhaeghe, J. P. (2005). Effects of explicit reading strategies instruction and peer tutoring on second and fifth graders' reading comprehension and self-efficacy perceptions. *Journal of Experimental Education, 73,* 291-329.

Vedder, P. (1985). *Cooperative learning: A study on processes and effects of cooperation between primary school children*. Dissertation: University of Groningen.

Vedder, P., & Horenzcyk, G. (2006). Acculturation and the school. In D. L. Sam & J. W. Berry (Eds.), *Psychology of acculturation; International perspectives* (pp. 419-438). Cambridge, UK: 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.

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., Ing, M., Kersting, N., & Nemer, K. M. (2006). Help seeking in cooperative learning groups. In S. A. Karabenick & R. S. Newman (Eds.), *Help seeking in academic setting: Goals, groups, and contexts* (pp. 45-88). Mahwah, NJ: Erlbaum.

Webb, N. M., & Mastergeorge, A. M. (2003). The development of students' helping behavior and learning in peer-directed small groups. *Cognition and Instruction, 21,* 361-428.

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

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 α = .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 α = .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.

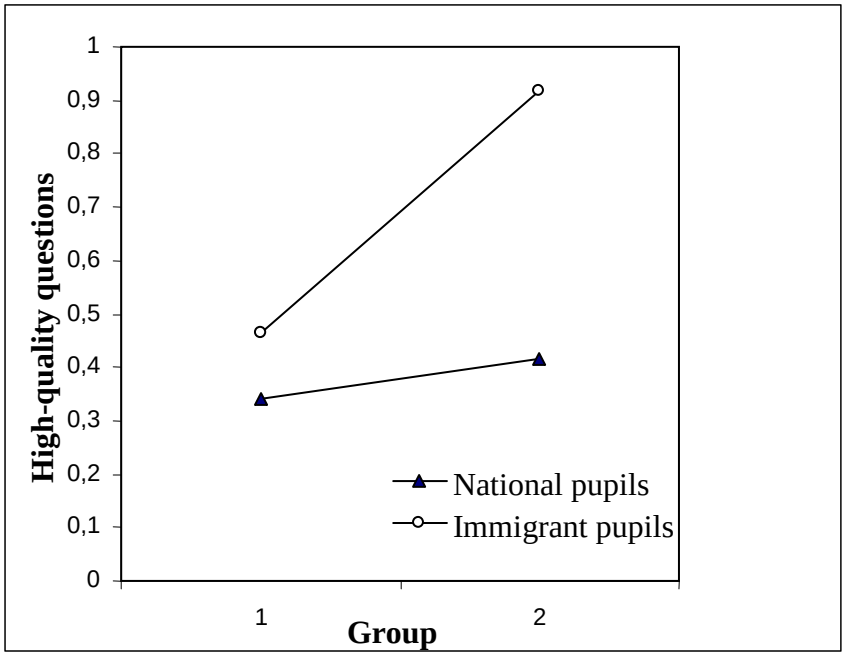

Figure 1. *Interaction of group (experimental or control condition) and ethnicity with the number of high-quality question markers*

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 Psycholinguistics*, *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.

CHAPTER 5

The impact of a cooperative learning curriculum on pupils' social status development and interethnic bias at multi-ethnic elementary schools

Abstract

In this study we investigated popularity and perceived non-cooperativeness in multi-ethnic elementary schools. Subjects were a subsample of study 1. 94 Pupils (26 teams) from five multi-ethnic elementary schools participated in a structured cooperative learning (SCL) curriculum of 11 lessons. Both the teachers and pupils had no prior knowledge of CL skills. The results show that SCL time increased popularity and decreased perceived non-cooperativeness across ethnic background. In addition, prior knowledge of CL skills enhanced the popularity of immigrant pupils and decreased differences in perceived non-cooperativeness between immigrant and national pupils. Importantly, SCL time only raised popularity and decreased perceived non-cooperativeness within ethnically heterogeneous teams. This last result extends the notion that enduring interethnic contact is fruitful for interethnic friendships.

Key words: cooperative learning experience; popularity; perceived non-cooperativeness; interethnic bias; multi-ethnic elementary schools

1. Introduction

Research revealed that the segregation of groups decreases the intergroup relations (Sherif, White & Harvey, 1955). Bettencourt, Dorr, Charlton and Hume (2001) and Dembo and McAuliffe (1987) showed that mere *perception* of distinguishable groups suffices to increase both *intra*group cooperation and *inter*group competition. Various studies have demonstrated that ethnicity is one of the most powerful facilitators of the perception of distinguishable groups (Garza & Santos, 1991; Kirchmeyer, 1993; Nessdale, Maass, Griffiths & Durkin, 2003) and is related to social status differences (e.g., Warring, Johnson, Maruyama & Johnson, 1985). In this study we use the term interethnic bias to refer to situations in which individuals favor people with the same ethnicity over people with a different ethnicity.

Most of the studies into interethnic bias have been carried out in an experimental setting with adult subjects. Phinney, Ferguson, and Tate (1997) argued that - although experimental research has provided important insights - the real challenge lies in the reduction of interethnic bias in the educational setting. In a similar vein, Dixon, Durrheim, and Tredoux (2005) stated that there is a gap between interethnic bias as measured in the experimental setting and how interethnic bias is measured in the field (e.g. in an educational setting). We agree that interethnic bias in an educational setting demands more scientific attention. In the Netherlands, interethnic bias occurs on a daily basis in a great number of multi-ethnic schools, most of which are located in the densely populated Western part of the country (Gijsberts, 2004). One of the major challenges these schools face is how to promote friendship and cooperation among students with different ethnic backgrounds. Gijsberts (2004) emphasized that the incline in the last decades in the number of immigrants has been accompanied by increased segregation, especially in the large cities. Increased segregation in urban areas is a widespread phenomenon that takes place in many countries, like the USA (see for a recent study, Shelton & Richeson, 2005) and Great Britain (Dixon et al., 2005). This phenomenon calls for research into interethnic bias in the multi-ethnic setting. A number of researchers have claimed that interethnic exposure time is an effective means to decrease interethnic bias (e.g., Eller & Abrams, 2004; McGlothlin & Killen, 2005). One aim of this study is to extend these claims. Our study differs from the studies of Eller and Abrams and McGlothlin and Killen on two grounds. Firstly, the two earlier mentioned studies investigated interethnic bias by assessing friendships, we studied interethnic bias by measuring interethnic popularity and perceived interethnic non-cooperativeness. Secondly, we investigated interethnic bias in a structured cooperative learning (SCL) setting rather than in a direct teaching setting. A SCL setting is defined here as an educational method in which pupils are placed in small groups (typically tetrads), and work on assignments that invite them to work together, and

have been trained how to give and receive verbal help, following Webb and her colleagues (Webb & Farivar, 1994; Webb, Troper & Fall, 1995).

In addition to interethnic bias, we studied how an SCL experience is related to a change in the intragroup social status (as measured with popularity and perceived non-cooperativeness). Several studies have found positive relations between time spent in a SCL setting and intragroup cooperation (Gillies & Ashman, 1997; Johnson & Johnson, 1994) and popularity (Wright, Giammarino & Parad, 1986).

In the remainder of this introduction we present the contact hypothesis as our theoretical framework, explore to what extent this hypothesis is supported by earlier studies in naturalistic educational settings, and explain how we investigated it in this study.

1.1. Theoretical background: The intergroup contact hypothesis

Allport (1954) proposed the intergroup contact hypothesis to explain interethnic bias. The contact hypothesis states that grouping people with different ethnic backgrounds is not enough to oppose bias. Interethnic bias will only be countered when four criteria are met. These are: cooperation instead of competition, equal status, common goals, and support of authorities and institutions (Allport, 1954; Van Dick et al., 2004). A meta-analysis carried out by Pettigrew and Tropp (2006) provided support for the importance of the four criteria as specified by Allport. However, Pettigrew and Tropp also demonstrated that the four earlier mentioned criteria are not essential for a reduction in interethnic bias. Rather, their presence facilitates positive interethnic relations. Pettigrew and Tropp asserted that it is not the presence of the four conditions, but the exposure time to ethnically distinct groups that is essential for a decrease in bias. That is, the more people from different ethnic groups get to know each other, the more they are inclined to like each other. As such, the contact hypothesis is interpreted as a longitudinal model, in which a fifth criterion, the opportunity to let people become friends, is the core feature (see also Pettigrew, 1998). Other studies have found support for this notion (e.g., Eller & Abrams, 2004).

1.2. Operationalizing interethnic bias

In the preceding section we mentioned that most studies investigated interethnic bias by asking pupils whether or not they think they can become friends with someone from a different ethnic background (e.g. McGlothlin & Killen, 2005). Few studies have investigated interethnic bias in a multicultural SCL context (e.g. Slavin & Cooper, 1999), Warring et al. (1985) did show that SCL intensified the number of interethnic activities of pupils (school-related activities or activities at home). Johnson, Johnson, and Tiffany (1984) demonstrated that SCL strengthened

interethnic acceptation and support. Little is known about the influence of SCL experiences on the perceived within teams interethnic non-cooperativeness and popularity. This is remarkable since the perception of having cooperative and popular team peers is likely to affect group productivity (Gillies & Ashman, 1997; Johnson & Johnson, 1994).

1.2.1. Popularity

Popularity is usually measured by asking pupils to nominate those pupils in the classroom whom they like most and least. The popularity status of a given student is then calculated by subtracting the standardized unpopular scores from the standardized popular scores. This nomination method has some drawbacks. Firstly, some researchers maintain that the use of nominations gives a distorted impression of pupils' popularity status, since pupils most often only think about who they like most (for a discussion see Maassen & Verschueren, 2005). Secondly, there is evidence to suggest that what researchers define as popularity is not the same as what pupils understand it to be: that is, the traditional operationalization of popularity is argued to lack ecological validity (see Košir & Pečjak, 2005 and Babad, 2001 for a more detailed discussion). Babad argued that the 'classic' method to infer popularity from pupils' ratings about which peers they like the most is an indirect measure, since only the pupils' personal liking and disliking of classroom peers is measured. He proposes a more direct and valid measure of popularity status, which Babad coined judgmental sociometry. Judgmental sociometry refers to the procedure in which pupils are asked to nominate those classroom peers whom they perceive to be the most representative of a social construct. Babad's study suggested that assessing the degree to which pupils are seen as well liked by everyone is a more valid operationalization of popularity.

An American study by Coie, Dodge, and Copotelli (1982) suggested that immigrant students in general are less popular than white pupils since they form a minority group (see also Kistner, Metzler, Gatlin & Risi, 1993). We argue that with prolonged exposure to SCL the popularity of immigrant pupils increases as compared to that of national pupils.

1.2.2. Perceived non-cooperativeness

This is another way to assess interethnic bias. Pupils are asked to nominate team members whom they perceive to be non-cooperative during SCL. Research has shown that SCL time is positively related to a rise in pupils' cooperativeness (Gillies & Ashman, 1997; Johnson & Johnson, 1994). An American study by Hallinan and Teixeira (1987) demonstrated that black pupils were more positive towards other pupils in their team than were white pupils. Other studies have revealed that a SCL experience can boost the popularity of immigrant pupils and decrease the difference between national and immigrant pupils regarding their interethnic cooperativeness (e.g., Slavin & Cooper, 1999).

2. Hypotheses

On the basis of the preceding section, we test three hypotheses in this paper. Firstly, we attempt to corroborate the findings of other researchers (e.g., Gillies & Ashman, 1997, and Wright, Giammarino & Parad, 1986) that SCL time is positively related to pupils' perceived popularity and negatively related to pupils' perceived non-cooperativeness. Secondly, based on Slavin and Cooper's (1999) study we hypothesize that a SCL experience augments the popularity of immigrant pupils and decreases the difference in perceived non-cooperativeness between national and immigrant pupils. Thirdly, we hypothesize that a SCL experience heightens the popularity within ethnically heterogeneous teams and lowers the perceived non-cooperativeness. In order to do so we contrast ethnically heterogeneous teams and ethnically homogeneous teams. A difference between these two types of teams clarifies whether or not the salience of ethnicity diminishes as a function of SCL time. We hypothesize that popularity increases as a function of SCL time whereas perceived non-cooperativeness decreases as a function of SCL time. This holds in the ethnically heterogeneous teams only.

3. Method

3.1. Sample

A SCL curriculum of 11 lessons was carried out in the 5th grade of five multi-ethnic elementary classrooms (i.e. classes with more than 25% immigrant pupils). The first two lessons of this curriculum covered a SCL training in which pupils were instructed in the use of basic SCL rules and helping behavior. During lesson three to 11 pupils worked in teams on math group assignments. 26 Teams participated in this study, consisting of three to four pupils each, amassing 94 pupils (10-12 years old; 43 national, 51 immigrant pupils; 51 boys and 43 girls). See Table 1 for an overview. This sample reported here draws from the same sample as Chapter 2, 3 and 4 -see also paragraph 5 of Chapter 1 of this thesis, entitled: *Overview of the thesis and hypotheses*. The teams consisted of pupils with comparable mathematical and linguistic skills (determined on the basis of class grades) and with roughly the same age. 18 Teams were heterogeneous in ethnicity and eight teams were homogeneous in ethnicity (either all national pupils, or either all immigrant pupils). The composition of the teams remained fixed throughout the SCL curriculum. All teachers indicated implementing a direct teaching method. Additionally, both the teachers and their pupils reported to have no prior knowledge of CL skills.

Table 1

Sample characteristics and scores on the social status questionnaire

Ethnicity	Number of teams[1]	Gender	Popularity at T1 (*SD*)	Popularity at T2 (*SD*)	Perceived non-cooperativeness at T1 (*SD*)	Perceived non-cooperativeness at T2 (*SD*)
National: 43		Boys: 25 Girls: 18	20.48 (12.51)	23.65 (11.80)	.49 (.60)	.19 (.28)
Immigrant: 51		Boys: 26 Girls: 25	18.56 (11.54)	19.75 (11.74)	.19 (.31)	.10 (.19)
Total	**26**					

3.2. Instrumentation

3.2.1. Popularity

The popularity scale was filled in twice by all pupils: at the start of the SCL curriculum (T1) and at the end (T2). Pupils were required to rate their team members as perceived by the whole class on the behavioral characteristic: ‘‘is well liked by everyone’’. Scores were averaged per pupil, excluding their own scores.

Assessment of the psychometric properties of instruments that aim to measure popularity is notoriously difficult (for a discussion see Terry, 2000). A great many studies use multiple measurements of popularity, as is the case in the present study. A compelling question regarding multiple measurements is whether the test-retest stability is satisfactory. That is, whether students’ scores at the second measurement of popularity can be accurately predicted on the basis of the scores obtained at the first measurement. An extensive literature review by Cilessen, Bukowski, and Haselager (2000) found that popularity categories showed satisfactory short-term stability. In addition, Jiang and Cilessen (2005) demonstrated in a meta-analysis that continuous popularity inventories (like popularity) also have good test-retest reliability and are more stabile than categorical types of popularity classification.

3.2.2. Perceived non-cooperativeness

From lesson four onwards pupils filled in a checklist at the end of every lesson about how well they implemented basic SCL rules and rules on giving and receiving help that they were taught in a SCL training that preceded the SCL math curriculum (see also *Procedure*). All pupils completed eight checklists. The pupils were required on this checklist to nominate team members

who did not implement the SCL rules by writing down the name(s) of these team members. For every lesson we recorded the number of times that a pupil was nominated as non-cooperative by his or her team peers.

3.3. Procedure

The SCL curriculum consisted of 11 lessons, one hour each. The five participating teachers were first instructed by the first author in a mini workshop of two hours how to teach in a SCL setting. Then the teachers taught the pupils rules for effective SCL in a training of two lessons. In lesson 1 basic rules of SCL were introduced to the pupils (‘‘everyone cooperates’’, ‘‘everyone listens to each other’’, “everyone shares their knowledge and opinions’’, and ‘‘check whether everyone agrees’’). These rules were practiced in an exercise, requiring pupils to build a bridge between their tables that could bear a small weight. In lesson 2, pupils were taught rules about giving and receiving help, which were adapted from studies carried out by Webb and her colleagues (Webb & Farivar, 1994; Webb et al., 1995). These rules included for example ‘‘ask precise questions’’ and ‘‘give help when needed’’. Subsequently, pupils practiced the SCL rules in a cooperative math assignment. During lesson three to 11, pupils completed similar cooperative math assignments in fixed teams, under supervision of the teacher. In each lesson two authentic math assignments with a common theme (e.g., the zoo) had to be solved by the pupils. Authentic math assignments are mathematical tasks with a strong narrative structure that are embedded in contexts familiar to the children and to which multiple solutions are possible. We used these assignments because research has demonstrated that assignments with multiple solutions stimulate pupils’ motivation to cooperate (e.g., Chizhik, 2001; Cohen, 1994). Pupils were assured that their job consisted of understanding rather than completing the assignments.

3.4. Analytical perspective

The hypothesis that SCL time is positively related to pupils’ popularity and negatively related to pupils’ perceived non-cooperativeness is analyzed individually. Nevertheless, since individual scores are not truly independent from each other in a SCL setting, an explorative analysis of the relationship of SCL time with popularity and perceived non-cooperativeness at the group level is also incorporated. We performed analyses at the group level in an attempt to corroborate the findings we found at the individual level. Due to the small sample size, the relationship of helping behavior with math post-test scores can not be evaluated with a multilevel approach. Inspired by earlier studies using a similar approach (Gillies & Ashman, 2000; Webb &

Farivar, 1994), we conducted analyses at the group level by aggregating individual scores for each team. Because of the small sample size, nonparametric tests were carried out.

The hypothesis that SCL time only affects popularity and perceived non-cooperativeness within ethnically heterogeneous teams is analyzed at the group level. Due to the fact that the data are non-parametric and the number of teams in the present study is too small to accommodate a repeated measures design, we analyzed the effect of SCL time on popularity for homogeneous and heterogeneous teams separately, using non-parametric tests.

4. Results

4.1. Preliminary analyses

A Pearson's correlation test revealed that there was no statistically reliable correlation between the averaged perceived non-cooperativeness and popularity.

4.2. Main results

4.2.1. Hypothesis 1

With respect to popularity, we found that pupils generally rated their fellow team members as more popular at the end of the SCL curriculum, Wilks' $F(1,93) = 5.37$, $p < .03$, $\eta^2 = .06$. Regarding perceived non-cooperativeness, we found that the frequency with which pupils nominated fellow team members as non-cooperative decreased as a function of SCL time, Wilks' $F(7,87) = 5.63$, $p < .001$, $\eta^2 = .31$.

Analysis at the group level. Regarding popularity, we found that SCL time positively influenced the popularity scores of teams, $Z(26) = -2.07$, $p < .04$. Thus, team members gave higher scores to each other at the end of the SCL curriculum as compared to the start of the SCL curriculum. With respect to the Perceived non-cooperativeness, we combined the nominations on the checklists to create three new group-level variables: T1 (averaged nomination on the first three measurements), T2 (averaged nomination on the fourth to sixth measurement) and T3 (averaged nomination for the last three measurements). A Friedman test showed that the non-cooperativeness nominations decreased as a function of time, $\chi^2(26) = 10.64$, $df = 2$, $p < .006$.

4.2.2. Hypothesis 2

With respect to popularity, a paired samples test revealed an effect for national pupils, Wilks' $F(1,50) = 4.86$, $p < .04$, $\eta^2 = .09$. National pupils were liked more at the end of the SCL curriculum than at the beginning. No such effect occurred for immigrant pupils.

Regarding the perceived non-cooperativeness, a repeated measures test was performed with ethnicity as independent variable and nominations from the perceived non-cooperativeness checklist as dependent variable (T1 to T8). The analysis revealed a significant effect, Wilks' $F(7,84) = 2.50$, $p < .03$, $\eta^2 = .17$. The difference between national and immigrant pupils in perceived non-cooperativeness decreased as a function of SCL time. The difference between national and immigrant pupils was significant at the start of the SCL curriculum, $t(92) = 2.97$, $p < .005$, with national pupils receiving more non-cooperativeness nominations than immigrant pupils. At the end of the curriculum, the difference between the perceived non-cooperativeness nominations of national pupils and immigrant pupils was no longer significant, $t(92) = 1.75$, $p > .08$. National and immigrant pupils both became more cooperative in the perception of their fellow team members, but the perceived non-cooperativeness of national pupils decreased more quickly than that of immigrant pupils.

4.2.3. Hypothesis 3

Ethnicity of each team was recoded into (1) homogeneous (all pupils national or all pupils immigrant), and (2) heterogeneous (one or more immigrant pupils combined with national pupils).

Regarding popularity, we found a trend in favor of our hypothesis that the popularity within ethnically heterogeneous teams is higher at the end of the SCL curriculum as compared to the start of the CL curriculum, $Z(18) = -1.86$, $p < .07$. No effect of SCL time on the popularity within ethnically homogeneous teams was found, $Z(8) = -.56$, $p > .57$.

With respect to perceived non-cooperativeness, a Friedman test revealed that the ethnically heterogeneous teams perceived less non-cooperativeness in their team at the end of the SCL curriculum than at the beginning, $\chi^2(18) = 11.29$, $df = 2$, $p < .005$. For ethnically homogeneous teams the perceived non-cooperativeness did not change as a function of SCL time, $\chi^2(8) = .96$, $df = 2$, $p > .61$.

4.3. Summary of findings

In this study we investigated the development of interethnic bias among pupils from multi-ethnic elementary schools during a SCL curriculum. We hypothesized that the time team members spent working in their team is positively related to their popularity and negatively related to their perceived non-cooperativeness. We further hypothesized that SCL time increases popularity and levels off the difference in perceived non-cooperativeness between national and immigrant pupils. Lastly, we expected that SCL time increases the popularity and lowered the perceived non-cooperativeness within ethnically heterogeneous teams, but not within ethnically

homogeneous teams. The results show that SCL time augmented the popularity scores and decreased non-cooperativeness nominations both at the individual and the group level. At the individual level we found that the SCL curriculum augmented the popularity of immigrant pupils and decreased the difference in perceived non-cooperativeness between national and immigrant pupils. The perceived non-cooperativeness of national pupils showed a steeper decrease than that of immigrant pupils. Lastly, SCL time lowered the perceived non-cooperativeness within ethnically heterogeneous teams and tended to increase the popularity.

5. Discussion

The finding that pupils rated their fellow team members as more popular and cooperative with increasing SCL time is in accordance with earlier findings (Gillies & Ashman, 1997; Johnson & Johnson, 1994; Wright et al., 1986). These findings highlight the positive impact of SCL on pupils' social status. There is a whole strand of research that suggests that SCL is a more effective teaching method than direct teaching, not only regarding social skills, but also with respect to academic performance (for overviews see Qin, Johnson & Johnson, 1995; Rohrbeck, Ginsburg-Block, Fantuzzo & Miller, 2003).

Implementing SCL on a regular basis was a new experience for both the teachers and the pupils who participated in this study. Nevertheless, this study shows that SCL can decrease interethnic bias in multi-ethnic teams even with minimal prior knowledge of CL skills. The results demonstrate that there was a steady decline in team members' perception of the non-cooperativeness of pupils with a different ethnicity with increasing SCL time. Even more positive effects of SCL on interethnic relations are to be expected with more experience in, and knowledge of, SCL.

We found that national pupils were perceived to be more popular at the end of the SCL curriculum. Additionally, the perceived cooperativeness of national pupils increased more quickly than that of the immigrant pupils: they were rated as less cooperative at the start of the SCL curriculum. This finding relates to the suggestion of Oetzler (1998) that national pupils have more difficulty to work in teams, possibly because they have a more individualistic learning style. This is also in keeping with the findings of Hallinan and Teixeira (1987), who found that black pupils in the USA had a generally friendlier attitude towards other pupils in their team than had national pupils. Our study suggests that SCL experiences can counter the tendency of national pupils to work alone.

The results further showed that composing multi-ethnic teams is more effective for the reduction of interethnic bias than composing ethnically homogeneous teams: only in the multi-ethnic teams did the popularity increase and the perceived non-cooperativeness decrease with

more CL experiences. These results confirm the revised contact hypothesis, as proposed by Pettigrew (1998), that the development of interethnic friendship reduces interethnic bias. We showed that not only friendship intensifies with increasing exposure to an ethnically heterogeneous group, but also that the interethnic popularity and perceived interethnic cooperativeness are boosted with increased exposure time. This finding needs to be qualified with regard to the term 'ethnicity'. A study by Kistner et al. (1993) asserted that research into interethnic bias in multicultural classrooms should not only focus on the students' ethnic status in terms of majority or minority per se, but that it also necessitates a focus on the relative majority / minority ratio in a specific classroom. In this respect, Bellmore, Witkow, Graham, and Juvonen (2004) provided evidence that pupils' relative ethnic status is related to maladjustment: pupils who were a majority in their class but a minority in the society showed more maladaptive behavior than pupils who were a minority both in the classroom and the society. Since the number of multi-ethnic schools where the majority of the pupils has an immigrant background is on the rise (Gijsberts, 2004), this assertion is becoming increasingly more important in the educational setting.

5.1. Conclusion

Allport's contact hypothesis has been criticized as containing too many 'exceptions to the rule'. That is, research has revealed many qualifying conditions that have obscured the originally transparent hypothesis (Dixon et al., 2005; Eller & Abrams, 2004). Also, the contact hypothesis does not explain why and how cooperation, equal status, common goals, and support of authorities and institutions decrease bias; it only describes when it does so (Pettigrew, 1998). Our study suggests that Pettigrew's (1998) longitudinal interpretation of the contact hypothesis provides a better explanation for the interethnic bias in ethnically diverse classrooms. In addition, our findings show that SCL may provide a solution to interethnic problems at multi-ethnic elementary schools.

References

Allport, G. W. (1954). *The nature of prejudice*. Oxford, England: Addison-Wesley.

Babad, E. (2001). On the conception and measurement of popularity: More facts and some straight conclusions. *Social Psychology of Education*, *5*, 3-30.

Bellmore, A. D., Witkow, M. R., Graham, S., & Juvonen, J. (2004). Beyond the individual: The impact of ethnic context and classroom behavioral norms on victims' adjustment. *Developmental Psychology*, *40*, 1159-1172.

Bettencourt, B. A., Dorr, N., Charlton, K., & Hume, D. L. (2001). Status differences and in-group bias: A meta-analytic examination of the effects of status stability, status legitimacy, and group permeability. *Psychological Bulletin*, *127*, 520-542.

Chizhik, A. W. (2001). Equity and status in group collaboration: Learning through explanations depends on task characteristics. *Social Psychology of Education*, *5*, 179-200.

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

Coie, J. D., Dodge, K. A., & Copotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology*, *18*, 557-570.

Dembo, M. H., & McAuliffe, T. J. (1987). Effects of perceived ability and grade status on social interaction and influence in cooperative groups. *Journal of Educational Psychology*, *79*, 415-423.

Dixon, J., Durrheim, K., & Tredoux, C. (2005). Beyond the optimal contact strategy: A reality check for the contact hypothesis. *American Psychologist*, *60*, 697-711.

Eller, A., & Abrams, D. (2004). Come together: longitudinal comparisons of Pettigrew's reformulated intergroup contact model and the common ingroup identity model in Anglo-French and Mexican-American contexts. *European Journal of Social Psychology*, *34*, 229-256.

Garza, R. T., & Santos, S. J. (1991). Ingroup / outgroup balance and interdependent interethnic behavior. *Journal of Experimental Social Psychology*, *27*, 124-137.

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.

Gijsberts, M. (2004). *Ethnic minorities and integration: Outlook for the future*. The Hague, Netherlands: Social and Cultural Planning Office (SCP).

Hallinan, M. T., & Teixeira, R. A. (1987). Opportunities and Constraints: Black-White Differences in the Formation of Interracial Friendships. *Child Development, 58*, 1358-71.

Jiang, X. L., & Cilessen, A. H. N. (2005). Stability of continuous measures of sociometric status: A meta-analysis. *Developmental Review*, *25*, 1-25.

Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperative, competitive, and individualistic learning*. Englewood Cliffs, NJ: Prentice-Hall.

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

Kistner, J., Metzler, A., Gatlin, D., & Risi, S. (1993). Classroom racial proportions and children's peer relations: Race and gender effects. *Journal of Educational Psychology, 85*, 446-452.

Košir, K., & Pečjak, S. (2005). Sociometry as a method for investigating peer relationships: What does it actually measure? *Educational Research, 47*, 127-144.

Maassen, G. H., & Verschueren, K. (2005). A two-dimensional ratings-based procedure for sociometric status determination as an alternative to the Asher and Dodge system. *Merrill-Palmer Quarterly*, *51*, 192-212.

McGlothlin, H., & Killen, M. (2005). Children's perceptions of intergroup and intragroup similarity and the role of social experience. *Applied Developmental Psychology, 26*, 680-698.

Oetzel, J. G. (1998). Explaining individual communication processes in homogeneous and heterogeneous groups through individualism-collectivism and self-construal. *Human Communication Research*, *25*, 202-224.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65-85.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*, 751-783.

Phinney, J. S., Ferguson, D. L., & Tate, J. D. (1997). Intergroup attitudes among ethnic minority adolescents: A causal model. *Child Development, 68*, 955-969.

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.

Shelton, J. N., & Richeson, J. A. (2005). Intergroup contact and pluralistic ignorance. *Journal of Personality and Social Psychology, 88*, 91-107.

Sherif, M., White, B. J., & Harvey, O. J. (1955). Status in experimentally produced groups. *American Journal of Sociology*, *60*, 370-379.

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

Terry, R. (2000). Recent advances in measurement theory and the use of sociometric techniques. In A. H. N. Cillessen, & W. M. Bukowski (Eds.), *Recent advances in the measurement of acceptance and rejection in the peer system* (pp. 75–93). San Francisco, CA: Jossey-Bass.

Van Dick, R., Wagner, U., Pettigrew, T. F., Christ, O., Wolf, C., Petzel, T., Castro, V. S., & Jackson, J. S. (2004). Role of perceived importance in intergroup contact. *Journal of Personality and Social Psychology*, *87*, 211-227.

Warring, D., Johnson, D. W., Maruyama, G., & Johnson, R. (1985). Impact of different types of cooperative learning on cross-ethnic and cross-sex relationships. *Journal of Educational Psychology*, *77*, 53-59.

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

Wright, J. C., Giammarino, M., & Parad, H. W. (1986). Social status in small groups: Individual-group similarity and the social "misfit". *Journal of Personality and Social Psychology, 50*, 523-536.

CHAPTER 6

General conclusion and discussion

1. Introduction

In this thesis we carried out a study into the effectiveness of a CL math curriculum. Following Webb and Palincsar (1996), we distinguished four domains as regards the effectiveness of CL. We investigated the impact of teacher stimulation and two pupil background characteristics (ethnicity and prior knowledge) on math post-test scores and motivation for CL (Chapter 2), helping behavior (Chapter 3), and linguistic proficiency (Chapter 4) of 10-12 year old pupils from multi-ethnic classrooms. In Chapter 5, we studied how pupils' experience with CL influenced their popularity and perceived non-cooperativeness in general and how the CL experience affected ethnic differences in popularity and perceived non-cooperativeness in particular. We start this concluding chapter by linking the findings from the four different studies and extracting their key conclusions. Subsequently, we discuss the implications of these findings for multicultural elementary schools.

1.1 The hypotheses put to the test

The hypotheses that were investigated are depicted in Figure 2, Chapter 1. The first set of relations we investigated concerns the relationship between teacher stimulation and the pupils' math post-test scores and motivation for CL. In particular, we examined how these relationships were affected by pupils' ethnicity and their prior knowledge (both mathematically and linguistically). We hypothesized that teacher stimulation (experimental condition) raises the pupils' math post-test scores more when compared to the control condition (where pupils were left to fend for themselves), especially the math post-test scores of pupils with low prior math knowledge and of immigrant pupils. We found partial confirmation for this hypothesis: the math post-test scores of the pupils in the experimental condition were raised when compared with pupils' math post-test scores in the control condition. However, this only applied to the national pupils: the immigrant pupils did not have higher math gains in the experimental condition when compared with the control condition. This finding conflicts with Webb and Farivar's (1994) conclusion that immigrant pupils whose high-quality helping behavior is stimulated achieve higher learning gains, when compared with pupils whose high-quality helping behavior is not stimulated. In line with Gillies and Ashman (2000), we did demonstrate that teams with low prior math knowledge had higher math post-test scores in the experimental condition. We further hypothesized that pupils in the experimental condition are more motivated to cooperate than are

their counterparts in the control condition, especially immigrant pupils and the pupils with low prior math knowledge. The results were not clear-cut. We found that immigrant pupils with low prior math knowledge whose high-quality helping behavior was stimulated were more motivated to cooperate, which is in line with studies by Johnson and Johnson (2003) and Gillies and Ashman (1997). Unexpectedly, the immigrant pupils with high prior math knowledge reported higher motivation to cooperate in the control condition compared with their counterparts in the experimental condition. Post hoc analyses suggested that these pupils were linguistically less proficient, when compared with the national pupils with high prior math knowledge. This might have undermined their motivation to work in a CL setting where the pupils were required to participate actively in productive peer discussions.

In the second set of hypotheses we examined the quality of the verbal helping behavior of the pupils. We hypothesized that high-quality helping behavior (both solicited and unsolicited helping behavior) is positively related to learning gains. The data revealed that unsolicited helping behavior (tutoring behavior) was positively related to learning gains. This is in line with Topping (2005). The tutors were responsible for the positive effect of tutoring behavior on learning gains. We did not corroborate the findings of Webb and Mastergeorge (2003) who found a positive relationship between high-quality help and high-quality help application with learning gains. Remarkably, we found a negative relationship between asking for an explanation and learning gains. We further hypothesized that a possible advantage of national over immigrant pupils in the use of tutoring behavior and high-quality solicited verbal helping behavior is related to more limited linguistic proficiency of immigrant pupils. This explorative hypothesis was supported by the data: National pupils provided more tutor actions than immigrant pupils and this was associated with lower linguistic proficiency of immigrant pupils. Additionally, we hypothesized that pupils in the experimental condition use more high-quality solicited verbal helping behavior than the pupils in the control condition. We found no support for this hypothesis: There was no difference in pupils' use of high-quality helping behavior between the experimental condition and the control condition. This finding does not match with earlier studies that have suggested that stimulating pupils' high-quality helping behavior augments high-quality helping behavior (Fuchs, Fuchs, Kazdan & Allen, 1999). The last hypothesis in the second set stated that immigrant pupils and pupils with low prior math knowledge display more high-quality helping behavior in the experimental condition compared with the control condition. Furthermore, national pupils with low prior math knowledge were hypothesized to display more high-quality helping behavior in the experimental condition as compared to immigrant pupils with low prior math knowledge. The data did not support the hypothesis: Pupils with low prior math knowledge and immigrant pupils did not display more high-quality helping behavior in the experimental condition than in the control condition. Thus, we could not corroborate earlier

findings of Gillies and Ashman (1997; 2000) and Webb and Farivar (1994). We did find that the national pupils in the control condition made more use of low-quality helping behavior compared with the experimental condition.

The third set of hypotheses regarded the impact of teacher stimulation and pupils' ethnicity on their math-related linguistic proficiency. Firstly, we hypothesized that the pupils' math-related linguistic proficiency is positively related to their math post-test scores. The data revealed a negative relation between the math-related linguistic proficiency and the math post-test scores for immigrant pupils in the control condition, which is tentative support for the hypothesis. This finding is in line with other studies that have found that linguistic proficiency is related to academic development (e.g., Abedi & Lord, 2001; Cardelle-Elewar, 1992). Furthermore, we hypothesized that the pupils' math-related linguistic proficiency is higher in the experimental condition than in the control condition. The data supported this hypothesis, which is in line with earlier studies, like the study by Calderón et al. (1998). We also hypothesized that the math-related linguistic proficiency of immigrant pupils in the experimental condition is stimulated more than that of national pupils in the experimental condition. Indeed, we found that the peer talk of immigrant pupils in the experimental condition was characterized by a higher math-related linguistic proficiency as compared to the peer talk of immigrant pupils in the control condition. This finding extends the findings by Webb and Farivar's (1994) study, which showed that teachers who stimulated the pupils' use of high-quality helping behavior augmented the high-quality helping behavior and math gains of immigrant pupils.

The fourth set of hypotheses concerned the effect of CL experience on pupils' popularity and perceived non-cooperativeness. We hypothesized that the team members' popularity ratings increase and their perceived non-cooperativeness decreases in function of the time pupils spend working in their team. This hypothesis was confirmed: pupils perceived their fellow team members to be more popular and less non-cooperative at the end of the CL curriculum than at the beginning of the CL curriculum. This result confirms findings from earlier studies that have demonstrated that experience in CL strengthens the intragroup relations (Furrer & Skinner, 2003; Gillies & Ashman, 1997; Johnson & Johnson, 1994; Wright, Giammarino & Parad, 1986). Furthermore, we hypothesized that the popularity of immigrant pupils is increased and their perceived non-cooperativeness decreased after a CL experience. The data confirmed the first part of the hypothesis regarding the popularity and partly supported the second part regarding the perceived non-cooperativeness. In line with Levy, Kaplan, and Patrick (2004) and Prater, Bruhl, and Serna (1998) we found that the perceived non-cooperativeness of immigrant pupils decreased as a function of CL time. Unexpectedly, we found that the positive effect of CL time on the perceived non-cooperativeness was larger for national pupils than for immigrant pupils. Finally, we hypothesized that the popularity of the ethnically heterogeneous teams is stimulated and the

perceived non-cooperativeness is decreased by a CL experience. The data supported this hypothesis: We found that CL time decreased the perceived non-cooperativeness within ethnically heterogeneous teams and there was a trend for a positive effect of CL time on the popularity within ethnically heterogeneous teams. These findings extend the results from earlier studies that have demonstrated the positive effect of CL time on interethnic friendships (e.g., Slavin & Cooper, 1999).

2. General discussion

The following conclusions are drawn from the earlier mentioned findings: 1) the pupils' use of high-quality helping behavior needs to be supervised by the teacher to guarantee successful CL, 2) a CL experience strengthens pupils' social and academic performance, 3) there is a discrepancy between teachers' self-reported behavior and observed teacher stimulation, 4) ethnic background alone is not enough to predict social and academic performance in CL, and 5) heterogeneous grouping positively impacts on interethnic relations. Below, these conclusions are discussed in more detail.

2.1. Pupils' use of high-quality helping behavior needs to be supervised by the teacher to guarantee successful CL

Whereas Webb and Farivar (1994) found that stimulating pupils' use of high-quality helping behavior increased the performance of immigrant pupils, we found that the national pupils displayed higher learning gains as compared to the immigrant pupils. We found that immigrant pupils were less motivated to cooperate in the experimental condition than in the control condition. Why? In Chapter 3 we showed that immigrant pupils were less active during CL in terms of team conversations in the experimental condition. We also showed both in Chapters 2 and 3 that the linguistic performance of immigrant pupils was lower than that of the national pupils. Thus, limited linguistic proficiency may be argued to be the reason for the fact that immigrant pupils did not perform better in the experimental condition. Webb and Farivar (1994) did not report the linguistic proficiency of the immigrant pupils in their study. Perhaps this is because they did not score lower than the national pupils.

With respect to pupils' use of helping behavior, Chapter 2 and 3 showed that the teachers did not adequately stimulate pupils' high-quality helping behavior in the experimental condition. Thus, pupils did not learn to use more high-quality helping behavior, which may also be an explanation why we did not find a positive relation between pupils' high-quality helping behavior and their math post-test scores. The fact that teachers could not stimulate pupils' high-quality

helping behavior is reminiscent of Webb, Nemer, and Fall's (2006) study, which showed that pupils tend to mimic the behavior that the teacher displays. Nevertheless, the teachers in the experimental condition were able to decrease pupils' use of low-quality helping behavior (giving answers only) when compared with the control condition. Webb et al. (2006) provided an explanation for this finding. They suggested that teachers are more inclined to use corrective feedback in their teacher-student interactions during CL, because they are more familiar with this type of interaction. Consequently, they are effective in reducing this type of helping behavior in the pupil teams when asked to do so (as was the case in the experimental condition).

2.2. CL experiences strengthen pupils' social and academic performance

Two reasons are put forward here to explain the phenomenon that teacher stimulation did not result in more use of high-quality helping behavior by the pupils. The first is that the CL curriculum was too short for the pupils to integrate the high-quality helping behavior in their peer interactions. Other studies have demonstrated that CL experience of the pupils enhances academic performance (e.g., Gillies & Ashman, 1997). The majority of the pupils in the sample had no prior knowledge of CL skills. If the pupils in the experimental condition had been more experienced in CL, the difference in the quality of helping behavior between the pupils in the experimental condition and the pupils in the control condition might have been more striking. Nevertheless, the CL curriculum was long enough to bring about a positive change in the pupils' social status and skills. This suggests that the presence of high-quality helping behavior is not required to increase social status and skills. In addition to the relative short duration of the CL curriculum, another reason for the fact that the teachers could not boost the high-quality helping behavior of the pupils may be the minimal prior knowledge of CL skills and CL training of the participating teachers. Gillies (2003) suggested that most teachers lack the skills to implement CL successfully. It is likely that it was a lack of knowledge on the part of the teachers about how to stimulate pupils' use of high-quality helping behaviors that hindered the use of high-quality helping behavior by the pupils. The fact that the teachers brought down the frequency of pupils' use of low-quality helping behavior (asking for answers, and giving outcomes only) indicates that teachers can influence pupils' helping behavior in a relatively short time. Teachers with more experience in teaching pupils how to use high-quality helping behavior might have induced an advantage in high-quality helping behavior in the experimental condition over the control condition. The assertion that the teachers lacked the necessary CL experience necessitates a qualification of the theoretical model as proposed in Chapter 1, Figure 2. In addition to the pupils' prior knowledge of CL skills, we contend that the prior knowledge of CL skills of the

teacher is of equal importance. To test this, in future studies a closer look at the teacher' prior knowledge of CL skills is warranted (see also Webb et al., 2006).

2.3. Discrepancy between teachers' self-reported behavior and the observed teacher stimulation

In Chapters 2, 3, and 4 we manipulated teacher stimulation during CL by creating an experimental and a control condition. To ascertain whether the experimental and the control condition actually differed with respect to teacher stimulation, we checked the treatment integrity. Teacher stimulation was assessed in two ways. Firstly, teachers were required to indicate on a series of checklists how much they felt they were stimulating pupils' high-quality helping behavior. Secondly, the degree to which the teachers encouraged pupils' high-quality helping behavior was judged by the researcher and an independent observer by coding a number of videotaped CL lessons. The treatment integrity check revealed an interesting fact. The teachers reported implementing more high-quality helping in the experimental condition during CL when compared with the control condition. On the other hand, the two observers judged the teachers as using more whole-class reflections after the group work in the experimental condition compared with the control condition. This finding supports the explanation that the teachers did not have the know-how to successfully stimulate pupils' use of high-quality helping behavior. This result reflects findings from earlier studies (Sharan, 1990; Vedder & Veendrick, 2003), which suggested that there is a difference between what teachers think they are capable of and what they are actually doing. The fact that the teachers were not observably stimulating pupils' high-quality helping behavior during group work coincides with the finding that there was no difference between the experimental and the control condition with respect to pupils' use of high-quality helping behavior.

2.4. Ethnicity alone is not enough to predict social and academic performance in CL

In this thesis we found evidence that prior math knowledge and linguistic proficiency are better predictors of math post-test scores and high-quality helping behavior than ethnicity. Regarding linguistic proficiency, Chapters 2 and 3 showed that the immigrant pupils were considerably less linguistically proficient than national pupils. We suggest that the limited linguistic proficiency of immigrant pupils inhibited their active participation in the team. This suggestion is based on the following two findings: 1) in Chapter 2 we showed that a lower motivation for CL of immigrant pupils with high math prior math knowledge in the experimental condition was associated with a more limited linguistic proficiency as compared to national pupils with high prior math knowledge, and 2) in Chapter 3 we found that the degree to which

immigrant pupils assumed the role of tutor depended at least in part on their linguistic proficiency.

In addition to linguistic proficiency, prior math knowledge was the second contributor to differences between national and immigrant pupils in math post-test scores and helping behavior. Chapter 2 showed that national teams with high prior math knowledge had higher math scores in the experimental condition than national teams with low prior math knowledge. On the other hand, the math scores of immigrant teams with high prior math knowledge did not differ from immigrant teams with low prior math knowledge in the experimental condition. Furthermore, in Chapter 3 we found that immigrant pupils with medium-to-high prior math knowledge gave less high-quality help and displayed less low-quality constructive activity than national pupils with medium-to-high prior math knowledge. We argued that this finding is related to the poorer linguistic proficiency of the immigrant pupils. The requirement in the experimental condition to use specific high-quality helping behavior in peer talk hindered the performance of the immigrant pupils with medium-to-high prior math knowledge and undermined their motivation for CL. In the control condition they were not restricted in any way and thus had a higher motivation for CL. The opposite applied to immigrant pupils with low prior math knowledge. They had higher scores and higher CL motivation in the experimental condition than in the control condition, although their linguistic proficiency was not higher than that of the immigrant pupils with medium-to-high prior math knowledge. It could be that the immigrant pupils with low prior math knowledge were simply not able to comprehend the assignment on their own, in contrast to the immigrant pupils with medium-to-high prior math knowledge.

2.5. Ethnic heterogeneous grouping positively affects interethnic relations

In Chapter 5 we showed that the perceived non-cooperativeness is decreased and the popularity of ethnically heterogeneous teams is increased after a CL experience. This finding confirms Pettigrew's assertion that interethnic bias is reduced after prolonged interethnic exposure only. We assessed interethnic bias with a questionnaire on social status and skills, not by measuring friendships. Other studies have revealed that a CL experience affects interethnic social status development in the same manner as it affects the development of interethnic friendships (e.g., Warring, Johnson, Maruyama & Johnson, 1985). Thus, prolonged contact of team members with different ethnic backgrounds facilitates the interethnic friendships and consequently lowers the interethnic bias in these teams.

3. Limitations and guidelines for future research

This thesis had a number of limitations that need to be taken into account in future studies. First of all, as stated before, the CL curriculum was probably too short to find differences in the application of high-quality helping behavior by both the teachers and the pupils. Practical constraints were responsible for the length of the CL curriculum. Most notably the implementation of a longer, more intensive investigation was prevented by the limited time that the teachers had at their disposal and the high workload they experienced, particularly in multi-ethnic schools (Tesser & Iedema, 2001). We concur with other researchers (Cohen, 1994; Webb et al., 2006) that effective CL is a long-term process: it needs to be implemented over an extended time to assess its full effectiveness. Nevertheless, the gains in the pupils' social skills and status as well as math and linguistic performance in the short term testify to its potential.

A second limitation regards the small sample size. Due to this, multi-level analyses were not performed as they would have provided inaccurate results. Therefore, we analyzed the findings both at the individual level and at the team level by aggregating the individual variables. In any case larger samples are advisable to appropriately analyze the effectiveness of group work. However, this is only half of the story: the other half concerns randomization procedures. In this thesis, randomization was implemented at the classroom level. We were of the opinion that the limited resources of the teachers (time, prior knowledge of CL skills) would make it difficult to let them randomly apply structured or unstructured CL to the different teams within their classroom. Randomization at the team level would probably have decreased the chance of having a biased sample.

A third limitation regards the term 'ethnicity'. A study by Kistner et al. (1993) asserted that research into interethnic bias in multicultural classrooms should not only focus on ethnic status in terms of majority or minority per se, but also required that attention should be paid to the relative majority / minority ratio in a specific classroom. We only measured ethnicity as a dichotomous variable. In future studies it would be more informative to use both pupils' ethnicity in the classroom and the society at large. Furthermore, a number of Dutch studies have revealed that there are a number of academically distinguishable ethnic groups in the Netherlands (Tesser & Iedema, 2001). Thus, it would be interesting to investigate the impact of CL on pupils with different ethnic backgrounds.

The fourth limitation regards the assessment of the teacher background characteristics. We randomly assigned the teachers to either the control or the experimental condition. After being assigned, a workshop was held to instruct the teachers how to carry out effective CL without regard to their teaching styles. We did not adequately control for the teaching style that teachers used prior to the CL curriculum. Earlier research has suggested that teaching style is

hard to change (Gill, Ashton, and Algina, 2004; Webb et al., 2006). The study of Webb et al. (2006) resembles the CL study carried out in this thesis. They introduced CL to teachers who had no prior knowledge of CL skills and instructed them to target the high-quality helping behavior of their pupils (10 to 12 year olds). Their results showed that after an intensive training, the peer interactions were characterized predominantly by low-quality helping behavior -instead of high-quality helping behavior. These researchers concluded that teachers need specific training in questioning students and subsequently how to use the students' feedback to guide their joint problem-solving process. This is in keeping with our findings: that teachers themselves need to learn how to ask and give high-quality feedback before teaching their pupils in the use of high-quality helping behavior. Thus, we recommend that in future studies more attention is paid to the teaching style and what aspects of the teaching style need to be revised to assure that the group work is best served.

Samenvatting

Hoe kan de leerkracht het leerproces van leerlingen tijdens een coöperatief leren (CL) curriculum op multiculturele basisscholen optimaliseren? Moet de leerkracht de leerlingen zoveel mogelijk aansturen bij het groepswerken, of ze juist de ruimte geven zelfstandig te werken? In dit proefschrift is getracht een antwoord te vinden op deze maatschappelijk relevante vraag. We hebben dit gedaan door één grote studie uit te voeren naar de effectiviteit van gestructureerde versus ongestructureerde CL. Gestructureerde CL wordt hier opgevat als een onderwijsvorm waarbij de leerkracht het verbale hulpgedrag van hoge kwaliteit van de leerlingen - medeleerlingen vragen om uitleg, uitleg geven aan medeleerlingen en het toepassen van de gekregen uitleg- zoveel mogelijk stimuleert (experimentele conditie). We contrasteerden deze vorm van CL met een ongestructureerde vorm van CL, waarin de leerkrachten zich zo min mogelijk bemoeiden met het verbale hulpgedrag van leerlingen (controle conditie). Klassen werden random toegewezen aan één van beide condities.

Om de effectiviteit van gestructureerde CL vast te stellen werden vier sets hypotheses geformuleerd. De eerste set had betrekking op de academische leerwinst van leerlingen. Er werd onderzocht of leerlingen meer leerwinst boekten en meer gemotiveerd waren om samen te werken in de experimentele conditie dan leerlingen in de controle conditie. Daarbij verwachtten we dat allochtone leerlingen en leerlingen met weinig rekenkundige voorkennis meer leerwinst zouden boeken en een hogere CL motivatie zouden hebben in de experimentele conditie dan in de controle conditie.

De tweede set betrof de kwaliteit van het verbale hulpgedrag van de leerlingen. Er werd onderzocht of leerlingen in de experimentele conditie meer verbaal hulpgedrag van hoge kwaliteit in hun interacties zouden laten zien en of leerlingen die meer gebruikmaakten van hulpgedrag van hoge kwaliteit ook meer leerwinst zouden boeken. Hierbij werd specifiek bekeken of allochtone leerlingen en leerlingen met weinig rekenkundige voorkennis meer verbaal hulpgedrag van hoge kwaliteit hadden in de experimentele conditie dan in de controle conditie.

De derde set had betrekking op de kwaliteit van het rekenkundige taalgebruik van de leerlingen. Er werd onderzocht of het rekenkundige taalgebruik van leerlingen positief samenhing met hun rekenprestaties, of de leerkracht het gebruik van rekenkundige taal door de leerlingen zou kunnen bevorderen en of allochtone leerlingen meer vooruitgang in hun rekenkundige taalgebruik zouden kunnen boeken dan nationale leerlingen.

De vierde set hypothesen tenslotte betrof de ontwikkeling van hun sociale status en sociale vaardigheden. Er werd onderzocht wat de invloed van CL ervaring is op de sociale statusontwikkeling (gemeten als populariteit en waargenomen coöperatief gedrag) van leerlingen en hoe de ervaring met CL de interetnische relaties beïnvloedt.

Om eerdergenoemde onderzoeksvragen te onderzoeken werd een CL curriculum ontwikkeld voor groep 7 leerlingen. CL wordt hier opgevat als een onderwijsvorm waarin leerlingen in kleine groepen zelfstandig werken aan een duidelijk omschreven groepstaak, zonder directe begeleiding van de leerkracht. De basis voor het CL curriculum was een bestaand rekenprogramma, te weten Pluspunt. Er werd gekozen voor een rekencurriculum, omdat het bij rekenen relatief gemakkelijk is om opgaven te creëren die leerlingen aanmoedigen om samen te werken. In de Pluspunt-lesmethode wordt gebruikgemaakt van realistische rekenopgaven, hetgeen betekent dat de rekenopdrachten uit de opgaven ingebed zijn in concrete contexten die bekend zijn bij de leerlingen. Er werden 19 rekenopgaven aangepast voor het CL curriculum; daarnaast werd een rekentoets ontwikkeld op basis van deze 19 rekenopgaven. De leerlingen moesten per les 2 rekenopgaven maken: De duur van het CL curriculum was 9 lessen. Hiernaast werd 1 rekenopgave gebruikt als oefening in een CL training van 2 lessen die de leerlingen kregen voorafgaande aan het CL curriculum en waarin ze de vaardigheden voor effectief CL werd uitgelegd. Na afloop van het CL curriculum maakten alle leerlingen individueel een rekentoets.

Voor de dataverzameling werd een groot aantal scholen eerst aangeschreven en vervolgens gebeld. Uiteindelijk namen 10 leerkrachten deel aan het onderzoek (met in totaal 172 leerlingen onder hun hoede). Vanwege praktische overwegingen vond de dataverzameling plaats in twee ronden. De gegevens werden verzameld met een vragenlijst over de motivatie van leerlingen om samen te werken, CITO toetsgegevens (alleen de onderdelen ‘rekenen algemeen’, ‘begrijpend lezen’ en ‘woordenschat’), afname midden groep 7, een rekentoets gebaseerd op de opgaven uit het CL curriculum en een sociale status vragenlijst. Leerlingen werden op basis van de nationaliteit van hun ouders bestempeld als autochtoon (één of beide ouders van Nederlandse komaf), of als allochtoon (beide ouders van allochtone afkomst).

Met betrekking tot de eerste set hypothesen laten de resultaten zien dat de autochtone leerlingen meer leerwinst boekten in de experimentele conditie, vergeleken met de controle conditie. Verder bleek dat teams met voornamelijk allochtone leerlingen alleen beter scoorden in de experimentele conditie indien zij laag tot gemiddeld presteerden op rekengebied: Hoogpresterende allochtone teams scoorden beter in de controle conditie. Bovendien was de CL motivatie van de hoogpresterende allochtone teams lager in de experimentele conditie dan in de controle conditie. We vonden aanwijzingen dat het feit dat de rekenwinst van allochtone leerlingen in de experimentele conditie niet hoger was dan die van allochtone leerlingen in de controle conditie te maken had met een beperktere taalvaardigheid van de allochtone leerlingen ten opzichte van de autochtone leerlingen. Door een beperktere taalvaardigheid voelden de allochtone leerlingen zich wellicht minder op hun gemak in de experimentele conditie, die werd gekenmerkt door een actieve stimulatie van hun verbale hulpgedrag van hoge kwaliteit. In het

verlengde hiervan suggereren de resultaten dat de beperktere taalvaardigheid van de hoogpresterende allochtone leerlingen een verklaring kan zijn voor een lagere CL motivatie in de experimentele conditie dan in de controle conditie. Het feit dat ze verplicht werden om verbaal hulpgedrag van hoge kwaliteit in hun interacties te integreren kan hun motivatie om actief te participeren in hun team hebben ondermijnd, met als indirect gevolg dat hun leerwinst niet hoger was dan die van de allochtone leerlingen in de controle conditie.

Wat betreft de tweede set hypothesen, waarin de kwaliteit van het verbale hulpgedrag werd onderzocht, stelden we vast dat de leerkrachten niet succesvol waren in het verhogen van het hoge kwaliteit hulpgedrag van de leerlingen, maar wel in het verminderen van het lage kwaliteit hulpgedrag, vooral dat van de autochtone leerlingen. Daarnaast vonden we dat het gebruik van hulpgedrag van hoge kwaliteit door leerlingen positief samenhing met hun rekenwinst en dat de allochtone leerlingen in het algemeen minder hulp gaven dan autochtone leerlingen: dit was gerelateerd aan een beperktere taalvaardigheid van allochtone leerlingen. Het feit dat de leerkrachten niet bij machte waren het hulpgedrag van hoge kwaliteit van de leerlingen te stimuleren werd toegeschreven aan het feit dat acht van de tien leerkrachten geen ervaring hadden met het implementeren van effectief CL. Het feit dat de leerkrachten wel in staat waren de frequentie te verlagen waarmee leerlingen verbaal hulpgedrag van lage kwaliteit hanteerden geeft aan dat de leerkrachten wel degelijk het hulpgedrag van leerlingen konden beïnvloeden op relatief korte termijn. Indien de leerkrachten meer ervaring hadden in het stimuleren van het hulpgedrag van hoge kwaliteit van de leerlingen, dan was er wellicht wel een verschil tussen de experimentele en de controle conditie opgetreden, ten faveure van de experimentele conditie.

Met betrekking tot de derde set hypothesen, betreffende de rekenkundige taalvaardigheid van de leerlingen, werden alle veronderstellingen grotendeels bevestigd: We vonden dat de leerlingen een hogere kwaliteit van rekenkundig taalgebruik hadden indien de leerkrachten hun gebruik van verbaal hulpgedrag van hoge kwaliteit aanmoedigden en dat voor de allochtone leerlingen de toename in rekenkundig taalgebruik in de experimentele ten opzichte van de controle conditie groter was dan voor de autochtone leerlingen. Bovendien vonden we een verband tussen rekenkundig taalgebruik en rekenwinst, maar alleen voor de allochtone leerlingen: een lage kwaliteit van rekenkundig taalgebruik door allochtone leerlingen hing negatief samen met hun rekenwinst. Opvallend was dat het rekenkundige taalgebruik van leerlingen in de experimentele conditie zowel meer taalkundige elementen van lage kwaliteit als hoge kwaliteit omvatte. Kortom, er werd meer gepraat in de experimentele conditie. Een verklaring hiervoor kan zijn dat er in de experimentele conditie een differentiatie optrad in het taalgebruik: de taalkundige elementen van lage kwaliteit hadden betrekking op organisatorische zaken (bijvoorbeeld het verdelen van taken; wie doet wat) en de taalkundige elementen van hoge kwaliteit werden ingezet om gezamenlijk de rekenopdrachten op te lossen.

Tenslotte vonden we ter beantwoording van de vierde set hypothesen, die betrekking hadden op de ontwikkeling van sociale status en sociale vaardigheden, dat ervaring met CL de algehele sociale status binnen teams verhoogde. Daarnaast bleek de populariteit van de autochtone leerlingen meer toe te nemen met meer CL ervaring dan die van de allochtone leerlingen en bleek het verschil tussen de teamleden in het waargenomen coöperatieve gedrag kleiner te worden met toenemende CL ervaring. Als laatste vonden we dat de populariteit en het waargenomen coöperatieve gedrag alleen toenamen in etnisch heterogene teams met toenemende CL ervaring. De leerlingen hadden bijna allemaal geen ervaring hadden met CL alvorens het onderzoek begon. Toch was het CL curriculum lang genoeg om een verbetering van de interetnische relaties teweeg te brengen. Dit betekent dat CL niet alleen een sterke invloed kan hebben op de academische vaardigheden van teams in etnisch gemengde klassen, maar ook op hun sociale vaardigheden. We vonden ook dat de autochtone leerlingen sneller als coöperatief werden beoordeeld met het toenemen van CL tijd dan de allochtone leerlingen. Als reden hiervoor werd aangedragen dat de autochtone leerlingen meer individualistisch zijn ingesteld. Daardoor werden ze voorafgaande aan het CL curriculum als minder coöperatief gezien. Er zijn echter kanttekeningen te plaatsen bij het begrip ‘allochtoon’: het is wellicht een te grove maat om etnische verschillen in multiculturele klassen te onderzoeken. Wellicht is het meer verhelderend indien ook rekening wordt gehouden met de etnische achtergrond die dominant is in de klas waar de leerlingen inzitten om zo het fenomeen van een ‘double identity’ (zowel een etnische minderheid in de klas als in de samenleving) te ondervangen.

Al met al kunnen we stellen dat de invloed van de leerkracht tijdens CL op de leerwinst van leerlingen wordt beïnvloed door de samenstelling van de teams wat betreft etnische achtergrond en rekenkundige voorkennis. Echter, de onderhavige dissertatie toont aan dat taalkundige voorkennis en voorkennis van CL belangrijke variabelen zijn die mede de leerwinst van de individuele leerlingen en de teams als geheel bepalen. Hiernaast suggereren de bevindingen in deze dissertatie dat de CL voorkennis van de leerkrachten ook een belangrijke voorspeller is van de effectiviteit van CL.

Hoewel de data voor deze bevindingen in de onderwijspraktijk zijn verzameld, is de onderhavige dissertatie geen handvat voor praktisch gebruik. Desalniettemin geven de resultaten wel ruimte tot enige praktische overwegingen. Ten eerste dienen leerkrachten in ogenschouw te nemen dat effectief samenwerken betekent dat het gedoseerd moet worden ingevoerd en toegepast. Het dient aanbeveling in elk team iemand aan te wijzen die de andere teamleden uitdaagt antwoorden en uitleg te geven. Om ervoor te zorgen dat iedereen hier baat bij heeft moet ieder groepslid deze taak op zich nemen. Hierbij dient de leerkracht vooral rekening te houden met de taalvaardigheid van de allochtone teamleden: diegenen die een lage taalvaardigheid hebben dienen extra aandacht te krijgen. Het belang van de taalvaardigheid kan worden

geïllustreerd met de bevinding in deze dissertatie dat de allochtone leerlingen minder hulp gaven dan autochtone leerlingen, wat aantoonbaar geassocieerd was met een lagere taalvaardigheid. Tenslotte dient de leerkracht zich er rekenschap van te geven dat het introduceren van CL niet moeilijk is, het effectief toepassen ervan echter wel, niet alleen voor de leerlingen, maar ook voor de leerkrachten. Deze dissertatie laat immers zien dat er zelfs met een uitgebreid lesprotocol een discrepantie ontstaat tussen hoe de leerkrachten hun eigen CL activiteiten waarnemen en hoe onafhankelijke observatoren dit doen. Derhalve is het essentieel dat leerkrachten pas beginnen met de implementatie van CL indien ze ten eerste worden getraind in het realistisch inschatten van hoe het samenwerken bij leerlingen verloopt (door de juiste vragen te stellen en door observatie) en ten tweede door een realistische inschatting te maken van hoe ze met deze kennis effectieve feedback kunnen geven aan de leerlingen.

Biography

Michiel Bastiaan Oortwijn graduated in November 2001 at the University of Amsterdam where he studied Psychology in the specialization Experimental Psychology. From August 2002 to August 2006, Michiel Oortwijn was appointed as a PhD student with the Centre for the Study of Education and Instruction, at Leiden University on a project entitled: 'Productive interaction in cooperative learning; the influence of students' ethnocultural background'. As of September 2006 he is working as a junior researcher within the Centre for the Study of Education and Instruction. In addition, he is participating in a project called: 'Talents centre', set up by Prof. Dr. Paul Vedder and the Johan de Witt school group in The Hague. The aim of this project is to reduce the educational disadvantage of pupils at multi-ethnic elementary and high schools.