(Insert here the logo Universiteit Gent Faculteit Psychologie en Pedagogische Wetenschappen)

Peer Tutoring and Strategy Instruction:

The Effects on Chinese Pupils' Reading Skills and Attitudes

Ling Li

Promoter: Prof. Dr. Martin Valcke

Proefschrift ingediend tot het behalen van de academische graad van Doctor in de Pedagogische Wetenschappen

2006

前言

正如人生历程中无数其它重要的经历，我曾经无数次地期盼这一刻的到来。窗外清风徐来，月光如水。追忆着似水年华，结识过的人和经历过的事，恍若隔日，一幕幕仍清晰浮现眼前。抚今追昔，人生似乎是一条斜踏而就的小径。谁会想到自己会被命运的大手笔抛入到这个美丽而陌生的国度，悄然度过这数年青涩的时光？回望这几年与生活的激情碰撞，曾几何时已悄悄收起了少年仗剑行的轻狂，从尖锐到圆滑，从特立独行到中庸。是欣慰？是悲哀？这个中的滋味已无暇细细回味。所幸饱尝了象牙塔的安逸和菁菁校园的灵秀之气，如今终于凝结成这本书。

“蓬生麻中，不扶而直。”我要真诚感谢我的导师，以及斧正我论文写作的教授们，感谢他们一路宝贵的建议和无私的帮助。在艰辛的学术之路求索，没有他们的提携是不可想象的。我也要对我们心理和教育学院的院长一直以来的宽容和支持表示深深的谢意。感谢给我灵感和激情的统计学教授们！只身为莘莘学子中的一员，他们的博学和人格魅力一直深深地感染着我，重燃了我青葱岁月时的激情和永不放弃的决心！

感谢教育系的新老同事给我创造的第二个温暖舒适的“家”。他们的工作热忱和成就提供了我成长的坐标，使我没有“胡乱地生长”。还有许多萍水相逢的中外朋友们，也借此方寸之地对他们道一声感激之语，感谢他们和我一起见证了许多快乐难忘的时光。

跨越万水千山，我要寄去对北京师范大学的申继亮教授、北京朝阳和海淀区的八所小学以及朝阳教科所的深深的谢意。他们是这本书的幕后英雄。是他们给了我实验园地的土壤和阳光雨露，以及多年无私的支持。同时还有身在北京曾经为我工作过的助手李平平和孙峰，以及多年来的老同学和老朋友们，在这里一并致谢！

感恩我远隔重洋的家人们！自我去京城求学，以及来比再续学业的十余载中，和他们总是聚少离多，相处的日子屈指可数。面对他们自始至终的支持和关爱，以及对我从未动摇过的信心，我有难以释怀的愧疚。随我海外经历的日日加深，对他们的思念也日复一日地积淀在心中。无论何时何地，家总是漂泊不定的航船最温暖的港湾。

人生的境遇，无论顺流逆流，都是收获的一笔财富。一个人静坐时，心中时时升起对生活的感激，感激它给了我如此丰富的经历，让我实现人生岁月中一次次的蜕变。前方未知的漫漫长路，不论是伴着行云流水的悠然，还是乘风破浪的快意，真诚将是我永不淡去的底色。

这是我的数年研究工作的结晶，却只是我漫长学术生涯的第一个稚嫩的脚印……

李玲

2006 年 5 月于比利时根特大学

Preface

As the most important event in my life, for more than once have I imagined the coming of this moment. Looking through the window, the little street in front is bathed in tender moonlight and the big trees are waving in a light breeze. The whole surroundings look so familiar, but also so strange. All those happenings in the past years are just like a movie from yesterday, still so close and so clear in front of my eyes. Looking back, life is full of coincidences. Never before have I dreamed that I would be brought by the magical hand of life to here, a scenic city and prestigious university in Europe, and spend a few unforgettable years of my life. The more than one thousand days to me is like water to a fish. Too cold or too warm, is only known by itself. Fortunately, here finally comes this dissertation.

I would like to give my whole-hearted acknowledgements to my promoter Prof. Valcke and professors in my guidance committee, for their support in the past years and insightful comments and constructive suggestions on the thesis. Without their help, it is unimaginable to pull together my pieces of work till this phase. A lot of thanks to our Dean, for his support and tolerance. Without his help, my Belgian journey could have been longer and harder. This work is also partly the result of help I have got from professors in my statistical courses. The inspiration and passion that I have gained in their classes have driven me till the end of this work.

I want to give my sincere acknowledgements to Dr. Jean-Pierre Verhaeghe. His knowledge, generosity and constant help are fully appreciated. During all these years of work, this comrade has always been a reliable tutor for me side by side.

Thanks to all the old and new colleagues in the Vakgroep Onderwijskunde. Their encouragement and sincere concern are deeply appreciated. The warm surroundings and promising research atmosphere have given me the feeling of a second home in daily life as well in professional development.

Thanks to Prof. Shen Jiliang from Beijing Normal University and to the school principals, directors, teachers and lovely students from the eight primary schools in Beijing for hosting my intervention for more than two years. They are the unknown heroes behind this book and for years they were the earth and sunlight of my fieldwork. And also thanks to Li Pingping and Sun Feng, the two wonderful assistants who have worked with me in China. On this little piece of sheet, I also want to send a lot of wishes to my old friends, classmates, and colleagues in China. Thinking of them always gives me such a warm feeling and a vivid image of the motherland.

In my Chinese manner, I would like to give my deepest gratitude to my family at the end: to my parents, my brothers and their families on the other side of the Globe. Since I left my home at 18 for Beijing to start my university study, for more than one decade I have only spent countable days with them. During all these years, their understanding, confidence and eternal trust have accompanied me and pulled me through various difficult periods of my life. They are my deepest roots and an eternal home that I can always return to.

This is the conclusion of my dissertation, but just a first step to a new journey. Tomorrow might be vague and uncertain. Deep in my heart there is always a flickering whisper: be fit, be nice and keep smiling!

Ling Li

May 2006 at Ghent University, Belgium

Content

General Introduction

The rhythm of the modern society development and the exponential accumulation of knowledge in science have put a great pressure and challenge on the mission of education, as well as on educators and students at various levels. "We are drown in information and starving for knowledge."[1] How to cultivate independent and effective readers is the mission and a big challenge for reading instruction in schools. In the present study, we report the findings of an intervention to improve students' reading comprehension through teacher facilitated reading strategy instruction and peer tutoring practice. The present study is partly a replication of a Flemish study (Van Keer, 2002, 2004; Van Keer & Verhaeghe, 2003, 2005), which is based on a similar Dutch intervention program (Fukkink *et al.*, 1997) in the Netherlands. The Flemish intervention was quite successful in improving 2nd and 5th grade students' reading comprehension through reading strategies instruction. Yet, the results with respect to the additional value added by having students practice the reading strategies in peer tutoring dyads were ambiguous. As will be explained further in this introductory chapter, both the successes of the Flemish study as well as the questions that remain provided reasons for a replication study with some modifications and extensions.

In this introductory chapter, we will first explain the important concepts involved in both the Flemish and the Chinese study. Then we will briefly summarize the results from the Flemish study and the research questions that will be addressed in our study. To highlight the relevance of these research questions for Chinese education, we will then point out a series of problems in common reading

[1] Quote from Rutherford D. Roger.

instruction practice in China, and indicate how these problems can be dealt with by introducing the two main elements from our intervention program. Then, the hypotheses guiding our study will be phrased and schemes showing how the research questions were addressed by the different research activities we have undertaken and the analyses we performed, will be introduced. Finally, the structure of the thesis will be described.

1. Main concepts and theoretical grounding

The intervention that is evaluated in our study combines two lines of research: reading strategies instruction research and research on peer tutoring. In the next two paragraphs each of both will be discussed first. The link between both angles can be found in the concept of metacognition, which is described right after.

1.1. Reading strategies

Strategies are mental operations or techniques used to solve problems and to enhance performance (Alexander & Jetton, 2000). *Reading* strategies are methods or means that individuals use to have a better grasp on the meaning of reading materials. They are essentially related to the goals of reading and to how to achieve those goals. The knowledge of reading strategies includes declarative, procedural and conditional knowledge. With declarative knowledge, children have a conceptual understanding of the functions and purposes of a repertoire of strategies. With procedural knowledge, they know how to use strategies step by step and apply them with the requisite actions. With conditional knowledge, children begin to understand when and why strategies are effective. The conditional knowledge is fundamental for children's spontaneous transfer of appropriate strategies to different situations.

Reading strategies can be seen as belonging to the broader category of learning strategies, and therefore share a number of characteristics (Paris & Lindauer, 1982): (a) They are deliberate actions performed to attain particular goals; (b) they are

based on self-generation and invention, rather than rule following; (c) they invoke cognitive skills and motivational will; (d) they include socially assisted tactics for problem-solving tasks; and (f) they involve automation and transferring to a variety of tasks (Paris & Winograd, 1990).

To be an independent learner requires that the students foresee the learning goals, make good planning, monitor and regulate the learning process and engage in self-reflection and evaluation of the learning results (Zimmerman, 1998). Goal setting, planning, monitoring, regulating, reflecting and evaluating are metacognitive acts. Their object is not the learning content as such but the cognitive strategies that are used to master the learning content (e.g., making a scheme, rehearsing, *etc*.). Not only can reading strategies be used for learning purposes, within reading strategies a similar distinction between cognitive reading strategies (directly applied to the text) and metacognitive reading strategies can be made.

In the present study students were given direct instruction in six reading strategies: (a) activating prior knowledge, (b) finding the meaning of unfamiliar words and expressions, (c) making predictions, (d) finding main ideas of texts, (e) monitoring and regulating comprehension of the texts, and (f) distinguishing different types of texts. The selection of these reading strategies is based on a lot of other reading research (e.g., Alexander, Graham, & Harris, 1997; Anderson & Roit, 1993; De Corte *et al*, 2001; Klinger & Vaughn, 1996; Palincsar & Brown, 1984; Pressley, *et al*., 1992). Taking this set of reading strategies as a whole, a number of characteristics can be distinguished.

First, the set contains both metacognitive strategies (e.g., (e) monitoring and regulating one's own understanding) and cognitive strategies (the rest). The cognitive strategies are useful for solving specific problems during the reading process. These cognitive strategies are distinguished from specific behaviors during reading. For example, Alexander (2006) includes questioning, clarifying,

and elaborating as reading strategies. We think that questioning, clarifying and elaborating are specific behaviors that could be performed when applying almost any of the six aforementioned strategies. Therefore they are not to be included in the strategy list. They are – in a sense – from a lower level than the six strategies listed above. In our program, one would find behaviors such as questioning, clarifying and elaborating as part of the steps that students are encouraged to take when applying and practicing the reading strategies.

Secondly, the series of strategies follows a more or less chronological order, from what one can do *before reading* (e.g., activating prior knowledge by talking about what the reader already knows about the topic, classifying the type of the text), going to what can be done *during reading* (e.g., make sure one understands the meaning of words and expressions,), till what to do *after reading* (e.g., finding the main ideas of the text). This chronology is easy for children to remember.

Third, in the cognitive operations that are invoked by the strategies, there is some ordering in complexity, from simpler strategies and cognitive operations, such as decoding and finding the meaning of new words, to more complex ones such as summarizing, reasoning and inferring to find the meaning of the text. This order facilitates the instruction by the teachers and mastering by students.

When comparing the six reading strategies with the six metacognitive acts distinguished above, one will find no explicit reference to goal setting, planning, reflecting, and evaluating. The focus on monitoring and regulating – as far as metacognitive strategies are concerned – is related to the broad variety of reading situations or contexts to which the reading strategies in our program are meant to be applicable. We did not want to focus on reading within the context of study only. Therefore we choose reading strategies that are widely applicable, for example also when reading for fun, for the joy of the imagination that is provoked by a story or the beauty of the phrasing. In such contexts, metacognitive acts such as goal setting, planning or evaluation are less explicitly present or needed than in a context in which studying or information gathering is the main purpose. It is clear that other

programs that do focus on reading for information or study purposes would emphasize these other metacognitive acts as well and would also include a number of additional cognitive strategies that are quite complementary to the ones that were taught in our intervention program.

Most research on reading strategies is done in western countries, involving English-speaking students and teachers using English as the language of instruction (Palincsar & Brown, 1984; Pressley *et al*., 1992), or some other western language. In 1997, the *Elementary School Journal* devoted a special issue on implementations of reading strategy instruction in elementary schools in the United States. The present study is based on a Flemish study (Van Keer, 2004; Van Keer & Verhaeghe, 2003, 2005) involving regular Flemish primary school classrooms with Dutch speaking pupils and Dutch as the language of instruction.

For the purpose of the present study the intervention program used in the Flemish study has been translated and adapted for use in regular Chinese primary school classrooms in Mainland China. The aim of the program was to improve Chinese pupils' reading comprehension when reading Chinese texts. When transferring interventions to a different culture, especially in the field of reading, both cultural and linguistic issues have to be tackled. Languages such as English or Dutch belong to the Indo-European language system, and have a phoneme-based writing system, which is totally different from a language with a morpheme-based writing system such as Chinese (McBride-Chang, *et al*., 2003; Packard, 2000). Apart from this very obvious difference, there are also differences with respect to the way words and more complex concepts are formed. In Chinese language, more complicated words are often built from simpler and more common used morphemes, while in Western languages, the morphological knowledge is often related to the knowledge of Greek or Latin roots or knowledge about suffixes that on their own have no meaning (Carlisle, 1988; Champion, 1997). In Chinese, these morphemes can be flexibly combined to form more complicated concepts. In comparison with Western languages such as English, Chinese is more analytic and

it is easier to derive the meaning of a new concept from the individual morphemes (Packard, 2000). This may facilitate Chinese beginners' reading process. Therefore, taking into account the language differences, it has to be verified whether reading strategies that have proven to be effective in English or Dutch would also be useful in improving reading comprehension in Chinese. Cultural differences may have consequences for the content of the reading materials considered to be appropriate or appealing for the pupils of the age levels involved in our intervention program. They may also affect the way teaching methods affect students' learning. These and other related aspects, as well as the adaptations that were made to the intervention program to make it fit to the Chinese linguistic, cultural and educational context will be discussed in more detail in Chapter 1 and some of the following chapters.

1.2. Peer tutoring

Peer tutoring (PT) is a specific form of peer-mediated instruction in which one student serves as an instructional agent (tutor) towards another student, the tutee (Topping & Ehly, 1998). It is a specific form of cooperative learning. With other forms of *cooperative learning* and with the wide diversity of learning activities known as *collaborative learning*, it shares the common feature that the individual learning process is based on and embedded in the interaction between learners. With other forms of *cooperative learning* (and in contrast to forms of collaborative learning), it shares the feature that these interaction processes are well-scaffolded by means of materials that are highly pre-structured by teachers (or the program designers) and the tasks students get are rather well-structured tasks. The learning processes focus more on the mastery of pre-defined concepts or skills rather than on the co-construction of open skills (Strijbos, 2005). Different from many other forms of cooperative learning, peer tutoring is characterized by (1) the specific type of role taking (tutor vs. tutee), (2) the way it is structurally embedded in the curriculum and classroom organization with frequent (e.g., weekly) tutoring sessions scheduled, (3) the way the learning and interaction processes are

scaffolded (Topping & Ehly, 1998; Topping, 1998). In our case, a set of pre-designed assignment cards was used, specifying the main steps of the target reading strategy and main generic questions that help to apply that reading strategy during reading. In addition, tutors were given a series of lessons on how to become an effective tutor.

Two categories of PT can be distinguished according to the age difference between tutor and tutee: "same-age tutoring" (SA) and "cross-age tutoring" (CA). "Reciprocal same-age tutoring" (RSA) is a specific form of SA in which the tutor/tutee role is exchanged regularly so that both students have equal chances to act as a tutor (Fantuzzo *et al.*, 1992). In this study, two types of peer tutoring are examined: CA (with a 5th grade tutor tutoring a 2nd grade tutee) and RSA (dyads formed classmates, in either 5th grade or 2nd grade).

1.3. Metacognition

Since one of the reading strategies in our intervention program is metacognitive in nature, and the other strategies are to be considered as tools for enhancing comprehension through the regulation of the reading process, metacognition is a key notion in our study.

Metacognition is knowledge about knowledge or thinking about thinking. It is defined by Flavell as *"knowledge and cognition about cognitive phenomena"* (Flavell 1979, p. 906). According to Myers & Paris (1978), metacognition relates to individuals, tasks, goals and strategy. It includes planning, monitoring, regulation and reflection activities during cognitive processes (Garner & Alexander, 1989) and it decides what, when and how strategies should be used (Paris, Lipson, & Wixon, 1983). Although there has been disagreement on the exact definition of metacognition, two broad classes of metacognition are often referred to: (a) knowledge about one's knowledge, processes, and cognitive and affective states; and (b) strategies to monitor and regulate that knowledge, those processes and those states (Hacker, 1998; Jacobs & Paris, 1987). The term metacognitive

awareness belongs to the first broad class. It refers to the awareness of one's cognitive state or processe; on a less articulated level, e.g., the awareness of being a competent reader.

Metacognitive skills are considered to help children becoming independent readers who can monitor and regulate their own reading processes, independent from teachers who point out and clarify difficult parts of reading materials. Including metacognitive strategies in reading instruction probably increases chances for continuous growth in reading comprehension, independent from teachers. But whether metacognitive strategies really have played a role in any improvement of reading comprehension that is observed, is not easily assessed. However, long-term effects on reading comprehension can be considered as a good indicator (albeit indirect) that indeed metacognitive skills have played a role.

In peer tutoring, the tutors' main task is to monitor and regulate the tutees' reading process. Tutors ask questions to check if tutees understand the text or master the reading strategies well and give corrections or explanations when necessary. Our hypothesis is that this "teaching experience" gives tutors unique opportunities in practicing metacognitive skills. Since monitoring and regulating someone else's behavior is probably easier than monitoring one's own behavior, a tutor can more easily acquire metacognitive skills and then transfer them to his/her own reading. Peers are close in age, and they use similar languages. They often do not hesitate to ask or answer questions. It is from this perspective that in the present study peer tutoring is considered to be equally beneficiary for tutors as it is for tutees.

In our study, we measured metacognition through two means. In the observational study reported in Chapter 2, observation and videotape recording was used. We selected dyads from the CA and RSA groups, videotaped a number of their reading sessions and analyzed the interactions within the dyads, focusing on indications for the use of (metacognitive) reading strategies. In the main study (including all students and all conditions) the use of reading strategies and metacognitive

awareness was measured by means of a self-report questionnaire. Students were asked “off-line” to report on their (“on-line”) awareness of reading tasks, their reading skills or how well they perform in reading. Other items in the self-report questionnaire were used as indicators for the use of reading strategies.

1.4. Theoretical ground – self-regulated learning

The two elements of intervention could be integrated into the theoretical framework of self-regulated learning. Self-regulated learning is a metacognitively, motivationally and behaviorally active learning process (Zimmerman, 1986). This process is a cyclical loop. At any point the process could be influenced by the past results or experience. Self-regulated learning depends on metacognition, self-monitoring and regulative skills that are adapted to the problem context and use various strategies to solve problems and reach the goals; it also requires motivation and volitional efforts.

Self-regulated learning is not a system of its own. It is contextually situated in a specific environment. It is also an individual’s learning process towards pre-set goals. In his social cognitive theory, Bandura (1986) stated that human functioning involves interaction among three resources: behavioral, environmental and personal variables. Zimmerman (1994, 1998) inherited the framework of social cognitive theory and broke the three resources into six areas where self-regulatory processes can be applied: motives, methods, time, outcome, physical environment and social environment. Self-regulated learning exerts efforts from motivation and behaviors and uses metacognitive knowledge to achieve learning goals.

The two elements of our intervention program, reading strategy instruction and peer tutoring practice can be integrated into in the theoretical framework of self-regulated learning. Reading is an important source and method for learning. As is mentioned above, although reading should not only be regarded in the context of learning, when talking about self-regulated reading comprehension and self-regulated learning, basically the same concepts apply. In both cases, cognitive

strategies are needed to solve specific problems, e.g., to find the meaning of an unfamiliar words or concept. The adequate use of these cognitive strategies requires that one is aware of the problem (monitoring of one's own understanding is therefore needed), that one knows the cognitive strategies (declarative knowledge), but also how (procedural knowledge) and when to use them (conditional knowledge), so that one can decide in an effective way which strategy to use (regulating one's own reading / learning process). Also the other phases in the metacognitive process that are usually distinguished (goal setting, planning, evaluation, reflection) principally apply to the reading process. That is particularly the case when reading takes place within the context of studying.

The second element--peer tutoring can be considered as an intermediate phase between dependent and independent learning. Self-regulated learning also includes "social forms of learning such as modeling, guidance and feedback from peers, coaches, and teachers" (Zimmerman, 2002, p. 1). It develops initially from social resources to self-resources, from observation and emulative study to self-control and self-regulation (Schunk, 2002). Independent learners grow from inter-dependent sources. We could not expect students to be independent from the very beginning. They learn to be independent by observation, by imitating others, by interdependent learning. Learning by teaching or learning with peers is an intermediate phase. From teaching, they can grow into more effective learners. By working together or learning from each other, students can learn to become independent learners. Both instruction by the teacher and cooperative learning with peers or older students are integrated in our intervention program.

We choose the framework of self-regulated learning, because (1) it advocates both cognitive and metacognitive strategies to achieve learning goals, which is also one of the key elements in our study; and (2) it emphasizes the bilateral, dynamic and cyclic interactive relationship between the social environment and individuals. In this sense it has a better ecological validity than other learning theories, such as information processing theory. In our study, this ecological element was primarily

taken into account by studying the effects of our intervention program within the natural context of existing class groups within schools, with regular school teachers as the primary facilitators of students' learning. It was also taken into account by including indicators of the family environment that may be of influence on students' learning.

In addition to self-regulated learning theory, we will also refer to other, complimentary theoretical bases, such as information processing theory and Vygotsky's theory about *Zone of Proximal Development* (Vygotsky, 1978) to explain particular findings, such as difference between CA and RSA dyad interaction processes, the benefits of CA tutees from cross-age tutoring practice, etc.

2. A replication of a Flemish study, with some modifications and extensions

Our intervention program is based on an original Dutch program (Fukkink *et al.*, 1997) and is partly a replication of similar study with Flemish primary school pupils (Van Keer, 2002, 2004; Van Keer & Verhaeghe, 2003, 2005). The Flemish study has found significant positive effects of reading strategy instruction on students' reading comprehension. The results appeared to be clearer for the fifth grade than for the second grade. In fact for the second grade, significant effects were only found for students in the CA-condition and on the short term (on the posttest); for students who had practiced the reading strategies individually under the teacher's supervision or in RSA dyads no differences with students in the control condition were found (Van Keer & Verhaeghe, 2005). For the fifth grade, significant short-term effects were found for all three experimental conditions. In addition, in Grade 5 it was also found that CA tutors made significantly more long-term gains than students in the RSA tutoring groups (Van Keer, 2004). When also the effects on the (self-reported) use of strategies and the incidence of negative self-efficacy related thoughts during reading are taken into account, the Flemish

study appears to indicate that the combination of direct instruction of the reading strategies by the teacher and practice in CA dyads is the most powerful learning environment for both second and fifth graders.

Our main reasons for replicating this study in China were:

(1) to find out whether reading strategies instruction would bring comparable benefits to Chinese primary school pupils with regard to their Chinese language reading comprehension skills and related variables such as the use of reading strategies, metacognitive awareness and reading attitudes; if so, introducing reading strategies instruction into Chinese schools would be a proper way to bring up more independent readers who are capable of using effective methods to solve reading problems and achieve good comprehension results. In this way a contribution would be made to solving a number of existing problems in Chinese reading instruction;

(2) to find out whether in the context of Chinese primary education peer tutoring (in any of the two forms that were tried out) would create any additional value for the students, either with respect to the their reading comprehension skills or with respect to the use of reading strategies, their attitudes toward reading or how they think about their reading competency;

(3) to find evidence for the hypothesis that cross-age peer tutoring – more particularly for the tutors – provides a more powerful learning environment for practicing metacognitive skills that enable them to become more independent, self-regulating readers than reciprocal same age tutoring.

Similar to the Flemish study, we explored the treatment effects on students' reading comprehension growth on the short term and in the long run, both in 2nd

and 5th grade. In addition, and also similar to the Flemish study, a number of related questions were explored: effects on students' use of reading strategies, reading attitude and metacognitive awareness.

The intervention took place in natural classroom settings, organized on a weekly basis and lasted for a whole school year. It involved eight primary schools from two districts of Beijing, about 900 second grade and 750 fifth grade pupils, coming from a wide variety of social backgrounds.

All students in the experiment received direct instruction on reading strategies. In addition, in two of the experimental conditions, these strategies were further practiced in cross-age (CA) and reciprocal same-age (RSA) dyads. In the third experimental condition, students practiced those reading strategies by themselves, with the teachers' guidance or help.

Regular Chinese language teachers were trained to carry out the different teaching tasks: training tutors, teaching reading strategies, supervising the initial classroom practice, and coaching the independent practice or peer tutoring sessions. These teachers were intensively trained before the start of the program and frequently observed and coached throughout the intervention period. They were provided with teachers' manuals, lesson scenarios, teaching assistance materials, children reading books, and checklists (outlines) of each teaching lesson in the program. In every school, discussions with all participating teachers from that school were frequently organized to refresh important tasks or exchange useful experiences.

A distinct feature of our research is the observational study that was carried out to gain a deeper insight in the processes that lead to the treatment effects. This observational study involves an in-depth investigation of the interaction processes within dyads, and how these processes differ across different types of tutoring dyads, and across students' initial reading level and gender.

3. Problems in reading instruction in China[1]

With our study we mean to address a number of problems in Chinese reading instruction. These problems mainly relate to the following characteristics commonly found in Chinese reading instruction practice: (1) It focuses too much on overt reading activities and ignores teaching students' efficient skills to understand the deeper meaning of reading materials. Not enough attention is given to more efficient ways to achieve a deeper understanding. (2) Reading performance or "reading with expression" is used as the most important indicator of students' reading ability in primary language class. The latest Chinese Curriculum Standard (Chinese Ministry of Education, 2001) summarizes the goal of teaching Chinese as to "cultivate patriotic feelings and socialist morality" and to "grow love for the mother tongue". The most direct way to show the "feelings", "emotions" and "love" is to read with expression, which is shown in changes in speed, tones or accents. (3) The question-answer and the knowledge infiltration style is commonly used in reading instruction. Through teachers' questioning and induction, students are often led to the "road" that is laid by the teacher in advance and are guided to give the expected answers (Cheng, 2006; Qi, 2004). (4) The Chinese curriculum standard requires that students use different "methods" to read, but doesn't give any explanation on what these "methods" are or nor does it give any suggestion on how to meet this requirement. In reality, it is up to students themselves to discover some effective reading strategies spontaneously. This situation is far from favorable to create independent readers. It separates reading practice from reading comprehension, effective methods from deeper understanding. Students lack awareness and concrete skills in monitoring their reading process to detect difficulties, and in solving specific reading problems.

[1] For more elaborations on this topic, please read Chapter 1.

Chinese schools are crowded with students. In big cities, the class size in primary schools ranges from 30 to 60. The traditional teacher-centered way of giving instruction more and more exposes its problems, such as a lack of attention to students' needs in classrooms, lacking of interactions with students, and lack of individualized learning opportunities (Qi, 2004). The official curriculum encourages social interaction and cooperative learning among students in the classroom. However, most teachers lack experience or useful guidance to organize cooperative learning. Teachers sometimes organize small-group cooperative learning in language classes, usually with 3-5 students with mixed abilities in a group and each group assigned a few questions to answer together. Theses cooperative reading activities generally are result oriented--to answer the questions assigned to the different groups, but do not provide enough scaffolding to students nor are students taught how to think and solve problems (Chen, 2004). Not all students in the groups are mobilized and the group usually relies on one or two good students to find out the answers and to report to the whole class to "support" the situation (Lu, 2006), while other students in the meantime got a "free-ride". Some activities are so called "cooperative", but in fact, "there is no change in organization, nor clearance, concreteness, reality or insight in content" (Cheng, 2004, p. 6). Zhong (2006) summarized the problems of cooperative learning in Chinese classrooms as full of spontaneity from teachers, lack of general design of the learning situation, participation from only a few students, and lack of effectiveness in learning results.

All in all, changes are needed to improve the situation in Chinese reading instruction. In our intervention program, reading strategies and peer tutoring are introduced in language instruction to lay the foundations for a deeper understanding of the reading content. Better test results are not the ultimate goal, but an indication that students are well on their way to reach the goal of becoming independent, self-regulated readers who seek to understand meaning in texts and solve the problems they meet while doing so.

4. Hypotheses

The aforementioned research questions relate to the effects of the intervention, having adjusted for the effects of students' social background, gender, and other potential confounders. The effects that were investigated include effects on reading comprehension growth, use of reading strategies, metacognitive awareness with respect to task requirements, reading performance and reading competence, and reading attitudes. For each of these "families of effects", specific hypotheses are phrased in every chapter. They generally follow the structure that was adopted for the hypotheses with regard to the effects on students' reading comprehension growth:

(1) Reading strategy instruction significantly promotes Chinese students' reading comprehension development in Chinese language. At the posttest, all the experimental groups will show significantly more growth in reading comprehension than the control groups.

(2) In addition to the positive effects of reading strategy instruction, there is a significant extra effect from practicing the reading strategies in peer tutoring dyads. At the posttest, both CA and RSA students will demonstrate significantly more growth in reading comprehension than RS students.

(3) The additional growth in reading comprehension is significantly larger for students in CA condition than in the RSA condition.

(4) Students in the experimental groups keep the gains they made during the intervention for at least six months. At the retention test, all three experimental conditions still show a significantly larger growth than the control groups; students involved in peer tutoring practice (CA and RSA groups) show significantly more growth than

those in RS groups, and 2nd graders in CA groups show significant more growth than 2nd graders in RSA groups.

5. Theoretical and operational scheme

In this section we present two schemes — a theoretical scheme and an operational scheme. In the theoretical scheme, the relationship among the main elements, concepts and outcomes in our study are shown. Within the framework of self-regulated learning, the two components of the classroom intervention are specified: reading strategy instruction and peer-tutoring practice. Expected outputs are shown on the right hand side of the scheme, adjusting for the effects of various confounders related to student characteristics, their family background, their classes and schools. The scheme indicates that it is hypothesized that the intervention program has an effect on students' metacognitive skills, which – in turn – enables them to perform better when confronted with a reading comprehension task. Associated with this better reading comprehension performance are effects on students' metacognitive awareness and students' reading attitudes. No specific hypotheses about the nature or direction of these associations are phrased.

The operational scheme shows how the different elements identified in the first scheme, were measured, how the resulting data were analyzed and in which chapter the results are reported.

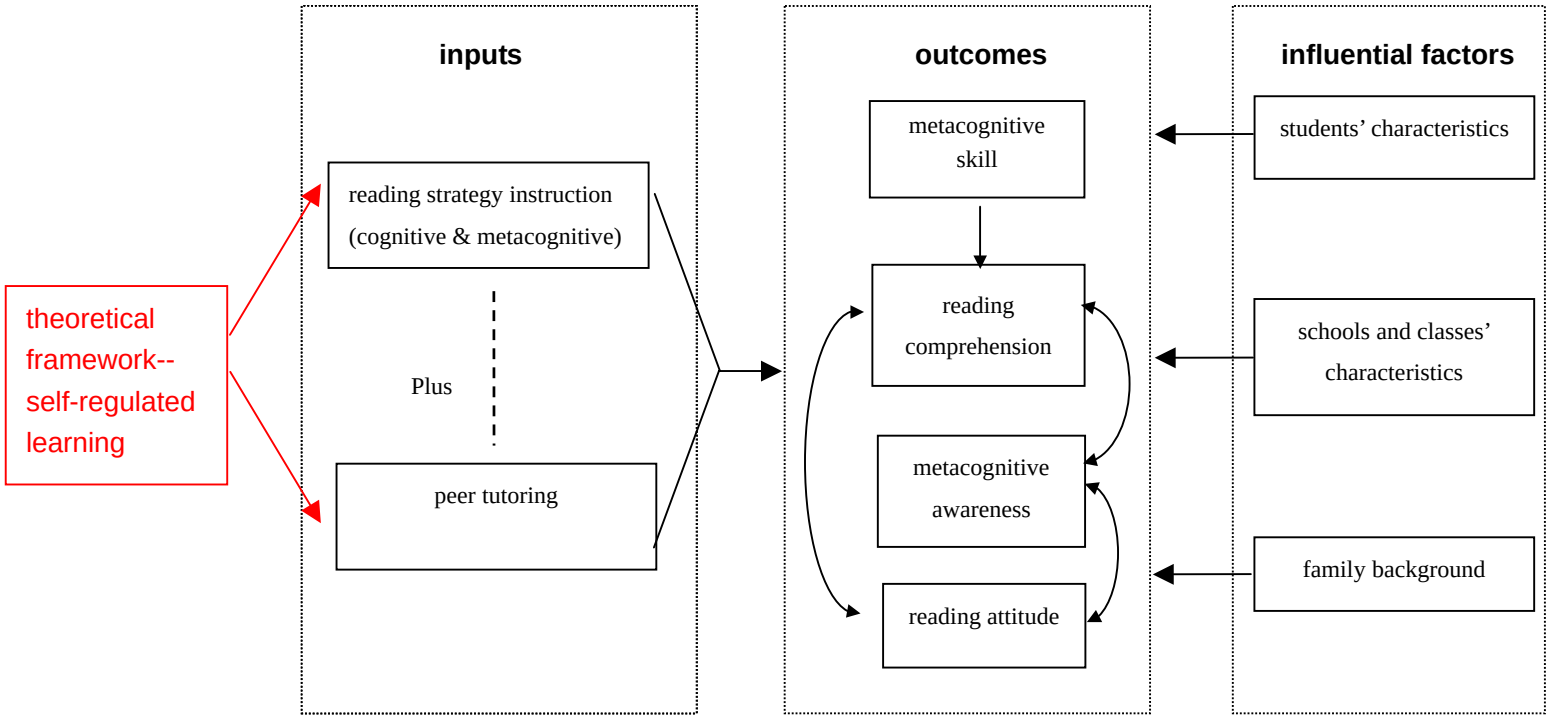
As one can see, different types of instruments and "measurements" (including narratives, observation, observation coding scheme, reading comprehension tests and questionnaires) were used and different types of analyses were performed. Some of these analyses only had an instrumental function, e.g., for creating measures or variables out of the raw data. This was the case for the IRT-calibration of the reading comprehension tests, the exploratory factor analyses and the confirmatory factor analyses using structural equation modeling. As one can see in

the scheme, no arrows are drawn directly from the boxes representing those analyses to the output section of the scheme, but only to the more central multilevel repeated measures analyses box. Since the focus of the study is on the substantive results, the purely instrumental analyses will only be briefly reported.

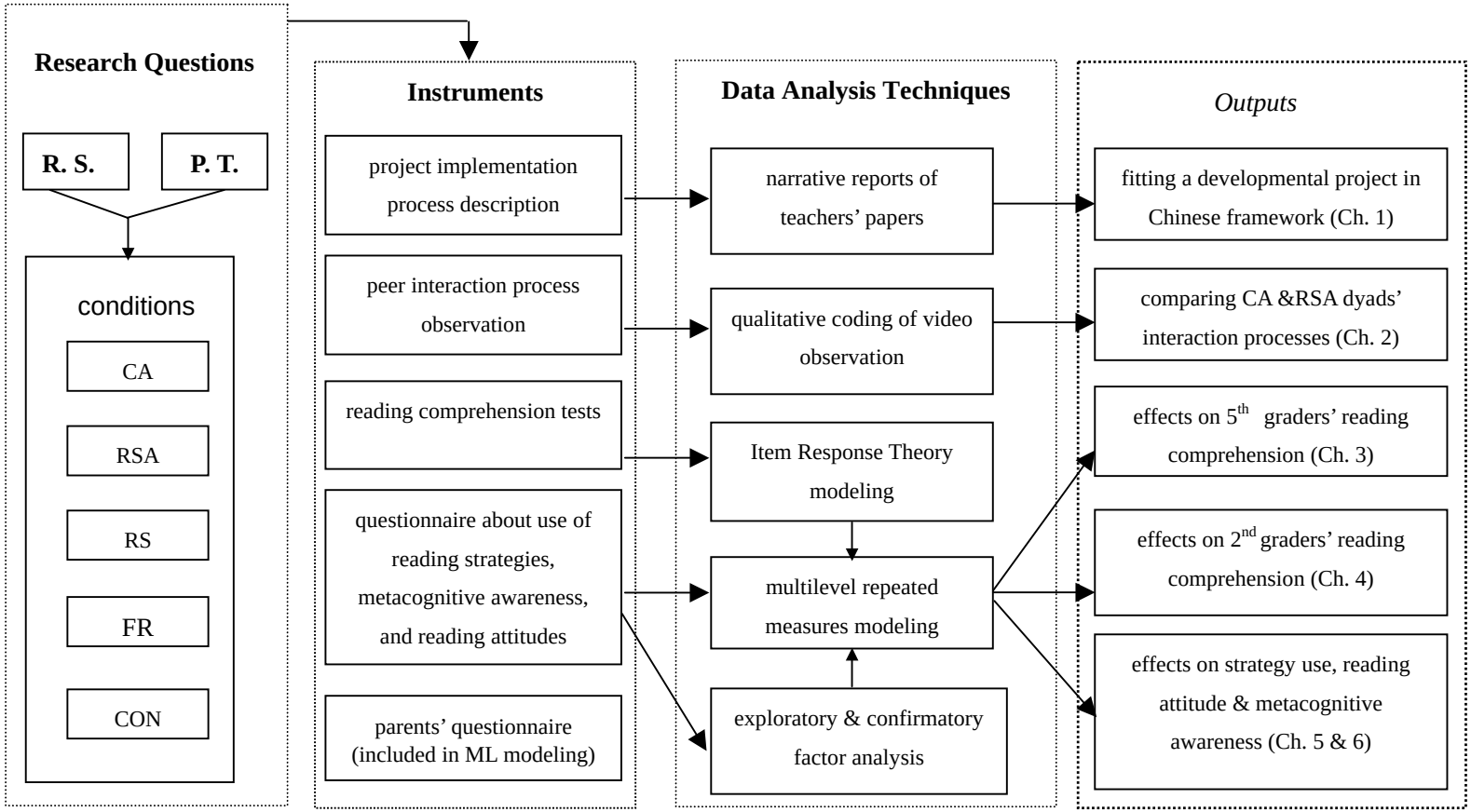
6. Structure of the thesis

The last column of the second scheme also shows the structure of the thesis. In the Introduction chapter, the general picture of the whole study is drawn. In Chapter 1, we describe the challenges we met during the implementation of the intervention and our solutions. In Chapter 2, we take an in-depth look at the within-dyads' interaction processes during the peer-tutoring sessions and compare the differences between CA and RSA dyads with regard to metacognitive processes. Chapter 3, 4, 5 & 6 report the intervention results of students' growth in reading comprehensions, use of reading strategies, reading attitude and metacognitive awareness. These results cover both fifth and second grade. The last chapter provides a general discussion. It synthesizes the findings in the study, presents comments and conclusions, and discusses the limitations of the study and possible directions for future research.

Theoretical Scheme



Operational Scheme



References

Alexander, P. A. (2006). *Psychology in Learning and Instruction*. Upper saddle river, NJ: Pearson.

Alexander, P. A., Graham, S., & Harris, K. (1998). A perspective on strategy research: Progress and prospects [Special issue]. *Educational Psychology Review*, 10, 129-154.

Anderson, V. & Roit, M. (1993). Planning and implementing collaborative strategy instruction for delayed readers in Grades 6-10. *The Elementary School Journal*, 94, 2, 121-137.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Carlisle, J. F. (1988). Knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders. *Applied Psycholinguistics*, 9, 247-266.

Champion, A. (1997). Knowledge of suffixed words: A comparison of reading disabled and nondisabled readers. *Annals of Dyslexia*, 47, 29-55.

Chen, J. S. (2004). Guan jian shi jiao yu guan nian de zhuan bian: Ke cheng gai ge zhong chu xian de ji ge wen ti [The key is to change ideas about teaching: A few questions arising in the teaching reforms], *Ban Zhu Ren* [*The Headmaster]*, 11, 12-14.

Cheng, L. Y. (2004). Ke cheng gai ge zhong yu wen ke tang jiao xue bi duan li xi. [The analysis of weak points in Chinese classroom instruction with examples during the curriculum reform]. *Zhong Xue Yu Wen: Jiao Shi Ban* [*Middle School Chinese: Teachers' Edition*], 11, 5-8.

Chinese Ministry of Education (2001). *Quan ri zhi jiu nian yi wu jiao yu yu wen ke cheng biao zhun* [*Full-time Nine-Year Compulsory Education Chinese Curriculum Standards*], Chinese Ministry of Education.

De Corte, E., Verschaffel, L., Van de Ven, A. (2001). Improving text comprehension strategies in upper primary school children: a design experiment. *British Journal of Educational Psychology, 74, 4, 531-560.*

Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment. *Journal of Educational Psychology*, 84, 331-339.

Flavell, J. H. (1979). Metacognition and cognitive monitoring : A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906-911.

Fukkink, R., Van der Linden, S., Vosse, A., & Vaessen, K. (1997). *Stap Door! Handleiding* [*Step out! Manua.*]. Utrecht, The Netherlands: Sardes.

Garner, R., & Alexander, P. A. (1989). Metacognition: Answered and unanswered questions. *Educational Psychologist*, 24(2), 143-158.

Hacker, J, D. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp.1-24). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Hacker, J, D., Dunlosky, J., Graesser, A. C. (1998). Metacognition in educational theory and practice. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Jacobs, J. E. & Paris, S. G. (1987). Children's metacognition about reading : issues in definition, measurement, and instrument. *Educational Psychologsist*, 22(3&4), 255-278.

Lu, Q. (2006). 'Zhen shui wu xiang' shi yu wen—guan yu dang qian yu wen ke tang jiao xue de si kao. ['The real water without fragrance is Chinese'--thoughts about the present Chinese instruction in classroom], *Yu Wen Jiao Xue Tong Xun* [*Journal of Chinese Teaching*], No.2, 38-39.

McBride-Chang, C., Shu, H., Zhou, A., Wat C. P., & Wagner, R. K. (2003). Morphological awareness uniquely predicts young children's Chinese character recognition, *Journal of Educational Psychology*, 95, 4, 743-751.

Myers, M. & Paris, S.G. (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology*, 70, 680-690.

Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge, England: Cambridge University Press.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension fostering and monitoring activities. Cognition and Instruction, 1, 117-175.

Paris, S. G., Lipson, M. Y., & Wixon, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8, 293-316.

Paris, S.G., & Lindauer, B.K.(1982). The development of cognitive skills during childhood. In B. Wolman (Ed.), *Handbook of developmental psychology* (pp.333-349). Englewood Cliffs, NJ: Prentice-Hall.

Paris, S.G., & Winograd, P.W.(1990). How metacognition can promote academic learning and instruction. In B. J. Jones & L. Idol (Eds.), D*imensions of thinking and cognitive instruction* (pp.15-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: transactional instruction of reading comprehension strategies, *The Elementary School Journal*, 92, 5, 513-555.

Qi, Z. H. (2004). Dang qian yu wen ke tang jiao xue zhong cun zai de liang ge wen ti [Two problems in present Chinese instruction in classroom], *Ning Xia Jiao Yu* [*Education in Ning Xia*], 5, 8.

Schunk, D. H. (2002). Social cognitive theory and self-regulated learning. In D. H. Schunk, & B. J. Zimmerman (Eds.). *Self-regulated learning and academic*

achievemente: Theoretical perspectivess (pp. 125-152). Mahwah, NJ: Erlbaum.

Strijbos, J. W. (2005). *The effect of roles on computer-supported collaborative learning.* Proefschrift ter verkrijging van doctor aan de Open Universiteit Nederland. Datawyse boek, Maastricht.

Topping, K. J. & Ehly, S. (1998). Introduction to peer-assisted learning. In K. J. Topping & S. Ehly (Eds.), *Peer-assisted Learning* (pp. 1-23). Mahwah, NJ: Erlbaum.

Topping, K. J. (1998). Effective tutoring in American Reads. A reply to Waski. *The Reading Teacher*, 52(1), 42-50.

Van Keer, H. (2002). *Reading strategies instruction and peer tutoring in primary school. A quasi-experimental study of the effects on students' reading comprehension and metacognition.* Proefschrift ingediend tot het behalen van de academische graad van Doctor in de Pedagogische Wetenschappen (Ghent University), Promotor: Prof. Dr. Jean Pierre Verhaeghe.

Van Keer, H. (2004). Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring. *British Journal of Educational Psychology*, 74, 37-70.

Van Keer, H., & Verhaeghe, J. P. (2003). Effects of explicit reading strategies instruction and peer tutoring in second and fifth grades. In E. De Corte, L. Verschafeel, N. Entwistle, & J. Van Merrienboer (Eds.), *Powerful learning environment: Unraveling basic components and dimensions* (pp. 211-228). Oxford: Elsevier.

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. *The Journal of Experimental Education,* 73(4), 291-329.

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

Zhong, Y. D. (2006). Shi xian jiao xue shuang ying de xue xi ce lue—he zuo xue xi [*Learning strategy to achieve win-win result--cooperative learning*], *Wu Li Jiao Shi* [*Physics Teacher]*, 27, 1, 19-21.

Zimmerman, B. J. (1994). Dimensions of academic self-regulation: A conceptual framework for education. In D. H. Schunk, & B. J. Zimmerman (Eds.). *Self-regulation of learning and performance: Issues and educational applications* (pp. 3-21). Hillsdale, NJ: Lawrence Erlbaum Associates.

Zimmerman, B. J. (1998). Developing self-fullfilling cycles of academic regulation: An analysis of exemplary instructional models. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 1-19). New York: Guilford Press.

Zimmerman, B. J. (1986). Development of self-regulated learning: Which are the key subprocesses ? *Contemporary Educational Psychology*, 16, 307-313.

Chapter 1

Fitting a Developmental Project in Chinese Educational Framework: Challenges, Solutions and Results*

Abstract

This article reports the implementation of an educational intervention in Chinese primary schools in Beijing (China). Reading strategies instruction and peer tutoring were the two elements implemented to improve Chinese students' reading comprehension abilities. Assessment of students' reading comprehension and teachers' written comments were used to illustrate the results of the intervention. It is argued that (1) it is very crucial to form a good partnership with the local educational authority and adapt the project to the educational framework and culture; (2) to train local teachers to carry out the project, both short-term intensive training and long-term scaffolding during the implementation are needed; (3) in order to keep teachers motivated and ensure the intervention quality and treatment fidelity, a great effort is needed to give constant evaluation, feedback and encouragement to the teachers involved.

Key words: education development China primary schools reading strategy peer tutoring

* This chapter is based on an article with the same title submitted to the *International Journal of Educational Development* for publication.

1. Introduction

For most developmental projects done in developing countries, there is no smooth implementation process. As is reflected by many developmental projects in education, one of the biggest difficulties is the resistance from the host countries towards complete acceptance of the design and teaching approach (Gu, 2005). This resistance may come from the local educational authority, because the main goals of the projects don't fully fit the goals set by the educational policy makers in the host countries; it can also be due to the gap between Western innovations and local realities; or the gap between practice-oriented teachers vs. academic researchers. Sometimes teachers could hardly see the meaning of the studies for their classroom practice. Or it is simply the consequence of insufficient training of the teachers carrying out the projects. This article describes an educational development project in eight primary schools in Beijing, China. Reading strategies instruction and peer tutoring were the two innovative elements implemented to improve Chinese students' reading comprehension skills. Assessments of students' reading comprehension and teachers' written comments were used to evaluate the results of the intervention. It is argued that recognizing the *resistance* and gaps and trying to adapt the project to the local educational and cultural framework are the solutions to various challenges. In addition, a long-term effort is needed to guarantee the quality of the implementation processes.

1.1. The usefulness of reading strategy and peer tutoring in Chinese primary schools

With our study we mean to address a number of problems in Chinese reading instruction[1]. These problems mainly relate to the following characteristics commonly found in Chinese reading instruction practice:

[1] For an illustration of typical Chinese reading lessons, please read the transcriptions of two exemplary lessons in the Appendix.

(1) It focuses too much on overt reading activities and ignores teaching students' efficient skills to understand the deeper meaning of reading materials. A lot of teachers still have a strong belief in the very famous Chinese saying "shu du bai bian, qi yi zi xian" (if books are read for a hundred times, their meaning will be shown automatically) and they tend to ignore the importance of deep understanding of the reading materials, as well as effective methods to achieve such understanding. A lot of different "ways of reading" have been organized by teachers, such as individual reading, silent reading, collective reading, small-group reading, whole-class reading, role-reading (students taking the role of different characters), and performance reading (Cao, 2004). Many teachers assume that reading comprehension is a natural consequence of lots of reading practice. If students read texts aloud for enough number of times or if they could memorize them, they are assumed to understand the content. Whether there are more efficient ways to achieve a deeper understanding, is not the main concern of teachers.

(2) Reading performance or "reading with expression" is used as the most important indicator of students' understanding in primary school language class. The latest Chinese Curriculum Standard (Chinese Ministry of Education, 2001) summarizes the goal of Chinese teaching as to "cultivate patriotic feelings and socialist morality" and to "grow love for the mother tongue". The most direct way to show "feelings", "emotions" and "love" is to read with expression and to show the characters' emotion or mood in the text with explicit performance, such as change in speed, tones or accent. Since it is easy for teachers to demonstrate and for students to imitate, the role of expressive reading is often dominant in language instruction. As one teacher wrote:

> *Since "training" is deleted in the Chinese Curriculum Standard, people [teachers] fear to mention training and don't dare to given any training. Instead, [they] talk about humanity, during the whole lesson let students feel the texts again and again, and talk about what they have felt [from the*

> *texts] and ignore any training and application [of skills] in language (Lu, 2006, p. 39)*[1].

(3) The question-answer and the knowledge infiltration style are still quite typical in the reading instruction. The common practice in Chinese reading instruction is the so-called "carrying questions in your mind to read ". "Presetting the lesson scenarios" is still a typical way for teachers to prepare for reading instruction (Ding, 2005). Through questioning and induction, students are often guided to the road that is laid by the teacher in advance (Cheng, 2006; Qi, 2004) and teachers lack the skills to cope with students' spontaneous responses.

(4) On the policy level, though, it is stated in the Chinese Curriculum Standard (2001) that the general goal of reading instruction is to "gain independent reading ability" and "use different methods to read". The Chinese Curriculum Standard (2001) however, doesn't give any explanations on what these "methods" are nor any guidance on how to meet this requirement. The task of defining effective means to reach theses general goals and have them mastered by students is left to schools or teachers. However, as most teachers are used to teaching contents and are not trained to teach strategies (Peterson & Swing, 1983), in reality it is up to students to discover reading strategies spontaneously themselves in reading activities. Unfortunately, only limited number of students discover such reading strategies by themselves. According to a vast body of research, reading strategies could be mastered by students through explicit instruction and these strategies could be transferred to their general reading activities (Anderson & Roit, 1993; Jacobs & Paris, 1987; Osman & Hannafin, 1992; Perkins & Salomon, 1988; Pressley & Van Meter, 1993; Sawyer, Graham & Harris, 1992; Pressley & Goodchild, *et al*., 1989; Pressley & El-Dinary, *et al*., 1992). Not different from many other countries, the Chinese situation is far from favorable to cultivate independent readers. It separates reading practice from reading comprehension,

[1] Translation from Chinese.

effective methods from deeper understanding. Students lack awareness and concrete skills in monitoring their reading process and detecting difficulties, and in solving specific reading problems.

Chinese schools are crowded with students. In big cities, the class size in primary schools ranges from 30 to 60. The teacher-centered traditional instruction has more and more exposed its problems, such as a lack of attention to students' needs in classrooms, too few interactions with students, and a shortage of individualized learning opportunities (Qi, 2004). The official curriculum encourages social interaction and cooperative learning among students in classroom. However, most teachers lack the experience or useful guidance to organize cooperative learning. Teachers sometimes organize small-group cooperative learning in language classes, usually with 3-5 students with mixed abilities in a group and each group assigned a few questions to answer together. Theses cooperative reading activities generally are result oriented. Students are expected to answer the questions assigned to their groups, but are not provided with enough scaffolding nor are taught how to think and solve problems (Chen, 2004). Not all students in the groups are mobilized and the group usually relies on one or two good students to find out the answers and to report to the whole class to "support" the situation (Lu, 2006), while other students in the meantime got "free-ride". Some activities are so called "cooperative", but in fact, "there is no change in organization, nor clearance, concreteness, reality or insight in content" (Cheng, 2004, p. 6). Zhong (2006) summarized the problems of cooperative learning in Chinese classrooms as full of spontaneity from teachers, lack of general design of the learning situation, participation from only a few students, and lack of effectiveness in learning results.

All in all, changes are needed to improve the situation in Chinese reading instruction. In our intervention program, reading strategies and peer tutoring are introduced in language instruction to lay the foundations for a deeper understanding of the reading content. Better results on reading comprehension tests are not the ultimate goal, but an indication that students are well on their way to

reach the goal of becoming independent, self-regulated readers who seek to understand meaning of the texts they read and solve the problems during the course.

Although teacher facilitated reading strategy instruction and peer-tutoring practice have been well-established in Western research fields and in some school curriculum, it is hardly practiced in China and the combined effect of the two elements has not been studied in China. Among various intervention programs in schools to improve students' reading and mathematics results, peer tutoring is the most cost-effective and has the most favorable input-output ratio (Levin, Glass & Meister, 1987). It makes use of natural human resources from students and empowers their self-regulated learning ability.

Our educational project also fits teachers' professional development. As is required by the Beijing educational authority, teachers need to have at least 36 hours of in-service training in each semester (Beijing Educational Committee, 2000). Teachers are also encouraged to participate in various kinds of educational research projects and their research experience has an impact on their evaluation and further career development. In this sense, our educational project also meets the needs of teachers' career development.

1.2. Design and experimental conditions

The study in China is a replication of a similar Flemish study (Van Keer, 2004; Van Keer & Verhaeghe, 2005). It involves two lines of intervention — reading strategy instruction and peer tutoring. Six reading strategies were taught to students: (a) activating prior knowledge, (b) finding the meaning of unfamiliar words and expressions, (c) making predictions, (d) finding main ideas of texts, (e) monitoring and regulating reading comprehension, and (f) distinguishing different types of texts. These reading strategies were chosen in accordance with many of other studies in reading instruction (De Corte *et al*, 2001; Klinger & Vaughn, 1996; Pressley *et al*., 1992; Palincsar & Brown, 1984). This series of reading strategies

encompasses the following features: (1) the strategies are organized according to the general sequence of reading activities, before reading, during reading and after reading; (2) they are organized from simple strategies, such as decoding and finding the meaning of new words, to more complex ones such as summarizing, reasoning and inferring meaning *between lines*; (3) the series contain both cognitive strategies (making predictions and summarizing main ideas of the paragraphs or text) and metacognitive strategies (e.g., monitoring and regulating understanding).

The other line of intervention is peer tutoring (PT). Peer tutoring is a specific form of peer-mediated instruction in which one student serves as a helper or tutor for one another student, the tutee (Topping & Ehly, 1998; Topping, 1998). We included two categories in the intervention: “cross-age tutoring” (CA, a fifth grade tutor tutoring a second grade tutee) and “reciprocal same-age tutoring”(RSA, two classmates taking turns to tutor each other, either in the fifth or second grade).

In the intervention, we interweaved these two lines of intervention in one program that was organized on a weekly basis for a whole school year. All students in the experiment received instruction on reading strategies. In CA and RSA groups, students were matched into dyads to read together. Because most of Chinese schools lack reading resources for students, to facilitate the research, approx. 2,500 children’s books were purchased for students in both grades and they were circulated regularly among the participating schools. To evaluate the different innovation elements, five conditions were set up: reading strategy instruction plus cross-age tutoring group (CA), reading strategy instruction plus reciprocal same-age tutoring group (RSA), reading strategy group (RS) and control group (CON). Another free reading group (FR) was set up as a second control group to test the effect of the new books. Students in all three experimental conditions got a series of lessons in the six reading strategies. All spent a comparable amount of time in practicing the reading strategies. But in CA and RSA groups, this practice mainly took place in peer tutoring dyads. In RS groups, this practice was done

individually in whole class settings. Students in the control groups spent a comparable amount of time in other Chinese language related activities.

1.3. Investment and resource allocation

This project was funded by the Special Research Fund from Ghent University, Belgium. The costs of the implementation included hiring a project coordinator and assistants, organizing teacher training and assessing the results of intervention. In addition, the costs also included purchasing the 2,500 books and various materials related to the intervention. The schools in the intervention invested with teachers' manpower and about one hour of extracurricular time per week. The project coordinator gave a full support to everything related to the implementation of the project. At the end of the one-year intervention, she also informed all schools briefly about the intervention effects. Students were assessed for three times: right before and right after the intervention, and half a year after intervention was stopped.

During this one and a half year period, the implementation and evaluation of the project went fairly smooth. However, we were confronted with various challenges.

2. Challenges and solutions in implementing the intervention

2.1. Challenge 1: to land the educational intervention on the local platform

As a project funded by a Belgian University in a Chinese educational setting, it was almost impossible to recruit Chinese primary schools directly. No Chinese schools were familiar with the Flemish university nor did they recognize its authority in educational research. Although in China it is beneficial for primary schools to participate in educational research and research experience of schools and teachers is considered to be one of the important indicators of school quality, this research experience has to be recognized by the local educational authority. In the city of

Beijing, there is a hierarchical educational administration, from city to district educational committee, to the district education and research institute and primary schools, as is shown in Figure 1. With regard to educational research, there is a close research partnership between the district educational research institute and universities doing research in education, e.g., Beijing Normal University. The district education and research institute depends on university professors or researchers to organize research and write reports; and universities need support from the district education and research institute to recruit schools in research projects. In our situation, to land the project in Beijing successfully, we needed to fit it in the existing research hierarchy and networks first. Therefore, a research partnership with Beijing Normal University was formed, based on common research interest. To make our developmental project recognized by the district education and research institute and to attract schools to participate, we reached an agreement with professors from Beijing Normal University to integrate our project as a self-funded subproject of a *National Key Project* in the 10th Five-Year Plan in Educational Science *"Education and development - integrative psychological research on creative intellectuals"*. Therefore, with the help of Beijing Normal University, we made contact with the district education and research institute to recruit and select schools. Once the "back-door" contact with the primary schools was formed, the Chinese project coordinator hired by Ghent University could start to discuss research plans with school principals and teachers and prepare for the intervention.

The local educational hierarchy and the partnership with Chinese primary schools are represented in the Figure 1. On the left hand side is the hierarchy of the educational administration; on the right hand side of the figure, the broken line with an arrow indicates the "back-door" contact from Ghent University to the local Chinese primary schools.

Eight schools were finally chosen for the study. About 900 second graders from 24 classes, 739 fifth graders from 20 classes, and 36 teachers were involved. 70% of

the students are from working class families, with a family monthly income of 3,000 - 4,000 RMB (or 375-500 USD). About 10% of the students come from "city new immigrants" families (Chinese farmers moving from countryside to big cities).

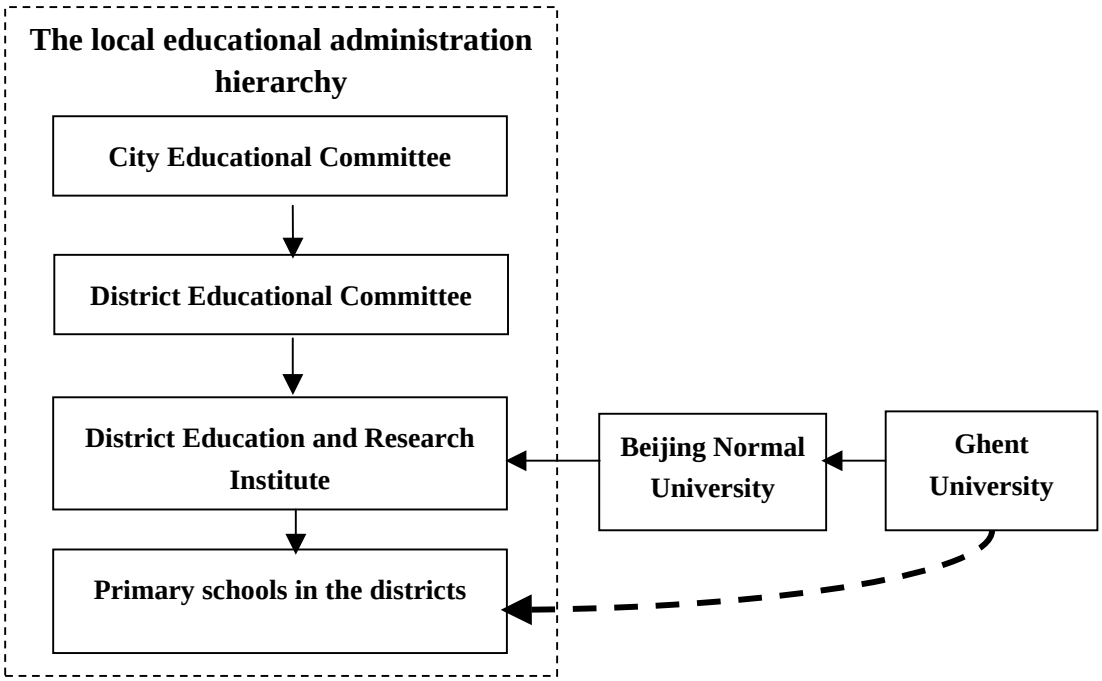


Figure 1. The local educational hierarchy and research partnership between Ghent University and Chinese primary schools

Once schools agreed to participate in the study, practical arrangements with regard to time were to be made. As the consequence of years of centralization in education, there is a tight control on school curriculum and teaching hours. Schools within the same district of Beijing have more or less the same subjects, textbooks, and fixed teaching hours per week. They are also evaluated from time to time by uniform testing organized by higher-level educational authorities. Schools hardly have their own autonomy in choosing teaching content. This situation is very different from that of Flemish schools in Belgium, where schools can choose their own way of

organizing teaching as long as at the end of Grade 6 the majority of the students meet the Flemish educational standards. As a consequence, Flemish schools have the autonomy to include reading strategies instruction and peer tutoring into their regular curriculum and organize the intervention during school hours. In Beijing, the only time during which we could organize the intervention is extracurricular hours, between 3:30-5:00pm. That is the period after school hours and before parents could come to fetch their children after work. All students stay at school during these extracurricular hours, either participating in various activities, such as arts, sports or literature appreciation, or simply finishing their homework under the supervision of their teachers. Each of experimental schools chose a fixed day of the week and allocated one hour to carry out the experiment.

In this way, we finally landed the project firmly on the local educational framework.

2.2. Challenge 2: forms and quality of teacher training

The intervention was carried out by the regular Chinese language teachers. Before the intervention started, experimental teachers needed to be mentally and technically prepared for their job in the intervention. Teacher training was organized through a three-day intensive training. During the training, the main goals and the two elements of the intervention were introduced to the teachers. The curriculum and critical issues were explained thoroughly to all the participants. In addition, teachers were also given full support with teaching materials. Teachers' manuals were written out in detail, explaining how each lesson was designed and taught. Big posters with highlights of reading strategies or tutor training to be hung on the walls of each classroom were demonstrated and discussed with the teachers. Explanation was given on how to use them to facilitate teaching and students' learning. A well-designed students' exercise book, including "reading strategy cards" (with steps of each reading strategy and main questions to be answered), administrative cards (to keep record of reading materials read and reading

strategies practiced) and selected reading materials were discussed with the teachers. Key lessons on tutor training and reading strategy instruction were demonstrated live using role-playing. Finally, group discussion was organized to give the teachers opportunities to express their opinions about the project, their perceived difficulties and suggestions.

The advantages of this type of intensive training were that it was very comprehensively and systematically done and it fitted the teachers' schedule. The big disadvantage was that it only gave a rather superficial impression of teachers' understanding of the project. By the time they were to do the job, the teachers seemed to have forgotten everything or they felt that they haven't got enough support and preparation to carry out the job. Generally speaking, it didn't seem to be that easy for primary school teachers to absorb the aims, purposes or main tasks of a new research so quickly in the short-term training. From the teachers' reactions, it became clear that within the scope of the teachers' experience and knowledge, it was difficult for teachers to take up their roles and carry out the tasks right at the beginning of the project. It was observed very often that teachers feel that they are left on their own "precisely at the time when they are most in need of a guide" (Bitan-Friedlandera *et al.*, 2004, p. 608).

To cope with such situation, besides the short-term training, a long-term regular coaching and support of the experimental teachers was needed. During the whole duration of the intervention, the project coordinator has worked with the teachers side by side and given them help and support whenever necessary. Every week, whenever there were new teaching contents, either in tutor training or reading strategy instruction, she organized a half an hour meeting at noon with all the experimental teachers from the same school, highlighting the main contents, the difficult parts and potential problems. Useful experience or good examples from other schools were also exchanged and communicated to the teachers. Sometimes the coordinator also gave demonstration lessons and taught in some classes when she was invited to do so. She also took turns to observe teachers' teaching and

organization in each of the classes and gave feedback, suggestions or reinforcement to support the teachers' work.

The long-term support and working together was a time-consuming and tiring job, but according to us, it was the only way to ensure a good quality teacher training and to guarantee that the teachers did what they were expected to do. The side-by-side working relationship and partnership also kept teachers motivated in carrying out the program.

2.3. Challenge 3: teaching the content

With regard to the implementation of reading strategies instruction, teachers were expected to carry out the program with their students. However, as Peterson & Swing (1983) pointed out, most teachers are not typically trained to teach reading strategies. Some of them also felt very insecure with coaching dyads on their interaction during peer tutoring. One of the main challenges when training and supporting the teachers was how to deal with their feelings of insecurity with respect to the content of the instruction.

Teachers had to guide students into the use of reading strategies in an inviting and creative way, while sticking to the basic principles of reading strategy instruction. In China, with its long tradition of centralization in education, teachers are used to teaching based on textbooks and teaching manuals. For more than two decades, there was only one version of Chinese textbooks for all primary schools across the country and uniform manuals to assist teachers on how to teach Chinese. For ordinary schoolteachers, it is a real challenge to carry out extra tasks that are not related to their textbooks or instruction manuals. Therefore, a good mental orientation and practical teaching preparation were needed.

Teachers in the peer tutoring conditions were expected to prepare and coach tutors with respect to interaction skills and pedagogical techniques. These tasks were beyond their daily teaching routines and therefore even a bigger challenge for them.

They had never had any social skills training experience and they felt insecure about this part of job. They had to make a major shift in their role as a teacher, from instructor and classroom manager to coacher, helper and observer of students working in dyad.

To solve these problems and to support teachers with their specific tasks, well-written teachers' manuals with goals, steps, important issue reminders, examples of questions to be asked in the classroom, exemplary reading materials for instruction were distributed to teachers during the teacher training. Teachers who felt insecure about specific tasks were suggested to follow the manuals' instruction. Those teachers who felt confident with the understanding of the intervention, got the freedom to build on their inspiration and experience and teach the content in the way that fitted their students best.

In the whole process of professional development that teachers went through, a number of phases can be distinguished:

(1) Content introduction: it was mainly done at the starting phase of the project, during teacher training. The introduction of the reading strategies includes declarative, procedural and conditional knowledge of the reading strategies. These include names, purposes, mainly procedures of the strategies and good timing (when to use in the reading process), their functions in text comprehension and so on. A detailed teachers' manual was provided and demonstrations of how to teach specific lessons were organized with teachers. These demonstrations were mainly done in the week before a new reading strategy was introduced. Based on training and teachers' reading of the manual, a small group discussion within each school with all teachers involved in the experiment was organized. It helped teachers to refresh the teaching content to pay special attention to certain critical points or potential problems involved in teaching or students' learning were stressed by the coordinator.

(2) Information exchange: teachers within the same schools spontaneously formed little collaborative learning groups and exchanged their ideas, experience and problems. The coordinator also brought useful experience from other schools and showed good examples of videotaped teaching lessons. Both experience and potential problems in the classroom were communicated among all the schools.

(3) Integrated reflection: this mainly occurred in the later phase of the intervention. After all the reading strategies were introduced and practiced, students needed to develop a comprehensive concept of the reading strategies and learn to use them flexibly during reading. During the last phase of the experiment, we asked teachers to organize discussions with students and give students an overall picture about the relationships among the six reading strategies. The first and third strategies are related, because they both involve imagination, guessing or prediction. The first strategy — "activating prior knowledge" is an open guess on what the text is about, based on the title, content or pictures (if there are any) in the book and students' prior knowledge about the topic. The third strategy — "making predictions" is to predict what will happen next in the text, based on what has been read. It is more content restricted *guessing* compared to Strategy No.1. The second, fourth and sixth reading strategies belong to the same group, because they are all content related comprehension. These strategies are about understanding new vocabulary, expressions, specific sentences, paragraphs or main ideas; or it is about classifying different types of texts. The relationship between metacognitive strategy (Strategy No.5) and cognitive strategies (the other five strategies) was also pointed out clearly to the students. Metacognitive strategy in fact is above all the other cognitive strategies and it can invoke cognitive strategies to solve specific problems during reading.

Soon after the start of the project, with a lot of feedback, demonstration, reassurance and praise, teachers felt gradually relaxed about their roles and tasks in the intervention. Some teachers tried to link the existing knowledge of students with the new teaching tasks. For example, reading Strategy No. 6, "distinguishing

different types of texts", was a completely new reading strategy that was never covered in the Chinese curriculum in primary schools. But from strategic reading point of view, it is important and useful for students to have some ideas about different types of texts (e.g., stories or narratives, scientific explanatory or comments), so that they can adapt their reading to characteristics of different texts and grasp the main ideas more efficiently. Students did meet different types of texts, but never were asked before to pay attention to their characteristics. Reading Strategy No.6 was a challenge for both teachers and students. After discussing with the project coordinator, one teacher in the second grade, instead of following the teaching scenarios described in the teachers' manual, taught in a way that she thought would be easier for her students to follow. She recalled together with the students various types of texts they had learnt from the Chinese textbook and discussed similarities and differences among them. Then she classified these texts into three types and used a few words to described the characteristics of each type. After the students formed some ideas how each type was like, she then presented a few short different types of texts and asked students to classify them. This way of linking new teaching content to existing knowledge was not described in the teachers' manual, but turned out to be quite successful in the strategy instruction. This teaching example was soon spread around and shared by other teachers in the experiment. This is just an example to show how teachers in the experiment as a group to tackle problems or challenges during the intervention.

2.4. Challenge 4: treatment fidelity and keeping teachers motivated

Treatment fidelity is always a crucial element in any educational intervention study. Our intervention lasted for a whole school year. For schools, it was a big investment both in terms of time and human resources. Chinese teachers are hard workers and they stay at school for more than 9 hours a day and have a heavy teaching load. Since the class size of most Chinese classes is 30-50, teachers also have a heavy load of homework to grade. Most teachers teaching Chinese language are also the head of the class and it is also their responsibilities to tackle any

problems with or from students and to communicate with parents. Under such circumstances, it is very important to have a good quality control of teachers' work during the treatment and to ensure treatment fidelity; on the other hand, we also tried to improve their efficiency and keep them motivated during the whole process.

During the treatment, the project coordinator has visited the schools almost 5 times a week to organize discussions with teachers about how to organize the forthcoming strategy introduction lesson, to observe and videotape how they did in classroom, and to give feedback or reinforcement about their practice.

Pre-designed checklists of each lesson with the outline of the classroom organization were distributed to teachers. They were used not only as a checklist to check for treatment fidelity, but also as an outline to remind teachers of the main elements of the teaching. Teachers were also asked to report their spontaneous practices that were different from the checklists.

Classroom observation and videotape recording were also means to evaluate the teaching quality and fidelity of the intervention. Every 2 weeks, each teacher was observed either by the coordinator or one of the two well-trained assistants. Feedback was given orally to teachers about how well they were doing with organizing teaching or coaching of students. Checklists from teachers were compared to classroom video recording. The match was above 90%. In this way, we were reassured of the quality of classroom practice. This regular evaluation also prevented teachers from slipping away from the experimental design and strengthened the bond between the coordinator and the teachers.

After almost one whole year of working together with the teachers, we were happy with the intervention fidelity and quality.

3. Results

The results of the intervention will be presented both in quantitative and qualitative forms. For the quantitative part, we present the results on the fifth graders' reading comprehension. Since similar results are found in both grades, here we will just describe briefly the results in the second grade and point out the differences between the two grades. And for the qualitative part, we present teachers' written comments and reflections on the project.

3.1. The experimental results

A multilevel repeated measures approach was used to assess the short-term and long-term effects of the intervention, taking into account the nested structure of students within classes and within schools. Because no significant differences among the three experimental groups were found at pre-, post- or retention test, for this chapter we chose to report the findings in a simpler way, and merged RS, CA, RSA conditions into one experimental condition. The same was done for the two control groups CON and FR, which were merged into one control group. The fixed part of the multilevel regression models are presented in the Table 1.

Model 0, 1 and 2 illustrate the model building process of the multilevel regression analysis. Model 0 is used to as a starting point to build more complex models. In Model 1, we first adjusted for the differences among the conditions with regard to confounding variables such as students' gender, their reading attitude before the intervention, mothers' average diploma, class average reading level before the intervention, and schools' prestige (whether they are city or district *key schools* or just ordinary schools). Then in Model 2, we enter the dummy variable *Experiment* indicating the experimental condition. We are mainly interested in the short-term intervention effect at the posttest and long-term effect at the retention test. The coefficient for *experiment*posttest* indicates a significant intervention effect, with an effect size of 0.23. The coefficient for *experiment*retention test* indicates a

significant long-term effect, with a larger effect size of 0.53. Figure 2 illustrates clearly the difference between the two conditions in their reading comprehension growth.

Table 1. Repeated measures multilevel analyses on 5th graders' reading comprehension

	Model 0	Model 1	Model 2
Fixed effects			
Cons (Intercept)	-0.444	-0.639	-0.664
SE	*0.096*	*0.052*	*0.075*
Posttest	0.522	0.518	0.396
SE	*0.043*	*0.046*	*0.061*
Retention test	0.488	0.598	0.401
SE	*0.059*	*0.068*	*0.089*
Girl		0.167	0.161
SE		*0.052*	*0.053*
Reading attitude pretest		0.313	0.304
SE		*0.038*	*0.039*
Mother's average diploma		0.119	0.114
SE		*0.025*	*0.025*
Mean class reading pretest		0.218	0.425
SE		*0.106*	*0.129*
City key school		0.550	0.645
SE		*0.072*	*0.080*
District key school		0.109	0.166
SE		*0.077*	*0.089*
Experimental conditions			
Experiment			-0.047
SE			*0.077*
Experiment * posttest			0.230
SE			*0.073*
Experiment * retention test			0.527
SE			*0.093*
-2log(likelihood)	4121.274	3517.397	3487.428
N (students)	1942	1718	1718

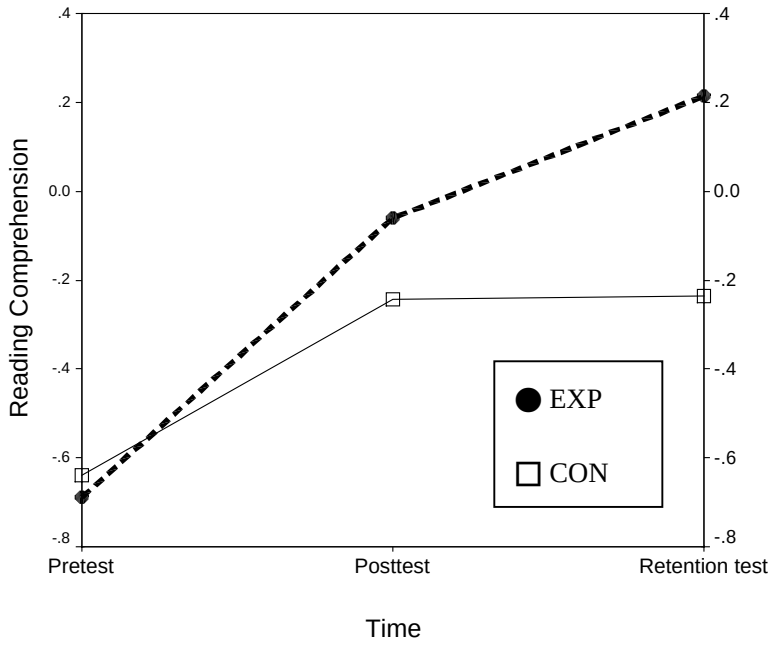


Figure 2. The growth of reading comprehension of fifth grade students in experimental vs. control group

The results in the second grade were similar to those in the fifth grade. The only difference between Grade 2 and Grade 5 is that in Grade 2 CA did significantly better than RSA in keeping the gains in reading comprehension for six months after the intervention was concluded. For the detailed results in Grade 2, please read the report in Chapter 4.

3.2. Sustainability of the influence of the intervention in general

3.2.1. Students' growth in reading comprehension

From Figure 2, we can see that even after the intervention was concluded, students in the experimental condition had the tendency to make bigger progress in reading comprehension, compared to the control group. We can infer that the instruction of

reading strategies, has a long-lasting influence on the development of students' reading comprehension ability.

3.2.2. Teachers' enthusiasm towards the intervention and their comments

Teachers were encouraged to write papers to give free comments or reflection on the intervention process and these articles were recognized by their schools and the District Education and Research Institute as the end products of their research participation. About 90% of the teachers handed in their papers. In general, they were enthusiastic about the intervention, esp. the series of reading strategies. We captured partly their reaction towards reading strategies and peer-tutoring as follows[1]:

> *Teachers taught reading strategies to students and these strategies made students master a way of reading and thinking, which could apply to other reading activities and learning activities* (written by a teacher in Grade 2 from an ordinary school).
>
> *[Strategy—making prediction] inspires students' imagination and encourages students to talk about their own ideas and it also helps with their writing skills* (written by a teacher in Grade 5 from a *city key school*).
>
> *Under the instruction of reading strategies, students improve their ability in comprehension, summary and reasoning and these are crucial elements in reading processes* (written by a teacher in Grade 5 from a city key school).
>
> *Peer tutoring makes students learn to read with others…. It helps students to reach their best level… and also improves the communication and interpersonal relationship among students* (written by a teacher in Grade 5 from an ordinary school*).*

[1] The following paragraphs are translated from Chinese.

> *Especially for the students with lower achievement, it [peer tutoring] gives them an experience that they never had before.…Instead of being helped, now they could also help other students. This [experience] has greatly improved their self-confidence* (written by a teacher in Grade 5 from a *district key school*).

3.2.3. Reactions at the school level

All school principals in the experiment were proud to be part of the project of "reading strategies and peer-tutoring". We have got full support from the school principals and directors in charge of Chinese teaching. They helped to run the project smoothly. Many schools listed the project as one of their biggest research actions on their research agenda. After we have finished all the research action in the experimental schools, we were still contacted by some of them to ask for support to carry on the instruction of reading strategies in other grades. Two schools demonstrated the reading strategy instruction project to parents on the open school days. Peer tutoring was used as one of the features to attract more students to enroll in their school. After the project, news was spread among primary schools. We also got enthusiastic request from other schools.

4. Discussion

So far, most of the research mainly focused on the results of innovations in schools, but omitted or only briefly mentioned the implementation process and how it would influence the results. The results of educational interventions are linked to how well we could adapt the intervention plan to the local platform of education; they are also inherently linked to the preparation and implementation process. More emphasis should be given to how to run the projects properly according to the educational design and how to guarantee the fidelity of the treatment.

Working with teachers is not such an easy thing, taking into account the differences in working or thinking style between academicians and teachers. The side-by-side experience not only helped us to have a better understanding of teachers' problems and difficulties, but also gave us opportunities to learn from them and their teaching activities. This experience helped to build up a healthy and stable relationship and laid a solid basis for the good results of the intervention.

From a theoretical point of view, this study helps to answer a series of questions about the effectiveness of peer tutoring across cultures, esp. in the Chinese educational system. Our results have shown that there is a big gain for students who received reading strategies instruction. This result is consistent with similar studies that have been done in the West during the past decades. The effect of reading strategies instruction and practice on students' reading comprehension seems to be independent of language and culture. And it is also effective in improving students' reading skills and reinforcing their gains in the long run.

From the results of students' reading comprehension assessment, we can conclude that all experimental groups did better than the control group, but the extra value added by peer-tutoring practice was not as much as it was hypothesized. In addition, more than 80% of the articles written by the teachers in our project reflect on reading strategies and less interest was given to peer tutoring practice. This could be due to the physical restrictions (e.g., small-sized classrooms and rather big class sizes) in Chinese schools in carrying out peer tutoring. It is also possible that the advantages of peer tutoring for students' reading were not fully appreciated by teachers.

From a practical point of view, this small scale innovation can serve as a probe stone for larger scale educational reforms. As is mentioned at the beginning of the chapter, Chinese reading instruction has shown its weakness in fostering students' reading comprehension development. There is also a gap between the goal of reading instruction stated in curriculum standards and the practice in schools. Our

study has shown beneficial effects on students' reading comprehension growth. Since it was carried out as a quasi-experimental study in natural classrooms, the ecological validity also holds. It can be used as an empirical basis for more innovation in reading strategy instruction. The next step will be to generalize this study into more schools or try to integrate reading strategy instruction as part of the regular Chinese reading curriculum and make more students benefit from the practice.

5. Conclusions

The developmental project was landed successfully in the Chinese local educational research framework and scheduled in schools' extracurricular hours. We also did a big investment of time and manpower to improve treatment fidelity. Besides a well-prepared training and sufficient well-written teaching materials, regular evaluation of classroom practice played an important role. Working together with teachers, discussing with them about their problems and difficulties helped to keep teachers motivated and grow their confidence and proficiency. The project achieved not only a short-term, but also a long-term effect on students' reading comprehension development.

Appendix 1:

Chinese teaching in classroom—a 2nd grade lesson transcription
Source: young teachers' reading instruction competition in Guang Zhou City
http://www.thjy.edu.cn/zhouweihua/article/1015/632810919293750000.aspx

Date: 2006-4-14 Place: Jiu Bu Qian Primary School

Teacher's affiliation: Xian Lie Middle Road Primary School

Text: Three Sons (the first lesson on the text)
(Note: T-teacher S-student)

T: If we go to Disneyland, we need an actor's certificate. We must get it from reading well.

(Students started to read the text.)

T: Read these words

(Students read the new words. The teacher praised students for their reading. He/she used a word game to let students read more new words.)

T: The whole class read the sentences.

*(Students **read aloud together** the sentences, from the first one on. The teacher corrected students' pronunciation, when necessary).*

T: **Girls read** the fourth sentence; **boys continue** the next. Please invite your **deskmate** (the student who share the same desk) **to read** Sentence 4. Read the last sentence. Now we **perform the reading**. What are the ways you would like to read the text?

S: Reading per paragraph, reading with a loud voice, reading with a low voice, reading with expressions…

T: Find the paragraphs of the text and read!

T: How many paragraphs are there? Who are in the text?

S: 13 paragraphs. There are three Mamas and one grandpa. There are three sons.

T: **Read silently** from paragraph 1 to 6. Why did the first Mama praise her son? Why did you add accent to the word "not only"?

(Students read again.)

T: Use "not only… but also" to make sentences.

T: What is "smart" like? Who want to read this paragraph? Please stand up and **read aloud**.

(Some students started to read this part aloud.)

T: Evaluate their reading.

(Some other students started to give comments: "when reading the word throat, their voice was too low"; "it didn't sound beautiful"…)

(The teacher praised the courage of the students who gave comments.)

T: Why did another mama say nothing? Why did you read (that part) without any expression?

S: Because the son has nothing special?

T: Maybe because this mother is very strict with her son. Therefore she didn't say anything. Instead, she hides her love in her heart. Her way to express love is different. Now perform Mamas' words. Who wants to be the first Mama, the second, the third? …

*(Finally teachers gave homework: **read more!**)*

(Evaluation from an expert teacher: The teacher could guide students to read with expressions, and dig out the author's feelings from between the lines. He/She did well.)

(The End)

Appendix 2:

Chinese teaching in classroom—a 5th grade lesson transcription

Source:
http://222.16.82.130/siwuliunianjiyuwen/Article/3602/632809942820781250.aspx
Title: The Falling Peanuts
(Note: T-teacher S-student)

T: Students, do you like to guess a riddle?

S (together): yes.

T: Now I will tell you a riddle and you guess what kind of plant it is. “Pieces of beard in sand and earth; self-made houses and self settle-down; above the earth only flowers no fruits; under the earth only fruits no flowers. ”

S: Peanuts!

T: *(Show the projection of peanuts)* peanuts are also called falling peanuts, because when the flowers fall, they enter the earth and turn into nuts. That’s why they are also called falling peanuts. The author of the text is Xu Dishan, a novelist and prose writer. In his childhood, his father once talked about peanuts and it left a deep impression on him. He decided to be a man like peanut, a person that is useful for others. Therefore, he used “falling peanut” as his pen name. **What did his father talk about peanuts? How did he talk about it? Let’s read the text.**

(The teacher wrote the title Peanut on the blackboard. Students read the title together.)

T: **Chose freely the way you like to read the text** and think: what has the author written about the topic?

S (after reading): Growing peanuts, harvesting peanuts, tasting peanuts and talking about peanuts.

(Teacher wrote down the students' answers on the blackboard.)

T: How did the whole family talk about the merits of peanuts?

S1: The sister said: "it tastes well."

S2: The brother said: "it can produce oil." "I" said: "it is cheap, and everybody likes to eat it."

S3: The father said: "the most precious thing is that its nuts are hidden under the earth."

T: The merits that the brothers and sisters talked about are quite obvious, but the point that the father talked out is often overlooked by common people. How many times did the father give comments to peanuts? What are the differences between these comments?

S: Two times. The first time, he compared peanuts with peaches, megranate and apples; the second time he told us the philosophy of being.

T: Read aloud together the first comment of the father and think: "how did the father explain the merit of peanuts through comparison?"

S: The father felt that peaches, megranate and apples were hung high on the branches of trees and showed themselves off, but peanut just hided under the earth.

T: Your answer is very good. What are the key words stressing the merits of peanut?

S1: I think "hidden in the earth", "bright read and tender green", "hung high on the braches" are the key words.

S2: I think, in addition, "low under the ground".

(The teacher wrote down these words and expression on the blackboard.)

T: Who wants to read?

(One student read.)

T: What do you think of his reading?

S1: I think he stressed the key words while reading.

S2: I think he read with full expression and fluency.

S3: He performed with different tones how peaches, megranate and apples show off.

T: Do you also want to read with expression?

S (together): Yes.

T: Good. Let' s read together with expressions.

(Students read all together with expression.)

T: Peaches, megranate and apples are with bright red and tender green, but peanuts hide their nuts under the earth, waiting for people to dig out and make use of them. Therefore, what kind of merits did you feel about peanuts?

S1: I feel the silent sacrifice spirit of peanuts.

S2: I feel the unselfish sacrifice spirit of peanuts.

S3: I feel the spirits of peanuts far away from fame and interests, keeping its nature without much ornaments.

T: The father used peanuts to educate his children to be what kind of persons? Underline these sentences. Who wants to read the father's second comment?

(One student read this part.)

T: What kind of person did the father want his children to be when he used peanuts as an example to educate them?

S1: The father used peanuts to educate his children to be a useful person for others.

S2: The father used peanuts to educate his children to be a person away from fame and interests, sacrificing themselves silently.

T: How did “I” (the author) understand the father’s words? Underline the part with “~~”.

S: “I” think a person should be useful, not just be descent but useless for others.

T: How do you understand the author’s word? Give an example from real life.

S1: Some persons look very handsome, but don’t have any skills.

S2: Once I saw a very beautiful aunt eating a banana while walking. After finishing eating, she just threw the skin of banana on the street. What if any old grandpa or grandma passes by and steps on the banana skins and falls?

S4: Some persons, although they look ugly, are willing to help others.

S5: My neighbor is a cleaning worker. He gets up everyday before day breaks to do cleaning. I think he is a useful person.

T: In fact, in our life, there are many persons like the common cleaning workers, who in his working post sacrifice themselves silently. They are the persons like falling peanuts. Are there any persons like falling peanuts? Give examples.

S1: Yes, teachers sacrifice themselves silently in their work for their students.

S2: The cleaning workers. To clean the streets, they would rather sacrifice their youth.

S3: Candles are great. They sacrifice themselves in exchange for the brightness for us.

S4: The spring silkworms. For beautiful clothes for human beings, they spin out silk till the end of their lives.

S5: The sun! Because your unselfish sacrifice light and heat, the Earth is full of life and various colors.

T: You students have said so well. There are so many persons and things in our life like the falling peanuts, if you pay attention to them.

T: In this text, from planting peanuts to harvesting peanuts and tasting peanuts, how long does it take and how much space does it take in the text? From tasting peanuts to talking about peanuts, how long does it take and how much space does it take in the text? Why did the author arrange like this?

S1: The author wrote like this in order to stress that we should be persons like falling peanuts.

S2: The text emphasizes *talking about peanuts*, in order to tell us the philosophy of life.

T: This text is mainly about talking about peanuts. The father's comments are written in detail. This could stress the central topic of the article. The other parts are written in brief, but they lay a basis for the main content of the article. They can't be omitted. When we write, we should be clear about the central topic, so that the central topic could be emphasized and the difference between main and minor content is clear.

T: This text tells us a deep philosophy through describing peanuts: to be a useful person, one should not be a person descent (good-looking) at the surface and useless for others. This method is "metaphorization by describing an object". In our surroundings, there are so many things that make us think, and let us understand a deep philosophy. Such as: a chalk is a very, very common little thing. In order to let our students acquire knowledge, it doesn't hesitate to sacrifice. A candle lightens

others, but destroys itself… we can learn to use this method when writing compositions.

T: After having learnt this text, what do you want to say? You can write down your ideas and communicate (with other classmates) within your group.

T: Who want to volunteer (to tell us your idea)?

S1: I want to say: “I want to learn from peanuts.”

S2: To be a person, we should not just care about the surface. We should be useful.

S3: There is a saying: “I would rather equip myself with knowledge, rather than make up with jewelry.” I think it is a good saying.

S4: I have learnt to this methods such as, emphasizing the main topic, making clear distinction between the main and the minor topics, and metaphorizing through an object. I think my composition level will become better.

S5: I think the father’s comments are not quite right. Peaches, megranate and apples hang high up on the branches. They are not showing off themselves. These are their growing characteristics. But why did the father say like that? Here the father was just making a metaphor for contrasting, in order to emphasize its silence sacrifice spirit. Of course, no metaphor is perfect.

T: Modern society not only needs someone like peanuts to sacrifice themselves silently, but also need someone like peaches, megranate and apples who dares to demonstrate his/her capacity.

(The End)

References:

Anderson, V. & Roit, M. (1993). Planning and implementing collaborative strategy instruction for delayed readers in Grades 6-10. *The Elementary School Journal*, 94, 2, 121-137.

Baldwin, C. M. (2002): Teaching self-reflection and meta-cognition: *ECT A Review of General Semantics*, 59, 1, 71-81.

Beijing Educational Committee (2002). *Bei Jing shi jiao yu wei yuan hui guan yu you er yuan, zhong xiao xue zhong deng zhi ye xue xiao jiao shi "shi wu" shi qi ji xu jiao yu gong zuo de jian yi* [*Beijing Educational Committee's suggestions on teachers' continuous education in kindergarten, primary, secondary and middle level professional schools during "the 10th Five-Year Period"*]. Beijing Educational Committee.

Bitan-Friedlandera N., Dreyfusa A., Milgromb Z. (2004). Types of "teachers in training": the reactions of primary school science teachers when confronted with the task of implementing an innovation. *Teaching and Teacher Education*, 20, 607 − 619.

Borkowski, J. G. (1985). Signs of intelligence: Strategy generalization and metacognition. In S. R. Yussen (Eds.). *The growth and reflection in Children*, pp. 105-144. New York: Academic Press.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markmen (Eds.). *Handbook of Child Psychology*, *Vol. 3*, pp. 77-166. New York: Wiley.

Brown, A. L., Campione, J. C., & Day J. D. (1981). Learning to learn: On training students to learn from texts. *Educational Researcher*, 10, 14-21.

Cao, X. B.(2004). Ji qu, zhi yi, shi yi, sheng hua: Yu wen ke tang jiao xue si bu qu [Activating interest, raising questions, explaining questions and sublimation: The tetralogy in Chinese instruction in classroom], Zhong Xiao Xue Xiao Zhang [*Principals in Primary Secondary Schools*], 11, 54.

Chen, J. S. (2004). Guan jian shi jiao yu guan nian de zhuan bian: Ke cheng gai ge zhong chu xian de ji ge wen ti [The key is to change ideas about teaching: A few questions arising in the teaching reforms], *Ban Zhu Ren* [*The Headmaster]*, 11, 12-14.

Cheng, L. Y. (2004). Ke cheng gai ge zhong yu wen ke tang jiao xue bi duan li xi [The analysis of weak points in Chinese classroom instruction with examples during the curriculum reform]. *Zhong Xue Yu Wen: Jiao Shi Ban* [*Middle School Chinese: Teachers'Edition*], 11, 5-8.

Chinese Ministry of Education (2001). *Quan ri zhi jiu nian yi wu jiao yu yu wen ke cheng biao zhun* [*Full-time Nine-Year Compulsory Education Chinese Curriculum Standards*], Chinese Ministry of Education.

De Corte, E., Verschaffel, L., Van de Ven, A. (2001). Improving text comprehension strategies in upper primary school children: a design experiment. *British Journal of Educational Psychology, 74, 4, 531-560.*

Ding, S. H. (2005). Dai zhe xue sheng yi shi zou jin jiao cai, huai chuai ke cheng yi shi zou chu jiao cai : Dui dang qian yu wen ke cheng jiao xue yu an she ji de si kao. [Carrying the awareness on students to step into the textbook, and carrying the awareness of curriculum to step out the textbook: Thoughts in present Chinese classroom teaching presetting], *Jiao Yu Xian Dai Hua* [*Modernization of Education*], 5, 95.

Gu, Q. (2005). The perception gap in cross-cultural training: an investigation of British Council English language teaching projects in China. *International Journal of Educational Development*, 25, 287-304.

Jacobs, J. E. & Paris, S. G. (1987). Children's metacognition about reading : issues in definition, measurement, and instrument. *Educational Psychologsist*, 22(3&4), 255-278.

Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities for students with learning disabilities who use English as a second language. *The Elementary School Journal*, 96, 3, 275-293.

Levin, H. M., Glass, G. V., & Meister, G. R. (1987). A cost-effectiveness analysis of computer-assisted instruction. *Evaluation Review*, 11(1), 50-72.

Lu, Q. (2006). 'Zhen shui wu xiang' shi yu wen—guan yu dang qian yu wen ke tang jiao xue de si kao. ['The real water without fragrance is Chinese'--thoughts about the present Chinese instruction in classroom], *Yu Wen Jiao Xue Tong Xun* [*Journal of Chinese Teaching*], No.2, 38-39.

Osman, M. E., & Hannafin, M. J. (1992). Metacognition research and theory: Analysis and implications for instructional design. *Educational Technology, Research and Development*, 40, 2, 83-99.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*,1, 117-175.

Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve Children's reading awareness and comprehension. *Journal of Educational Psychology*, 76, 1239-1252.

Perkins, D. N., & Salamon, G. (1988). Teaching for transfer. *Educational leadership*, 46(3), 22-32.

Peterson, P.L., & Swing, S.R.(1983). Problems in classroom implementation of cognitive strategy instruction. In Pressley, M. & Levine, J.R. (Eds.).

Cognitive strategy research: Educational Applications, pp. 267-287. New York: Springer-Verlag.

Pressley, M., & Van Meter, P. (1993). Memory strategies: Natural development and use following instruction. In R. Pasnak & M. L. Horne (Eds.), *Emerging themes in cognitive development: Vol. 2, Competencies* (pp. 129-165). New York: Springer-Verlag.

Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation : transactional instruction of reading comprehension strategies, *The Elementary School Journal*, 92, 5, 513-555.

Pressley, M., Goodchild, E., Fleet, J., Zajchowski, R., & Evans, E. D. (1989). The challenges of classroom strategy instruction. *The Elementary School Journal*, 89, 301-342.

Qi, Z. H. (2004). Dang qian yu wen ke tang jiao xue zhong cun zai de liang ge wen ti [Two problems in present Chinese instruction in classroom], *Ning Xia Jiao Yu* [*Education in Ning Xia*], 5, 8.

Sawyer, R. J., Graham, S., & Harris, K. R. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities. *Journal of Educational Psychology*, 84(3), 340-352.

Topping, K. J. & Ehly, S. (1998). Introduction to peer-assisted learning. In K.J.Topping & S. Ehly (Eds.), *Peer-assisted learning* (pp. 1-23). Mahwah, NJ: Erlbaum.

Topping, K. J. (1998). Effective tutoring in American Reads. A reply to Waski. *The Reading Teacher*, 52(1), 42-50.

Zhong, Y. D. (2006). Shi xian jiao xue shuang ying de xue xi ce lue—he zuo xue xi [learning strategy to achieve win-win result--cooperative learning], *Wu Li Jiao Shi* [*Physics Teachers*], 27, 1, 19-21.

Chapter 2

Fostering Metacognition Through Peer Tutoring: A Comparison of Cross-Age and Reciprocal Same-Age Dyads' Interaction Processes*

Abstract

Within the framework of a school year long class wide intervention program, once a week students practiced previously taught reading strategies in peer tutoring dyads. Based on classroom observations, two types of dyads are compared: cross-age tutoring dyads (CA), in which a 5th grader tutors a 2nd grader, and reciprocal same-age tutoring dyads (RSA), with either a 5th grader or a 2nd grader tutoring a classmate. Forty videotaped sessions were time based coded and analyzed using multilevel analysis, with one minute sequences (level 1) within sessions (level 2) as levels. Coding and analysis focused on indicators of metacognitive activity in the dyads. Results show that CA tutors demonstrated significantly more monitoring behaviours, strategy use and information exchange than RSA-tutors. The findings indicate that CA peer tutoring provides more opportunities for practicing metacognitive skills, more particularly for the tutor.

Keywords: reading strategy peer tutoring metacognition peer interaction

* This chapter is based on an article with the same title submitted to *Learning and Instruction* for publication.

1. Introduction

1.1. Peer tutoring

Peer tutoring (PT) is a specific form of peer-mediated instruction in which one student serves as an instructional agent (tutor) towards another student (tutee). Different from more occasional instances of other helping behaviors among students, peer tutoring is characterized by its role taking (tutor vs. tutee) during cooperative learning and it is structurally embedded in the curriculum and classroom organization with frequent (or weekly) scheduled tutoring sessions. To ensure the quality of peer tutoring, some scaffolding, such as structured reading materials are provided (Topping & Ehly, 1998; Topping, 1998b). Two categories of PT can be distinguished according to the age difference between tutor and tutee: "same-age tutoring" (SA) and "cross-age tutoring" (CA). "Reciprocal same-age tutoring" (RSA) is a specific form of SA in which the tutor/tutee role is exchanged regularly so that both students have the chance to act as tutor (Fantuzzo *et al.*, 1992). In this article, we focus on the interaction processes occurring in CA and RSA dyads with 2nd and 5th graders practicing reading strategies.

1.2. Different growth for CA and RSA dyads

The purpose of this study is to investigate the hypothesis that, compared to RSA, interactions in CA dyads offer more opportunities for tutors to engage in metacognitive activities such as monitoring or regulating the tutee's reading behaviors. If so, such differences could help explain the findings from Van Keer (2004), Van Keer and Verhaeghe (2003, 2005) that practicing reading strategies in CA dyads leads to more significant gains in reading comprehension for tutors, particularly on the long-term, than practicing reading strategies in RSA dyads. The underlying reasoning then would be that monitoring and regulating a younger tutee's reading behavior is a helpful step in the acquisition of metacognitive skills needed for monitoring and regulating one's own reading behavior.

Looking at applications for higher order skills learning, much of the research on PT involves types of RSA (e.g., Fuchs *et al*., 1999; King, Staffieri, & Adelgais, 1999). Using RSA, the distinction between tutee-learning and tutor-learning is hard to make. As far as cross-age or non-reciprocal same age tutoring is concerned, most research focuses on the tutees, for whom significant and quite large effects are reported in various reviews (e.g., Maheady, 1998; Topping & Lindsay, 1992). Tutors have often been assumed to be "fully skilled peers" (e.g., Bloom, 1984), and relatively less attention has been given to their progress. Existing results on the gains of tutors appear to be somewhat ambiguous. Sometimes tutors are reported to make academic gains comparable to the tutees (e.g., Miller, Barbetta & Heron, 1994) or even higher (Topping, 1987). However, when higher thinking skills were to be practiced in a peer-tutoring program, significant gains were only found for tutees and not for tutors (Topping & Bryce, 2004)

The present study builds on the research by Van Keer (2004), Van Keer and Verhaeghe (2003, 2005) who investigated the effects of a year-long program in which 2nd and 5th grade students practiced previously taught reading strategies during either individual seatwork, RSA or CA peer tutoring sessions. All students in the intervention were scaffolded by means of "assignment cards" that provided the necessary structuring (Cohen, Kulik, & Kulik, 1982), yet allowed for a high degree of self-managed learning. To improve the quality of the tutors' helping behavior, a tutor preparation program was included (Bentz & Fuchs, 1996; Fukkink *et al*., 1997). For the 2nd grade tutees in the CA-condition significant but rather small short-term effects on reading comprehension were found. The effects did not last long after the intervention was finished. For the 5th grade CA-tutors larger, moderate short-term effects were found (Van Keer & Verhaeghe, 2003). Most remarkably, these 5th grade CA tutors kept on growing in reading comprehension till at least six months after finishing the intervention. Rather large effect sizes were observed at the retention test (Van Keer, 2004). Although for 5th graders in the RSA-condition, similar effects were found at the posttest, no continuous growth

in reading comprehension was observed once the intervention was concluded. No effects on reading comprehension at all were found for 2nd graders in the RSA condition (Van Keer & Verhaeghe, 2003). Applying a quite similar program in Chinese primary schools, very similar results were obtained by Li and Verhaeghe (2005). Moreover, it was also found that 5th graders self-reported use of reading strategies was somewhat higher for CA than for RSA-students (Li & Verhaeghe, 2005).

It seems quite obvious that 2nd grade tutees in CA dyads tutored by elder 5th grade tutors make bigger progress than 2nd graders who are tutored by a classmate in RSA dyads. By natural developmental growth, 5th graders can be assumed to have a deeper understanding of the texts and to have better help-giving skills than 2nd graders. However, it is less obvious why 5th graders tutoring a younger 2nd grade tutee, therefore using reading materials that are below the tutors' reading level, make equal progress in reading comprehension as 5th graders tutoring a classmate, using reading materials that are appropriate for the reading level of both. It is even less obvious why tutoring a younger tutee could lead to continuous growth in reading comprehension for at least 6 months after the intervention was stopped.

At the content level, this finding would appear to contradict Vygotsky's theory about the Zone of Proximal Development (Shayer, 2003; Vygotsky, 1978). According to his theory, the potential development is only fostered by being guided by someone with a higher development level, such as an adult or a more capable peer. Coaching someone below one's own level is not likely to foster his *Proximal Development*. It also seems to differ from existing findings that children of the same age or similar ability would benefit more from working together than children with different age or ability (e.g., Duran & Gauvain, 1993; Manion & Alexander, 1997).

1.3. Research questions and hypotheses

With respect to these CA - RSA differences in reading comprehension growth a number of possible explanations have been given. For example, some suggested that CA dyads are characterized by more tutor authority because of age difference (Fantuzzo *et al.,* 1989), more efficiency in using reading time and better quality of interaction (Topping, 1998a, 2000), more role distinction between tutor and tutee in the CA condition (Van Keer, 2004) or more cognitive as well as affective modeling (Juel, 1996).

Apart from the two factors mentioned by Juel (1996), the explanations suggested so far all refer to aspects of the tutoring setting. Such setting characteristics might yield a quite plausible explanation for short-term differences in reading comprehension growth, especially in *tutees.* However, it is difficult to see how they could act as a direct cause for long-term differences, i.e. differences that are still there or even have become (more) apparent six months after the last peer tutoring session, and – more particularly – how they could account for long-term effects in *tutors.* For example, differences in off-task behavior could eventually explain short-term differences between CA and RSA dyads. However, if CA and RSA dyads would make equal extra progress on the short-term (despite differences in off-task behavior), it would be difficult to see how differences in off-task behavior during the past peer tutoring sessions, could be the reason for a subsequent loss of gain in the long run or a subsequent decrease in long-term growth in reading comprehension achievement in RSA dyads only, particularly since between the second (short-term) measurement and the third (long-term) measurement no peer tutoring nor other practice of reading strategies was done at all.

Long-term effects appear to indicate that some sustainable change has occurred in the students, i.e. a change that is not that easily washed away by time and/or new (eventually negative) experiences with respect to reading. At this point, two (complementary) explanations or families of explanations come into view. One

would stress the importance of affective factors, such as students' self-confidence or academic self-concept with respect to reading, or their attitudes towards reading. The other one would focus on the possibility that some students have acquired more competence than others in becoming an independent, self-regulating reader. The latter hypothesis refers to the acquisition of metacognitive skills such as self-monitoring and the ability to chose and to apply appropriate cognitive strategies to solve problems during reading.

It is well conceivable that acting as a tutor enhances a 5th grade student's academic self-concept and self-confidence towards reading. This might more particularly be the case for the less able readers in 5th grade, who – in Van Keer (2004)'s program as well as in the present study – got exactly the same opportunities to act as tutors as their higher achieving classmates. Having the experience of being of help to a younger student surely can have a positive influence on the tutors' academic self-concept. But the already high achieving 5th graders also improved their reading comprehension achievement – Van Keer (2004) did not find any interaction effect with initial reading level.

At this point, the second hypothesis comes more to the foreground. In this hypothesis it is assumed that tutors, by monitoring and regulating their tutee's reading, by checking their tutee's comprehension and by being in the lead when choices about the most appropriate cognitive strategy have to be made, acquire metacognitive skill that they can apply to their own reading as well. Similar transfers from co-operative learning to individual task performance have been shown in other domains (e.g., Blaye *et al*., 1991; Manion & Alexander, 1997). So, in this hypothesis, a key role is attributed to the opportunities students have to engage in metacognitive activities during interactions within peer tutoring dyads. It should be noted here that the term *metacognitive activity* not only refers to self-directed metacognition, but also refers to the tutor's behavior when planning, monitoring, regulating, and evaluating the tutee's reading behavior and understanding. A key assumption in this hypothesis is that monitoring and

regulating someone else's reading and understanding is a very useful step to enhance self-monitoring and self-regulating. And that is probably the case because monitoring and regulating someone else's behavior and understanding is easier – and therefore a good field for practice – than monitoring and regulating one's own behavior and understanding. In some way, this hypothesis reflects the old saying that *teaching is the best way to learn* something. Another key element in this hypothesis is then that CA dyads would offer more opportunities for the tutor to engage in metacognitive activities than RSA dyads, and therefore would lead to more growth in metacognitive skill in tutors. When confronted with a reading task themselves, tutors would benefit from this better skill and achieve more on the reading comprehension test.

To check this latter hypothesis, we would need to examine the interaction processes within dyads, looking for indications of metacognitive activity. The underlying research question then is: Could long-term differences between CA and RSA tutors' growth in reading comprehension be explained by differences in opportunities for metacognitive activity during peer tutoring dyad interaction? Investigating whether indeed CA and RSA dyads differ in the opportunities they create for tutors to practice metacognitive skills, would only be a first step in answering that question. A second step then would be to investigate whether tutors indeed tend to show more self-monitoring and self-regulation when doing a reading task themselves. And, finally, if such differences are found, it should be investigated whether they indeed account for long-term differences in reading comprehension growth.

The present study focuses on the first step. We more particularly have looked at differences between CA and RSA dyads with respect to such tutoring behaviors as monitoring and regulating the tutee's reading behavior (Baker & Brown, 1984; Dole, 2000), and choosing reading strategies to promote comprehension, check understanding or solve problems. In addition, we also investigated differences between the two types of dyads with respect to off-task behavior. Based on the

general hypothesis on the mediating role of metacognition explained above, it was more particularly predicted that, compared to RSA dyads,

(1) Following tutee's reading mistakes or reading difficulties, significantly more corrections and help from the part of the tutor would be observed in CA dyads;

(2) Following tutee's reading mistakes or reading difficulties, significantly more corrections and help from the part of the tutor, involving the use of or referring to reading strategies would be observed in CA dyads;

(3) Including non-reading episodes as well, in the interaction between tutor and tutee more overall use of reading strategies would be observed in CA dyads; this would more particularly be true for the tutor separately, but also for the tutee;

(4) Significantly higher frequencies of spontaneous tutor questions and explanations involving the use of or referring to reading strategies would be observed in CA dyads;

(5) A significantly higher overall frequency of explaining, questioning and answering behavior would be found in CA-dyads;

(6) Significantly higher frequencies of exchanges clarifying relationships in a text or involving logical inferences from the text and exchanges clarifying how one can find a meaning or solve a problem would be observed in CA dyads;

(7) No significantly different amounts off-task behavior would be observed in CA dyads.

The first prediction follows from the hypothesis that CA tutors have more opportunities to monitor the tutee's reading. The second prediction specifies corrections in which the tutor points out how the tutee could avoid the kind of mistake he or she makes or solve the reading difficulty by looking at the part of the

Chinese character that gives a clue about its pronunciation or – in other cases – the part that gives a clue about the meaning. So this kind of corrections is strategy oriented and implies a higher level of metacognitive activity from the part of the tutor. The third prediction expands the second one to other types of interactions, including non-reading episodes. The more discussions about the text refer to or show use of reading strategies, the higher the level of metacognitive activity, so we assume, because it shows that students consciously check their (or each other's) understanding of the text or think about how they can find the meaning or solve some problem. Of course, this is only at a rather low level as long as this strategy use is purely based on a duty-full use of the assignment card. Therefore, in the fourth prediction a distinction is made between spontaneous strategy-related questions and explanation, raised by the tutor on the one hand, and questions picked from the assignment card on the other hand. In a similar way, a distinction is made between the overall frequency of explanations, questions, and answers (fifth prediction) and those exchanges between tutor and tutee involving "why"- en "how"-level exchanges (sixth prediction). Asking or explaining *why* one thinks that certain things in the story happen or why some character in the story would do or think or feel something expresses a conscious effort to understand the story or to check the tutee's understanding, and can therefore be considered as an indication of metacognitive activity. Asking or explaining *how* one can know why something happens in the story or *how* one can find out the meaning of something, clearly refers to metacognitive regulative behavior. But the overall frequency of explanations, questions and answers on itself, irrespective of its level or the specific use of one of the six reading strategies, also provides an indication of some reflection or some efforts to make the tutee reflect on what has been read.

The seventh prediction follows from the aforementioned distinction between setting characteristics and sustainable changes in the students as causes for differences in reading comprehension growth. Since the two peer tutoring

conditions do not appear to differ in short-term reading comprehension growth, no differences in situational characteristics such as off task behavior is expected.

2. Method

2.1. Intervention and design of the study

2.1.1. Design

The present study is part of a broader quasi-experimental intervention study that took place in natural classroom settings, involving 36 classes, and with 785 2nd grade students, 739 5th grade students and 36 teachers out of eight primary schools in Beijing, China. The intervention combines both reading strategy instruction by regular Chinese language teachers and practice in peer tutoring dyads. Six reading strategies were explicitly taught: (a) activating prior knowledge, (b) finding the meaning of unfamiliar words and expressions, (c) making predictions, (d) finding the main idea, (e) monitoring and regulating comprehension of the text, and (f) distinguishing different types of texts.

For the strategy instruction and practice a sandwich model was applied. Each strategy was first introduced during one lesson period. This lesson involved direct instruction of that particular reading strategy, with explanations of the declarative (main concept and function), procedural (how to use it step by step) and conditional knowledge (when and where to use it) about the strategy. To deepen students' comprehension of the "how-to-do-it" aspects, the use of each reading strategy was modeled by the teacher by thinking aloud, followed by some teacher-led class wide practice of the reading strategy on selected texts, applying the appropriate assignment card. Such an assignment card synthesizes the main steps of a particular reading strategy using general, text-independent questions as *"What are the main characters in the story?"* or *"What is told in the paragraph I just read?"* For each strategy one assignment card was designed, several copies of which were provided to every student.

After this introduction, two sessions of independent practice was scheduled. In view of this practice, students could choose from books that were circulated among the experimental groups. Independent practice was scaffolded by means of the aforementioned assignment cards. Students were asked to write down the answers on the copy of the assignment card they were actually using.

After the introduction and the individual practice sessions, further practice was organized. For two conditions this involved peer tutoring in either CA dyads or RSA dyads. After three reading strategies had been introduced and practiced, for a number of sessions students were asked to practice all three of them flexibly, with the help of integrative strategy cards. After all six reading strategies had been introduced and practiced separately, another series of integration sessions was organized.

Before the start of the peer tutoring sessions a series of six lessons was organized in which the students who would take the role of tutor were trained in *"how to become a good tutor"*. These lessons discussed topics such as *how to make my tutee feel at ease*, *how to give positive feedback to my tutee*, *how to react on mistakes by my tutee*, *how to cope with own mistakes,* as well as aspects related to the management of a tutoring session, such as starting and ending a session, keeping record of the texts that were read and the assignment cards used, *etc*. These lessons were taught using explanation, modelling, role-play and discussion.

Students in the experimental conditions were provided with reading materials such as children's reading books and magazines. Administration cards were also attached to assignment book to keep track of reading materials and reading strategies being covered. Vivid cartoon posters highlighting the main content of the reading strategies and the interaction skills discussed in the tutor training were displayed on the classroom walls.

The intervention was scheduled during after-school extra curricular activities, for an average of 60 minutes per week and lasted for a whole school year (25 weeks).

The lessons were taught by the students' regular Chinese language teachers. Those teachers were trained intensively for three working days two weeks before the start of the program and were provided with a detailed written manual on how to teach reading strategies, how to prepare and coach the tutors, and how to coach students with their strategic reading. During the whole course of the intervention, the author went to visit the participating schools alternatively every week, whenever the introductory lessons or the peer tutoring sessions were organized. All of the teachers involved participated in a short discussion of half an hour that was organized prior to every introductory lesson, to highlight the main points of that day's teaching. Special attention was given to the potential problems that might arise during the tutor training or the strategy instruction. During these weekly meetings teachers were encouraged to share useful experiences, difficulties or puzzles. The fidelity of the treatment was evaluated through two kinds of means: direct observation and teachers' self-report on checklists of main procedures related to the strategy teaching. A pre-designed outline of each strategy-teaching lesson was distributed to teachers at the beginning of the intervention, also serving as a checklist for themselves. Teachers were asked to evaluate their teaching using the checklist and hand it in right after the end of the lesson on that day. Observations were made by the coordinator and her two research assistants. Feedback was given to the teachers subsequently, based on both the observations and the checklists.

2.1.2. Peer tutoring sessions

Dyads in the CA condition were formed by matching their relative reading comprehension level, e.g., a good reader from the 5th grade forming a dyad with a good reader from the 2nd grade. RSA dyads were formed with two classmates with similar reading comprehension levels.

Both CA and RSA peer tutoring sessions, typically involved the following "episodes": briefing by the teacher, getting ready (such as choosing reading books,

getting the assignment book, *etc*.), reading, answering questions and filling in an assignment card, filling in the administration card, having a little chat after reading; and, at the end a short "debriefing" by the teacher in which a few general remarks were made about the peer tutoring activities of the day. The middle episodes were not necessarily run through in a linear way. The tutee's reading could be interrupted by the tutor for some discussion based (or not) on the assignment card, e.g. to check the tutee's understanding of the text, after which the reading would continue. It was up to the tutor to decide whether or not, when, how and why to do so. It was also possible that students engaged in two (or sometimes three) *"reading cycles"* within one session, i.e. after a text was read and the assignment card was "done", they could start with another text and another copy of the assignment card. In that case some episodes, such as *getting ready* or *chat after reading* might occur more than once within one session. As a consequence, reading cycles could differ in duration.

2.2. Sample

A sub-sample of 28 dyads (11 from CA dyads and 17 from RSA dyads) was selected, based on the students' pretest reading comprehension results. To control for the effect of initial reading level, both high and low achieving dyads were selected. It should be noted however, that not dyads but *reading cycles* (or *"observations"*) *within peer tutoring sessions* were taken as the sampling unit. For the present study, 40 videotaped *reading cycles (observations)* were analyzed. They lasted on average 26 minutes (SD = 7.7, min = 12 min. 15 sec, max = 44 min. 20 sec.). An overview of the sample of reading cycles, divided according to students' achievement levels and experimental conditions is presented in Table 1.

Table 1. Overview of analyzed reading cycles (observations)

Conditions / Reading level	Same-age G2	Same-age G 5	Cross-age	Total
High	5	8	8	21
Low	5	6	8	19
Total	10	14	16	40

2.3. Video coding method and scheme

The digital videotapes were transferred into mpeg-files and coded with the interface of a software package called *Catmovie 4* (Wild, 2001). Catmovie allows dichotomous, multiple-response codling and could store the coded results automatically in an SPSS data file.

A coding scheme was developed, taking into account the following steps and principles. First, each video recorded peer tutoring session, was segmented in *"reading cycles"* and each cycle was segmented in *"episodes"* according to the main focus of what was going on in the dyad: listening to the teacher's briefing, getting ready, reading, filling the assignment card, filling in the administration card, having a chat after reading, listening to the teacher's debriefing. Second, a time-based coding was chosen, with five-second intervals as the basic unit. Third, multiple coding was applied. For every interval the type of episode was coded and more specific codes were used to describe who was doing what, in which way or at what level. In this way a hierarchical category system was developed, relating categories describing tutors and tutees' behaviors to the "episodes" in which it took place. Initially the coding system was developed based on a number of theoretical assumptions, but this scheme was gradually expanded and enriched by including

categories for new emerging behaviors as the coding progressed. The goal leading this development was to find indications of monitoring or regulating behaviors and other strategic behaviors within the dyads. This resulted in a coding scheme that can be described briefly as follows:

For reading episodes, a set of categories indicates whether a tutee reads or makes mistakes in reading, misinterprets the meaning of a word or sentence, hesitates because of some difficulties, corrects himself/herself, or asks the tutor for explanation. Another set of categories indicates whether the tutors monitors or listens to the tutee's reading, corrects the tutee's mistakes or asks questions to check the tutee's understanding. Furthermore, off-task behaviors, such as being absentminded, or chatting with other dyads about unrelated topics were coded.

Special attention was given to the analysis of the discourse in dyads involving questioning and answering or giving explanations. In this analysis we focused on whether reading strategies were being used or referred to. With respect to the use of reading strategies, a distinction was made between picking questions from the assignment card versus asking questions spontaneously to monitor tutees' understanding. The latter behavior was considered to reflect a higher level of metacognitive skill than the former.

To complement the quantitative analysis based on the coding, excerpts of the dialogue protocols (in appendix) were selected to demonstrate students' use of the six reading strategies and other variables captured in the coding scheme.

In order to capture other indications of metacognitive processes that would not be reflected by the use of any of the six strategies, further distinctions were made based on the content of the exchanges between tutor and tutee. King (1991) listed four types of explanations: (a) clarification of a concept or analysis of a situation, (b) elaborated description of how to do something, (c) a rationale for using particular strategy and (d) reasons for why the attempt is (un)successful. In our coding scheme we only distinguish between three types or levels, but we apply

them to explaining, questioning and answering, by tutor or tutee: "what", "how", and "why". "What" questions and explanations are comparable to the King's type (a) clarification. They are on the information demand and providence level. When having to answer a "what"-question, the tutee's task is to retrieve information from the text or from previous knowledge. "Why"-level explanations and questions reflect or ask for some reasoning and logical thinking, relating different elements of information in the text. The "how" level is comparative to King's type (b). King's types (c) en (d) are not reflected as such in our coding scheme. They refer to further metacognitive elaborations on the use of reading strategies. If any such an explanation or questions would occur in the observations we made, it would have been coded as a "how"-level question or explanation.

2.4. Reliability of the coding

Two Chinese master students majoring in educational sciences and blind about the research design and research hypotheses were hired to code the observations, after an intensive training on the coding scheme. Coding agreement of above .85 was reached before the actual coding started. During the coding they were closely supervised by the first author. The database resulting from the coding was checked independently by the second author on consistency among the variables, for example: (a) for any record there should be indications of type of episode and specific activities from both tutor and tutee, (2) whenever a question occurs, it should be indicated who raised the question, what type of question it was – spontaneous or a question from an assignment card – and whether the question involved any use of reading strategies; (3) whenever an interval was coded as *answer*, a question a few records backwards was expected; it should also be coded who gave the answer, whether it was correct, and whether the answer was related to any of the reading strategies. Whenever inconsistencies were found or suspected, that part of the coding was sent back to the first author for further check with the videotape and eventual correction. In such a case also the neighboring two minutes of coding were checked with the videotape.

3. Statistical Analysis

3.1. Structure of the database

Initially, the coding resulted in one data file per coded *reading cycle*. In those data files 5-second intervals constitute the basic coding unit. For each coding unit (interval of 5 seconds) there was one record containing a number of dichotomous or polytomous codes, according to the variables in the coding scheme. For a reading cycle of 25 minutes such a data file would contain 25 x 12 = 300 records. Merging the 40 separate data files yielded a dataset of thousands of records. Before any statistical analyses were done, a number of data management steps were taken.

Firstly, only three relevant episodes, *"reading", "doing the assignment card"* and *"chat after reading"* were selected for further analysis. For our purposes the most relevant interactions took place in the two former categories of episodes. However, some spontaneous reading content related conversation was also noticed in *"chat-after-reading"* episodes, which were therefore included in the analysis.

Secondly, to reduce the database into a more manageable size, records were aggregated into larger sequences of one minute. Since 5-second intervals constituted the original coding unit, this aggregation turned the original dichotomous or polytomous codes into continuous variables ranging from 0 to 12, indicating how frequently (= times per minute) a specific behavior was observed. An immediate advantage of this aggregation is that results of further analyses would be easy to interpret, more particularly as frequencies, expressed in terms of "times per minute". A very important advantage was that much of the sequences of consecutive behaviors (e.g. tutor – tutee interactions) were now captured in one record. This would facilitate the analysis of how tutor and tutee behaviors co-vary. Because of the above-mentioned selection of target episodes, and since the start of a new episode not always coincided with the start of another full minute, incomplete sequences occurred in the database, i.e. sequences constituting less than

12 intervals. To adjust for such unequal numbers of intervals, the number of intervals in a sequence was used to weigh the sequences.

A last step in preparing the database consisted of creating some broader categories or second order variables, e.g. the total frequency of reading strategy use (summing across the six strategies and across questioning, answering, explaining), the frequency of "what", "why" and "how" level exchange of information across instances of questioning, answering and explaining, *etc*. At this phase the dependent variables were created (see the *results* section).

3.2. Multilevel analysis

To analyze the dataset, 2-level analyses with *one minute sequences* (level 1) nested within *reading cycles* (level 2) were performed using MLwin 2.0. Because of the low videotaped sessions / dyad ratio (40 / 28), we didn't include dyad as a third level. This implies that characteristics that principally are dyad characteristics (such as being a CA or a RSA dyad, tutor's gender, *etc*.) are entered at level 2. Using a multilevel structure, differences in duration of the reading cycles are automatically accounted for.

A series of models were built with different dependent variables, according to our hypotheses. For each dependent variable the tutor and the tutee's gender, and their prior reading level were first introduced as explanatory variables at the *reading cycle* level (level 2). These are predictors that do not vary across sequences within a given reading cycle. Then, relevant behavioral variables at the sequence level, e.g. number of reading mistakes or difficulties, number of wrong or incomplete answers, *etc*. were entered in the model to adjust for the effect of tutee behavior on the tutor (and vice versa) in subsequent intervals. Type of episode was also entered at this level. As one can easily see, these are variables that can vary across the one-minute-sequences within a given reading cycle. Overviews of the counts resp. overall average frequencies and standard deviations for the predictor variables that proved to be significant are shown in Tables 2 and 3.

Table 2. Distribution of tutor gender across conditions

	Type of peer tutoring dyads			Total
Tutor gender	Same-age grade 2	Same-age grade 5	Cross-age	
Girl	1	9	8	18
Boy	9	5	8	22
Total	10	14	16	40

Table 3. Means (average frequencies) and standard deviations of predictors

Predictor	Description	M	SD	Max.
Assign. Card	Proportion of minutes within sequence spent on doing the assignment card	0.25	0.40	1.0
MIS&DIF	# reading mistakes or difficulties by tutee per minute	0.64	1.21	10.0
RS-tutee	# intervals/minute with reading strategy use by tutee	0.70	1.39	10.0
RS-tutor	# intervals/minute with reading strategy use by tutor	0.88	1.64	11.0
QRS-tutor	# tutor questions with reading strategy use per minute	0.50	1.06	11.0
Off-task tutee	# intervals/min. in which tutee was off-task	0.88	2.00	12.0
Off-task tutor	# intervals/min. in which tutor was off-task	0.60	1.69	12.0

N (sequences) = 1015. Sequences are weighed by n of intervals / sequence. Minimum = 0 for all variables.

In Table 3 it can be seen that (after having excluded time spent by teachers on briefing and debriefing and time spent by students in getting ready or pure administrative tasks) one quarter of the time was spent on episodes coded as *"doing the assignment card"*. With only a very little amount of time spent on *"chat after reading"*, almost three quarters of the time was spent on episodes coded as *"reading"*. Within the reading episodes, however, not all time is spent on reading as such. The actual reading is interrupted by corrections, or questions to

check tutees' understanding, *etc*. For all other variables it can be seen that averaging for the whole reading cycle and across reading cycles, less than 5 seconds (= 1 interval) per minute was spent on the behavior described by the variable. But as the figures for the observed maximum score indicate, there were instances in which the behavior could take almost a whole minute (50 or 55 seconds, or 10 to 11 intervals). Since the coding was time-based (using 5-second intervals as the original basis) no distinction can be made between "incidence" and "duration". So, a score of 10 could indicate that the behavior occurred 10 times during that sequence, or just once, but lasting for 50 seconds. For behaviors such as "off task" an interpretation in terms of duration is more appropriate; for other behaviors, such as "reading mistake" it is clear from the videotapes that an interpretation in terms of incidence is more appropriate. So, the mean of 0.64 for *MIS&DIF* can be read as: reading mistakes and difficulties with reading on the average occurred less than once per minute (or about 3 times every 5 minutes).

To facilitate the interpretation of the coefficients for the other variables, all continuous predictors were centered around their overall mean.

Finally, dummy variables for the experimental conditions were entered in the multilevel model to test the aforementioned hypotheses. It should be noted that the experimental conditions in this study are coded in a hierarchical way. The main distinction we make here is between CA and RSA dyads. This is done by entering a dummy variable CA, using RSA as the reference category. Having entered this dummy in the model, the beta-coefficient in the tables will indicate how much higher or lower the frequency for the dependent variable on the average is in CA dyads, compared to the average frequency in RSA dyads (having controlled for the effects of previously entered predictors). Therefore, in the tables this dummy will be referred to as "CA dyad". Within the reference category (RSA dyads) a further distinction is made between 2nd grade RSA dyads and 5th grade RSA dyads. Here, the 5th grade RSA dyad is taken as the reference category. Having already entered the CA-dummy in the model, the beta-coefficient for this second dummy will show

how much the 2nd grade RSA dyads on the average differ from the 5th grade RSA dyads (having controlled for previously entered effects). Therefore this second dummy is referred to as "2nd grade RSA dyad". When both dummies are entered in the model, the coefficient for the CA-dummy shows the difference between CA-dyads and 5th grade RSA dyads. However, for reasons of parsimony, the RSA-dummy is only kept in the model if proven to be significant. The CA dummy is always kept in the model, even if not significant, because otherwise the coefficient for the RSA-2 dummy becomes less well interpretable.

For multilevel analysis, several ways of calculating effect size have been suggested (Snijders & Bosker, 1999). In the present study, for dummy coded predictors the beta-coefficient is divided by the standard deviation at the appropriate level. For the CA- and RSA-effects, this appropriate level would be level 2. The SD at that level was calculated by taking the square root of the variance for level 2 in the null-model. Effect sizes are only reported in the case of significant effects.

Apart from fixed effects, random effects are explored as well. Significant higher-level random effects indicate that the effect of a certain predictor varies across higher-level categories. In our case these higher-level categories are reading cycles. For example, it is possible that the relationship between the frequency of mistakes made or difficulties experienced while reading on the one hand and the dependent variable on the other hand varies across reading cycles (and thus varies between dyads). Significant lower level random effects indicate complex variances (i.e. non-homogeneity of the dependent variable variance across levels of an independent variable). This is for example the case when sequences characterized by high frequencies of reading mistakes show more variation in the dependent variable than sequences characterized by low frequencies of reading mistakes. Since it does not make sense to allow level 2 predictors to vary randomly at level 1, for such predictors significant random effects at level 2 also indicate complex variance. Examples will be discussed in the results section.

4. Results

Results are reported in five sections. First, we focus on monitoring behaviors, corrections and help provided by tutors when tutees made reading mistakes or experienced reading difficulties. Both corrections and help with and without use of reading strategies were analyzed. The second section presents results about tutors and tutees' overall use of reading strategies, irrespective of the context within which reading strategies were used, e.g., during questioning, answering or explaining, *etc*. In the third section we focus on spontaneous use of reading strategies by the tutor, either when asking questions or giving explanations. In the fourth section we discuss the results for three different levels that were distinguished in questioning, answering or explaining and that we called the "what"-, "why"- and "how"-levels. Finally, we checked for differences in off-task behaviors between CA and RSA conditions. Overall means and standard deviations for the dependent variables are shown in Table 4 (on the next page). The interpretation of the figures is similar as for Table 3.

It should be noted that for all multilevel results presented here, a number of general interpretation rules apply. The number of cases reported does not indicate the number of students, but the total numbers of sequences (or minutes) analyzed. Since the sequences are weighted to adjust for unequal numbers of intervals within sequences, the coefficient reported with the constant ("cons") in the null-model gives the overall average number of intervals the dependent variable was observed per minute. As explained above, in some cases this should be interpreted as duration, but for most dependent variables it can be interpreted as a frequency, i.e. as the number of times this behavior on the average occurs per minute.

When dichotomous predictors are entered the model, the constant reflects the average frequency or duration the dependent variable is observed per minute in dyads belonging to the reference category.

Table 4. Means and standard deviations for the dependent variables

Depend. var.	Description	M	SD	Max.
COR-RS	# corrections / minute made by tutor, involving use of or referring to reading strategies	0.03	0.21	3.0
COR-NRS	# corrections / minute made by tutor, not involving any use or any reference to reading strategies	0.41	1.01	8.0
RS-tutor	# intervals/minute with reading strategy use by tutor	0.88	1.64	11.0
RS-tutee	# intervals/minute with reading strategy use by tutee	0.70	1.38	10.0
Spon-QRS	# intervals/minute in which tutor asks spontaneous questions to tutee, involving the use of or referring to reading strategies	0.43	1.01	11.0
Expl_RS	# intervals/minute in which tutor gives explanation to tutee, involving the use of or referring to reading strategies	0.38	1.08	10.0
WHAT	# intervals/minute spent on WHAT-level questions or explanations within the dyad	1.28	2.08	11.0
WHY	# intervals/minute spent on WHY-level questions or explanations within the dyad	0.14	0.72	11.0
HOW	# intervals/minute spent on HOW-level questions or explanations within the dyad	0.06	0.38	6.0
OFFTASK	# intervals/minute in which dyad was off-task	0.90	1.89	12.0
Offtask-tutor	# intervals/minute in which tutor was off-task	0.59	1.69	12.0
Offtask-tutee	# intervals/minute in which tutee was off-task	0.88	2.00	12.0

N (sequences) = 1015.
Sequences are weighed by n of intervals per sequence. Minimum = 0 for all variables.

4.1. Tutors' correction/help following reading mistakes and difficulties

A total of 425 instances (intervals) were observed in which tutors gave correction of help when tutees had reading mistakes or difficulties. When these instances are related to the instances in which reading mistakes were made or in which some hesitation indicating a reading difficulty was observed, then it is found that on the average 69.6% of the mistakes or difficulties are followed by a correction.[1] Only

[1] This interpretation is based on the assumption that in general a reading mistake or difficulty wouldn't take longer than five seconds, and that within each five seconds interval no more than one mistake or difficulty is observed. Each interval coded as *reading mistake* or *difficulty*, can be counted as one mistake or difficulty.

7% of the corrections involved the use of reading strategies. Therefore, on the average only 4.9 % of the reading mistakes or difficulties were followed by some correction that incorporated the use of a reading strategy. (These figures can also be derived from Tables 3 and 4.)

How these general figures are affected by the type of peer tutoring and other predictor variables is shown in Tables 5 and 6, which presents the results of the multilevel analyses. Not surprisingly the frequency of corrections (with or without reading strategies use) appears to be influenced by the frequency of reading mistakes and difficulties. The more mistakes or difficulties were observed, the more corrections were provided. This is true for both types of corrections.

Table 5. Tutor corrections using reading strategies in response to tutee's reading mistakes and difficulties. Results of multilevel analyses

	Model 0	Model 1	Model 2	Model 3
Fixed effects				
Cons (Intercept)	0.030	0.029	0.013	0.020
SE	*0.009*	*0.009*	*0.004*	*0.009*
Mis&Diff.		0.034	0.032	0.035
SE		*0.012*	*0.012*	*0.011*
CA dyad			0.041	0.025
SE			*0.018*	*0.011*
Random effects				
Reading cycle				
Cons / Cons	0.002	0.002	0.001	0.001
SE	*0.001*	*0.001*	*0.001*	*0.001*
Cons / Mis&Diff.				0.002
SE				*0.001*
Mis&Diff. / Mis&Diff.				0.002
SE				*0.001*
Sequence				
Cons / Cons	0.476	0.459	0.459	0.432
SE	*0.191*	*0.186*	*0.187*	*0.020*
-2log(lh)	1202.39	1166.62	1165.25	1110.24
N (sequences)	1015	1015	1015	1015

Weighted by N_intervals in Sequence
Notes: MIS&DIFF is reading mistakes or difficulties by tutee.

A significant CA-effect was observed and this too was the case for both corrections with or without reference to reading strategies. Having controlled for the frequency of reading mistakes and difficulties, the frequency with which CA-tutors provided corrective feedback involving reading strategies was on the average 0.041 per minute higher than in RSA dyads (see model 2 in Table 5). So, relative to the frequency of reading mistakes or difficulties, four times more strategy based corrective feedback was given in CA dyads (frequency = 0.0533 corrections / minute) than in RSA dyads (frequency = 0.0125 corrections / minute). The effect size is very big (ES = 0.99).

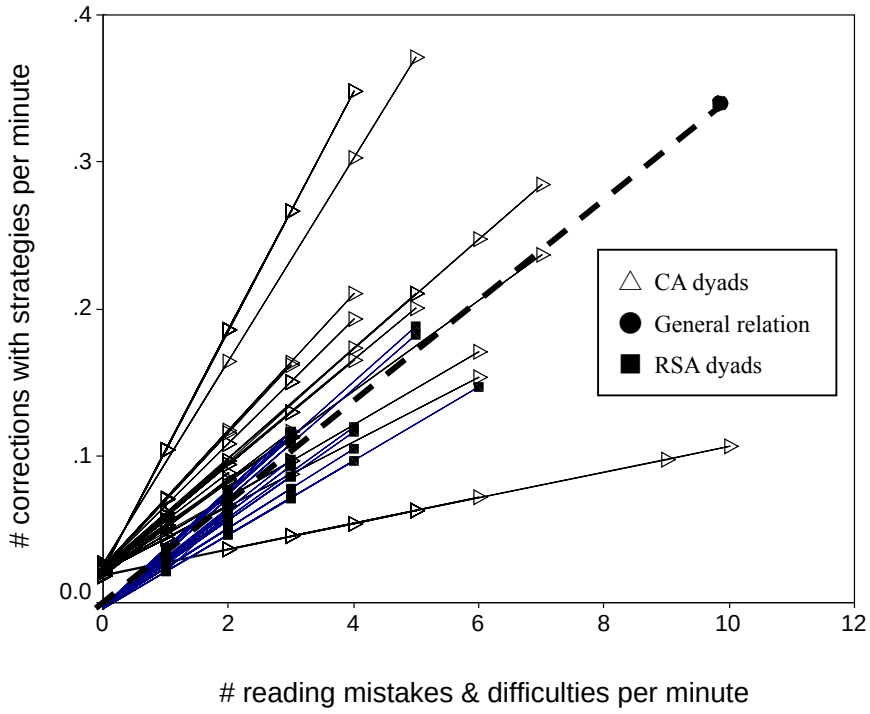


Fig. 1. Relationship between frequency of reading mistakes or difficulties and frequency of corrections with strategies: variation among observation cycles (and dyads) and between CA and RSA dyads

However, when allowing for a random effect at level 2 for the frequency of reading mistakes and difficulties (see Model 3 in Table 5), it becomes clear that the relationship between reading mistakes or difficulties and strategy-based feedback varies quite a lot across reading cycles and dyads. For this random effect, both the linear and the quadratic parameter are significant. The positive linear parameter indicates that the variance in the frequency of corrections becomes larger as the frequency of reading mistakes or difficulties is higher. Obviously, differences in the way the tutor copes with the tutee's reading mistakes and difficulties become more visible as the tutee is a less skilled reader and makes more mistakes. Modeling this random effect caused the fixed effect for CA to become weaker but still significant ($\beta = 0.025$, SE = 0.0113, $p < .05$). The effect size also went down, but is still rather big (ES = 0.61). A representation of how the relationship between frequency of reading mistakes or difficulties and tutor corrections using reading strategies varies across reading cycles and dyads, is shown in Fig. 1. One can see that for most observations made in CA dyads, the slope of the regression line is much steeper than in observations in RSA dyads. This reflects the finding that the relationship between number of reading mistakes and strategy-based feedback is stronger for observations made in CA dyads. There is one notable exception, the much flatter line at the bottom of the picture, representing an observation in a CA dyad. Dummying this case out revealed no significant difference between this one reading cycle and the other ones, neither with respect to the intercept, nor with respect to the slope.

A moderate (ES = 0.45), but still significant difference between observations in CA and RSA dyads was found with respect to tutors' corrections *not* involving any use of nor reference to reading strategies (see Table 6, Model 2). For CA dyads the frequency of non-strategy based corrections is 0.301 corrections per minute higher than for RSA dyads. That is about twice as high as the average frequency found in RSA dyads (0.289 + 0.301 vs. 0.289 corrections / minute).

Again, when allowing for level 2 random effects for the frequency of reading mistakes and difficulties, the CA dyad effect dropped (see Model 3 in Table 6). The CA effect becomes now only marginally significant ($\beta = 0.0698$, SE = 0.0360, $p = 0.053$) and small (ES = 0.10). Figure 2 shows a representation for the random effect for frequency of reading mistakes and difficulties. Comparison with Fig. 1 reveals how the differences between observations in CA and in RSA dyads are much smaller when considering tutors' corrections not reflecting any use of reading strategy.

Table 6. Tutor corrections without reading strategy in response to tutee's reading mistakes and difficulties. Results of multilevel analysis

	Model 0	Model 1	Model 2	Model 3
Fixed effects				
Cons (Intercept)	0.449	0.409	0.289	0.314
SE	*0.110*	*0.047*	*0.049*	*0.038*
MIS&DIFF		0.613	0.612	0.487
SE		*0.057*	*0.057*	*0.060*
CA dyad			0.301	0.070
SE			*0.099*	*0.036*
Random effects				
Reading cycle				
Cons / Cons	0.4561	0.089	0.067	0.050
SE	*0.274*	*0.034*	*0.025*	*0.009*
Cons / Mis&Diff.				0.079
SE				*0.013*
Mis&Diff. / Mis&Diff.				0.121
SE				*0.021*
Sequence				
Cons / Cons	7.260	3.030	3.030	1.953
SE	*1.589*	*0.537*	*0.537*	*0.505*
-2log(lh)	4022.06	3111.61	3105.39	2690.20
N (sequences)	1015	1015	1015	1015

Weighted by N_intervals in Sequence.
Notes: READING%, proportion of intervals within sequence that are part of a reading episode; MIS&DIFF is reading mistakes or difficulties by tutees.

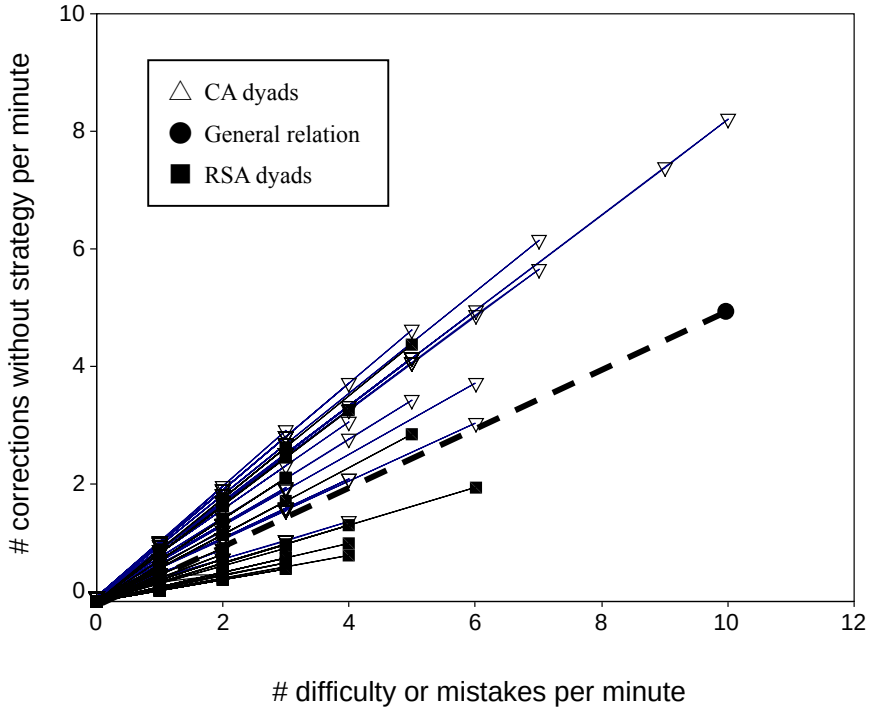


Fig. 2. Relationship between frequency of reading mistakes or difficulties and frequency of corrections without strategies: variation among observation cycles (and dyads) and between CA and RSA dyads

4.2. Use of reading strategies

4.2.1. Overview of reading strategy use

Students' use of reading strategies was coded whenever the main questions from the assignment card were picked and asked or whenever the main content of any reading strategy was reflected in a student' explanation or spontaneous question. For example, a question such as *"Who do you think is the main character in the story?"* and its answer or explanation was coded as an indicator for the use of reading strategy (d) *"finding the main idea of the text"*. A question such as *"What do you think will come next?"* and any sensible answer to it were coded as using reading strategy (c) *"making predictions"*. Later on, a more general variable was created by summing frequencies across all reading strategies and types of

exchanges in which they were observed (explanations, questions, and answers). Table 8 reports the multilevel analysis results for strategy use by tutors, and Table 9 reports similar results for strategy use by tutees and the total strategy use within the dyads. First, we present an overview of the different types of interactions in which any use of reading strategies was found (Table 7). This overview shows that, across all dyads, in about 14% of the interaction time within the dyads (i.e. when the tutee was not actually reading) reading strategy use was observed. For almost 70% of that time, these interactions reflected a typical questioning-answering pattern, sometimes elaborated with somewhat longer explanations given by the tutor. Tutee questions accounted for less than 6 % of the interaction time in which reading strategy use was observed.

Table 7. Overview of use of reading strategies during episodes of Reading, Doing Assignment Card and Chat-After-Reading

Intervals coded	With RS		Without RS		Total N of intervals	
	Counts	Percent	Counts	Percent	Count	Percent
Tutor explanation	368	24.39	39	0.39	407	3.52
Tutor question	493	31.34	13	0.13	506	4.38
Tutee answer	593	37.70	39	0.39	632	5.47
Tutee question	91	5.79	7	0.07	98	0.85
Dyads discussion	13	0.83	79	0.79	92	0.80
Non-verbal (others)	15	0.95	9812	98.23	9827	84.99
Total num. of intervals	1573	100.00	9989	100.00%	11562	100.00
Percent	14%		86%		100%	

4.2.2. Tutor's overall use of reading strategies

In CA dyads the use of reading strategies was observed more than 1.3 times per minute on the average. That is almost 2.5 times higher than in RSA dyads, for which a frequency of 0.554 times per minute was observed (Table 8). The effect size (ES) equals 1.02, which is very large. Within the RSA-condition the difference

between 2nd and 5th grade is not significant. There is a quite small (ES = 0.12) but significant positive effect for the strategy use by tutees. Our analysis does not allow making any causal inference with regard to this relationship. It is possible that tutor questions or explanations referring to some reading strategy caused the tutee to refer to a reading strategy when answering, but it is also possible that some of the tutor explanations referring to the use of a reading strategy followed a tutee question related to that strategy.

Table 8. Tutors' total use of reading strategies. Results of multilevel analysis

	Model 0	Model 1	Model 2
Fixed effects			
Cons (Intercept)	0.868	0.842	0.554
SE	*0.126*	*0.122*	*0.080*
MIS&DIFF		-0.134	-0.150
SE		*0.029*	*0.033*
Reading strategy use by tutee		0.401	0.416
SE		*0.060*	*0.060*
CA dyad			0.752
SE			*0.230*
Random effects			
Reading cycle			
Cons / Cons	0.545	0.512	0.319
SE	*0.148*	*0.181*	*0.125*
Cons / RS_tutee		0.134	0.074
SE		*0.050*	*0.048*
RS_tutee / RS_tutee		0.076	0.074
SE		*0.019*	*0.018*
Sequence			
Cons / Cons	23.386	18.250	18.240
SE	*4.846*	*4.064*	*4.059*
-2log(lh)	5184.95	4945.72	4936.05
N (sequences)	1015	1015	1015

Weighted by N_intervals in Sequence.

Notes:MIS&DIFF is reading mistakes or difficulties by tutees.

The frequency of the tutee's mistakes and difficulties during reading is negatively related with total use of reading. One might think that is because reading mistakes

and difficulties can only occur during reading episodes whereas the use of reading strategies is more likely to occur when doing the assignment card. But entering the dummies for type of episode didn't yield any significant effect.

Another possible explanation relates to the tutee's reading fluency. With less fluently reading tutees probably a lot of the time and energy is absorbed by correcting reading mistakes, leaving less time and attention for checking and discussing understanding and for the use of reading strategies going with that. As a result, in the case of a higher frequency of reading mistakes, less use of reading strategies would be observed. Anyhow, the relationship with the frequency of the tutee's reading mistakes and difficulties is very weak (ES = 0.04).

4.2.3. Tutees' use of reading strategies

No significant CA-effect was shown with regard to tutees' use of reading strategies (Table 9). Again, as for the tutor's use of reading strategies, there is a significant but very weak (ES = 0.02) negative relationship with the frequency of reading mistakes and difficulties by the tutee. But now, complex variance was detected. The more reading mistakes are made, the less variation is found with regard to the tutee's use of reading strategies. This gives somewhat more credit to the aforementioned hypothesis about the role of reading fluency.

When the tutor is a boy, tutees tend to use significantly less reading strategies (ES = 0.51). As could be expected, there is a significantly positive effect of the tutors' use of reading strategies. This effect is also rather small (ES = 0.18).

Table 9. Tutees' and total use of reading strategies during peer tutoring. Results of multilevel analysis

	Tutees' RS use		Total RS use in dyad (tutor & tutee)		
	Model 0	**Model1**	**Model 0**	**Model 1**	**Model 2**
Fixed effects					
Cons (Intercept)	0.731	0.958	1.482	1.907	1.450
SE	*0.088*	*0.112*	*0.170*	*0.275*	*0.263*
Tutor boy		-0.250		-0.773	-0.685
SE		*0.104*		*0.326*	*0.290*
MIS&DIFF		-0.073		-0.356	-0.362
SE		*0.019*		*0.062*	*0.063*
Reading strategy use by tutor		0.481			
SE		*0.068*			
CA dyad		-0.064			1.015
SE		*0.096*			*0.314*
Random effects					
Reading cycle					
Cons / Cons	0.236	0.114	0.972	0.846	0.614
SE	*0.068*	*0.031*	*0.227*	*0.217*	*0.166*
Cons / RS_tutor		0.090			
SE		*0.028*			
RS_tutor / RS_tutor		0.113			
SE		*0.033*			
Sequence					
Cons / Cons	19.172	13.300	48.138	43.483	43.546
SE	*3.084*	*2.441*		*6.131*	*6.124*
Cons / MIS&DIFF		-2.705		-7.160	-6.987
SE		*0.950*		*2.001*	*1.946*
MIS&DIFF / MIS&DIFF		0.730		1.715	1.623
SE		*0.305*		*0.620*	*0.598*
-2log(lh)	4968.77	4627.39	5913.000	5829.12	5822.48
N (sequences)	1015	1015	1015	1015	1015

Weighted by N_intervals in Sequence.

Notes: MIS&DIFF is reading mistakes or difficulties by tutee.

4.2.4. Overall use of reading strategies in the dyads

With respect to the overall use of reading strategies (by either tutor or tutee) within the dyads, Table 9 shows a significant and strong CA-effect (ES = 1.03). Again,

more strategy use was observed when the tutor was a girl (ES = 0.70). Also the weak but significant negative relationship with the frequency of the tutee's reading mistakes and difficulties was found again (ES = 0.06).

4.3. Tutor's spontaneous use of reading strategies

So far no distinction was made between the use of reading strategies by going through the assignment card and the spontaneous use of reading strategies when asking questions or giving explanations. Spontaneous use of reading strategies was observed when a tutor tried to help a tutee to find the answer to a question they were discussing and, while doing so, elaborated or phrased own questions without the help of the assignment card.

Table 10. Spontaneous strategy use in the tutor's questioning. Results of multilevel analysis

Fixed effects	**Model 0**	**Model 1**	**Model 2**
Cons (Intercept)	0.421	0.432	0.280
SE	*0.079*	*0.086*	*0.095*
MIS&DIFF		-0.091	-0.101
SE		*0.037*	*0.037*
CA dyad			0.370
SE			*0.141*
Random effects			
Reading cycle			
Cons / Cons	0.214	0.264	0.197
SE	*0.055*	*0.066*	*0.051*
Cons / MIS&DIFF		-0.058	-0.041
SE		*0.022*	*0.019*
MIS&DIFF / MIS&DIFF		0.026	0.025
SE		*0.012*	*0.012*
Sequence			
Cons / Cons	8.785	6.565	6.573
SE	*0.398*	*0.369*	*0.371*
Cons / MIS&DIFF		-1.910	-1.913
SE		*0.239*	*0.240*
MIS&DIFF / MIS&DIFF		1.171	1.179
SE		*0.225*	*0.228*
-2log(lh)	4190.92	4083.12	4079.04
N (sequences)	1015	1015	1015

Weighted by N_intervals in Sequence
Notes: MIS&DIFF is reading mistakes or difficulties by tutees; l; Off-task are total off-task behaviors by dyads.

It was also observed when the tutor elaborated on some helpful clues in the text or in the Chinese character in order to explain to the tutee how to avoid some reading mistakes or solve some reading difficulties. The spontaneous use of reading strategies is to be considered as a higher and more desirable level of practicing reading strategies than merely going through the assignment cards.

The spontaneous use of reading strategies by tutor while asking questions (Table 10) was observed 2.3 times more in CA dyads than in RSA dyads. This difference is significant and quite big (ES = 0.80). No significant differences were found between grade 2 and grade 5 RSA dyads. A significant, but weak negative relationship was found with the frequency of the tutee's reading mistakes and difficulties, which showed significant random effects at both the sequence and the reading cycle level.

Table 11. Spontaneous strategy use in the tutor's explanations. Results of multilevel analysis

	Model 0	Model 1	Model 2
Fixed effects			
Cons (Intercept)	0.362	0.364	0.182
SE	*0.073*	*0.072*	*0.062*
MIS&DIFF		-0.066	-0.076
SE		*0.022*	*0.021*
Tutor questions involving reading strategies		0.114	0.102
SE		*0.046*	*0.044*
CA dyad			0.454
SE			*0.136*
Random effects			
Reading cycle			
Cons / Cons	0.171	0.165	0.116
SE	*0.051*	*0.046*	*0.034*
Sequence			
Cons / Cons	11.132	10.927	10.930
SE	*3.062*	*2.994*	*2.994*
-2log(lh)	4421.80	4402.36	4396.95
N (sequences)	1015	1015	1015

Weighted by N_intervals in Sequence MIS&DIFF is reading mistakes or difficulties by tutees; QRS_Tutor is strategic questions asked by tutors.

Table 11 shows that CA tutors also gave significantly more explanations involving the use of strategies. The effect size is very high (ES = 1.10). The frequency of explanations with use of reading strategies is 3.5 times higher in CA dyads than in RSA dyads. It is also higher for tutors for whom more frequent use of questions involving reading strategy use was observed.

4.4. Overall frequency of questions, answers and explanations at the three levels

Table 12. Multilevel analysis of within dyad interactions at the "what"-level

	Model 0	Model 1	Model 2
Cons (Intercept)	1.308	1.728	1.334
SE	*0.148*	*0.238*	*0.243*
MIS&DIFF		-0.292	-0.298
SE		*0.051*	*0.051*
Tutor boy		-0.758	-0.682
SE		*0.281*	*0.254*
CA dyad			0.868
SE			*0.268*
Random effects			
Reading cycle			
Cons / Cons	0.716	0.631	0.441
SE	*0.156*	*0.166*	*0.134*
Sequence			
Cons / Cons	40.602	37.362	37.387
SE	*5.932*	*5.306*	*5.301*
Cons / Cons		-5.662	-5.558
SE		*1.693*	*1.634*
Cons / Cons		1.134	1.087
SE		*0.493*	*0.468*
-2log(lh)	5737.23	5658.04	5651.94
N (sequences)	1015	1015	1015

Weighted by N_intervals in Sequence.

Notes: MIS&DIFF is reading mistakes or difficulties by tutees.

So far, the within dyad interactions were analyzed from the angle of reading strategies use. The analysis presented in the present section, takes a different perspective, involving the distinction between three levels of exchange (explanations, questions, answers) between tutor and tutee: the "what"-, "why"- and "how"-levels. Looking at the overall frequency of the within dyad interactions, significant differences between CA and RSA dyads were only found for exchanges at the "what"-level, i.e. interactions in which tutor and tutee discuss what happens in the text they have been reading (Table 12).

The effect size is again very high (ES = 1.03). A significant effect for tutor's gender was found. Dyads with a girl tutor had significantly more exchanges (ES = 0.81) than boy tutor. A significant, but weak negative relationship was found with the frequency of the tutee's reading mistakes and difficulties (ES = 0.06), for which complex variance was found, too. The less reading mistakes and difficulties were observed, the more variation was found with respect to the overall frequency of "what"-level exchanges between tutor and tutee. Contrary to our hypotheses, no significant CA-effects were found for "*why*"- and "*how*"-level exchanges between tutor and tutee.

4.5. Off-task behavior

Table 13 reports the results for off-task behavior for tutors and tutees separately. With regard to the frequency of off-task behavior no significant differences were found between CA and RSA dyads on either tutors or tutees (Table 13). Compared to 5th grade RSA dyads however, significantly more off-task behavior was observed in 2nd grade RSA students who – at the time of the observation – assumed the role of tutee ($\beta = 0.408$, SE = 0.205, $\chi 2(1) = 3.950$, $p < .05$). The effect size is rather moderate (ES = 0.48). It should be noted that the reference category now is 5th grade RSA dyads. Contrast effects show no significant differences between CA and 2nd grade RSA dyads nor between 2nd grade RSA dyads on the one hand and the combination of CA and 5th grade RSA dyads on the other hand.

Table 13. Multilevel analysis on off-task behavior in tutors and tutees

Dependent	**Tutor's off-task behavior**			**Tutee's off-task behavior**		
Fixed effects	**Model 0**	**Model 1**	**Model 2**	**Model 0**	**Model 1**	**Model 2**
Cons (Intercept)	0.559	0.458	0.515	0.905	0.891	0.751
SE	*0.153*	*0.180*	*0.208*	*0.146*	*0.089*	*0.136*
Doing assignment card					1.432	1.413
SE					*0.146*	*0.147*
High reading level		-0.385	-0.390			
SE		*0.191*	*0.188*			
Tutor boy		0.557	0.499			
SE		*0.192*	*0.206*			
MIS&DIFF					-0.091	-0.098
SE					*0.029*	*0.030*
Offtask_tutee		0.269	0.267			
SE		*0.039*	*0.039*			
Offtask_tutor					0.406	0.391
SE					*0.090*	*0.088*
CA dyad			-0.143			0.093
SE			*0.217*			*0.181*
RSA 2nd grade dyad			0.146			0.408
SE			*0.270*			*0.205*
Random effects						
Observation						
Cons / Cons	0.855	0.309	0.296	0.714	0.165	0.140
SE	*0.210*	*0.081*	*0.078*	*0.191*	*0.065*	*0.058*
Cons / Offtask_tutor					0.016	0.007
SE					*0.044*	*0.040*
Offtask_tor / Offtask_tor					0.102	0.098
SE					*0.049*	*0.047*
Sequence						
Cons / Cons	21.882	13.788	13.750	37.171	24.861	24.814
SE	*0.991*	*1.429*	*1.426*	*1.683*	*1.162*	*1.160*
Cons / MIS&DIFF					-5.838	-5.786
SE					*0.458*	*0.460*
MIS&DIFF / MIS&DIFF					1.476	1.460
SE					*0.193*	*0.195*
Cons / offtask_tutee		2.055	2.037			
SE		*0.682*	*0.680*			
offtask_tee/ offtask_tee		1.638	1.653			
SE		*0.529*	*0.529*			
-2log(lh)	5131.16	4719.85	4719.02	5649.97	5213.60	5212.10
N (sequences)	1015	1015	1015	1015	1015	1015

Weighted by N_intervals in Sequence.
Notes: MIS&DIFF is reading mistakes or difficulties by tutees; High is that the dyad is at high reading level; RSA_2 is the RSA in grade 2; offtask_tutee and offtask_tutor are off-task behaviors by tutor and tutee respectively.

Furthermore, significantly more tutee off-task behavior was observed during *"doing the assignment card"*-episodes (ES = 0.23) and in sequences in which the tutor was also observed to be off-task. However, the latter relationship appeared to be quite weak (ES = 0.06). It also appeared to vary across observations, indicating that in some dyads (or in some sessions) students were less influenced by their partner's off-task behavior than in other dyads.

Taking the tutor's off-task behavior as the dependent, the reverse effect (of tutee's off-task behavior on tutor's off-task behavior) was found to be equally significant and equally weak (ES = 0.06). So, no indications were found to decide about the direction of the influence. Analyzing the tutor's off-task behavior, no influence from type of episode was found, but other predictors appeared to play a role. Boy tutors for example, were more frequently observed off-task than girl tutors (ES = 0.54). Tutors with a lower reading level were more prone to be off-task than higher achieving tutors (ES = 0.42).

From the analysis that was done on the tutors' and the tutees' off-task behavior separately (see Table 13), we know that the tutee played a bigger role in this overall effect. But in contrast to the effect size found for the effect on the tutee's off-task behavior separately, the effect size found for the dyad as whole is much larger (ES = 0.75). Again contrast effects show no significant differences between 2nd grade RSA dyads on the one hand and the combination of CA and 5th grade RSA dyads on the other hand, nor between CA and 2nd grade RSA dyads.

Table 14 reports the results for off-task behavior in dyads as a whole. For this analysis, off-task was scored whenever off-task behavior was observed, for either the tutor or the tutee. No significant differences were found between CA and RSA dyads, but significantly more off-task behavior was observed in 2nd grade RSA dyads.

Table 14. Results of multilevel analyses on Off-task behavior within dyads

	Model 0	Model 1	Model 2
Fixed effects			
Cons (Intercept)	0.922	0.313	0.223
SE	*0.142*	*0.174*	*0.191*
Tutor boy		0.704	0.518
SE		*0.197*	*0.200*
Tutee boy		0.417	0.348
SE		*0.197*	*0.186*
MIS&DIFF		-0.135	-0.138
SE		*0.024*	*0.025*
CA dyad			0.190
SE			*0.219*
RSA 2nd grade dyad			0.620
SE			*0.271*
Random effects			
Reading cycle			
Cons / Cons	0.679	0.282	0.232
SE	*0.180*	*0.086*	*0.074*
Sequence			
Cons / Cons	32.829	32.391	32.421
SE	*1.487*	*1.475*	*1.476*
Cons / MIS&DIFF		-4.452	-4.485
SE		*0.339*	*0.327*
MIS&DIFF / MIS&DIFF		0.609	0.617
SE		*0.076*	*0.073*
-2log(lh)	5525.09	5468.68	5466.46
N (sequences)	1015	1015	1015

Weighted by N_intervals in Sequence

Notes: MIS&DIFF is reading mistakes or difficulties by tutees

Tutors' gender played an important role in the off-task behavior observed in the dyad. Significantly more off-task behavior was observed in dyads with boy tutors (ES = 0.63). The gender of the tutee also played a role, but the effect is less big (ES = 0.42) than the effect of the tutor's gender. Still looking at each dyad as a whole, a marginally significant effect (not reported in Table 14) was found for the students'

initial reading level. Somewhat more off-task behavior was observed in low achieving dyads (β = -0.334, SE = 0.179, p = .062, ES = 0.40).

4.6. Summary of findings

Table 15 presents a summary of the findings.

Table 15. Summary of significant findings from the multilevel analyses

Indicator of metacognitive activity	Adjusted average frequency in		Ratio of adj. freq.	Effect size
	RSA dyads	CA dyads	CA / RSA	
Corrections / help by tutor involving reading strategy use	0.020	0.045	2.25	0.61
Corrections / help by tutor not involving reading strategy use	0.314	0.384	1.22	0.10
Overall reading strategy use by tutor	0.554	1.306	2.36	1.02
Spontaneous use of reading strategies in tutor questions	0.280	0.650	2.32	0.80
Spontaneous use of reading strategies in tutor explanations	0.182	0.636	3.49	1.10
Overall frequency of "what"-level exchanges between tutor and tutee	1.334	2.202	1.65	1.03

For five indicators of metacognitive activity within peer tutoring dyads, significant differences between CA and RSA dyads were found. These indicators include (1) reading strategy related corrections and help by the tutor following reading mistakes or difficulties by the tutee, (2) overall reading strategy use by the tutor of by the dyads, (3) spontaneous strategy use by the tutor when asking questions, (4) spontaneous strategy use by the tutor when giving explanations and (5) the overall frequency of "what"-level exchanges between tutor and tutee. The differences between CA and RSA dyads are quite big. This is not only shown by the effect size measures that range from 0.61 to 1.10, but also in the ratio of the adjusted average frequencies in CA vs. RSA dyads. The most extreme example is the tutors' spontaneous use of reading strategies while giving explanations. This was observed

almost 3.5 times more in CA dyads than in RSA dyads. Other relevant behaviors were observed more than twice as much in CA dyads than in RSA dyads.

In addition to these five indicators, a significant and large CA-effect was also found for the overall reading strategy use within the dyad as a whole. A marginally significant CA-effect was found for giving help or making corrections without use of reading strategies.

5. Discussion

5.1. CA peer tutoring and metacognition

Although in the past decades much has been reported on the effects of peer tutoring and its different forms, little is known about the mechanisms or the peer tutoring interaction processes that promote these outcomes (Ginsburg-Block & Fantuzzo, 1997). In our study we have tried to shed some light on these interaction processes, by identifying instances in which metacognitive activity was apparent, either from the part of the tutor or the tutee. It was expected that more such instances would be found in CA dyad sessions. The focus was on acts that could be seen as indicators for (self-) monitoring or (self-) regulating behavior.

Interestingly enough, the highest differences between CA and RSA dyads were found for what could be considered to be the strongest indicators of metacognitive activity: spontaneous use of reading strategies (either when giving explanations or when asking questions), overall use of reading strategies, use of reading strategies when giving help or correcting reading mistakes. For weaker indicators, such as the overall frequency of exchanges between tutor and tutee about what they had read in the text, and giving help or making corrections without reading strategy use much smaller differences (in terms of ratio of adjusted frequencies) were found.

Spontaneous use of reading strategies indicates that the students have internalized the strategies. It also implies a decision to use a strategy. So, both active

monitoring and regulation can be inferred. This is much less sure when the questions are read from the assignment card. Those assignment cards are typical scaffolding tools. The whole idea is that students can leave them behind as soon as possible. From the observational data, there is quite some evidence that indeed students became able to tutor without having to rely on the assignment cards all the time.

According to Topping (1998b) the deeper and more frequent the questions and explanations from either of the students in the dyad are, the richer the interaction is and the better the joint understanding is likely to be. From this perspective, it is already a positive finding that the total frequency of exchange was observed to be 1.65 times higher in CA-dyads. In our analysis, the aspect of "deeper" exchange was taken up by identifying "why"- and "how"-level exchanges. For these two variables no differences between CA and RSA dyads were found. This can be attributed to the fact that these two levels were not frequently observed at all. Because there is only little variance on these two levels, it is not surprising that we could hardly find any significant difference between the two conditions. The fact that most of the questions and explanations are on the information identification ("what") level might be due to the questions the assignment cards. They are mainly on the "what"-level, e.g., *"What kind of text do you think it is?"* or *"What does the story mainly tell us?"* There is only one "why" question printed on an assignment card: *"Why do you think they do that?"* It is quite well possible that higher frequencies of higher-level exchanges would have occurred, had these types of questions had a more prominent place in the reading strategy instruction and in the assignment cards.

Quite interesting, and in accordance to our expectations, is the finding that most of the indicators for which a significant different between CA and RSA dyads were found, relate to the tutor's behavior. With respect to the tutee's overall strategy use for example, no significant differences between CA and RSA dyads were found.

So, the better opportunities for engaging in metacognitive activities that are offered by CA tutoring mainly hold for the tutor, much less for the tutee.

Also striking is the fact that for none of these indicators any difference between 5th and 2nd grade RSA dyads was found. It is not the age of the tutor by itself that matters, so it appears, but the age difference between tutor and tutee.

The findings of present study might offer some explanation for the results obtained by Van Keer and Verhaeghe (2003), indicating that CA tutors could benefit more from peer tutoring than RSA tutors. The reading tasks that CA tutors face may seem to contradict Vygotsky's theory if we simply look at the relatively easy materials tutors have to read. But the role of CA tutors is more than understanding the contents of the texts. Their role mainly lies in monitoring and regulating the reading process of the tutees. Tutors have to plan and monitor and choose the right timing and appropriate opportunities to intervene in their tutee's reading process. And they are also constantly confronted with questions or difficulties or mistakes from tutees and are forced to explain or put things in new perspectives, in a way that suits the tutee's level of understanding. For most tutors this is a quite demanding task. Therefore, at the metacognitive level, the findings with respect to the CA tutors' development is in consistency with Vygotsky's Zone of Proximal Development theory.

The present finding that CA tutors demonstrated more monitoring and regulating behaviors than RSA tutors can also be viewed from still another perspective. As noted before, with respect to the reading task by itself, the main difference between CA and RSA tutors is to be found in the difficulty level of the reading materials they deal with. According to information processing theory (e.g., Winne, 2001; Biemiller *et al.*, 1998), the possibility to engage in metacognitive activities is related to the capacity of working memory and how much of the working memory can be allocated to metacognitive activities. The higher the relative difficulty level of the reading materials is, the less working memory could be allocated to

monitoring and regulating activities. This explanation parallel the "shared limited capacity" theory in linguistic (Chall, 1983; Chen, Lau, Yung, 1993; Norman & Bobrow, 1975; Perfetti & Hogaboam, 1975; Rumelhart, 1978). This theory states that within the limited capacity of reading process, if fairly large processing capacity is devoted to lower reading process, less available capacity will be devoted to higher-order processes. However, if atomization is achieved for the lower order process, more capacity could be allocated to the higher-order process—which in our case is metacognitive activities. In the case of CA, the relatively less demanding reading task could allow tutors to allocate more working memory or processing capacity to monitoring and regulating the tutees' reading. As a consequence, CA tutors can benefit more from their tutoring experience and internalize better the metacognitive strategies they practice while tutoring their tutee (Webb *et al.*, 1995)

5.2. Using multilevel analysis to analyze observational data

To disclose the interaction processes within peer tutoring dyads working on a reading assignment, the present study combined a qualitative data collection approach with quantitative analysis techniques. For the first phase of the analysis, a time-based coding of the observations was chosen rather than an event-based coding. It has to be recognized that compared to event-based coding, time-based coding causes some loss of information. It was felt however that the advantages with respect to possibilities for subsequent quantitative analyses outweighed the disadvantages of time-based coding.

Using multilevel analysis it was possible to perform analyses that *"keeps observations together that belong together"*. This principle was applied at two levels. By aggregating the original small time units into larger one-minute sequences, behaviors that from a time perspective are close to each other, such as questions and subsequent answers, were kept together and their co-variation could be analyzed. By nesting these sequences into the larger units of "reading cycles",

the nested structure of the observational data and the associated phenomenon of auto-correlation could be taken into account. It is important to do so. Ignoring auto-correlation in the data could lead to "incorrect estimates of precision and standard error", and "increased risk of finding differences and relationships where none exist" (Jones & Bullen, 1994; Cliff & Ord, 1981; Skinner, Holt & Smith, 1989). In our analysis, we used a two-level structure – sequences nested in reading cycles. We argue that sequences within one videotaped session are more alike than sequences in other sessions, because of the specific characteristics of the students (dyads) involved, the reading materials being used, and the specific task they have been working on.

With respect to the analyses that were performed, a number of limitations in the present study should be mentioned. First of all, a number of possibly interesting explanatory variables was not explored yet. For example, the assignment card the students were actually working on during the videotaped session was coded, but not used as an explanatory variable. Preliminary analyses had shown no much difference; therefore it was decided not to take this variable to the multilevel analyses. For the present study, when defining our dependent variables we also did not make any distinction between the reading strategies that were distinguished in the program, but instead summed across all strategies. As a result, we get a more general picture, assuming equal effects across the different assignments students have been working on in the course of the school year.

Principally the database we used would have made it possible to perform repeated measures analyses. This could be done by entering sequence as an explanatory variable and let it vary randomly across reading cycles. Such an analysis would have permitted to investigate whether the value of the dependent variable changes over time within the course of a session or reading cycle. This kind of analysis however would not meet our research questions. Our interest was in the global amount of opportunities for metacognitive activity created within peer tutoring sessions and the differences between CA and RSA dyads with respect to that.

When exactly those opportunities emerged most – in the beginning of a session, in the middle, or at the end – was not exactly the focus of our interest. Of more interest to our research would have been to disclose differences in evolutions during the course of the whole intervention. For example, it can be expected that the use of reading strategies during peer tutoring evolves between the start of the intervention (the very first peer tutoring sessions) and the end of it. Some growth curve can be expected. And then it becomes interesting to see whether these growth curves differ according to the condition (CA versus RSA). This could be interesting for two reasons. First of all, such a picture of the evolution over time, would give an indication of the relative positions at the beginning of the treatment. Secondly, it would have given a better idea of the students' position at the end. In the present analysis, we do not have a real baseline, at least not with respect to the dependent variables as such. To compensate for that, we matched the dyads with respect to initial reading comprehension achievement level – and we also entered this variable as predictor in all analyses, assuming that initial reading level correlates well with the variable of interest. No doubt, a repeated measures analysis covering the whole time span of the treatment would have yielded a more precise picture. In principle, such an analysis can be done by nesting the reading cycles within a time variable indicating phases in the treatment (e.g., month 1, month 2, *etc*.). Or the variable indicating the assignment card the students are working on could be used as a proxy time indication. By using the same time-variable as a predictor the evolution over time could be analyzed. We did not endeavor into such an analysis because of the rather limit number of videotaped and coded sessions we had available at present. The analyses we performed instead, yield an overall picture of the differences between observations in CA and RSA dyads, irrespective of any time perspective. The results nonetheless appear to be promising.

6. Conclusions

For five indicators of metacognitive activity during peer tutoring sessions significant and large effects were found between observations in CA dyads and observations in RSA dyads. The largest effect sizes were found for the strongest indicators: tutors' spontaneous use of reading strategies when explaining something to their tutees or when asking questions, and giving reading strategy based help when the tutee is making a reading mistake or experiences some difficulty. For weaker indicators, such as the overall amount of exchanges between tutor and tutee (apart from correcting mistakes), less big differences were found. Most of the indicators for which significant differences were found between CA and RSA dyads involved tutor's behavior only. With respect to the tutee's behavior no CA – RSA differences were found. For the significant indicators no differences were found between second and fifth grade RSA dyads.

The results support the hypothesis that – compared to RSA tutoring – CA peer tutoring creates more opportunities for tutors to engage in activities that enhance their metacognitive skills. Compared to tutoring a classmate, it appears that tutoring a younger student is a more effective learning environment for practicing monitoring and regulative skills that help tutors become self-regulating, independent readers themselves.

Appendix: dyads dialogue excerpts

Reading strategy 1:

Tutor: from the title and pictures, what can you think?

Tutee: I can think of a stork.

Tutor: and?

Tutee: little river, flowers.

Tutor: And what else?

Tutee: And water plants, little fish, lobsters, and shrimp.

Tutor: And anything else? (Waiting for reply.) Nothing else. Ok, let's start to read.

Tutor: from the word "a trick with a straw hat", what could you think of?

Tutee: straw…

Tutor: you read this title, what can you think that the article is about?

Tutee: they… they use straw hat to make a trap…

Reading strategy 2:

Tutor: I ask you the meaning of one word. What is the meaning of "disobey"?

Tutee: it is "protest". It means he asked you not to do that, but you definitely did that. That is "disobey".

Reading strategy 5:

Tutor: have you understood this part?

Tutee: (nodded) I have understood.

Tutor: ok, then you tell me what this part is about?

Tutee: This part is about a robber robbed something and ran away. He took the key and ran away from the backyard. He locked the door and threw the key into the big dustbin in the backyard.

Reading strategy 4 & 5:

Tutor: did you understand the article? Could you tell me the main idea of the article?

Tutee: It was mainly that tornado came and Dorothy was blown away by it.

Tutor: yes, very simple, isn't it? When did the story happen?

Tutee: it happened when Alice was sad.

Tutor: (laugh) then on which day did it happen?

Tutee: it happened on today, oh, "one day".

Tutor: where did it happen?

Tutee: it happened at their home.

Tutor: did it happen at their home?

Tutee: it happened in a room.

Tutor: and?

Tutee: it happened on big grassland in Kansas.

Tutor: (nodded) Mmm, the big place is the big grassland in Kansas. And what is the "smaller" place?

Tutee: the "smaller" place is their home.

Tutor: and anywhere else? Read this part "Uncle Henry sat on the staircase in front of the entrance"? Isn't that a place?

Tutee: (nodded) yes, and also in the sky.

Tutor: in the sky, isn't it? And anywhere else? Try to find the words indicating places in the text.

Tutee (reads): "She creamed loudly and ran into the basement." And also the basement.

Tutor: still anywhere else? (Both read through again together) it seemed nowhere else. Then who are the characters in the story?

Tutee: Uncle Henry, Aunty Emma and Alice.

Tutor: still someone else?

Tutee: and Tuo Tuo (the name of a dog)

Tutor: There is a Tuo Tuo, right? The story told about tornado blew them to the sky and finally they landed. Who is the main character?

Tutee: Alice.

Tutor: Then you tell me why was Alice blown up in the sky?

Tutee: because she didn't hide in the basement in time.

Tutor: why didn't she have time to hide in the basement?

Tutee: because Tuo Tuo ran out of it and Alice wanted to save Tuo Tuo and then she ran out and she didn't have time to hide in the basement.

References

Baker, L. & Brown, A. L. (1984). Metacognitive skills and reading. In P.D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.). *Handbook of reading research* (pp.353-394). New York: Longman.

Bentz, J. L., & Fuchs, L. S. (1996). Improving peers' helping behavior to students with learning disabilities during mathematics peer tutoring. *Learning Disability Quarterly,* 19, 202-215.

Biemiller, A., Shany, M., Inglis, A., & Meichenbaum, D. (1998). Factors influencing children's acquisition and demonstration of self-regulation on academic tasks. In D. H, Schunk & B. J. Zimmerman (Eds.). *Self-regulated learning: From teaching to self-reflective practice* (pp. 203-224). New York: Guildford.

Blaye, A., Light, P., Joiner, R., & Sheldon, S. (1991). Collaboration as a facilitator for planning and problem solving on a computer-based task. *British Journal of Developmental Psychology*, 9, 471 – 483.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13, 4-16.

Chall, J. S. (1983). Building theories of reading ability: on the relation between individual differences I cognitive skills and reading comprehension. *Cognition,* 9, 73-113.

Chen, M. J., Lau L. L., & Yung Y. F.(1993). Development of component skills in reading Chinese, *International Journal of Psychology*, 28(4), 481-507.

Cliff, A. D. & Ord, J. K. (1981). *Spatial processes, models and applications.* London: Pion.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: a meta-analysis of findings. *American Educational Research Journal*, 19, 237-248.

Dole, J. A. (2000). Explicit and implicit instruction in comprehension. In B. M. Taylor, M. F. Graves, & P. Van Den Broek (Eds.). *Reading for meaning. Fostering comprehension in the middle grades* (pp. 52-69). New York: Teachers College Press.

Duran, R. T. & Guavain, M. (1993). The role of age versus in peer collaboration during joint planning. *Journal of Experimental Child Psychology*, 55, 227-242.

Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: a component analysis. *Journal of Educational Psychology*, 84, 331-339.

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.

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. *The Elementary School Journal.* 99(3), 201-219.

Fukkink, R., Van der Linden, S., Vosse, A., & Vaessen, K. (1997). *Stap Door! Handleiding* [*Step out! Manual*]. Utrecht, The Netherlands: Sardes.

Ginsburg-Block, M., & Fantuzzo, J. (1997). Reciprocal peer tutoring: An analysis of "teacher" and "student" interactions as a function of training and experience. *School Psychology Quarterly*, 12(2), 134-149.

Jones, K. & Bullen, N. (1994). Contextual models of urban house prices: A comparison of fixed- and random-coefficient models developed by expansion, *Economic Geography,* 70(3), 252-272.

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly,* 31, 3, 263-289.

King, A. (1991). Effects of training in strategic questioning on children's problem solving performance. *Journal of Educational Psychology*, 83, 3, 307-317.

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

Li, L. & Verhaeghe, J. P. (2005). Reading strategy use, metacognitive awareness and reading attitudes: Effects of strategy instruction and peer tutoring. Paper presented at the 86th AERA Annual Meeting, Montréal, Canada.

Li, L., & Verhaeghe, J. P. (2005). The effect of explicit reading strategies instruction and peer tutoring on Chinese students' reading comprehension. Paper presented at the 11th Biennial Conference of European Association for Research on Learning and Instruction (Earli), Nicosia, Cyprus.

Maheady, L. (1998). Advantages and disadvantages of peer-assisted learning strategies. In K. Topping & S. Ehly (Eds.), *Peer-assisted learning* (pp. 45-65). Mahwah, New Jersey: Erlbaum.

Manion, V. & Alexander, J. M. The benefits of peer collaboration on strategy use, metacognitive causal attribution, and recall. *Journal of Experimental Child Psychology,* 67, 268-289.

Miller, A. D., Barbetta, P. M., & Heron, T. E. (1994). Start tutoring: designing, training, implementing, adapting, and evaluating tutoring programs for school and home settings. In R. Gardner III, D. M. Sainato, J.O. Cooper, T. E. Heron, W. L. Heward, J. Eshleman, & T. A. Grossi (Eds.). *Behaviour*

analysis in education: Focus on measurably superior instruction (pp. 265-282). Pacific Grove, California: Brooks/Cole Publishing Company.

Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processes, *Cognitive Psychology*, 7, 44-64.

Perfetti, C. A., & Hogaboam, T. (1975). Relationship between simple word decoding and reading comprehension skill, *Journal of Educational Psychology,* 67, 461-469.

Rumelhart, D. E. (1978). Toward an interactive model of reading. In S. Dornic (Ed.). *Attention and performance VI*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Shayer, M. (2003). Not just Piaget; not just Vygotsky, and certainly not Vygotsky as alternative to Piaget. *Learning and Instruction*, 13, 165-185.

Skinner, C. J, Holt, D., & Smith, T. F. (Eds,1989). *Analysis of data from complex surveys*. Chichester: John Wiley.

Snijders, T. & Bosker, R. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling.* London: Sage.

Topping, K. J. & Bryce, A. (2004). Cross-age tutoring of reading and thinking: influence on thinking skills. *Educational Psychology,* 24(5), 595-621.

Topping, K. J. & Ehly, S. (1998). Introduction to peer-assisted learning. In K. J. Topping & S. Ehly (Eds.). *Peer-assisted learning* (pp. 1-23). Mahwah, New Jersey: Erlbaum.

Topping, K. J. & Lindsay, G. (1992). Paired reading: a review of the literature. *Research Papers in Education*, 7(3), 1-50.

Topping, K. J. (1987). Peer tutored paired reading: Outcome data from 10 projects. *Educational Psychology*, 7(2), 133-144.

Topping, K. J. (1998a). Effective tutoring in American Reads. A reply to Waski. *The Reading Teacher*, 52(1), 42-50.

Topping, K. J. (1998b). Paired learning in literacy. In K. J. Topping & S. Ehly (Eds.), *Peer-assisted learning* (pp. 87-104). Mahwah, New Jersey: Erlbaum.

Van Keer, H. (2004). Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring. *British Journal of Educational Psychology*, 74, 1, 37-70.

Van Keer, H. (2002). *Reading strategies instruction and peer tutoring in primary school. A quasi-experimental study of the effects on students' reading comprehension and metacognition.* Proefschrift ingediend tot het behalen van de academische graad van Doctor in de Pedagogische Wetenschappen (Ghent University), Promotor: Prof. Dr. Jean Pierre Verhaeghe.

Van Keer, H., & Verhaeghe, J. P. (2003). Effects of explicit reading strategies instruction and peer tutoring in second and fifth graders. In E. De Corte, L. Verschaffel, N. Entwistle, & J. Van Merriënboer (Eds.), *Powerful learning environments: Unraveling basic components and dimensions* (pp. 221-228), Oxford: Elsevier.

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. *The Journal of Experimental Education,* 73(4), 291-329.

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

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

Wild, K. P. (2001). *Catmovie 4.08*. Freiburg (Germany), University of Freiburg.

Winne, P. H. (2002). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & H. S. Dale (Eds.). *Self-regulated learning and academic achievement: Theoretical perspectives*. Mahwah, NJ: Erlbaum.

Chapter 3

Effects of Reading Strategy Instruction and Peer Tutoring on 5th graders' Reading Comprehension: A Multilevel Repeated Measures Approach*

Abstract

This study examined the effects of explicit instruction of reading strategies (RS) and two types of peer tutoring, cross-age (CA) and reciprocal same-age tutoring (RSA) on Chinese 5th grade students' reading comprehension development. Having adjusting for the effects of potential confounders from students, schools and family environments, the results indicate that: (1) schools and family environments, significant short-term experimental effects were found on students' reading comprehension growth for CA groups, and long-term effects were found for all the experimental groups; (2) no significant differences were found among experimental groups in either posttest or retention test. The results were compared with those of similar studies and explanations for similarities and dissimilarities were sought.

Key Words: reading strategy peer tutoring multilevel analysis repeated measures

* This chapter is based on an article with the same title submitted to the *American Educational Research Journal* for publication.

1. Introduction

1.1. Peer tutoring

Peer tutoring (PT) is a specific form of peer-mediated instruction in which one student serves as an instructional agent (tutor) for another student (tutee) (Topping & Ehly, 1998; Topping, 1998). Different from the occasional instances of helping behaviors among students, peer tutoring is characterized by its distinct role-taking (tutor vs. tutee), structurally embedded curriculum with regular sessions (once or twice per week) and social skills training to the tutors to ensure the quality of peer interaction. Two categories of PT can be distinguished according to the age difference between tutor and tutee: "same-age tutoring" (SA) and "cross-age tutoring" (CA). To motivate both students in SA tutoring, a specific form "reciprocal same-age tutoring" (RSA) in which the tutor/tutee role is switched regularly is suggested and put into practice (Fantuzzo *et al*., 1992).

In the past decades, research literature has reported positive results for peer tutoring in a big variety of fields, such as mathematics, reading and writing (Cohen *et al*., 1982; Mathes & Fuchs, 1994; Mathes *et al*., 2001; Mellanby *et al*., 2001), social adjustment (Kamps *et al*, 1999), physical education, and sex education (Mellanby, 2001) and so forth. It has been shown that in classroom settings, peer tutoring has a positive effect on students' socialization and cognitive achievement (Steinberg *et al*., 1992; Wentzel, 1999). Better results were obtained than for unstructured cooperative learning or individual learning (Rohrbeck *et al.*, 2003). The particular combination of structure and student autonomy is probably one of its most powerful features.

From the perspective of learning opportunities, one of the prominent reasons to advocate peer tutoring is that interaction between students about texts provides dyads with opportunities for metacognitive exchange and modeling (Almasi, 1996; Fuchs *et al.*, 1997; Palincsar *et al.*, 1991). As is suggested by researchers,

children's failure to use appropriate strategies in cognitive tasks may be due to a lack of monitoring and regulation in their learning process, which to a great degree is due to a lack of appropriate external feedback (Bereiter & Scardamalia, 1987).

Peer tutoring creates a convenient opportunity to provide external feedback to students, much more than normal class teachers could do. A particular issue addressed here is whether the age of the tutor or tutor-tutee age difference makes a difference.

1.2. Reading strategies

Reading strategies (RS) can be defined as "operations over and above the process of reading which produce efficient, insightful comprehension of the content of reading texts (Pressley *et al*., 1989). Following Van Keer and Verhaeghe (2005), in the present study six reading strategies were taught to students directly: (a) activating prior knowledge, (b) finding the meaning of unfamiliar words and expressions, (c) making predictions, (d) finding main ideas of texts, (e) monitoring and regulating comprehension of the texts, and (f) distinguishing different types of texts. These reading strategies were chosen in accordance with other reading research (De Corte *et al*, 2001; Klinger & Vaughn, 1996; Pressley *et al*., 1992; Paslincsar & Brown, 1984). The whole set encompasses a number of features: (1) the strategies are organized chronically, showing what can be done before reading (e.g., activating prior knowledge and talking about what have been known about the topic), during reading (e.g., understanding words, expressions, paragraphs and the whole text) and after reading (e.g., classifying what types of texts were read); (2) they are organized from simple strategies, such as decoding and finding the meaning of new words, to more complex ones such as summarizing and reasoning; (3) the set contains both cognitive strategies (such as making predictions and summarizing main ideas of the paragraphs or text) and metacognitive strategies (e.g., monitoring and regulating students' own understanding).

Although a lot of studies have been done on students' reading strategies and peer tutoring respectively, research combining the two lines of research is still relatively sparse (Van Keer, 2004; Van Keer & Verhaeghe, 2005). Uniting these two lines of research yields the following advantages: (1) for primary school kids, mastering abstract strategies and actually applying them flexibly in their reading process is not an easy task, esp. for weak readers. Individualized help and scaffolding are necessary and PT is a very useful way to provide this; (2) having peers monitor another student's reading, reinforcing the use of reading strategies could be beneficial for both students, for the tutor as well as for the tutee; (3) cooperative reading in dyads is more than putting two students together and ask them to help each other. It needs concrete tasks and learning content. A structured learning environment to promote improved performance is the key to academic and cognitive growth (Klinger *et al.,* 1996). Reading strategies instruction provides students with declarative, procedural and conditional knowledge of reading strategies (Pressley *et al.*, 1992; Paris, Byrnes & Paris, 2001) and also with other scaffolding materials (such as assignment cards), which can support the dyads' cooperative learning processes.

1.3. Fostering metacognition in reading

As mentioned above, some of the RS that were included in the experimental intervention are metacognitive in nature. Metacognition is "knowing about knowing" or "thinking about thinking". It includes planning, monitoring, regulation and reflection activities during cognitive processes (Garner & Alexander, 1989) and it involves what, when and how strategies should be used (Paris, Lipson, & Wixon, 1983).

Metacognition helps to turn kids into independent readers who can monitor and regulate their own reading progress, independent from a teacher who points out and clarifies difficult parts of reading materials. Including MC strategies yield more chances for continuous growth in reading comprehension, independent from

teachers. In the present study we will evaluate this effect by assessing long-term effects on a retention test, several months after the end of the intervention. Whether metacognitive strategies play a role in the improvement of reading comprehension is not easily assessed. If long-term effects on reading comprehension were found, we would have a good indicator (albeit indirect) that indeed metacognitive strategies had played a role.

In peer tutoring, the tutors' main task is to monitor and give feedback on the tutees' reading. They also ask questions to check if tutees understand or master the reading strategies well and give corrections or explanations when necessary. The "teaching experience" gives tutors unique opportunities in practicing metacognitive skills. Since monitoring and regulating someone else's behavior is easier than monitoring one's own behavior, a tutor can more easily acquire metacognitive skills and transfer them to his/her own reading. Peers are close in age, use similar languages, and they will not hesitate to ask or answer questions. From this perspective that in the present study peer tutoring is considered to be equally beneficiary for tutors as it is for tutees.

Various tutoring programs in the late 20th century (e.g., Clay, 1985; Butler, 1991; Invernizzi, Juel, & Rosemary, 1997; Winter, 1987) used teachers, paraprofessionals or colleague students to help primary school children with their reading problems. Those programs mainly stressed the benefit of cognitive learning on tutees, as a remedy for regular reading practice. There are also other projects, using individual students with learning difficulties as tutors to help younger tutees, with a main focus on the individual tutors' benefits from tutoring experience (e.g., Schrader & Valus, 1990). In our study, the intervention focused on complete regular classes in schools, providing all students with opportunities to practice cognitive and metacognitive strategies.

1.4. Chinese classrooms

Most research on reading strategies or peer tutoring is done in western countries, involving English speaking students and teachers using English as the language of instruction, or some other western language. The present study is based on a Flemish study (Van Keer, 2004; Van Keer & Verhaeghe, 2005) involving regular Flemish primary school classrooms with Dutch speaking pupils and Dutch as the language of instruction. In that study a significant positive effect of reading strategy instruction was found on students' reading comprehension achievement. In addition cross age tutoring was shown to be more effective than same age tutoring and to yield larger long-term effects than any of the other conditions (Van Keer, 2004).

For the purpose of the present study the intervention program used in the Flemish study has been translated and adapted for use in regular Chinese primary school classrooms in Mainland China. The intervention took place during Chinese language classes. The aim of the program was to improve Chinese pupils' reading comprehension when reading Chinese texts.

When implementing interventions in different cultures, especially in the field of reading, both cultural and linguistic issues have to be tackled. Languages such as English or Dutch belong to the Indo-European language system. They are phoneme-based writing languages, which are totally different from a morpheme-based writing system such as Chinese. There are also other differences. For example, in Chinese, complex concepts are often formed with simple, commonly used morphemes, which is not the case in Western languages. It has to be verified whether reading strategies that are useful in English or Dutch would also be useful in improving reading comprehension in Chinese.

There is also a cultural issue. Though the Chinese educational system is changing fast, influences from traditional teaching styles still persist. Introducing such

practices as teachers thinking aloud in the classroom to model the use of reading strategies, or having students choose their own reading texts, or having students tutor other students (which are all elements in the intervention program) in Chinese classrooms is probably to be considered as a much bigger endeavor than the introduction of similar practices in European or American primary school classrooms. This is not only true for teachers, but also for students who have to get acquainted to this different way of working as well (and get back to the more traditional way when reading class is over). A contextual issue that should not be underestimated relates to class size in Chinese primary schools (which is much larger than in most European or American schools) and the fact that – unlike most Flemish and other western schools – most Chinese primary schools do not have a school library or a local public library from which reading books can be borrowed for use in reading classes. The adaptation of the Flemish program involved ways to cope with this quite different educational context. One element in this was the provision of attractive Chinese children's reading books to the experimental classrooms. This of course could raise a methodological issue. We had to make sure that improvements in students' reading comprehension could be interpreted as the result of the strategy instruction and the use of peer tutoring and should not attributed solely to the provision of attractive children's reading books.

Also the Chinese school culture is considerably different from the European situation. Officially, all schools are equal and there are no officially rankings among them. But in reality in big cities, there are three types of public schools: city key schools (or model schools at city level), district key schools and ordinary schools. These schools are classified unofficially by their prestige, students' performance in various district-wide or citywide exams, teachers' qualifications. These differences among schools were partly the consequence of naturally accumulated differences. For example, Chinese universities had the tradition of organizing schools to serve their personnel and these university-attached schools generally had a better performance than other schools in the same neighborhood.

As a consequence, more and more parents wanted to enroll their kids to such key schools. The key schools adopted a selective enrollment policy for children from outside the working unit or neighborhood, based on entrance exams and/or financial contributions by the parents. As a consequence, key schools differs from ordinary schools in their financial situation, basic infrastructure, students' starting level and teachers' bonus on top of their official salary. The latter enabled key schools to hire teachers with higher qualifications, which again enlarged the existing inequalities among schools.

1.5. Research questions

The present study focuses on the effects of the intervention on the fifth grade students involved in the study. The general question is whether the teaching of reading strategies, in combination or not with either cross age or reciprocal same age tutoring, in regular Chinese language classes in Chinese primary schools would replicate the Flemish results and enhance the students' Chinese language reading comprehension significantly.

Based on the results of previous research, the following hypotheses were stated:

(1) The explicit instruction of reading strategies significantly improves 5th graders' reading comprehension growth, in short-run as well as in the long run.

(2) In addition to the positive effects of reading strategy instruction, there are significant extra benefits from peer tutoring (both CA and RSA) on Chinese 5th graders' reading comprehension growth, both in the posttest and the retention tests results.

(3) Comparing the effects of CA and SA, significantly more improvement of Chinese 5th graders' reading comprehension is to be found in 5th graders who acted as a tutor in CA dyads.

2. Experimental design

2.1. Conditions

This intervention involved 8 schools and 18 classes, with about 650 students in 5th grade. Because most of Chinese schools lack reading resources for students, to facilitate the research, approx. 2,500 different Chinese children's books for the 2nd and 5th grade students were purchased and circulated regularly among the participating schools. To test all the possible experimental factors involved, 5 reading conditions were set up: reading strategy plus cross-age tutoring group (CA), reading strategy instruction plus reciprocal same-age tutoring group (RSA), reading strategy instruction plus individual practice (RS), control group (CON) and a free reading condition. The free reading condition (FR) was set up as a second control condition to test the effect of the reading books. In FR groups, students got reading books and they read individually while the intervention was done in other experimental groups. Students in RS group received teaching on the six reading strategies and applied them in their reading activities, under teachers' guidance. Students' in CA and RSA conditions practiced reading strategies with dyads, after receiving the same instruction on reading strategies. In CA, a student from 5th grade acted as a tutor for a student from 2nd grade (tutee). Dyads in CA were matched by the relative reading level. For example, a strong reader from the 5th grade tutored a strong reader from the 2nd grade. The same was true for the average and weak readers in both grades. In RSA, dyads were formed with students of similar reading levels and the roles of tutor and tutee alternated weekly. Every week, about 60 minutes of extracurricular time was allocated to the intervention. Control groups didn't get any of the books or intervention. When the intervention was being organized, students in the control groups were in various language related activities, such as Children's literature appreciation, little plays or doing Chinese homework under the supervision of their teachers. All students were kept in school during these hours before they were fetched by their parents after work.

2.2. The Intervention

The intervention was scheduled once a week during extracurricular hours (from 3.30pm to 5:00pm) for a whole school year. Reading strategies were taught in all the experimental groups. For the strategy instruction, a "sandwich" pattern was applied. Each strategy was first introduced to the students during one lesson period. Teachers first explained the declarative (the name and main concept), procedural (main steps) and conditional knowledge (when and where to use) of the strategy with a "thinking aloud" method. Then they introduced the assignment card, which synthesized the main steps and questions that helped to apply the strategy. These questions include "What are the main characters in the story?", "How have I understood the paragraph I just read?", "what will happen in the story?" and so on. Then a guided practice on the strategy with selected texts was organized. Students were asked to answer the questions in the assignment card, based on their understanding of the texts and feedback to their answers was given in the whole class. After the introduction lesson, two sessions of independent practice on the strategy were scheduled, using the assignment card as a scaffold. Students could choose freely the books that were circulating among the experimental groups to practice the reading strategy. After the introduction and individual practice sessions, further practice either individually (as in RS condition) or in dyads (as in CA or RSA conditions) were organized. After three reading strategies were practiced, for a number of sessions students were asked to practice all three of them flexibly, with the help of a synthesizing assignment card including the three strategies. After all the six strategies were introduced, other integration sessions were organized.

Because one of the key factors in the effect of peer tutoring program is tutor training (Hock *et al*., 2001; Klingner & Vaughn, 1996; Barron & Foot, 1991), in the RSA and CA groups, we organized a number of focused on tutoring sessions to train tutors on tutoring and social skills. These sessions focused on: "how to make

him/her relax during peer-tutoring sessions", "how to listen properly to my tutee", "how to deal with his or her mistakes and difficulties", "how to deal with my own mistakes", "how to plan and keep records of our reading sessions", etc. Most of these sessions were organized by role-taking plays or group discussion.

2.3. Treatment fidelity

Since our study replicated a similar program in the Dutch-speaking region of Belgium, all program materials were translated into Chinese and adapted to the Beijing local educational situation. These included the teachers' manuals, classroom posters highlighting the key elements from the reading strategies and tutor training, students' assignment books (with assignment cards and selected texts) and assessment instruments. Before the real intervention, a pilot study was carried out in 3 classes of 2 schools with students from a proceeding cohort. Half a month before the pretest data collection, all teachers in the experiment received a training related to the experiment. The training lasted for 3 full working days for a total of 24 hours. The trainings provided not only the fundamental goals of the intervention, but also detailed teaching contents and demonstration of a few important lessons in reading strategy instruction and tutor training. Teachers got their teaching manuals with detailed written instructions on how to teach each of the lessons. To avoid contamination of different conditions, teachers were provided with the necessary training and manuals matching their conditions.

During the intervention all through the year, the project coordinator and her assistants went to observe the experimental classes every week. Half an hour discussion with all the involved teachers was organized prior to every introductory lesson either on tutor training or reading strategy, to highlight its main content. Special attention was given to the potential problems that might arise in the classroom. Teachers were encouraged to talk about their experiences, difficulties or puzzles during the weekly meeting.

The fidelity of the treatment was evaluated through two kinds of means: direct observation and teachers' self-report on a checklist for each new introduction lesson. These checklists were pre-designed outlines of the main procedures that should be followed by the teachers during their instruction. At the same time, observations with video camera recording on these lessons were made by the coordinator and the two research assistants.

Treatment fidelity checks indicate that, teachers' self-report complied more than 90% of the content written on the checklist, with some variations according to specific teaching styles of some teachers. For example, during some reading strategy introduction lesson, 3 reading articles were recommended to be discussed in the whole class. Some teachers sometimes organized two thorough discussions in the classroom while keeping the last one for students' individual homework, taking into consideration students' independent study ability.

2.4. Instruments

Reading comprehension tests

In accordance with the age characteristics of our 11-13 years old students, we measured seven separate domains of students' reading comprehension: understanding specific words, phrases or sentences in the context; identifying specific information or details; summarizing main idea of paragraph(s) or whole text; recognizing the structure or genre of the text; understanding the logic or reasoning in a text; appreciating the mood, emotion of characters or author; recognizing the viewpoint of a character or the author. The content of the reading comprehension tests is in general agreement with most of the tests in reading in use (e.g., OECD, 2003).

The test was previously used in a similar Flemish study and was translated into Chinese. Like most reading comprehension tests, it provides a number of short stories followed by a varying number of multiple-choice questions to test students'

understanding of the texts. To adapt to the students' reading ability, a general form was administered first, and based on the results of the general test, a second easier or more difficult form was administered. This double-test (in total about 50 items) approach was organized on each of the measurement occasions –pretest, posttest and retention test. The pretest was organized before the intervention at the beginning of the school year, the posttest immediately upon the completion of the intervention and retention test half a year after the intervention.

Test items were calibrated by fitting a three-parameter logistic regression model, using the BILOG software (BILOGMG 3.0). A very good overall fit was obtained (Chi-square = 1333.4, df = 1534.0, $p > .90$). For any booklet the number of items with a bad fit ($p < .10$) is below the acceptable level of 10%. Reading ability scores were calculated, using Bayes' Modal (Maximum *a posteriori*) method, resulting in a unidimentional reading comprehension ability scale for all different booklets, which makes direct comparison from pretest to retention test scores possible.

Students' questionnaires on use of reading strategies and reading attitudes

Self-report questionnaires were used to measure students' use of reading strategies and reading attitudes, both at the pretest and the posttest. The *use of reading strategies* items mainly measures how often students use some strategies before, during or after reading. The items cover all the six reading strategies that were taught in the intervention. Exploratory factor analysis yielded one scale of general strategy use. The reading attitude scale has a positive and negative subscale and measures students' attitudes towards reading. All these scales were constructed from an exploratory principal component analysis with Varimax rotation and further tested by confirmatory factor analysis with an acceptable goodness of fit (RMSEA below 0.05). These analyses resulted in a score for the self-reported use of reading strategies and a score reflecting students' reading attitudes. More details

about these questionnaires are given in Chapter 4. Pretest scores of reading attitude and use of reading strategies were included in the analysis as covariates.

Parents questions and teachers' checklists

Data on students' social background were gathered by means of a parents' questionnaire. This background information included parents' socioeconomic status, parents' educational level, income and occupation (Sirin, 2005; Hauser, 1994). Besides, reading resources (e.g., number of books appropriate for children to read at home), and parents' reading habits (number of hours parents spend on reading; types of materials they read at home) were inquired. Information of teachers involved in the study was obtained from school archives. Teachers' information includes age, gender, initial diplomas and highest diploma, years of teaching. School directors were interviewed about these teachers' overall capability in teaching Chinese language.

2.5. Sample

624 5th graders participated in the study. They were from eight primary schools from two different districts in Beijing. These eight schools cover existing three main types of public schools in Beijing: city key schools, district key schools and ordinary schools. Almost all conditions occurred in all three types of schools. All the students share the same mother tongue, although some speak their dialects with some small variation in pronunciation. About 52% of the students are boys. Descriptive statistics with the size of groups, mean and standard deviations for the reading comprehension scores at the three measurement occasions are presented in Table 1.

Table 1. Means and Standard Deviations on the calibrated reading comprehension scores

Tests	Groups	N	Mean	SD
Pre-test	CON	266	-0.26	0.99
	FR	62	-0.20	0.77
	RS	56	-0.18	0.74
	RSA	119	-0.52	0.83
	CA	166	-0.52	0.94
	Total	669	-0.36	0.92
Post-test	CON	255	0.25	0.88
	FR	61	0.13	0.77
	RS	56	0.46	0.70
	RSA	114	0.08	0.81
	CA	160	0.04	0.86
	Total	646	0.17	0.84
Retention-test	CON	249	0.18	1.20
	FR	59	0.22	1.09
	RS	52	0.74	0.97
	RSA	115	0.06	1.07
	CA	149	0.06	1.13
	Total	624	0.18	1.15

3. Statistical Analysis

3.1. A multilevel repeated measures approach

In our study, the school based sampling generated spatially clustered data — students nested in classes and classes in schools. The pretest, posttest and retention test design generated chronologically clustered data, namely, with measurement occasions nested in students. These two main characteristics determine that multilevel repeated measures approach should be used to model the intervention effects. The advantage of repeated measures design in multilevel modeling is that it tolerates missing data on some measurement occasions (Hox, 2002), which is our

case. The effect sizes are calculated using the square root of the total variance in the null model as the denominator.

3.2.The model building process

The detailed model building process is reported in Table 2 in the Appendix. Two dummy time variables were used to indicate posttest and retention test measurement occasions. In the reference model (indicated as Model 0 in Table 2), the coefficient of the constant indicates the average reading comprehension score of students across all schools and classes at the pretest occasion. The coefficients of posttest and retention test indicate the average evolution of all students' reading scores from pretest to post-test, and from pretest to retention test respectively. In the random part of Model 0, the variance at level one (measurement occasion) turned into zero, as a logical consequence of allowing time variables to vary both on student and class level. In this way, individual growth differences and difference in growth between classes could be modeled. To keep the hierarchical structure of the models and for the convenience of interpretation, interaction effects are only kept when at least one main effect is significant.

In the next step, we entered 4 dummies indicating the different conditions, taking the control group (CON) as the reference category. We created interaction terms of these dummies with time dummies (posttest and retention test), to explore the effects of each experimental condition on students' growth in reading comprehension. Then students' social background variables, students' and teachers' gender and students' self-reported use of reading strategies, reading attitudes and metacognitive awareness at the baseline were entered as explanatory variables, in order to adjust for the initial differences among conditions. To test the influence of class average on experimental effects, average class reading comprehension scores at the baseline were also included in the model. At the random part of the multilevel models, random effects of significant variables in the fixed part were checked.

4. Results

In Model 0 in Table 2, the coefficients for posttest and retention test indicate that on the average, across all classes in all schools, students made a growth of 0.5 standard deviation in reading comprehension from pretest to posttest and about half a standard deviation from pretest to retention test. From the results we can infer that, taking all the students into consideration, there is on average no significant evolution from posttest to retention test on students' reading, either because no progress was captured by the retention test or because there was a differential growth rate among students that summed up to zero.

The initial reading level of students from different experimental conditions is indicated in Model 1. With the presence of interaction terms of the condition dummy variables with the dummies indicating the posttest and retention test, the main effect of the condition dummies indicate the initial level of each condition at the pretest occasion. The interaction terms of conditions with the measurement occasion dummies, indicate the experimental effects measured at posttest and retention test respectively. There were no significant differences among the different experimental conditions at the baseline measure. They all started more or less at the same level. By the end of the experiment, CA was significantly different from the control condition ($p = .011$) and RSA and RS were marginally significantly different from control condition ($p < .10$). However, there were no significant differences among the three experimental conditions. At the retention test, all of the three experimental groups were significantly different from the control group. The effect sizes range from 0.44 to 0.56. Again there were no significant differences among the three experimental conditions, although RSA showed somehow less positive results than the other two. The free reading condition, taken as a second control condition, didn't differ significantly from the pure control condition, neither at the posttest nor the retention test occasion.

In the following models, we adjusted the experimental effects for other possible influential variables, to adjust for the effects of initial differences among different conditions. There are three groups of variables we adjusted for: (1) characteristics of students, such as students' gender, their use of reading strategies and their initial attitude towards reading; (2) school and class characteristics, such as schools' prestige and initial average reading level in the classes; (3) family background, parents SES (e.g., parents' educational level, family monthly income), reading books at homes, parents' reading habits and so forth. All these influential variables are entered in the model as time independent covariates, assuming that these influential variables in general had a constant influence on the dependent variable both in time and across conditions.

Model 4 shows the results of final model. First of all, we see that school prestige played a very important role in students' average reading level. Students from "city key schools" on average have a significantly higher reading level than those from ordinary schools. No significant difference was found between "district key schools" and ordinary schools. The class mean for the pretest reading scores proved to have a significant positive effect, so did students' initial reading attitude. Although students' self-reported use of reading strategies appeared to have a small but significant positive influence on the average reading level, when students' positive reading attitude scores were put in the model, the coefficient for reading strategy use became insignificant. Students' initial positive reading attitude appeared to be a stronger predictor for students' reading level, compared to self-reported use of strategies. Girls in general have a higher achievement in reading comprehension than boys. The gender effect on the reading comprehension is quite constant and stable and no interaction was found between gender with experimental conditions or time. This means that boys and girls have a significantly different reading level and this difference persists during the intervention across different experimental conditions. So, although a number of significant confounders were found, their effects didn't differ across time or across

conditions. The detailed results from Model 0 to Model 4 and effect size were reported in Table 2 in Appendix.

The experimental effects in the final model

After have adjusted for the effects of influential factors from students, schools and classes, and family background, we can take a look at the experimental effects.

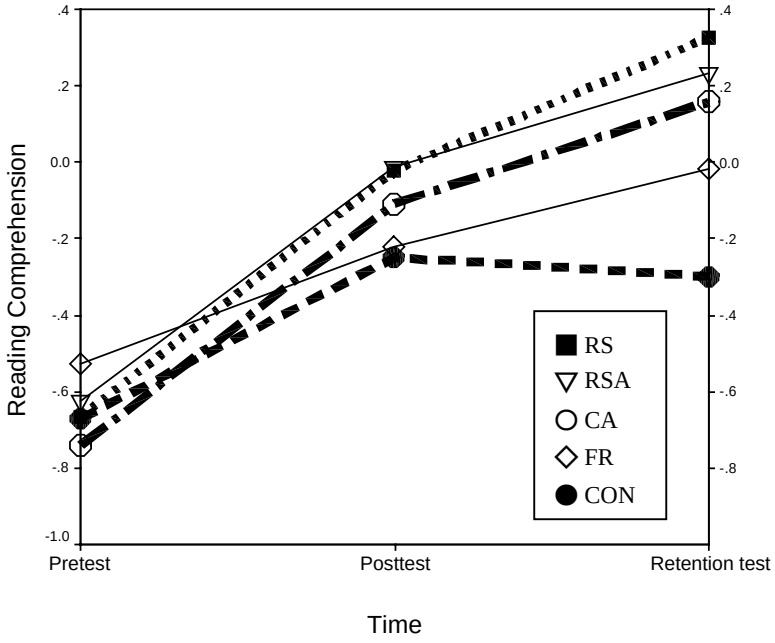


Fig. 1. Result of multilevel repeated measures analysis on the reading comprehension without adjusting for the effect of different initial reading levels across conditions

Compared to the raw model (Model 1), the experimental effects haven't changed much. In the initial model, both CA and RSA were significantly different from CON in the immediate intervention effect. In the final model, the intervention effect of CA persists, while RSA became insignificant, and RS became marginally

significant. Having a retrospective look at the modeling building process, it is after we adjusted for the influence of mothers' average diploma that RSA lost its experimental effect in the posttest. On the other hand, the long-term effects across different experimental conditions were quite stable. There are no significant differences among different experimental conditions or between control and free reading conditions.

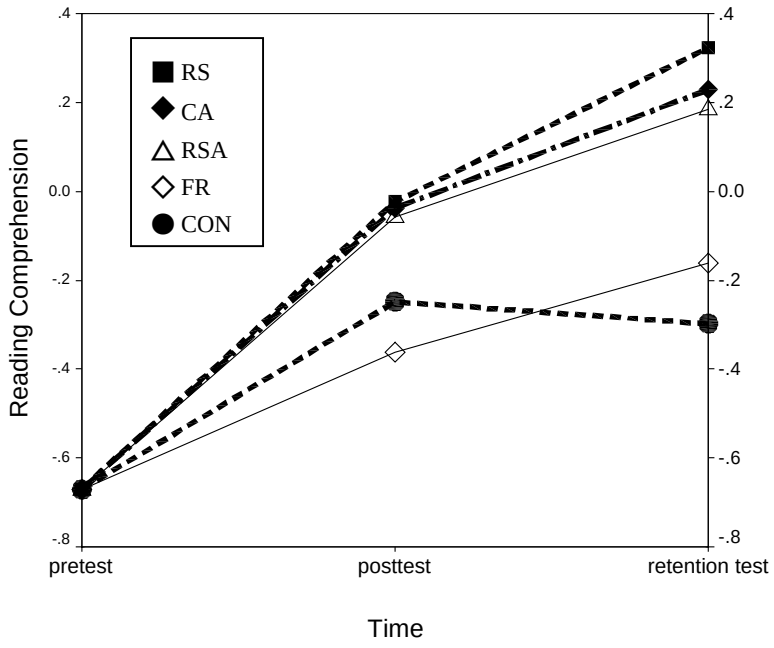


Fig. 2. Result of multilevel repeated measures analysis on the reading comprehension after adjusting for the effect of different initial reading levels across conditions

Figure 1 and 2 are equivalent in displaying the reading comprehension growth of all conditions. Figure 1 shows a graphical display of students' growth in reading comprehension for each of the conditions, adjusted for the effects of background variables and other confounders. Since there were no significant differences among the conditions at the baseline measurement, in Figure 2, the baseline reading scores

were set equal. Therefore, every condition starts at the same level and their evolutions across time are more clearly illustrated. In Figure 2, two groups of lines can be discerned. The three experimental conditions on the one hand, and the two control groups on the other hand. The most striking thing in Figure 2 is that students in the experiment groups kept on growing after the intervention was finished. Remarkable enough, this is also true for the free reading condition. But students in FR condition didn't do significantly better than students in the pure control condition.

5. Discussion

5.1. Adjusting for influence of potential confounders—school prestige and family backgrounds

Our intervention was carried out in natural classroom settings and students' reading results naturally have been influenced by their surrounding environments, including schools and families. The main confounders we have examined include school prestige (key school at district level or city level vs. ordinary schools), class average reading level at the baseline and students' family characteristics, mainly parents' SES indicators. Based on our results, the family SES indicators have a consistent influence on all students' reading growth over time, because no significant interaction effects were found between SES indicators with experimental conditions or with time. Therefore, only the main effects of the family influence confounders were included in the model. The *city key school* did significantly better on reading comprehension growth than the *district key school* and ordinary schools, even after adjusted for family characteristics and students' gender and reading attitudes. It is not surprising of course, since the key schools attracts the best teachers in primary schools and they also select the best students in the neighborhood. A significant effect for initial average class reading level was found across conditions and across time. Including the baseline class average

scores didn't make a big change the experimental results, as is expected, but did magnify the effects of RS and RSA conditions.

Main confounders related to family backgrounds are parents' SES indicators, such as family monthly income, parents' educational level, and also parents' reading habits, and early childhood literacy education. Among all the SES indicators, mothers' educational level has the most dominant influence on students' reading growth. Although fathers' educational level and family income were also significant in the model, they correlate positively with the mothers' educational level and became insignificant when mothers' educational level was included in the model. So was the case for other variables such as parents' own reading habits and early childhood rearing practice related to reading. The influence from mother's educational level reflects the potential educational resources that school children could get at home and it is also "one of the most stable aspects of SES because it is typically established at an early age and tends to remain the same over time" (Sirin, 2005, p.419). The correlation between mother's educational levels with income is 0.63, and father's educational level with income is 0.68. This may due to mothers' traditional role in rearing children in Chinese family, not only in nurturing, but also in education. Since the changing society and growing markets demand more on educational certificates, there is a growing competition among students on their study results. Chinese mothers are pushed to devote more time and energy to their children's study results and give more useful help on children' study difficulties. It strengthens the effect of mothers' educational level on their children's reading achievement.

5.2. Experimental effects in general and the difference between the three experimental groups

The results of multilevel repeated measures analyses showed a significant experimental effect for CA and a marginally significant effect ($p < .10$) for the RS and SA conditions at the posttest; significant long-term effects for all of the

experimental conditions at the retention test ($p < .05$). It indicates that the intervention successfully improved students' reading comprehension growth and this effect could still be traced at least 6 months after the intervention was finished. On the other hand, the fact that only the CA groups showed significant experimental effect at the posttest and all the three groups showed significant robust long-term effects six months later, seemed to suggest that it took some time for the intervention effect to be shown in all experimental conditions. The effect size increased from 0.40 at posttest to 0.60 at retention test. It also yields arguments for the hypothesis that the observed growth is to be attributed to growing metacognitive skills that was fostered by mastering the series of six reading strategies. The finding that no differences were found between the control and the FR condition shows that the experimental gains cannot be simply attributed to the new attractive reading materials or extra reading time that was created by the experimental intervention.

The differences between the three experimental conditions at both posttest and retention tests are small and insignificant. Different from our hypotheses, there was no extra value added from peer tutoring on top of the reading strategy instruction. It is suggested that what mattered most for Chinese students' reading comprehension growth was the teacher facilitated reading strategy instruction, while peer-tutoring practice, either CA or RSA tutoring, didn't show any value for Chinese students. It is also possible that the potential extra benefit from peer-tutoring practice was cancelled out by the inconvenience in organizing peer tutoring in the big-sized classes. Another explanation is that the teacher-led whole class reading activities or individual coaching from teachers in the RS classes have strengthened the benefits of reading strategies instruction on students' reading growth more than the peer-tutoring practice could ever have done, under the specific atmosphere of Chinese classrooms.

5.3. Similarity and difference between the Chinese and Flemish results

The fact that the intervention lasted for a whole school year and was conducted in regular classrooms with regular Chinese teachers adds to the ecological validity and the practical relevance of the study.

Compared to the results of the Flemish project in Flanders (Van Keer, 2004; Van Keer & Verhaeghe, 2003, 2005) in which the experimental design was quite similar, the main difference is that peer-tutoring practice didn't demonstrate extra effects on Chinese students' reading growth, while Flemish students seemed to have benefited from peer-tutoring practice, in addition to the benefit of the reading strategies instruction. The difference could be due to the differences in basic infrastructure in schools between the two countries, e.g., the average surface per student. Chinese classes on average have 40 students per class, while Flemish classes on average have 20 students per class. The only surface that Chinese students could use for reading is the fairly crowded classrooms, while Flemish classes usually could have extra reading rooms or corridors for reading activities during peer tutoring. The restrictions with respect to physical space may have jeopardized the benefits of peer tutoring. Similar differences could be found with respect to teacher-students ratio in the class. The bigger class size in Chinese schools also relatively decreased the amount of help or coaching that teachers could give to dyads, compared to the Flemish situation.

Thirdly, another possible reason for the differences between the Chinese and the Flemish results may be found in different perceptions on the value of peer tutoring for students' reading development and the feasibility of peer-tutoring practice in natural classrooms. In China, a lot of new policies have led to some educational reform in primary school education, but most of them focus on how to improve teachers' teaching and students' "independent reading ability" and "use of multiple methods to read" (The standard Chinese curriculum, 2001). Almost no attention, at

least not written in the official documents, is given to the value of cooperative or collaborative learning among students. This implicit belief in individual learning is deeply in line with the traditional Confucian ultimate goal of growing toward "self-perfection" and outside environment is not viewed as essential as inner search for knowledge (Li, 2003). Principles written in the Chinese Curriculum Standards are often used to evaluate teachers' teaching quality and decide on their further career development. Therefore, it is no surprise that most of teachers in the experimental groups were very enthusiastic about reading strategies instruction, while rather indifferent to peer-tutoring, even if they admitted the potential benefits of it for students' reading development. In practice, the low teacher-student ratio and the limited physical space also tend to harm the feasibility to organize peer tutoring. For them, a teacher-centered classroom organization seemed to be at least equally efficient as peer-tutoring activities.

By pointing to the cross-cultural differences in intervention effects, this study raises new research questions, e.g., what are the mechanisms behind the cross-cultural similarities and dissimilarities in the field of reading instruction? Do the same factors that account for the stability of individual differences within a cultural group, account for the differences across cultures?

6. Conclusions

To conclude, the intervention with two elements – reading strategy instruction and peer tutoring – has demonstrated significant positive experimental short-term effects and follow-up effects on Chinese students, potential confounders being adjusted for. Group comparison suggests that reading strategy instruction has significantly promoted students' reading comprehension growth, while peer-tutoring practice hasn't added extra value to Chinese students' reading. The results might be related to mainstream educational beliefs in China about learning and/or partly due to material conditions.

Appendix: Table 2. Repeated measures multilevel analyses on reading comprehension

	Model 0	Model1	Model 2	Model 3	Model 4	effect size
Fixed effects						
Cons (Intercept)	-0.444	-0.504	-0.423	-.457	-0.671	
SE	*0.096*	*0.116*	*0.101*	*0.085*	*0.077*	
Posttest	0.522	0.433	0.407	0.410	0.422	
SE	*0.043*	*0.060*	*0.064*	*0.064*	*0.066*	
Retention test	0.488	0.210	0.202	0.208	0.372	
SE	*0.059*	*0.090*	*0.096*	*0.096*	*0.097*	
Girl			0.156	0.159	0.158	
SE			*0.054*	*0.053*	*0.053*	
Reading attitude pretest			0.329	0.306	0.306	
SE			*0.040*	*0.039*	*0.039*	
Mother's diploma centered				0.134	0.115	
SE				*0.024*	*0.025*	
Mean class reading pretest					0.393	
SE					*0.127*	
City key school					0.648	
SE					*0.077*	
District key school					0.139	
SE					*0.146*	
Experimental conditions						
Reading of FR pretest		0.299	0.100	-0.049	0.143	
SE		*0.316*	*0.248*	*0.195*	*0.198*	
Reading of RS pretest		0.018	-0.008	-0.050	0.003	
SE		*0.155*	*0.144*	*0.137*	*0.124*	
Reading of RSA pretest		0.003	-0.145	-0.094	0.047	
SE		*0.169*	*0.153*	*0.135*	*0.109*	
Reading of CA pretest		0.108	0.053	-0.080	-0.071	
SE		*0.132*	*0.118*	*0.107*	*0.098*	
Reading of FR posttest		-0.124	-0.099	-0.100	-0.114	
SE		*0.144*	*0.147*	*0.147*	*0.150*	
Reading of RS posttest		0.170	0.204	0.197	0.225	
SE		*0.115*	*0.118*	*0.118*	*0.121*	
Reading RSA posttest		0.190	0.211	0.210	0.190	
SE		*0.104*	*0.108*	*0.110*	*0.114*	
Reading of CA posttest		0.202	0.220	0.219	0.211	0.19
SE		*0.088*	*0.090*	*0.091*	*0.094*	
Reading FR retention test		0.219	0.228	0.221	0.137	
SE		*0.237*	*0.240*	*0.239*	*0.206*	
Reading RS Retention test		0.558	0.608	0.582	0.621	0.56
SE		*0.144*	*0.145*	*0.146*	*0.152*	
Reading RSA retention test		0.423	0.456	0.454	0.484	0.44
SE		*0.142*	*0.150*	*0.153*	*0.154*	
Reading CA retention test		0.522	0.532	0.521	0.527	0.48
SE		*0.114*	*0.116*	*0.118*	*0.123*	

(To be continued)

Table 2. Repeated measures multilevel analyses on reading comprehension

	Model 0	Model1	Model 2	Model 3	Model 4	effect size
Random effects						
Class						
Cons / Cons	-0.153	0.150	0.084	0.043	0.007	
SE	*0.056*	*0.055*	*0.035*	*0.021*	*0.008*	
Posttest / Cons	-0.003	0.010	0.017	0.009	-0.013	
SE	*0.018*	*0.017*	*0.014*	*0.011*	*0.009*	
Posttest / Posttest	0.021	0.020	0.020	0.020	0.021	
SE	*0.011*	*0.011*	*0.011*	*0.012*	*0.012*	
Retention test / Cons	0.053	0.071	0.064	0.045	-0.014	
SE	*0.027*	*0.033*	*0.027*	*0.020*	*0.011*	
Retention test / Posttest	0.017	0.026	0.027	0.027	0.025	
SE	*0.012*	*0.015*	*0.016*	*0.016*	*0.014*	
Retention test / Retention t.	0.047	0.075	0.076	0.075	0.049	
SE	*0.022*	*0.031*	*0.032*	*0.032*	*0.023*	
Student						
Cons / Cons	0.684	0.685	0.621	0.604	0.604	
SE	*0.038*	*0.038*	*0.037*	*0.036*	*0.036*	
Posttest / Cons	-0.306	-0.306	-0.313	-0.310	-0.309	
SE	*0.025*	*0.025*	*0.026*	*0.026*	*0.026*	
Posttest / Posttest	0.455	0.450	0.455	0.456	0.455	
SE	*0.026*	*0.025*	*0.027*	*0.027*	*0.027*	
Retention test / Cons	-0.176	-0.177	-0.178	-0.185	-0.184	
SE	*0.028*	*0.027*	*0.028*	*0.027*	*0.027*	
Retention test / Posttest	0.287	0.273	0.265	0.265	0.265	
SE	*0.025*	*0.024*	*0.025*	*0.025*	*0.025*	
Retention test / Retention test	0.646	0.609	0.614	0.615	0.614	
SE	*0.037*	*0.035*	*0.037*	*0.038*	*0.037*	
Occasion						
Cons / Cons	0	0	0	0	0	
SE	*0*	*0*	*0*	*0*	*0*	
-2log(lh)	4121.27	4084.94	3535.66	3509.24	3408.07	
N (students)	1942	1942	1718	1718	1718	

References

Barron, A., & Foot, H. (1991). Peer tutoring and tutor training. *Educational Research,* 33, 174-185.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Butler, S. R. (1991). Reading program—Remedial, integrated and innovative. *Annals of Dyslexia,* 41, 119-127.

Clay, M. M. (1985). *The early detection of reading difficulties.* Portsmouth, New Hampshire: Heinemann.

De Corte, E., Verschaffel, L., Van de Ven, A. (2001). Improving text comprehension strategies in upper primary school children: a design experiment. *British Journal of Educational Psychology, 74, 4, 531-560.*

Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment. *Journal of Educational Psychology,* 84, 331-339.

Garner, R., & Alexander, P. A. (1989). Metacognition: answered and unanswered questions. *Educational Psychologist,* 24, 2, 143-158.

Hauser, R. M. (1994). Measuring socioeconomic status in studies of child development. *Child Development,* 65, 6, 1541-1545.

Hock, M. F., Kim A. Pulvers, K. A., Deshlez, D. D., & Schumaker, J. B. (2001). The effects of an after-school tutoring program on the academic performance of at risk students and students with LD. *Remedial and Special Education,* 22(3), 172-186.

Invernizzi, M., Juel, C., & Rosemary, C.A. (1997). A community volunteer tutorial that works. *The Reading Teacher,* 50, 304-311.

Kamps, D. M, Dugan, E., Potucek, J., Collins, A. (1999). Effects of cross-age peer tutoring networks among students with autism and general education students. *Journal of Behavioral Education,* 9, 2, 97-115.

Klingner, J. K., Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities for students with learning disabilities who use English as a second language. *The Elementary School Journal,* 96, 3, 275-293.

Li, J. (2003). U.S. and Chinese cultural beliefs about learning. *Journal of Educational Psychology,* 95, 2, 258-267.

Mellanby, A. R., Newcombe, R. G., Rees, J. & Tripp, J. H. (2001). A comparative study of peer-led and adult-led school sex education. *Health Education Research,* 16, 4, 481-49.

OECD (2003). *Learners for life—student approaches to learning results from PISA 2000*. OECD.

Palincsar, A. S., David, Y. M., Winn, J. A., & Stevens, D. D. (1991). Examining the context of strategy instruction. *Remedial and Special Instruction,* 12, 43-53.

Paris, S. G., Byrnes, J. P., Paris, A. H. (2001). Constructing theories, identities, and actions of self-regulated learners. In B. J., Zimmerman & D. H., Schunk (Eds). *Self-regulated learning and academic achievement: Theoretical perspectives*, 253-287.

Paris, S. G., Lipson, M. Y., & Wixon, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology,* 8, 293-316.

Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: transactional instruction of reading comprehension strategies, *The Elementary School Journal,* 92, 5, 513-555.

Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J., & Kurita, J. (1989). Strategies that improve children's memory and comprehension of text. *The Elementary School Journal*, 90, 3-32.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., Miller, T. R. (2003). Peer assisted learning intervention with elementary school students: a meta-analytic review. *Journal of Educational Psychology,* 95, 2, 240-257.

Schrader, B. & Valus, A. (1990). Disabled learners as able teachers: a cross-age tutoring project. *Academic Therapy*, 25, 589-597.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Review of Educational Research,* 75, 3, 417-453.

Steinberg, L., Dornbusch, S. M., & Brown, B. B. (1992). Ethnic differences in adolescent achievement: An ecological perspective. *American Psychologist,* 47, 723–729.

Topping, K. J. & Ehly, S. (1998). Introduction to peer-assisted learning. In K. J. Topping & S. Ehly (Eds.). *Peer-assisted learning* (pp. 1-23). Mahwah, New Jersey: Erlbaum.

Topping, K. J. (1998). Paired learning in literacy. In K.J. Topping & S. Ehly (Eds.), *Peer-assisted learning* (pp. 87-104). Mahwah, New Jersey: Erlbaum.

Van Keer, H. (2004). Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring. *British Journal of Educational Psychology*, 74, 2, 37-70.

Van Keer, H., & Verhaeghe, J. P. (2003). Effects of explicit reading strategies instruction and peer tutoring in second and fifth grades. In E. De Corte, L. Verschafeel, N. Entwistle, & J. Van Merrienboer (Eds.), *Powerful learning environment: Unraveling basic components and dimensions* (pp. 211-228). Oxford: Elsevier.

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. *The Journal of Experimental Education,* 73(4), 291-329.

Wentzel, K. R. (1999). Social-motivational processes and interpersonal relationships: Implications for understanding motivation at school. *Journal of Educational Psychology*, 91, 76–97.

Winter, S. (1987). Parents and pupils as remedial tutors, *Bulletin of the Hong Kong Psychological Society*, 18, 15-32.

Chapter 4

The Effects of Reading Strategy Instruction and Peer Tutoring on 2nd Graders' Reading Comprehension: A Multilevel Repeated Measures Approach*

Abstract

This study examined the effects of explicit instruction in reading strategies (RS) and two types of peer tutoring, cross-age (CA) and reciprocal same-age tutoring (RSA) on Chinese 2nd grade students' reading comprehension development. Having adjusted for the effects of potential confounders related to students, schools and family background, the results indicated: (1) significant experimental effects on students' reading comprehension growth in CA and RSA groups, and marginally significant effect in RS groups; (2) a long-term experimental effect only for the CA group. Although no significant extra effects were found for peer tutoring practice, CA in general demonstrated a bigger strength in improving 2nd graders' reading comprehension, compared to RSA and RS groups.

Key Words: reading strategy multilevel analysis cross-age tutoring reciprocal same-age tutoring repeated measures

* This chapter is based on an article with the same title submitted to the *Journal of Educational Psychology* for publication.

1. Introduction

1.1. Peer-tutoring

Peer tutoring (PT) is a specific form of student interaction based instruction in which one student acts as a tutor for another student the tutee (Topping & Ehly, 1998; Topping, 1998). In several ways, peer tutoring is quite different from occasional helping behaviors between students. First, in peer tutoring distinct roles (tutor vs. tutee) are assumed by the students. PT is also structurally embedded in the curriculum, focusing on specific topics or skills (e.g., in language or in mathematics) within the framework of a broader program, and with regular sessions (e.g., once or twice a week). To ensure the quality of peer interaction in the dyads, most often tutors are trained in advance and receive some social skills training before engaging in the actual peer tutoring. The interaction within the dyads is further scaffolded and structured by means of specific aids, instructions or assignments, that previously have been introduced to both tutors and tutees. So, students do not enter a peer tutoring session unprepared. They have a clear set of tasks to perform, and are well-prepared for it, both with respect to the content and with respect to the social aspects of their tasks.

Based on the age difference between tutor and tutee, two categories of PT can be distinguished: "same-age tutoring" (SA) and "cross-age tutoring" (CA). "Reciprocal same-age tutoring" (RSA) is a special form of SA in which the tutor/tutee role is switched systematically (Fantuzzo *et al*., 1992).

In the past decades, positive effects have been reported for peer tutoring in a wide variety of learning domains, such as mathematics, reading and writing (Cohen *et al*., 1982; Mathes & Fuchs, 1994; Mathes *et al*., 2001; Mellanby *et al*., 2001), social adjustment (Kamps *et al*, 1999), physical education, sex education (Mellanby, 2001), and so forth. Peer tutoring was shown to have positive effects on students' socialization and cognitive achievement in classroom settings (Steinberg *et al*., 1992; Wentzel, 1999) and it was shown to have better results than

unstructured cooperative learning or individual learning (Rohrbeck *et al.*, 2003). The particular combination of structure and student autonomy and interaction is probably one of its most powerful features.

One of the important reasons to advocate peer tutoring is that it creates opportunities for metacognitive exchange and modeling between students (Almasi, 1996; Fuchs *et al.*, 1997; Palincsar *et al.*, 1991). This probably makes peer tutoring into a particularly appropriate learning environment for the acquisition or further practice of problem solving strategies and self-regulated learning skills. As is suggested by researchers, children's failure to use appropriate strategies in cognitive tasks may be due to a lack of monitoring and regulation in their learning process. To overcome this, more individually tailored external feedback is needed (Scardamalia & Bereiter, 1986). Much more than teachers could do during whole class activities, peer tutoring creates excellent opportunities to provide such individual feedback to students.

One of the issues addressed in the present study is whether the age of the tutor or tutor-tutee age difference plays a role. Therefore the effects of peer tutoring in CA dyads are compared with the effects of peer tutoring in RSA dyads. We more particularly focus on the effects on second graders growth in reading comprehension. In the CA condition, every second grader is tutored by a fifth grade tutor once a week during a whole school year. In the RSA condition every second grader is tutored by another classmate with whom they exchange the tutor/tutee roles every week. In both conditions, the PT dyads' task is to practice a series of previously taught reading strategies. Obviously, fifth grade students are older and might therefore be expected to have better social skills. They also might be expected to have a deeper understanding of the strategies to be practiced. And, since the texts that are read, are at the level of the second grade tutee, they may be expected to have no difficulties with it themselves. So, from different perspectives, fifth grade tutors may be expected to be better tutors for second grade tutees than second grade students ever could be for each other. And thus, for second grade

students better results should be expected from CA tutoring than from RSA tutoring.

On the other hand, RSA offers second grade students the opportunity to take the tutor role themselves. As we have explained in Chapter 3, there are a number of reasons to believe that it is *being a tutor* that provides the best opportunities for practicing reading strategies and the metacognitive skills involved, rather than being a tutee. From this point of view, RSA would offer better opportunities for second graders to acquire or deepen their metacognitive skills and improve their reading comprehension.

But the question then of course is whether the reasoning that could be made for fifth graders, also holds for second graders. Analysis of the interactions within the peer tutoring dyads (see Chapter 2), not only indicated that CA dyads provide more opportunities for practicing meta-cognitive skills by the tutor, it also indicated that in CA dyads more corrective feedback was given to the tutee than in RSA dyads and that in CA dyads this feedback referred significantly more to one or more of the previously taught reading strategies. If indeed the teaching and practicing of reading strategies would lead to an improvement of the students' growth in reading comprehension, the previously presented findings of the interaction analysis (see Chapter 2) would suggest that second grade students would benefit more from CA tutoring than from RSA tutoring.

1.2. Reading strategies

Reading strategies can be defined as "operations over and above the process of reading which produce efficient, insightful comprehension of the content of reading texts" (Pressley *et al*., 1989). Based on a similar study with Flemish primary school pupils (Van Keer, 2004; Van Keer & Verhaeghe, 2005), in the present study students were given direct instruction in six reading strategies: (a) activating prior knowledge, (b) finding the meaning of unfamiliar words and expressions, (c) making predictions, (d) finding main ideas of texts, (e) monitoring

and regulating comprehension of the texts, and (f) distinguishing different types of texts. The selection of these reading strategies is founded on other reading research (De Corte *et al.*, 2001; Klinger & Vaughn, 1996; Pressley *et al.*, 1992; Palincsar & Brown, 1984). Taking this set of reading strategies as a whole, a number of characteristics can be distinguished. First, the series of strategies shows some chronological ordering, showing what can be done before reading (e.g., activating prior knowledge and talking about what have been known about the topic), during reading (e.g., understanding words, expressions, paragraphs and the whole text) and after reading (e.g., classifying what types of texts were read). Second, there is some ordering in complexity, starting from simple strategies, such as decoding and finding the meaning of new words, to more complex ones such as summarizing and reasoning. Third, the set contains both cognitive strategies (such as making predictions or summarizing main ideas of a paragraph or a text) and metacognitive strategies (e.g., monitoring and regulating one's own understanding).

A lot of studies have been devoted to both the importance of and the possibility to give instruction in reading strategies, and to the merits of peer tutoring as well. However, research combining these two lines of research is still relatively sparse (Van Keer, 2004; Van Keer & Verhaeghe, 2005), although the mutual benefits of such a combination is quite obvious. As has been argued above, peer tutoring probably provides a most appropriate learning environment for practicing the kind of skills that are taught in reading strategies instruction. Providing students with good opportunities to practice the reading strategies they have been taught is of course of the utmost importance when one wishes to investigate the relevance of reading strategies instruction. Without such opportunities no conclusions could ever be drawn on the usefulness of reading strategies instruction. So, when investigating the usefulness and relevance of reading strategies instruction, peer tutoring appears to be a logical choice.

For those primarily interested in the possibilities of peer tutoring, it is quite useful to take reading strategies as the content to focus on. Research in this domain is well

developed and a number of programs have been designed with a wide variety of ways to scaffold students' learning. This meets one of the key characteristics of peer tutoring; as a structured learning environment it provides scaffolding for both the content related and the social aspects of the learning activities within the dyads. The materials and procedures that provide scaffolding to students for practicing the reading strategies also help the tutors to structure their interaction with their tutees.

1.3. Fostering metacognition in reading

A number of the reading strategies in our intervention program can be considered as metacognitive strategies. Metacognition is *"knowing about knowing"* or *"thinking about thinking"*. It includes planning, monitoring, regulating and reflecting on cognitive processes (Garner & Alexander, 1989). Among a number of other things, metacognition is about deciding which, when and how strategies should be used (Paris, Lipson, & Wixon, 1983).

Metacognition helps kids to become independent readers who can monitor and regulate their own reading progress, independent from a teacher who points out and clarifies difficult parts in reading materials. It follows that when the metacognitive strategies included in the program really are adapted and applied by the students in such a way that metacognitive processes are fostered, a continuous growth in reading comprehension can be expected, even after the intervention has finished. Therefore, in the present study long-term effects on a retention test, several months after the end of the intervention, are assessed. Such long-term effects can be considered as a good (albeit indirect) indicator that indeed improvement in metacognitive processes forms the basis of improved growth in reading comprehension.

In peer tutoring, the tutors' main task is to monitor and give feedback to the tutees' reading. They also ask questions to check if the tutee understands or masters the reading strategies well and give corrections or explanations when necessary. In doing so, they scaffold and model the use of cognitive and metacognitive strategies

to their tutees. Whether this would really enhance their own and their tutees' capabilities to monitor and regulate their own reading behavior would of course depend on the quality of the tutoring. A very important element in this is related to the degree to which the tutor is successful in surpassing the boundaries implied by the scaffolding materials. In our program, "generic" assignment cards with general, text-independent questions were used as a means to scaffold the students' learning process (cfr. infra). When using these cards, the most obvious thing for a tutor to do, would be to read out the questions from the card, listening to the tutee's answer and then discuss it. However useful this may be, it is no more than a first step. As long as the students stick in a rather mechanical way to the questions printed on the assignment cards, it is doubtful whether what they do reflects any metacognitive activity at all (even when the questions on the card are supposed to reflect a meta-cognitive strategy). An important step forward is set, when the tutor – inspired as it may be by the questions on the card – starts to phrase questions of his own, in a more close response to the specific difficulties the tutee experiences or to the answers the tutee gives during the discussion. Only from this moment on, there is real modeling and metacognitive exchange. Looking at this from a developmental perspective, there are probably more chances for this to occur with fifth grade tutors than with second grade tutors.

Various tutoring programs in the late 20th century (e.g., Clay, 1985; Butler, 1991; Invernizzi, Juel, & Rosemary, 1997; Winter, 1987) used teachers, paraprofessionals or colleague students to help primary school children with reading problems. These programs mainly stressed the benefits of individual tutoring as a way to provide remedial teaching. There are also other projects, using individual students with learning difficulties as tutors to help younger tutees; with a main focus on the benefits individual tutors can gain from the tutoring experience (e.g., Schrader & Valus, 1990). In our study, the intervention focused on complete regular classes in schools, providing all students with opportunities to practice cognitive and metacognitive reading strategies.

1.4. Chinese classrooms

Most research on reading strategies or peer tutoring is done in western countries, involving English speaking students and teachers using English as the language of instruction, or some other western language. The present study is based on a Flemish study (Van Keer, 2004; Van Keer & Verhaeghe, 2003, 2005) involving regular Flemish primary school classrooms with Dutch speaking pupils and Dutch as the language of instruction. In that study, a significant positive effect was found for the second grade students who – in combination with reading strategy instruction by the teachers had practiced the reading strategy in CA dyads. For students in the other conditions, no effects on reading comprehension were found (Van Keer & Verhaeghe, 2003, 2005).

For the purpose of the present study the intervention program used in the Flemish study has been translated and adapted for use in regular Chinese primary school classrooms in Mainland China. The aim of the program was to improve Chinese pupils' reading comprehension when reading Chinese texts.

When implementing interventions in different cultures, especially in the field of reading, both cultural and linguistic issues have to be tackled. Languages such as English or Dutch belong to the Indo-European language system with a phoneme-based writing system, which is totally different from a morpheme-based writing system as is used in Chinese. There are also other differences. For example, in Chinese, complex concepts are often formed with simple, commonly used morphemes, which is not the case in Western languages. These differences might affect the reading processes, particularly when readers are confronted with unfamiliar words or expressions. It has to be verified whether reading strategies in English or Dutch would also be useful in improving reading comprehension in Chinese.

There is also a cultural issue. Though the Chinese educational system is changing fast, influences from traditional teaching styles still persist. Introducing such practices as teachers thinking aloud in the classroom to model the use of reading strategies, or having students choose their own reading texts, or having students tutor other students (which are all elements in the intervention program) in Chinese classrooms is probably to be considered as a much bigger endeavor than the introduction of similar practices in European or American primary school classrooms. This is not only true for teachers, but also for students who have to get acquainted to this different way of working as well (and get back to the more traditional way when reading class is over). A contextual issue that should not be underestimated relates to class size in Chinese primary schools (which is much larger than in most European or American schools) and the fact that – unlike most Flemish and other western schools – most Chinese primary schools do not have a school library or a local public library from which reading books can be borrowed for use in reading classes. The adaptation of the Flemish program involved ways to cope with this quite different educational context. One element in this was the provision of attractive Chinese children's reading books to the experimental classrooms.

Also the Chinese school culture is considerably different from the European situation. Officially, all schools are equal and there are no official school rankings. But unofficially, in big cities, three types of public schools are distinguished: *city level key schools* (or model schools at the city level), *district level key schools* and *ordinary* schools. These schools are classified unofficially by their prestige, students' performance in various district-wide or city-wide exams, and teachers' qualifications. Those differences among the schools have grown historically, probably originally building on partly naturally occurring differences in school populations, but later on strengthened by a growing practice of selective admission of students based on intellectual and financial criteria. Parents' financial contributions allowed these *key schools* to get better equipped. It also allowed them to raise the teachers' salaries above what neighboring schools could afford, which

in turn enabled these key schools to become more selective in hiring teachers. In this way a spiral of growing social segregation was set off in Chinese education in the big cities. For example, Chinese universities (just like many other *'working units'*) used to organize primary and secondary schools for the children of their personnel. These university-attached schools generally performed better in centrally organized exams than other schools in the neighborhood. This difference in outcomes was (and is) attributed to differences in school quality, and attracted more and more parents – also from outside the neighborhood - who tried to have their children enrolled in those schools. The growing demand for enrollment led these "key schools" to adopt the practice of organizing entrance exams and asking school fees. Children from the own working unit or neighborhood could not legally be refused, but for children from outside a policy a selective admission got installed. As a consequence, key schools differ from ordinary schools in their financial situation, their basic infrastructure, teachers' bonus on top of their official salary, teachers' qualifications and student performance. Because of these differences it cannot be excluded that *key schools* provide a better environment for the implementation of the intervention. Therefore, two key schools were included in the study (a city level and a district level key school) and close attention was given to the social mix of the students participating in the study and the potentially confounding effects of their social backgrounds.

1.5. Research questions and hypotheses

The research questions relate to the effects of the intervention, having adjusted for the effects of potential confounders. Both short-term effects (posttest, right after the conclusion of the intervention) and long-term effects (retention test, six months later) are investigated. Based on the previous findings and theoretical assumptions discussed above, the following hypotheses were formulated:

(1) Reading strategy instruction will significantly promote 2nd grade students' reading comprehension development. At the posttest, all the

experimental groups will show significantly more growth in reading comprehension than the control groups.

(2) In addition to the positive effects of reading strategy instruction on 2nd graders' reading comprehension growth, there will be a significant extra effect from practicing the reading strategies in peer tutoring dyads. At the posttest, both CA and RSA 2nd grade students will demonstrate significantly more growth in reading comprehension than RS 2nd grade students.

(3) The additional growth in reading comprehension will be significantly larger for 2nd graders in the CA condition than for 2nd graders in the RSA condition.

(4) Students in the experimental groups are expected to keep the gains they made during the intervention for at least six months. At the retention test, all three experimental conditions will still show a significantly bigger growth than the control groups, 2nd graders who have been involved in peer tutoring dyads will show significantly more growth than those who practiced the reading strategies individually, and 2nd graders who have had a 5th grade tutor will show more growth than 2nd graders who have practiced the reading strategies in RSA dyads.

2. Design of the study

The study can be characterized as a quasi-experimental study, carried out in natural classroom settings. About 900 second grade students from 24 classes in 8 schools in Beijing (and a comparable amount of fifth grade students from the same schools) participated.

2.1. Conditions

To test the effects of the two intervention elements, four conditions were set up: reading strategy instruction plus practice in cross-age tutoring dyads (CA), reading

strategy instruction plus practice in reciprocal same-age dyads (RSA), reading strategy instruction plus individual practice in whole classroom setting (RS), and control group (CON). Students in the RS groups received teaching on the six reading strategies and applied them individually in their reading activities, under the teachers' guidance. Students' in CA and RSA conditions practiced reading strategies in peer tutoring dyads, after receiving the same instruction on reading strategies. In CA, a student from 5th grade acted as a tutor for a student from 2nd grade (tutee). Dyads in CA were matched by the relative reading level. For example, a strong reader from the 5th grade tutored a strong reader from the 2nd grade. The same was true for the average and weak readers in both grades. In RSA, dyads were formed with students of similar reading levels and the role of tutor and tutee alternated weekly. Every week, about 60 minutes of extracurricular time was allocated to the intervention and it lasted for a whole school year. Control groups didn't get any intervention involving reading strategies instruction or peer tutoring. However, they spent a comparable amount of time on other Chinese language related extracurricular activities, such as children's literature appreciation, little plays or doing Chinese homework under the supervision of their teachers. All students were kept in school during these hours before they were fetched by their parents after work.

2.2. The Intervention

The intervention was scheduled once a week during extracurricular hours (from 3.30 pm to 5:00 pm) for a whole school year. Reading strategies were taught in all the experimental groups. For the strategy instruction, a "sandwich" pattern was applied. Each strategy was first introduced to the students during one lesson period. Teachers first explained the declarative (the name and main concept), procedural (main steps) and conditional knowledge (when and where to use) of the strategy using a "thinking aloud" method. Then they introduced the assignment card, which synthesized the main steps and questions that helped to apply the strategy. These questions include "W*hat are the main characters in the story?*" or "*How well have*

I understood the paragraph I just read?" or *"What will happen in the story?"* and so on. Then a guided practice on the strategy with selected texts was organized. Students were asked to answer the questions on the assignment card, based on their understanding of the texts. Feedback to their answers was given in the whole class. After the introduction lesson, two sessions of independent practice of the strategy were scheduled, using the assignment card to scaffold students' reading. Students could choose freely the books that were circulating among the experimental groups to practice the reading strategy. After the introduction and individual practice sessions, further practice either individually (as in the RS condition) or in dyads (as in the CA or RSA conditions) was organized. After three reading strategies had been practiced, for a number of sessions students were asked to practice all three of them flexibly, with the help of a synthesizing assignment card including the three strategies. After all the six strategies were introduced, other integration sessions were organized.

Because one of the key factors in the effect of peer tutoring programs is tutor training (Hock *et al*., 2001; Klingner & Vaughn, 1996; Barron & Foot, 1991), in the RSA and CA groups, we organized a number of sessions to train the tutors. These tutoring sessions include: "how to make him/her relax during peer-tutoring sessions", "how to listen properly to my tutee", "how to deal with his or her mistakes and difficulties", "how to deal with my own mistakes", "how to plan and keep records of our reading sessions", etc. Most of these sessions involved role-taking plays and group discussion.

While the intervention was organized, students in the control groups were in various language related extracurricular activities,

2.3. Treatment fidelity

Since it was a study replicating a similar program in the Dutch-speaking region of Belgium, all the materials were translated into Chinese and adapted to the local educational situation in Beijing. These included the teachers' manuals, classroom

posters highlighting the key concepts taught in reading strategies and tutor training, students' assignment books (with assignment cards and selected texts) and the assessment instruments. In the school year proceeding the main study, a pilot study was carried out in 3 classes from 2 schools. Half a month before the pretest data collection, all teachers in the experiment received a training related to the experiment. The training lasted for 3 full working days for a total of 24 hours. The trainings not only clarified the fundamental goals of the intervention, but also detailed teaching contents and demonstrations of some important lessons on reading strategy instruction and tutor training. Teachers got their teaching manuals with detailed written instructions on how to teach each of the lessons. To avoid contamination of different conditions, teachers were only provided with the necessary training and manuals matching their conditions.

During the intervention all through the year, the project coordinator and her assistants went to observe the experimental classes every week. Half an hour discussion with all the experimental teachers was organized prior to every introductory lesson either on tutor training or reading strategy, to highlight the main contents. Special attention was given to the potential problems that might arise in the classroom. Teachers were encouraged to talk about their experiences, difficulties or puzzles during weekly meetings.

The fidelity of the treatment was mainly evaluated through two kinds of means: direct observation and teachers' self-report on the checklists of each new introduction lesson. These checklists were pre-designed outlines of the main procedures that should be followed by the teachers during their instruction. At the same time, observations with video camera recording were made by the coordinator and the two research assistants.

Fidelity results indicate that, teachers' self-report checklist complied more than 90% of the content written on the checklist, with some variations according to specific teaching styles of some teachers. For example, during some reading strategy introduction lesson, three reading articles were recommended in the

checklist to be discussed in the whole class. Some teachers sometimes organized two thorough discussions in the classroom while keeping the last text for students' individual exercise, taking into consideration students' independent study ability.

2.4. Instruments

Reading comprehension tests

To measure reading comprehension ability, a test previously used in a similar Flemish study was translated into Chinese. Like most reading comprehension tests, it provides a number of short stories, each followed by a varying number of multiple-choice questions to test students' understanding of the text.

With the reading comprehension tests for 2nd grade students, we intended to measure several aspects of students' reading comprehension skills, e.g., finding out the meaning of specific words, phrases or sentences from the contexts; understanding the main ideas of paragraphs or the whole texts; recognizing the structure or genre; appreciating the mood or emotion of the characters in the stories, or inferring authors' purpose to write the texts, etc. The content of the reading comprehension tests is generally in agreement with the definition and content of most reading comprehension tests (e.g., OECD, 2000).

Three tests were administrated to measure students' reading growth over time: a pretest before the intervention at the beginning of Grade 2, a posttest immediately upon the completion of the intervention, and retention test half a year after the intervention.

In view of a repeated measures analysis, the three reading comprehension tests were calibrated following Item Response Theory. To assess the necessary overlap of the test items, additional data-collection was organized on several occasions, in which students were asked to take two consecutive tests. A three-parameter logistic regression model was applied to calibrate these three tests, using the BILOG software (BILOGMG 3.0). A very good overall fit was obtained (Chi-square =

620.0, SE = 669.0, p = .91). The number of items with a bad fit (p < .10) was below 10%. Theta-scores were calculated, using Bayes' Modal (Maximum *a posteriori*) method, resulting in unidimentional reading comprehension scale, with a mean of 0.90 and SD of 0.91. This unidimentional scale made it possible to estimate students' reading comprehension growth over one and a half year. A summary of descriptive statistics for the three measurement occasions is presented in Table 1.

Table 1. Descriptive statistics for each group on calibrated reading comprehension scores

Conditions		Pretest	Posttest	Retention test
Control	**Mean**	0.19	1.19	1.61
	SD	0.86	0.58	0.59
	N	330	306	289
RS	**Mean**	-0.09	1.15	1.40
	SD	0.96	0.56	0.65
	N	142	132	132
RSA	**Mean**	0.08	1.33	1.52
	SD	0.81	0.45	0.52
	N	211	191	193
CA	**Mean**	-0.15	1.15	1.51
	SD	0.95	0.50	0.55
	N	230	204	176
Total	**Mean**	0.037	1.21	1.53
	SD	0.90	0.53	0.58
	N	913	833	790

Students' questionnaire

Self-report questionnaires were used to measure students' use of reading strategies, metacognitive awareness related to reading and their reading attitude, both at the pretest and the posttest. Principle component analysis with Varimax rotation yielded one unidimentional scale for *use of reading strategies* and two different

subscales (with only positive items versus negative items) for *metacognitive awareness* and for *reading attitude*. Pretest scores (sum of the items) on each subscale were included as covariates to adjust for the effects of initial differences on these variables on students' reading comprehension growth.

Parents' questionnaire

Data on students' family background, including parents' educational level, income and occupation (Sirin, 2005; Hauser, 1994) were gathered through parents' questionnaire with enclosed letters, and they were included into the data analysis. This questionnaire also yielded data on parents reading behavior, parenting practices with respect to reading and children's reading behavior at home.

2.5. Sample

Eight primary schools were drawn from two different districts in Beijing. About 900 students from 24 classes were involved in the study. These 8 schools cover three main types of public schools in Beijing: city level key schools, district level key school and 6 randomly chosen ordinary schools. (More detailed information about the differences among these schools is given in Chapter 1). Most of the students speak standard Chinese. Only a small minority of students speak some dialects at home, involving some variation in pronunciation. Among the students, about 54% are boys. About 70% of the students were from working class families with a monthly income of 3,000 to 4,000 RMB (375 to 500 USD) or lower that.

2.6. Analysis

Given the nested structure of the data (students belong to classes, which are part of schools), a multilevel approach was opted for. Because of limitations of the data set, at this point only two levels were distinguished: students and classes. To investigate students' growth over time, and taking the availability of data on three measurement occasions into account, a repeated measures design was adopted. This was done by creating a level "time" (or "measurement occasion") within

students. So we have a three level structure with time nested in students and students nested in classes.

To model growth curves, time was not only used as a level indicator, but also entered as an explanatory variable in the model. Therefore two dummies were created, the first one indicating the growth from pretest to posttest, the second one indicating the growth from pretest to retention test. To investigate differences in growth between students and between classes, these dummies were allowed to vary randomly at the student and the classroom level. In doing so, the variance at the lowest level (time) turns structurally to zero. Differential growth over time related to other variables (e.g., gender or experimental conditions) was investigated by creating interaction terms with both time dummies. Time independent explanatory variables were entered in the model as well.

The total variance that can be calculated from the random part in the "null-model" is used as the basis for estimating the effect sizes of the experimental effects. This total variance is 0.65; the standard deviation is 0.81. To investigate how different effects relate to each other, models were built up gradually. The reported effect sizes for the experimental effects pertain to the final model (Model 4).

To avoid loss of data due to missing values on some of the covariates, we use dummies to indicate the missing cases. The coefficients for these dummies indicate whether these cases are significantly different from the cases with a value equal to the mean for that covariate (or from the reference category in the case of non-metric covariates). To facilitate interpretation of the coefficients and to avoid inflating the variance in case of interaction with other explanatory variables, all continuous covariates were centered on their overall means.

As explained in a previous section, four treatment conditions are distinguished: three experimental conditions and a control condition. For the treatment variable, the control condition serves as the reference category against which each of the three experimental conditions is compared. In addition, in order to verify the hypotheses, several contrast effects were specified, allowing for estimating (1) the

joint effect of the three experimental conditions in comparison with the control condition, (2) the joint effect of the two peer tutoring conditions (CA & RSA) in comparison with the reading-strategies-without-peer-tutoring condition (RS), (3) the difference between the two peer tutoring conditions (CA vs. RSA).

3. Results

Five models are reported in Table 2 in the Appendix. Model 0 is a model with only the constant and two time variables (posttest and retention test). The coefficient for the constant in Model 0 indicates the average reading comprehension scores at the pretest; it is not significantly different from 0. At the posttest the average is about 1.2 units (or 1.5 standard score) and at the retention test 1.5 units (or 1.85 standard score). Both coefficients are significantly different from zero. They indicate a significant overall growth in reading comprehension from pretest to retention test.

Since the main interest of the analysis is the experimental effects, dummies indicating different conditions and their interaction with time are entered first in the model, as is reported in the Model 1. These dummies are kept in the models even when their coefficients are not significant, whereas for other variables only significant effects are kept in the model and reported in Table 2.

Model 2, 3 and 4 report the results after having adjusted for the effects of students' gender, mothers' diploma, students' reading attitudes and metacognitive awareness, and school prestige. Model 2 shows significant short-term effects for the CA and RSA conditions, a marginally significant short effect ($p = .07$) for the RS condition, and a significant long-term effect for the CA condition. These effects are quite stable. Adjusting for the effects of other confounders in Models 3 and 4 does not change the coefficients for the experimental effects that much. The effect sizes are rather moderate. In Model 4, the largest effects were found for the CA condition, with an effect size of 0.44 at the posttest and 0.47 at the retention test.

In Table 2 the separate short-term experimental effects we just mentioned, are indicated by the coefficients for the interaction terms *CA*Posttest, RSA*Posttest* and *RS*Posttest*. Contrasting these three experimental effects jointly against the control condition reveals a very significant short term joint effect for the three conditions in which the students received reading strategies instruction ($\Delta\beta = 2.271$, $\chi^2_{(1)} = 30.873$, $p < .001$).

As can be seen from the coefficients for the *CA*retention test, RSA*retention test* and *RS*retention test* interaction effects in Table 2, significant long term effects are only observed for the CA condition. Six months after finishing the intervention, both RS and RSA students appeared to have lost much from what they had gained from the intervention. However, as can be seen from Graph 1, which is based on Model 4, this does not mean that in the end RS and RSA students have lost all of the extra growth in reading comprehension they had made during the intervention. Contrasting the joint effect of all three experimental conditions against the control condition confirms that there is an overall significant long term effect in favor of the students who received reading strategies instruction ($\Delta\beta = 3.415$, $\chi^2_{(1)} = 76.264$, $p < .001$), which is even somewhat larger than the joint effect at the short term. So, the explicit instruction of reading strategies appeared to have improved Chinese 2nd graders' growth in reading comprehension skills significantly at both the short and the long term.

Contrasting the joint effect for both peer tutoring conditions (CA and RSA) against the effect for the condition in which reading strategies instruction was not followed by practice in peer tutoring dyads (RS), reveals no significant difference at the posttest ($\Delta\beta = -0.083$, $\chi^2_{(1)} = 0.448$, $p > .05$), nor at the retention test ($\Delta\beta = -0.090$, $\chi^2_{(1)} = 0.570$, $p > .05$). At first sight, this means that practicing the reading strategies in peer tutoring dyads does not add any additional effect to the explicit teaching of those strategies. However, contrasting the CA condition against both other experimental conditions, reveals a significant long term effect in favor of the CA condition ($\Delta\beta = 0.228$, $\chi^2_{(1)} = 4.390$, $p < .05$). With respect to their long term

effects, there also appeared to be a marginally significant difference between both peer tutoring conditions (CA and RSA) in favor of the CA students ($\Delta\beta = 0.244$, $\chi^2_{(1)} = 3.756$, $p = .052$). No such long-term difference was found between the CA and the RS students. As for the short-term effects measured at the posttest, no significant differences were found between the experimental conditions.

The experimental effects reported so far, pertain to differences in growth. Significant differences in initial reading level (at the pretest) were found between the control condition and two of the experimental conditions, as is shown by the significant main effects for RS and CA in Table 2. In Figure 1 (on the next page) these initial differences have been adjusted for, so that the differences in growth become more apparent. What is also clearly shown in Figure 1 is that the CA condition keeps the gain it made, for several months after the intervention was finished. For CA the growth from posttest to retention test is not significantly different from the growth observed in the control condition. This was confirmed by contrasting the effects of CA*retention with CA*posttest ($\Delta\beta = 0.023$, $\chi^2_{(1)} = 0.172$, $p > .05$). For the other two experimental conditions, the partial loss of their short term gains is clear from the much less pronounced growth that can be observed between the second and the third measurement. Contrast effects show that in the case of RSA this growth is significantly smaller than the growth form posttest to retention test observed in the control condition ($\Delta\beta = -0.062$, $\chi^2_{(1)} = 0.092$, $p > .05$ for RS and $\Delta\beta = -0.133$, $\chi^2_{(1)} = 5.19$, $p < .05$ for RSA).

The significant covariates in the final model include students' gender and the interaction between gender and time. At the pretest, girls were about 0.20 points (or 0.24 standard scores) higher in reading comprehension scores than boys, but boys caught up partially thanks to an accelerated growth between pretest and posttest, as is indicated by the significant negative coefficient for girl*posttest. By the time of the retention test, the gap with the girls has become somewhat wider again, reducing the boys' catching up to a move that is only marginally significant,

leaving still 0.197 - 0.084 = 0.113 points (or 0.14 standard score) of difference between girls and boys.

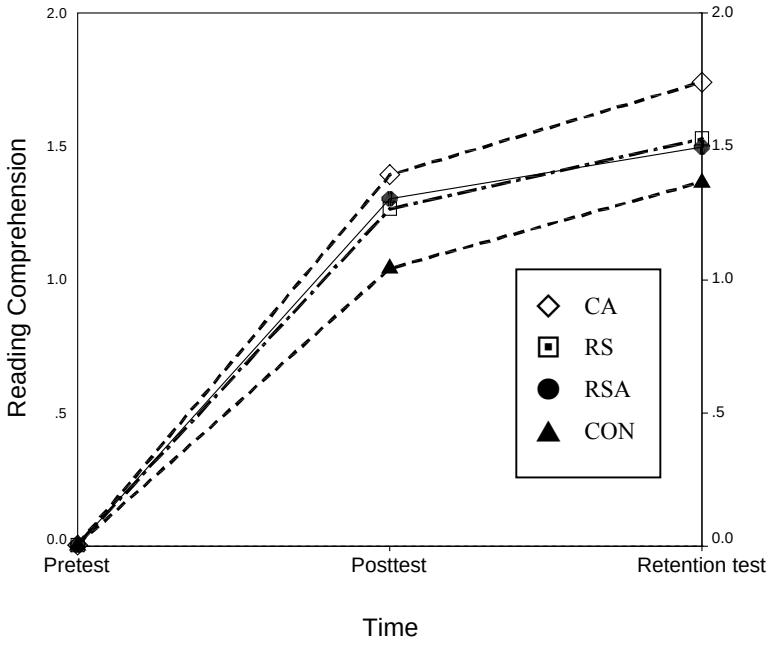


Fig. 1. The result of multilevel analysis on the experimental effects after adjusting for the differences in the initial reading levels across conditions

There is a significant negative association between reading comprehension and the initial level of negative metacognitive awareness; the initial level for positive reading attitude shows a positive relationship with students' reading comprehension scores. Entering these variables in the model, did not affect the coefficients for the experimental treatments much. That was also the case for the one social background variable that showed a significant association with reading comprehensions achievement, mother's diploma. Students in the one "City level key school" that was included in the study show a significantly larger growth in reading comprehension than students in ordinary schools. Such a difference was

not found for students in the less prestigious "*district level key school*". No significant interaction effects were found with experimental treatment and time. The treatment effects appear to be independent from the school's status and prestige.

4. Discussion

The results of multilevel repeated measures analysis, showing both significant short term and significant long term effects for the three experimental conditions jointly, generally confirm our first hypothesis about the expected improvement in reading comprehension growth following the explicit instruction of the six reading strategies. Students in all three conditions have received reading strategy instruction and practiced the strategies in reading assignments. The findings in the reading research literature with respect to the positive effects of reading strategies instruction also hold for Chinese primary school students in Chinese language classes. Interestingly enough, no interaction effects with school status were found. The effects of reading instruction (as well as the other treatment effects discussed further) are not different in the so-called *key schools* than in the ordinary schools. Students in the one *city level key school* that was included in the study however, on the average outperformed their peers in the other schools. But this higher performance level did not interfere with the treatment effects. No such interference was found for student's social background variables either. So, it appears that the findings hold for a wide variety of Chinese primary schools and students.

With respect to our second hypothesis that peer tutoring would yield extra value for students' reading comprehension growth, in addition to the benefit of reading strategy instruction, the results present a more complex picture. The hypothesis was not confirmed by the posttest results. There was no significant difference between the average short-term growth observed in the peer tutoring groups and the growth that on the average was made in the RS groups. However, the

differences in effect sizes for the three experimental conditions show a tendency that is in accordance with our hypotheses.

Looking at the retention test results, the findings are still different. The growth between pretest and retention test was significantly larger for the CA condition than for the other two experimental conditions jointly. Also, the difference between the CA and the RSA conditions was nearly significant. So, the question whether peer tutoring results in an additional acceleration of growth in reading comprehension, on top of the acceleration that can be attributed to the reading strategies instruction, cannot be answered with a simple yes or no. Whereas no such additional effect has been observed at the short term, on the long term, an additional effect has been observed, but only for the CA condition. In fact, the CA condition was the only experimental condition that showed a long-term growth that was significantly larger than the growth observed in the control condition. In the other two conditions the short-term gains that were made, were partially lost again. In contrast, CA students appeared to be able to keep the gains they had made during the intervention.

Since the intervention was stopped six months before the retention test, this long-term CA effect appears to indicate that the CA tutees have acquired a number of skills that allow for better reading comprehension independent from immediate help. These results fit the concept of self-regulated learning and also supported Vygotsky's the *Proximal Development Zone* theory (Vygotsky, 1978). Under the coaching of older tutors, CA tutees benefited more than students in other conditions in their reading skills and perhaps also metacognitive skills.

Comparison with the Flemish results

Compared to Flemish results in the 2nd grade, Chinese students seemed to have benefited more both from reading strategy instruction and peer tutoring practice, esp. from cross age tutoring. The first Flemish study didn't produce any significant intervention effects for the experimental conditions. The second Flemish study

only came out with a short-term intervention effect in CA tutees. Explanations from different angles can be suggested to explain this remarkable differences.

Firstly, we compared the age of the Chinese and Flemish students. According to the local regulation in Beijing, children have to be at least six years old when they first enter primary schools on September 1st. In the Flemish regulation, students have to enroll on the September 1st, if they are or will become six years old before the end of the year. This difference in official regulations lead to the results that Chinese students on average are three months older than their Flemish counterparts. This age difference was confirmed by the actual data of the students involved in both studies. According to research on the development of metacognition (Corsale & Ornstein, 1980), six to eight year-old students begin to monitor their cognitive processes with some strategies and they begin to understand the usefulness of the effective strategies for their performance on tasks. Ten-year-old students can already apply strategies spontaneously. So, age six to eight appears be a period of fast growth in metacognition and three months difference in age might make a difference when learning the reading strategies and mastering metacognitive skills. But this explanation doesn't sound very convincing, since the difference in age is only three months. Secondly, another reason could be found in language differences. As is introduced in the general introduction of this thesis, the Chinese writing system is morpheme based and more complicated concepts are usually a combination of simpler morphemes that have a meaning of their own. In many cases the meaning of the more complex concept can be derived from its constituent morphemes (McBride-Change *et al*., 2003). This feature makes Chinese language more analytic and transparent than Dutch when it comes to more complex words or concepts. Besides, Chinese has little or almost no rules in grammar. There are no tenses or different forms of verbs in sentence composition. At the starting stage, students' main obstacle in reading is related to vocabulary. Fortunately, most Chinese characters consist of two parts: one part, called the phonetic radical, indicates more or less the pronunciation, and the other part, called semantic radical, refers to the meaning of the character (Shu & Anderson, 1997). Getting to know

new characters from their morphological characteristics and try to use their phonetic and semantic radicals to remember its pronunciation and meaning is a common practice in language teaching. During our intervention, when teaching Strategy No.2 "enlarging vocabulary", we further strengthened these morphological awareness by encouraging students to guess (and then verify) the meaning of difficult words in reading materials. For example, 木 is a simple morpheme indicating trees or anything made of wood. When meeting a very difficult word such as 榛 (木 plus 秦, filbert), students should not feel bothered by the complicated looking of the character and only need to guess from its semantic radical that it is a kind of tree. Probably the combined effects of the morphological characteristics of Chinese language, its morpheme based writing system and the explicit reinforcement from the intervention have fostered Chinese students' analytic and meaning reconstruction skills, which in return have facilitated the Chinese reading process. Then it is also not surprising that, compared to other second graders, CA tutees have benefited most from working in tutoring dyads. It is very well conceivable that at a comparable young age, Flemish students cannot benefit that much yet from instruction in reading strategies and practice of metacognitive skills. With respect to the morphological structure of words (the ways words are formed), Dutch – as other Western languages – is a much more "abstract" language, less transparent for young children. Complex concept are very often expressed in long words with Latin and/or Greek roots; words take different forms (depending whether they are used as a verb, a noun, an adjective), for derived meanings words are formed with "abstract" prefixes and suffixes that by themselves have no meaning, etc. In daily life, in spoken language, Flemish children use these complex language forms intuitively. Because of the rather abstract level of these language features, systematic instruction in these aspects is postponed to a later age. At a young age, Flemish children have a little insight in the morphological structure of their language and therefore this morphological structure was not taught when teaching them the strategy related to vocabulary. In short, it could be argued that – because of its specific linguistic

features – for a language as Dutch, reading strategies become only useful at a somewhat later age than for a language such as Chinese. That could explain some of the differences in intervention effects that were found between Flemish and Chinese second graders.

Limitations of the study

One of the limitations of the study is that not all conditions started from the same level at the pretest. More particularly the CA students have an average pretest score that is significantly lower than the control group students. We should consider the possibility that the larger growth of the CA students can be explained as the result of some *natural* catching up, rather than the effect of the experimental treatment. Because the intervention was done in natural school settings, randomization was only feasible at the school level. Once schools were chosen, whole classes were assigned to different treatment conditions. Initially we presumed that within the same school, different classes from the same grade level would not differ significantly from each other. Classes were assigned to different treatment conditions after negotiating with school principals and teachers. Practical considerations played a role. For example, organizing CA necessitated close cooperation between a second and a fifth grade teacher, who had to mix their students every week for two class periods. And, since students were going to work in dyads, within each school the number of 2nd grade students participating in the CA-condition had to equal the number of 5th grade students in that condition. After the pretest it appeared that in the 2nd grade more relatively weaker classes were selected for the CA and for the RS conditions. Since every class and teacher had already been assigned to the different conditions, teacher training had been organized accordingly, and because we had to work within the framework of the after school extra curricular activities, it was not possible any more to adjust for the initial differences between conditions by shifting classes or students between conditions.

It might be argued that the accelerated growth shown in CA groups was at least partly due to some accidental lower pretest level for which they caught up in the year the intervention was carried out. But from our results, this explanation could be ruled out. RS students also started from a significantly lower pretest level than the control groups and they did not show similar long-term growth. What is more, CA demonstrated a steady growth and kept the experimental effects for 6 months after the intervention, whereas students in the other experimental conditions lost part of the short-term gains they had made. Therefore, we tend to conclude that the long-term gains shown for CA students are rather to be considered as a treatment effect. However, in future studies in natural classroom settings, close attention should given to the initial achievement levels of the classes and the possibilities to match class groups.

5. Conclusions

Explicit instruction in reading strategies that is scheduled weekly for a whole school year appears to have a positive effect on Chinese second grade students' growth in Chinese language reading comprehension. For students who practiced the reading strategies individually, under supervision of their teacher, and for students who had practice in a reciprocal same age peer tutoring dyad with a classmate, the effect is rather small and seems to last only as long as the treatment is given. For students who, in addition to the reading strategies instruction by their classroom teacher, were tutored by a fifth grader while practicing the strategies, a larger and much longer lasting effect was found. This long-term effect can be interpreted as an indication that those students have gained more skill in monitoring and regulating their own reading behavior.

Appendix: Table 2. Results of multilevel analyses on reading comprehension

	Model 0	Model1	Model 2	Model 3	Model 4	E.S.
Fixed effects						
Cons (Intercept)	0.019	0.192	0.955	0.110	0.005	
SE	*0.086*	*0.104*	*0.094*	*0.091*	*0.083*	
Posttest	1.163	0.985	1.035	1.041	1.036	
SE	*0.058*	*0.072*	*0.074*	*0.074*	*0.074*	
Retention test	1.470	1.333	1.364	1.368	1.361	
SE	*0.052*	*0.069*	*0.072*	*0.071*	*0.072*	
Girl			0.233	0.198	0.197	
SE			*0.052*	*0.051*	*0.051*	
Girl*posttest			-0.125	-0.128	-0.127	
SE			*0.047*	*0.047*	*0.047*	
Girl*retention test			-0.080	-0.084	-0.084	
SE			*0.044*	*0.044*	*0.044*	
Mothers' diploma			0.094	0.083	0.072	
SE			*0.014*	*0.014*	*0.014*	
Dummy missing diploma			-0.078	-0.078	-0.089	
SE			*0.046*	*0.043*	*0.043*	
Nega. awareness				-0.146	-0.135	
SE				*0.018*	*0.017*	
Dummy missing on awareness				0.069	-0.264	
SE				*0.020*	*0.214*	
Positive attitude				0.069	0.066	
SE				*0.020*	*0.020*	
Dummy for missing on attitude				0.054	0.024	
SE				*0.207*	*0.207*	
City key school					0.331	
SE					*0.055*	
District key school					-0.016	
SE					*0.056*	
Experimental conditions						
RS		-0.279	-0.282	-0.286	-0.247	
SE		*0.149*	*0.140*	*0.135*	*0.122*	
RSA		-0.140	-0.105	-0.154	-0.122	
SE		*0.151*	*0.139*	*0.133*	*0.118*	
CA		-0.403	-0.381	-0.377	-0.315	
SE		*0.133*	*0.124*	*0.120*	*0.108*	
SO * posttest		0.222	0.236	0.223	0.224	**0.28**
SE		*0.114*	*0.115*	*0.115*	*0.116*	
RSA * posttest		0.262	0.260	0.247	0.263	**0.33**
SE		*0.110*	*0.111*	*0.111*	*0.112*	
CA * posttest		0.346	0.362	0.360	0.351	**0.44**
SE		*0.100*	*0.101*	*0.101*	*0.102*	

To be continued

Table 2. Results of multilevel analyses on reading comprehension (Continued)

	Model 0	Model1	Model 2	Model 3	Model 4	E.S.
RS * Retention test		0.151	0.160	0.154	0.162	
SE		*0.109*	*0.110*	*0.110*	*0.111*	
RSA * retention test		0.127	0.125	0.115	0.130	
SE		*0.106*	*0.107*	*0.106*	*0.107*	
CA * retention test		0.358	0.370	0.367	0.374	**0.47**
SE		*0.097*	*0.098*	*0.097*	*0.098*	
Random effects						
Class						
Cons / Cons	0.159	0.152	0.103	0.095	0.059	
SE	*0.051*	*0.049*	*0.034*	*0.032*	*0.022*	
Posttest / Cons	-0.088	-0.081	-0.068	-0.067	-0.051	
SE	*0.032*	*0.029*	*0.025*	*0.024*	*0.019*	
Posttest / Posttest	0.068	0.057	0.056	0.057	0.056	
SE	*0.023*	*0.020*	*0.020*	*0.020*	*0.020*	
Retention test / Cons	-0.076	-0.074	-0.064	-0.062	-0.055	
SE	*0.028*	*0.028*	*0.024*	*0.023*	*0.020*	
Retention test / Posttest	0.052	0.050	0.050	0.050	0.049	
SE	*0.020*	*0.019*	*0.018*	*0.018*	*0.018*	
Retention test / Retentiont t.	0.053	0.054	0.054	0.053	0.054	
SE	*0.019*	*0.019*	*0.019*	*0.019*	*0.019*	
Student						
Cons / Cons	0.650	0.643	0.609	0.564	0.568	
SE	*0.031*	*0.031*	*0.029*	*0.027*	*0.027*	
Posttest / Cons	-0.447	-0.441	-0.427	-0.413	-0.415	
SE	*0.024*	*0.024*	*0.023*	*0.022*	*0.022*	
Posttest / Posttest	0.478	0.474	0.471	0.470	0.470	
SE	*0.023*	*0.023*	*0.023*	*0.023*	*0.023*	
Retention test / Cons	-0.397	-0.390	-0.381	-0.369	-0.371	
SE	*0.022*	*0.022*	*0.021*	*0.021*	*0.021*	
Retention test / Posttest	0.368	0.362	0.360	0.360	0.360	
SE	*0.020*	*0.020*	*0.020*	*0.020*	*0.020*	
Retention test / Retention test	0.418	0.411	0.410	0.410	0.411	
SE	*0.021*	*0.020*	*0.020*	*0.020*	*0.020*	
Occasion						
Cons / Cons	0	0	0	0	0	
SE	*0*	*0*	*0*	*0*	*0*	
-2log(lh)	3821.339	3790.253	3720.987	3626.798	3607.569	
N (cases)	2536	2536	2536	2536	2536	

References

Almasi, J. F. (1996). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussion of literature. *Reading Research Quarterly*, 30, 314-351.

Barron, A., & Foot, H. (1991). Peer tutoring and tutor training. *Educational Research,* 33, 174-185.

Bloom, B. S. (1984). The search for methods of group instruction as effective as one-to-one tutoring. *Educational Leadership*, 41(8), 4-17.

Butler, S. R. (1991). Reading program—Remedial, integrated and innovative. *Annals of Dyslexia,* 41, 119-127.

Clay, M. M. (1985). *The early detection of reading difficulties*. Portsmouth, New Hampshire: Heinemann.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: a meta-analysis of findings. *American Educational Research Journal*, 19, 237-248.

Corsale, K., & Ornstein, P. A. (1980). Developmental changes in children's use of semantic information in recall. *Journal of Experimental Child Psychology*, 30, 231-245.

De Corte, E., Verschaffel, L., Van de Ven, A. (2001). Improving text comprehension strategies in upper primary school children: a design experiment. *British Journal of Educational Psychology, 74, 4, 531-560.*

Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: a component analysis. *Journal of Educational Psychology*, 84, 331-339.

Fuchs, D., Fuchs, L. S., Mathes, P. G., Simmons, D. C. (1997). Peer-assisted learning strategies: making classrooms more responsive to diversity. *American Educational Research Journal,* 34, 174-206.

Garner, R., & Alexander, P. A. (1989). Metacognition: answered and unanswered questions. *Educational Psychologist,* 24(2), 143-158.

Hauser, R. M. (1994). Measuring socioeconomic status in studies of child development. *Child Development,* 65, 6, 1541-1545.

Hock, M. F., Pulvers, K. A., Deshlez, D. D., & Schumaker, J. B. (2001). The effects of an after-school tutoring program on the academic performance of at risk students and students with LD. *Remedial and Special Education,* 22(3), 172-186.

Invernizzi, M., Juel, C., & Rosemary, C. A. (1997). A community volunteer tutorial that works. *The Reading Teacher,* 50, 304-311.

Kamps, D. M., Dugan, E., Potucek, J., & Collins, A. (1999). Effects of cross-age peer tutoring networks among students with autism and general education students. *Journal of Behavioral Education,* 9, 2, 97-115.

Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities for students with learning disabilities who use English as a second language. *The Elementary School Journal,* 96, 3, 275-293.

Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review*, 23, 59-80.

Mathes, P. G., Torgesen, J. K., & Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal,* 38, 371-410.

McBride-Chang, C., Shu, H., Zhou, A., Wat C. P., & Wagner, R. K. (2003). Morphological awareness uniquely predicts young children's Chinese character recognition, *Journal of Educational Psychology*, 95, 4, 743-751.

Mellanby, A. R., Newcombe, R. G., Rees, & J. Tripp, J. H. (2001). A comparative study of peer-led and adult-led school sex education. *Health Education Research*, 16, 4, 481-49.

OECD (2003). *Learners for life—student approaches to learning results from PISA 2000*, OECD.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.

Palincsar, A. S., David, Y. M., Winn, J. A., & Stevens, D. D. (1991). Examining the context of strategy instruction. *Remedial and Special Instruction,* 12, 43-53.

Paris, S. G., Lipson, M. Y., & Wixon, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8, 293-316.

Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: transactional instruction of reading comprehension strategies, *The Elementary School Journal,* 92, 5, 513-555.

Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J., & Kurita, J. (1989). Strategies that improve children's memory and comprehension of text. *The Elementary School Journal,* 90, 3-32.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., Miller, T. R. (2003). Peer assisted learning intervention with elementary school students: a meta-analytic review. *Journal of Educational Psychology,* 95, 2, 240-257.

Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 778-803). New York: Macmillan.

Schrader, B. & Valus, A. (1990). Disabled learners as able teachers: a cross-age tutoring project. *Academic Therapy*, 25, 589-597.

Shu, H., & Anderson, R. C. (1997). Role of radical awareness in the character and word acquisition of Chinese children. *Reading Research Quarterly*, 32, 78-98.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Review of Educational Research,* 75, 3, 417-453.

Steinberg, L., Dornbusch, S. M., & Brown, B. B. (1992). Ethnic differences in adolescent achievement: An ecological perspective. *American Psychologist,* 47, 723–729.

Topping, K. J. & Ehly, S. (1998). Introduction to peer-assisted learning. In K. J. Topping & S. Ehly (Eds.). *Peer-assisted learning* (pp. 1-23). Mahwah, New Jersey: Erlbaum.

Topping, K. J. (1998). Paired learning in literacy. In K.J. Topping & S. Ehly (Eds.), *Peer-assisted learning* (pp. 87-104). Mahwah, New Jersey: Erlbaum.

Van Keer, H. (2004). Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring. *British Journal of Educational Psychology,* 74, 2, 37-70..

Van Keer, H., & Verhaeghe, J. P. (2003). Effects of explicit reading strategies instruction and peer tutoring in second and fifth grades. In E. De Corte, L. Verschafeel, N. Entwistle, & J. Van Merrienboer (Eds.), *Powerful learning environment: Unraveling basic components and dimensions* (pp. 211-228). Oxford: Elsevier.

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. *The Journal of Experimental Education,* 73(4), 291-329.

Vygotsky, L. S. (1978). Mental development of children and the process of learning (M. Lopez Morillas, Trans.). In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.). L. S. Vygotsky: Mind in society (pp. 7–8). Cambridge, MA: Harvard University Press (Original work published 1935).

Wentzel, K. R. (1999). Social-motivational processes and interpersonal relationships: Implications for understanding motivation at school. *Journal of Educational Psychology*, 91, 76–97.

Winter, S. (1987). Parents and pupils as remedial tutors, *Bulletin of the Hong Kong Psychological Society*, 18, 15-32.

Chapter 5

The Effects of Reading Strategy Instruction and Peer Tutoring on 5th Graders' Use of Reading Strategies, Metacognitive Awareness and Reading Attitude*

Abstract

In a one-year long intervention study, reading strategies were explicitly taught and practiced by students individually (RS), in Cross-Age peer tutoring dyads (CA) or in Reciprocal Same-Age peer tutoring dyads (RSA). About 670 5th grade pupils from 20 classes out of 8 Chinese primary schools and their parents were included in the study. Multilevel analysis with repeated measures approach is used to analyze the intervention effects. Having controlled for initial reading comprehension level, social background and other relevant characteristics, students in CA and RS conditions showed a significant increase in reading strategy use, positive metacognitive awareness and positive attitudes, compared to the control group.

Key words: reading strategy peer tutoring metacognitive awareness reading attitude

1. Introduction

This study is partly a replication of a Flemish study (Van Keer, 2004; Van Keer & Verhaeghe, 2005) and was conducted in eight primary schools in Beijing, China. The intervention program combines two lines of research in reading

* This chapter is based on an article with the same title submitted to the *British Journal of Educational Psychology* for publication.

comprehension learning. The first one stresses the importance of explicit instruction of reading strategies. Research in this line has shown that proficient readers monitor and regulate their comprehension while reading, using a flexible repertoire of cognitive and metacognitive strategies (e.g., Pressley & Allington, 1999). A simple exposure to a large amount of reading materials does not necessarily lead to the development of monitoring and regulating skills (Hartman, 2001). However, those skills could be mastered through explicit instruction in order to improve reading comprehension (Dole *et al.*, 1996; Pressley *et al.*, 1989; Raphael, 2000). Inspired by a number of existing programs (e.g., Brown *et al.*, 1996; Pressley *et al.*, 1992; Palincsar & Brown, 1984), lesson scenarios were developed to teach students six strategies: (a) activating prior knowledge, (b) finding the meaning of unfamiliar words and expressions, (c) making predictions, (d) finding the main idea, (e) monitoring and regulating comprehension of texts, and (f) distinguishing different types of texts. To scaffold students with practicing those strategies, a set of assignment cards were provided, each specifying a number of steps to be taken and generic questions to be answered. Those questions are not text-specific and could be extended by students to adapt to the reading context.

The second line of research on which the program is built, emphasizes the importance of interaction between students about texts (Almasi, 1996; Fuchs *et al.*, 1997), providing them with opportunities for metacognitive exchange and modeling (Palincsar *et al.*, 1991). Peer tutoring has been shown to have positive effects on both tutor and tutee, in a variety of curriculum areas, including reading (Cohen *et al.,* 1982; Mathes & Fuchs, 1994; Mathes *et al.*, 2001), but has seldom been applied to reading comprehension instruction specifically (Van Keer & Verhaeghe, 2005). The idea behind applying peer tutoring for acquiring metacognitive skills is that monitoring and regulating someone else's reading behavior is easier than monitoring and regulating one's own. Through practicing reading strategies with a tutee, a tutor can more easily acquire metacognitive skills that can be transferred into his/her own reading process.

1.1. Implementation of the intervention in China

The Flemish intervention program was translated and adapted for use in Chinese language classes in regular primary schools in Beijing, China. This involved the translation of lesson scenarios, reading materials, exercise materials for students and a teacher's manual into Chinese. Prior to the actual intervention study, a shorter-term pilot study was carried out in 3 classes from 2 of the experimental schools for the CA and RSA conditions. All the materials and testing were tried out in the pilot study with a preceding cohort of students.

To evaluate the effect of the intervention a quasi-experimental study was designed, involving 22 complete class groups from 8 Chinese primary schools, including ordinary schools, "district key schools", and "city key schools". Class groups were assigned to one of five different conditions, in such a way that each of the conditions occurred in each of these types of schools. Three experimental conditions were distinguished: reading strategies instruction with individual practice (RS), reading strategies instruction with practice in reciprocal same age peer tutoring dyads (RSA) and reading strategies instruction with practice in cross age peer tutoring dyads (CA).

Students in the RS, RSA and CA conditions received explicit instruction from their regular Chinese language teachers in the abovementioned six reading strategies. Each of these reading strategies took at least two sessions of teaching, including direct instruction, modeling of how to use a reading strategy, and teacher-led whole class practice. After the introduction of a particular reading strategy, students had two more lessons to practice the strategy, either independently (as in the RS condition) or in peer tutoring dyads (as in the CA and RSA conditions) before they moved on to the next reading strategy. From the third strategy on, additional review sessions were added to encourage students to apply freely all reading strategies learnt so far. The intervention lasted for a whole school year. On average,

each class in any of the experimental conditions spent 60 minutes per week on the program. For the RSA condition, students from the same class formed dyads and changed roles every week. In the CA condition, 5th graders tutored 2nd graders. Principally, the dyad composition did not change in the course of the program.

Teachers were provided with reading materials, students' assignment books and administration cards (to fill in which texts had been read and what strategy had been practiced). Prior to the implementation of the program, teachers in the experimental conditions received 24 hours of training in the theoretical background and the practical implementation of the program. Throughout the school year they were visited and coached weekly by the first author. Data were collected, through teachers' self-reports and independent classroom observation by one of two research assistants, to establish measures of treatment fidelity.

The intervention was scheduled during extra curricular activities that take place after school hours (from 3.30 pm to 5 pm) and are regularly organized 5 days a week in all of the schools involved in the intervention study. Students in the control group condition (CON) participated for a comparable amount of time in other Chinese language related extra curricular activities that did not involve any instruction in reading strategies nor any form of cooperative learning.

We presume that a wide diversity of attractive and interesting children's books is an important element in the students' reading growth. Since such books are not easily available in most of the Chinese primary schools, a total number of 2,500 children's books were bought and circulated among the experimental classes. These books suit students from different reading levels. To check for the effect of the provision of these books on students' change in reading attitude and growth in reading comprehension, a second control condition, the free reading (FR) condition was created. Students in the FR condition did not get explicit instruction of reading strategies nor any practice in peer tutoring. These students had free reading sessions parallel to the reading strategies instruction or peer-tutoring activities in

the experimental groups. Students in the strict control condition did not have access to any of the books.

1.2. Previous results

Previous analyses (Li & Verhaeghe, 2005) have shown significant positive effects of the intervention on students' reading achievement. Using a repeated measures design, multilevel analyses revealed both significant short-term effects as well as significant long-term effects 6 months after the conclusion of the intervention. Short-term effect size for the CA condition is 0.19 and the long-term effect sizes for the CA, RSA and RS conditions are 0.48, 0.44 and 0.56 respectively. Students in the experimental conditions clearly outperformed students from the control groups as well as students from the free reading condition.

The fact that in the FR condition no significant effect was found at all compared to control condition, seems rule out the possibility that the new reading books or more reading engagement have directly enhanced students' reading achievement. The rather large experimental long-term effects indicate that growth in reading achievement continued till at least six months after the intervention was concluded. This seems to imply that the acquisition of (metacognitive) reading strategies is responsible for the larger growth in reading comprehension in the three experimental conditions. However, the reading comprehension test results by themselves do not offer any direct explanation for the experimental results.

1.3. Research questions and hypotheses

To clarify the processes underlying the effects of the program on students' reading achievement, the present study therefore examines the effects on three aspects: students' use of reading strategies, students' metacognitive awareness during reading and students' attitude towards reading. The first aspect, the use of reading strategies, is directly related to the focus of the intervention. With the second variable we try to highlight a metacognitive process that is closely related to

student's monitoring behavior during reading ("Am I doing well?") as well as to their academic self-concept with regard to reading ("Am I capable in this?"). With the third variable we envision to capture a motivational element related to reading. Whereas as significant effects on the first two variables would give support to the hypothesis that the acquisition of reading strategies is the main reason for the extra growth in reading comprehension in students in the experimental conditions, positive effects on reading attitudes would highlight an alternative explanation, focusing on motivational factors.

In this study we try to find an answer to the following research questions:

(1) What is the effect of explicit instruction of reading strategies on 5th grade students' use of reading strategies, having adjusted for the effects of students' characteristics and family background?

(2) What is the effect of explicit instruction of reading strategies on 5th graders' attitudes towards reading, having adjusted for the effects of students' characteristics and family background? Do students have a more positive attitude towards reading after the intervention?

(3) What is the effect of the explicit instruction of reading strategies on 5th graders' metacognitive awareness? Does the program enhance students' metacognitive awareness in some degree, compare to students from control groups?

(4) Does the practice of reading strategies in peer tutoring dyads (either CA or RSA) add any extra effect on the use of reading strategies, on metacognitive awareness or reading attitudes?

(5) Do CA and RSA have significantly different intervention effects on the aforementioned measures?

The following hypotheses are stated:

(1) Explicit instruction of reading strategies improves significantly students' use of reading strategies. Compared to students in the control groups (CON and FR conditions) and having adjusted for the effects of students' characteristics and their family background, students in the experimental conditions will report significantly more use of reading strategies after intervention.

(2) By the end of the intervention, a significant increase in positive and/or decrease in negative attitude towards reading is expected for students in the experimental conditions or the FR condition, compared to control group students. No significant difference is expected between the experimental conditions and the FR condition.

(3) With respect to the students' metacognitive awareness during reading, a significant decrease in negative metacognitive awareness and/or increase in positive awareness is expected for students in the experimental groups, compared to students in the control and FR conditions.

(4) Peer tutoring practice adds extra value to the experimental effects of reading strategies instruction. That is, compared to students in the RS condition, by the end of the intervention students in CA and RSA conditions will report significantly more use of reading strategies, significantly more positive metacognitive awareness and they will show significantly more positive attitudes toward reading.

(5) The intervention effects on students' use of reading strategies, metacognitive awareness and reading attitudes are larger for CA students than for RSA students.

2. Method

2.1. Sample

A total of about 670 5th graders participated in this study, with about half of them in the control condition and the other half divided among the three experimental conditions. All students were enrolled in one of 8 Beijing primary schools, situated in different parts of the city and as a total serving a socially mixed public. In one third of the students' families, at least one of the parents has a university degree (bachelor, master or doctor); in another third the parents' highest educational attainment level is senior high school. Half of the families have a monthly income of no more than 3,000 RMB (360 USD), the average income being about 3,500 RMB (420 USD). The distribution of students with respect to their social background differed among schools, depending of the "type" of school as identified above, with "ordinary" schools mainly serving working class students, whereas relatively more middle class and upper middle class students were found in "district key schools" and "city key schools". About half of the students are girls. At the start of the study, 10% of the 5th graders were one or two years behind their age group.

2.2. Design

The study can be characterized as a quasi-experimental intervention study, taking place in natural classroom settings, with whole class groups assigned to one of the conditions, and using a pretest-posttest control group design. Data on students' strategy use, metacognitive awareness, and reading attitudes were collected prior to the start of the intervention and right after finishing it.

2.3. Instrumentation

Measurements on strategy use, metacognitive awareness and reading attitude were collected by questionnaires completed individually by the 5th grade students. A

total of 83 items with 5-point Likert scales were presented to the students both at the pretest and posttest. Response alternatives were phrased either in terms of how often a particular behavior happen (varying from "*always"* to "*never")* or to what extend a specific statement matches their own situation (from "*to a great extend"* to "*not at all"*).

For each subscale, we started with an exploratory principal component analysis with Varimax rotation for pretest and posttest data separately, which resulted in similar component structures for pretest and posttest data. To test whether the same structure holds for pretest and posttest, confirmatory factor analyses with recursive models were performed on each subscale, allowing for covariance between pretest and posttest residuals (Du Toit & Du Toit, 2001), by using the Amos 5.0 software package. The confirmatory factor analyses in general led to a satisfying fit (on all scales, RMSEA is lower than or very close to 0.05). Based on the confirmed models, average scores of all the items in the scales were calculated for further analyses. These averaged scores ranged from 1 to 5, reflecting the original scale of the items.

Use of reading strategies

The measure for use of reading strategies is based on items describing cognitive acts such as "I stop to think: what happened first, then next?" or "I find [the meaning of] an unknown word from the context". Using a variety of wordings, the items reflect mainly the six reading strategies that were taught to the students in the program. Based on principal components analysis with Varimax rotation and reliability analyses, an internally consistent scale was constructed, using 12 of the originally 27 items. Cronbach's alpha is 0.86 for the pretest data (N = 639) and 0.88 for the posttest data (N = 639). Confirmatory factor analysis in which the model was simultaneously tested for pretest and posttest data yielded a good fit (RMSEA = .048, LO90 = .044, HI90 = 0.053, N = 639). Standardized coefficients for the 12 items range from .45 to .72 for the pretest and from .45 to .73 (see Figure

6 in Appendix 2). For both pretest and posttest sum scores ranged from 1 to 5, with means of 3.68 (pretest) and 3.94 (posttest) and a SD of 0.80 (pretest) and 0.81 (posttest). For both pretest and posttest distributions were somewhat skewed (Sk = -0.63, SE = 0.097 for pretest and Sk = -0.81, SE = 0.097 for posttest).

Metacognitive awareness

The measure for *metacognitive awareness* focuses on students' awareness about themselves, the reading tasks and their performance in relation to reading tasks. Principal component analysis with Varimax rotation suggest a scale for *negative awareness* (based on 11 items such as: *"I think: I don't understand, this is too difficult for me"* or *"I think: I can do this well only because I get help"*), with a Cronbach's alpha of 0.85 for the pretest data (N = 565) and 0.86 for the posttest data (N = 616) respectively, and a scale for *positive awareness* (based on 7 items such as: *"I think: I can tell this story"*), with a internal consistency coefficient of 0.67 for the pretest (N = 654) and 0.69 for the posttest (N = 660).

Two types of items were initially entered in the principal component analysis. One type of items reflect some metacognitive awareness of one's own reading ability, e.g., *"I think: I don't understand, this is too difficult for me" (Q50)* or *"I think: I can tell this story" (Q39).* Other items reflect students' performance in relation to the reading tasks, e.g., *"I think: I cannot read this because I am not good in reading"(Q54)* or *"I think: I did well because I tried hard" (Q48).* Twenty items were initially entered in the analyses. Two items were dropped because of inconsistencies in loadings across pretest and posttest data. Interestingly enough, item Q52 *("I think: I can do this well, because I got help")* and item Q56 *("I think: I can read this well, because I am a lucky")* both load high on the negative component. According to the theory of the inverse compensatory scheme (Marsh, Cairns, Relich, Barnes & Debus, 1984; Nicholls, 1978), the external attribution of a success is indeed likely to reflect a low esteem of own one's own ability.

Confirmatory factor analysis with a recursive model was used for both pretest and posttest data simultaneously and yielded a good fit for the positive scale (RMSEA = .047, LO90 = .038, HI90 = .056, N = 623) and a fairly good fit for the negative scale (RMSEA = .053, LO90 = .048, HI90 = .058, N = 623). With the exception of one item that had a rather low loading, standardized coefficients all were above .40, with a maximum of .68 (Please see Figure 7 and 8 in Appendix 2).

The sum scores for positive metacognitive awareness are normally distributed, ranging from 1.14 to 5.00 (M = 3.45, SD = 0.72) for the pretest and from 1.29 to 5.00 (M = 3.59, SD = 0.74) for the posttest. The sum scores for negative metacognitive awareness have comparable ranges and standard deviations, but lower means (M = 2.16 for the pretest and M = 2.09 for the posttest) and the distributions are rather skewed (Sk = 0.563, SE = .098 for the pretest and Sk = 0.920, SE = .098 for the posttest).

As to the interpretation of the scales, it should be noted that the scores reflect the degree to which a student's mind is occupied with metacognitive awareness during reading. Positive and negative awareness do not exclude each other mutually ($r = .13$, $p < .001$, $N = 623$ for pretest and $r = .175$, $p < .001$, $N = 623$ for posttest). As can be seen in Fig. 1, a high incidence of positive metacognitive awareness can occur together with both a low and a high incidence of negative metacognitive awareness. On the other hand, a high incidence of negative awareness is more likely to occur with a high incidence of positive awareness and less likely to occur with a low incidence of positive awareness. This relationship shows that both scales cannot simply be subtracted one from the other in order to obtain a uni-dimensional metacognitive awareness scale.

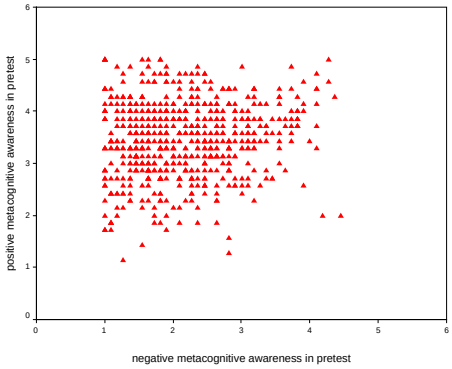


Fig. 1. Scatterplot of negative vs. positive metacognitive awareness in pretest

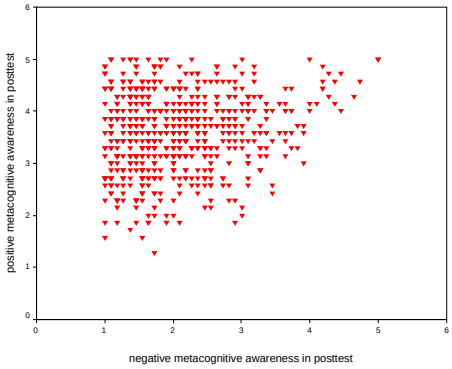


Fig. 2. Scatterplot of negative vs. positive metacognitive awareness in posttest

Reading Attitudes

To create a measure for attitude towards reading 27 items based on Aarnoutse (1996) were presented to the students. The results of a principal component analysis suggest a positively oriented subscale (15 items, alpha = .89 for both pretest and posttest with N = 603) and a negatively oriented subscale (10 items, alpha = .87 for both pretest and posttest, N = 603). Fitting the pretest and posttest data simultaneously, confirmatory factor analysis yielded a good fit for both the recursive model for the positive subscale (RMSEA = .051, LO90 = .047, HI90 = .054, N = 603) as well as for the model for the negative subscale (RMSEA = .040, LO90 = .033, HI90 = .046, with N = 603). Loadings are all above .40 with a maximum of .76 (see Figure 9 and 10 in the Appendix 2).

For the negative subscale sum scores ranged from 1.00 to 4.80 (M = 2.20, SD = 0.89) for the pretest and from 1.00 to 5.00 (M = 1.94, SD = .81) for the posttest. Distributions are rather skewed (Sk = 0.66, SE = 0.10 for the pretest and Sk = 1.06, SE = 0.10 for the posttest). The distribution of the positive subscale appears to mirror these characteristics, with sum scores ranging from 1.13 to 5.00 (M = 4.13, SD = 0.69, Sk = -1.204, SE = 0.10) for the pretest and from 1.33 to 5.00 (M = 4.25,

SD = 0.64, Sk = -1.324, SE = 0.10) for the posttest. Both scales (sum scores) correlate negatively with each other ($r = -.429$, $p < .001$, $N = 603$ for the pretest and $r = -.483$, $p < .001$, $N = 603$ for the posttest).

Family background

Data on students' family background were collected by means of a parents' questionnaire. Parents were asked to provide data on their educational level, profession, family monthly income, parents' reading behaviors, reading interests and attitudes, early childhood rearing practices with regard to literacy education and children's actual reading behavior at home. A high number of parents responded to the questionnaires, yielding high proportions of valid answers on the basic questions relevant for determining the family's SES (93% to 98%). Questions on early childhood rearing practices were least well answered, with proportions of valid answers ranging between 75% and 85%.

2.4. Multilevel analysis

Multilevel analysis was used to analyze the effects of the intervention on the measures described above. Since the same scales were used for pretest and posttest, a repeated measures multilevel design with three levels (measurement occasion nested in students, who in turn are nested in class groups) was used.

Prior to entering the experimental conditions in the model, the effects of several other variables were adjusted for: students' gender, various aspects of their family background, initial reading level, and initial values for related metacognitive or motivational variables. Continuous predictor variables were centered around their means to avoid variance inflation with related interaction terms and to facilitate interpretation of the regression coefficients.

To avoid reduction of the database and bias due to non-ignorable missing values on the independent variables and to preserve the comparability of nested models fitted in the course of the analyses, for any predictor entered in the model missing values

were treated by creating a dummy while on the centered variable the missing value was converted into zero (the "neutral center"). The purpose of creating these dummies is to assess whether the missing values on a certain predictor has a different prediction result from the non-missing ones. If this was not the case, the dummy was removed from the model and as a consequence, the missing values on the predictors were treated as average values.

In view of the parsimony principle in model building, only predictors with significant effects were retained in the model. An exception is made for the experimental conditions, which were entered at last in the model. To keep an overview of the distinct effects of the intervention, all of the four dummies for the experimental conditions were kept in the model, whether significant or not.

3. Results

Tables 1 to 5 (in Appendix 1) display the results of the repeated measures multilevel analyses on the five subscales. The traditionally reported null-model, in which only the constant is entered in the model, has been left out from the tables because it gives little or no information in the case of repeated measures designs. The tables start with a model 1 in which only a dummy time variable (measurement occasion) is entered in the equation. Since we only have two measurement occasions, only one dummy (*time2*) is used. The coefficient for the dummy *time2* initially reflects the overall evolution in the dependent variables from pretest to posttest From the moment predictors are entered in the model, the coefficient for *time2* reflects the evolution made by students belonging to the reference category (or in the case of continuous predictor variables, students with an average value on the predictors). From the moment the experimental conditions are entered in the model, this reference category is further reduced to the students belonging to the control group. Typical for analyses following a repeated measures design is the rather large random part in the tables. Since we want to explore

differences in evolution over time between students, a random effect for *time2* is modeled at the student level. In all tables presented here, this random effect is highly significant, indicating that indeed students differ in their evolutions over time. In all our models the linear part of that random effect (the *"co-variance"* between *time2* and the constant) is negative, indicating that for each of our dependent variables the variance between students becomes smaller over time. From the moment the dummy *time2* is allowed to vary randomly across students (level 2), the parameter for the random part at the measurement occasion level (level 1) is structurally turned to zero.

To explore differences in evolution over time between class groups, a random effect of *time2* is also modeled at the class level. Except for the variables with regard to attitude towards reading (Tables 4 and 5), for each of the dependent variables a significant coefficient is shown for at least one of the elements in the random part (the linear and/or the quadratic element) in at least model 1, indicating that indeed (at least prior to entering the predictor variables in the model) overall significant differences in evolution over time are observed between class groups.

The coefficients for the interaction terms *time2 * experimental condition* in the fixed part reflects the differential growth from time1 (pretest) to time2 (posttest) in each of the experimental conditions. Normally these interaction terms are preceded by the main effects for the experimental conditions (as in Table2). Once the interaction terms *time2*experimental condition* are entered in the model, these main effects reflect the differences between the experimental condition and the control condition at time1, i.e. at the pretest. However, for Table 1, 3, 4 and 5 these main effects were totally non-significant (reflecting no significant differences between the experimental conditions and the control conditions were found at the pretest) and their removal from the model had no effect at all on the coefficients for the interaction effects. Since there is still one main effect in the model *(time2),* the interaction effects remain well interpretable. The following part will cover the results for each of the dependent variables.

3.1. Use of reading strategies

As is reported in Table 1, generally speaking, the 5th graders report a significant increase (β= 0.2710, SD = 0.0662, $p < .001$) in the use of reading strategies, albeit with a moderate effect size (ES = 0.37). Having controlled for the effects of gender (girls report significantly more strategy use), parents' preferences with regard to children's literature, the incidence of parents' reading together with their children and the initial incidence of positive and negative awareness during reading as reported by the students, significant effects were found for each of the three experimental conditions. On average, RS and CA students appear to have greater gains in strategy use, followed by RSA students. The differences with the control condition are highly significant ($p < .001$).

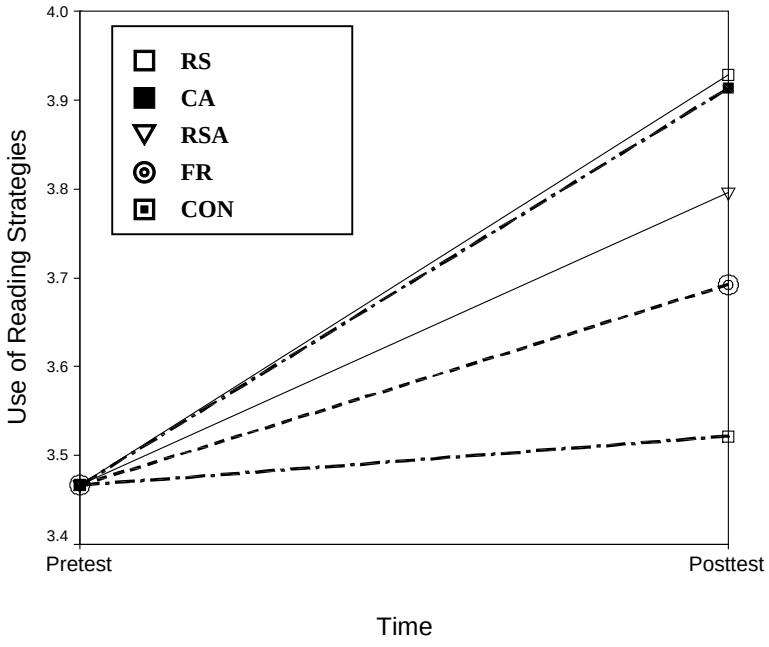


Figure 3. The evolution of reading strategy use from pretest to posttest across conditions

Effect sizes range from a very moderate 0.38 (SA) to a somehow larger 0.54 (CA) and 0.56 (RS). The three experimental conditions do not differ significantly from each other. For the free reading condition no significant differences from any of the other conditions, including the control group, were found. It is worth mentioning that the control group students didn't show any significant increase in their self-reported reading strategy use ($\beta = 0.06$, SE = 0.08, $p > .40$). The relative evolution of reading strategy use across different conditions is shown in Figure 3.

3.2. Metacognitive awareness during and right after reading

Overall, the occurrence of metacognitive awareness during or right after reading decreased over time. This decrease is remarkably large for positive metacognitive awareness ($\beta = -0.85030$, SE = 0.06014, $p < .001$, ES = 1.19, see table 2). It is only small and but marginally significant for negative metacognitive awareness ($\beta = -0.08144$, SE = 0.04459, $p = .068$, ES = 0.10, in Model 1 in Table 3). After having adjusted for the effect of the initial incidence of positive metacognitive awareness and other relevant predictors, the overall evolution in the incidence of negative metacognitive awareness becomes even significantly positive, i.e. on the average students report an increase in negative metacognitive awareness. The effect is however small ($\beta = 0.19750$, SE = 0.03624, $p < .001$, ES = 0.25, in Model 3 in Table 3). So, in general, with regard to metacognitive awareness during and after reading, students tend to show a negative evolution over time, with much less positive awareness and somewhat more negative awareness.

For the incidence of positive metacognitive awareness a significant difference was found between the evolutions observed in the control and the CA conditions ($\beta = 0.20890$, SE = 0.10360, $p = .044$). A marginally significant difference was found for the RS condition ($\beta = 0.26710$, SE = 0.13730, $p = .052$). It should be noted that these effects show positive coefficients, indicating that CA and RS students on average tend to have a (marginally) significantly smaller decrease in positive awareness during and right after reading, compared to the control group students.

With respect to negative metacognitive awareness, no difference in evolution was observed between the experimental conditions. They all follow the general trend of showing a small increase in negative awareness during or right after reading. Figure 4 displays the relative evolution of positive metacognitive awareness from pretest to posttest across different conditions.

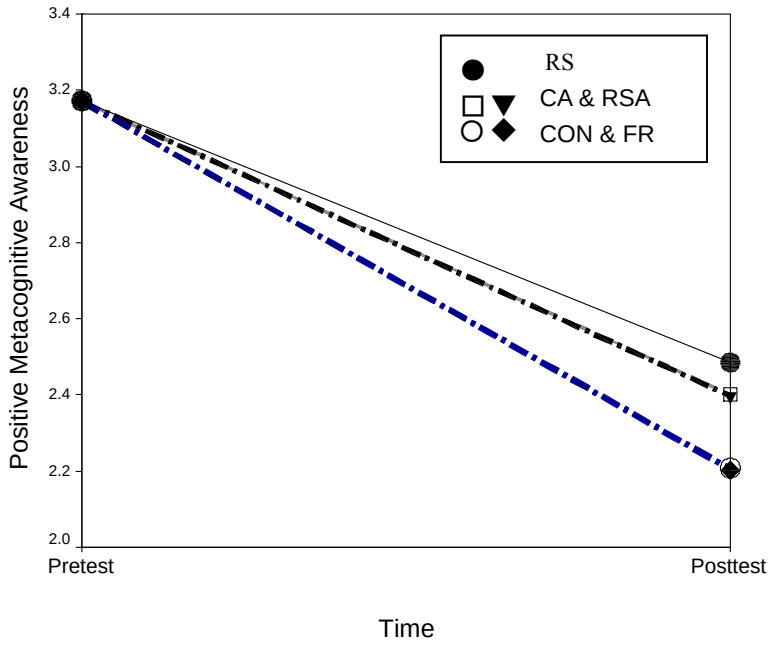


Figure 4. The evolution of positive metacognitive awareness from pretest to posttest across conditions

3.3. Reading attitudes

In general students' reading attitudes became significantly more positive over time. This evolution appears to be more pronounced for the scores on the negative subscale (β = -0.2753, SE = 0.0519, $p < .001$, ES = 0.31) than for the scores on the positive subscale (β = 0.1262, SE = 0.0385, $p < .01$, ES = 0.18).

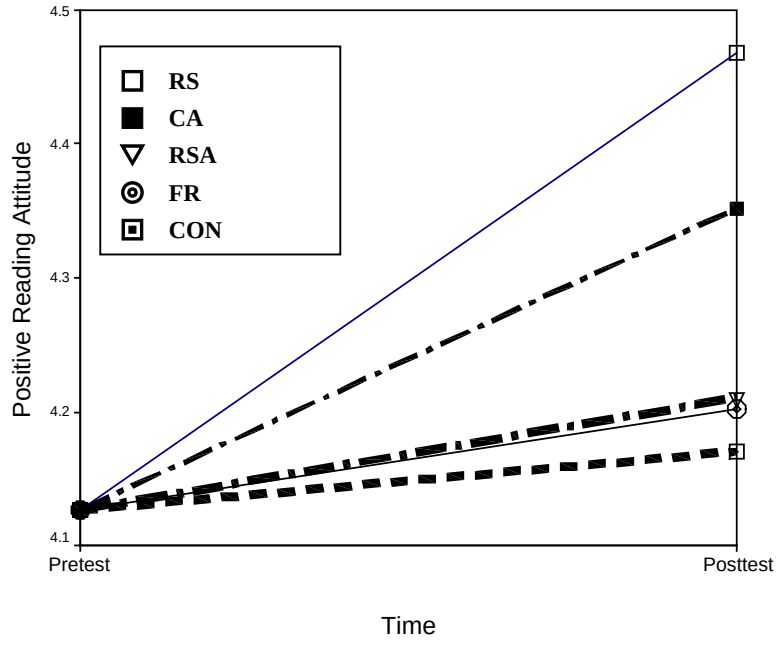


Figure 5. The evolution of positive attitudes from pretest to posttest across conditions

For the negative subscale no significant differences in evolution were found among the experimental conditions. For all experimental conditions the same general pattern is observed with a less negative attitude at the posttest. Results are different for the positive subscale. Whereas the control group students on the average showed almost no change in their attitudes towards reading (β= 0.04, SE = 0.05, p > .30) and a similar observation is made for the RSA students (β= 0.04, SE = 0.08, p > .60) and the students in the free reading condition (β= 0.03, SE = 0.10, p > .75), a positive evolution is shown for the RS students (β= 0.30, SE = 0.10, p < .01, ES = 0.43) and for CA students (β= 0.18, SE = 0.07, p < .02, ES = 0.26). The difference in evolution between CA and RSA is marginally significant ($\beta_1 - \beta_2 =$ 0.14, $\chi^2_{(1)} = 2.60$, p = .10), the difference between RSA and RS is significant at the .05 level ($\beta_1 - \beta_2 = 0.26$, $\chi^2_{(1)} = 4.81$, p < .05). The difference between CA and

RS is not significant. Figure 5 displays the relative evolution of positive attitudes from pretest to posttest across different conditions.

4. Discussion

Although questionnaires are not often a successful means to capture use of reading strategies and metacognitive processes (Veenman & Van Hout-Wolters, 2002), in this study both the results of the principal component analyses and the results for the experimental effects are quite consistent with theoretical expectations. The results show a clear positive effect of the intervention on self-reported strategy use. This is especially true for the CA and the RS conditions, and to a somewhat less degree for the RSA condition.

The finding that the impact of the intervention on students' self-reported strategy use is larger for the CA condition than for the RSA condition parallels with previously presented findings from an observational study of the interaction processes taking place in the peer tutoring dyads (see Chapter 2). This observational study showed a larger use of reading strategies and more opportunities for practicing metacognitive skills in CA than RSA dyads (Li & Verhaeghe, 2004). Since the peer tutoring component of the program is based on the hypothesis that this kind of text and reading-centered interaction between students would enhance the opportunities to practice reading strategies, it was not expected that the RS condition, in which the students practice the reading strategies individually in a whole class setting, would yield comparable effects on reading strategy use as the CA condition. As is explained in Chapter 3, it may due to the "self-perfection" effects (Li, 2003) among Chinese 5th graders in RS condition. Once these students received reading strategy instruction and were given equal reading environment to read, they could make equal progress as those in the CA condition. This particular finding as well as the other findings with respect to strategy use appears to confirm the outcomes for the effects on students' reading

comprehension achievement (see Chapter 3). This very strict parallelism supports the hypothesis that the improvement of students' reading comprehension achievement builds highly on the acquisition of the repertoire of reading strategies that was taught to the students. Also the finding that control group students and students in the free reading condition did not show any increase in reading strategy use supports this hypothesis.

However, the findings with respect to the effects on attitude towards reading, more particularly the positive subscale, indicate that the motivational hypothesis cannot be excluded either. Remarkably enough and quite contrary to our expectations, students in the free reading condition did not show any increase in positive attitude towards reading. In this respect no difference with the control group students was found. We call this finding remarkable because the free reading condition was explicitly created to test the hypothesis that the availability of new, attractive books would account for (at least some of) the experimental effect on reading achievement because of an increased motivation to read and a more positive attitude towards reading. It was expected that this motivational effect would be quite obvious in the free reading condition, because in this condition students read to enjoy themselves; there was no pressure on them to do anything with the books that would look like an assignment for school, as was the case in the experimental conditions. From the findings it is clear that the books by themselves didn't have any effect at all, not on students' attitudes towards reading, nor on their reading strategy use or any other of the dependent variables. This of course can explain why the free reading condition didn't have much effect on the students' reading comprehension achievement either.

Another remarkable finding is that more particularly that experimental condition that we would have expected to create the least positive changes in students' motivation for reading appeared to have had the largest positive effect on students' attitude towards reading, the *RS* condition. In that condition students get a teacher-led introduction of the reading strategies and practice them individually

under the help of their teacher in a whole class setting. It was expected that this condition would have been experienced as more boring than both peer tutoring conditions, in which students can interact with another student. Further exploration is needed to find out in what way this finding could relate to daily Chinese classroom practices.

A positive effect on students' attitude towards reading was also found for the peer tutoring condition. Interestingly enough, such an effect was found for students in the cross age tutoring condition, fifth graders tutoring second graders, whereas no such an effect was found for students in the same age peer tutoring condition, fifth graders interacting with one of their classmates. In this respect it should be reminded that cross age tutoring implied the fifth grade tutors to practice reading strategy use through helping younger students mastering these strategies. In order to do this, second grade level books appealing to the interests of these second grade tutees were used. In contrast, RSA students could read and enjoy books at their own reading level and of their own interest. Apparently, the opportunity to be a *real tutor*, responsible for a younger student outweighed the disadvantage of having to spend the practice periods on books that are way below one's own level. This difference between CA and RSA conditions with respect to the impact on the tutors' reading attitude parallels the differential impact on reading strategy use. Most probably they are jointly together responsible for the observed difference in impact on the tutors' reading achievement.

The results with respect to the effects on metacognitive awareness are not that easy to interpret. On the one hand, a high preoccupation with metacognitive awareness pertaining to one's own ability does not seem to reflect a great orientation towards self-regulation and strategy-use. In this sense, the general decrease in reported incidence of metacognitive awareness during and right after reading can be considered as a positive finding. On the other hand, a higher incidence of metacognitive awareness may reflect a greater metacognitive awareness, eventually leading to better self-regulation in a later phase. And of course, a higher

incidence of positive metacognitive awareness is likely to indicate a more positive academic self-concept. From this perspective, the finding of a smaller decrease of positive metacognitive awareness in the CA and RS conditions compared to control conditions is a positive finding. Again, the rather unexpected combination of positive effects for both in CA and RS conditions (but not in the RSA condition) co-occur the findings with regard to students' growth in reading comprehension achievement (see Chapter 2). Taking that co-occurrence into account, it appears that a relatively higher incidence of positive metacognitive awareness should indeed be interpreted as an indication of an improved academic self-concept for the students involved. It should also be noted that with respect to negative metacognitive awareness no large changes were observed. And no differences were found between the experimental conditions.

Apart from some effects on the self-reported use of reading strategies and the large effects on positive reading attitudes and metacognitive awareness, the effects reported here appear to be rather moderate or even small. That is not unusual in the case of naturalistic studies.

In the Flemish study (Van Keer, 2002), the same questionnaires were used to measure 5th graders' *use of reading strategies*, *metacognitive awareness* and *reading attitude*. No significant intervention effects were found with respect to use of reading strategies or reading attitude among Flemish 5th graders. In the first Flemish study, the only significant effect found was that RSA condition had significantly bigger decrease in negative metacognitive awareness than control condition at the retention test. This result is hard to interpret. In the second Flemish study, similar result was found for RSA condition at the posttest. Surprisingly, RSA also had a bigger decrease in negative metacognitive awareness than CA condition at the posttest. As is reported earlier in the result section, the results for RSA condition in Chinese study in general are less positive than CA and RS conditions. The Flemish results in RSA seem to be opposite to Chinese results in some way.

5. Conclusions

The results give clear indications that the explicit teaching of reading strategies had benefits for 5th grader' use of reading strategies and their reading attitudes. The findings support the hypothesis that the observed improvement of students' reading comprehension achievement can be attributed considerably to the acquisition of a repertoire of reading strategies and improved reading attitudes. But the potential role of motivational factors and an enhanced metacognitive awareness cannot be ruled out, although the effect sizes were less impressive.

The present findings do not support the hypothesis that peer tutoring practice increases the use of reading strategy, or improves metacognitive awareness and attitudes on top of the effect of explicit reading strategies instruction. In Chinese primary school classrooms, individual practice of reading strategies under the teachers' guidance seems to have comparable positive effects to practice in dyads. Between the two peer tutoring practice, RSA condition in general has less favorable results than CA condition.

Appendix:

Table 1. Multilevel analyses on self-reported reading strategy use

	Model 1	Model 2	Model 3	Model 4	E.S.
Fixed Effects					
Constant (Intercept)	3.69	3.61	3.46	3.47	
SE	*0.07*	*0.07*	*0.06*	*0.06*	
Time 2 (T2)	0.27	0.27	0.27	0.06	0.08
SE	*0.07*	*0.07*	*0.07*	*0.08*	
Girl		0.16	0.15	0.14	0.19
SE		*0.05*	*0.05*	*0.05*	
Parents' att. children's literature		0.16	0.13	0.13	0.11
SE		*0.04*	*0.04*	*0.04*	
Reading together with child		0.02	0.02	0.02	0.06
SE		*0.01*	*0.01*	*0.01*	
Initial reading compreh. level		0.08			0.10
SE		*0.03*			
Initial positive awareness			0.37	0.38	0.35
SE			*0.04*	*0.04*	
Initial negative awareness			-0.09	-0.10	0.09
SE			*0.03*	*0.03*	
Missing on awareness			0.29	0.27	0.15
SE			*0.16*	*0.16*	
Experimental conditions					
T2 * CA				0.39	0.54
SE				*0.11*	
T2 * SA				0.27	0.38
SE				*0.12*	
T2 * RS				0.41	0.56
SE				*0.14*	
T2 * FR				0.17	0.23
SE				*0.17*	
T2 * CON (= reference)				~~	
SE				~~	

(To be continued)

Table 1. Multilevel analysis on use of reading strategy use *(Continued)*

	Model 1	Model 2	Model 3	Model 4	E.S.
Random					
Class					
Cons / Cons	0.06	0.05	0.04	0.04	
SE	*0.03*	*0.02*	*0.02*	*0.02*	
T2 / Cons	-0.04	-0.03	-0.02	-0.02	
SE	*0.02*	*0.02*	*0.02*	*0.01*	
T2 / T2	0.06	0.06	0.06	0.04	
SE	*0.03*	*0.03*	*0.03*	*0.02*	
Student					
Cons / Cons	0.58	0.56	0.49	0.49	
SE	*0.03*	*0.03*	*0.04*	*0.04*	
T2 / Cons	-0.29	-0.30	-0.27	-0.27	
SE	*0.03*	*0.03*	*0.029*	*0.029*	
T2 / T2	0.62	0.62	0.62	0.61	
SE	*0.04*	*0.04*	*0.04*	*0.03*	
Pos. awareness / Cons			-0.12	-0.12	
SE			*0.03*	*0.03*	
Pos. awareness / T2			0.008	0.004	
SE			*0.029*	*0.03*	
Pos. awareness /Pos. awareness			0.11	0.11	
SE			*0.04*	*0.04*	
Occasion					
Cons / Cons	0	0	0	0	
SE	*0*	*0*	*0*	*0*	
-2log(lh)	2848.86	2829.89	2671.34	2652.84	
N (students)	639	639	639	639	

Table 2. Repeated measures multilevel analyses on positive metacognitive awareness

	Model 1	Model 2	Model 3	Model 4
Fixed				
Constant (Intercept)	3.43	3.17	3.12	3.21
SE	*0.05*	*0.10*	*0.09*	*0.10*
Time 2 (T2)	-0.85	-0.85	-0.85	-0.97
SE	*0.06*	*0.06*	*0.06*	*0.08*
Father's educational level		0.15	0.16	0.16
SE		*0.04*	*0.04*	*0.04*
Parents' att. children's literature		0.07	0.07	0.07
SE		*0.04*	*0.04*	*0.04*
Parents' attitude tw comics		0.05	0.05	0.05
SE		*0.02*	*0.02*	*0.02*
reading comprehension pretest		0.09	0.13	0.13
SE		*0.03*	*0.03*	*0.03*
Neg. Metacognitive awareness			0.15	0.15
SE			*0.03*	*0.03*
Experimental conditions				
Cross Age Tutoring				-0.12
SE				*0.08*
Same Age Tutoring				-0.10
SE				*0.09*
Reading Strategy				-0.22
SE				*0.12*
Free Reading				-0.10
SE				*0.12*
Control (= reference)				~~
SE				~~
T2 * Cross Age Tutoring				0.21
SE				*0.10*
T2 * Same Age Tutoring				0.19
SE				*0.13*
T2 * Reaing Strategy				0.27
SE				*0.14*
T2 * Free Reading				0.00
SE				*0.18*
T2 * Control (= reference)				~~
SE				~~

(To be continued)

Table 2. Repeated measures multilevel analysis on positive metacognitive awareness *(Continued)*

	Model 1	Model 2	Model 3	Model 4
Random				
Class				
Cons / cons	0.04	0.02	0.02	0.01
SE	*0.02*	*0.01*	*0.01*	*0.01*
T2 / cons	-0.03	-0.02	-0.02	-0.01
SE	*0.02*	*0.01*	*0.01*	*0.01*
T2 / T2	0.05	0.05	0.05	0.03
SE	*0.02*	*0.02*	*0.02*	*0.02*
Student				
Cons / cons	0.48	0.47	0.44	0.44
SE	*0.03*	*0.03*	*0.03*	*0.03*
T2 / cons	-0.28	-0.29	-0.29	-0.29
SE	*0.03*	*0.02*	*0.03*	*0.02*
T2 / T2	0.61	0.61	0.61	0.61
SE	*0.04*	*0.04*	*0.04*	*0.04*
Neg. metacognitive awareness / cons			-0.05	-0.06
SE			*0.02*	*0.02*
Neg metacognitive awareness / T2			0.01	0.01
SE			*0.02*	*0.02*
Neg metacognitive awareness / Neg metacognitive awareness			0.05	0.05
SE			*0.03*	*0.03*
Occasion				
Cons / cons	0	0	0	0
SE	*0*	*0*	*0*	*0*
-2log(lh)	2606.73	2564.68	2531.69	2523.23
N (students)	623	623	623	623

Table 3. Repeated measures multilevel analyses on negative metacognitive awareness

	Model 1	Model 2	Model 3	Model 4
Fixed				
Constant (Intercept)	2.17	2.34	1.99	1.99
SE	*0.05*	*0.06*	*0.03*	*0.03*
Time 2 (T2)	-0.08	-0.08	0.20	0.18
SE	*0.05*	*0.05*	*0.04*	*0.05*
Girl		-0.19		
SE		*0.05*		
Mother's diploma		-0.33	-0.15	-0.15
SE		*0.07*	*0.05*	*0.05*
Initial reading level			-0.16	-0.15
SE			*0.03*	*0.03*
Positive MC. awareness			0.19	0.19
SE			*0.02*	*0.02*
Negative reading attitude			0.46	0.46
SE			*0.02*	*0.02*
Experimental conditions				
T2 * Cross age tutoring				0.05
SE				*0.07*
T2 * Same age tutoring				-0.02
SE				*0.07*
T2 * Strategies Only				0.13
SE				*0.11*
T2 * Free reading				-0.06
SE				*0.10*
T2 * Control (= reference)				~~
SE				~~

(To be continued)

Table 3. Repeated measures multilevel analyses on negative metacognitive awareness *(continued)*

	Model 1	Model 2	Model 3	Model 4
Random				
Class				
Cons /cons	0.04	0.03	0.01	0.00
SE	*0.02*	*0.01*	*0.01*	*0.00*
T2 /cons	-0.01	-0.01	0.01	0.00
SE	*0.01*	*0.01*	*0.00*	*0.00*
T2 / T2	0.02	0.02	0.00	-0.00
SE	*0.01*	*0.01*	*0.01*	*0.005*
Student				
Cons /cons	0.57	0.54	0.35	0.35
SE	*0.03*	*0.03*	*0.02*	*0.02*
T2 /cons	-0.29	-0.28	-0.26	-0.26
SE	*0.03*	*0.03*	*0.02*	*0.02*
T2 / T2	0.59	0.59	0.54	0.54
SE	*0.03*	*0.03*	*0.03*	*0.03*
Occasion				
Cons /cons	0	0	0	0
SE	*0*	*0*	*0*	*0*
-2log(lh)	2717.57	2680.61	2224.53	2221.10
N (students)	623	623	623	623

Table 4. Repeated measures multilevel analyses on negative reading attitude

	Model 1	Model 2	Model3	Model 4	Model 5
Fixed Effects					
Constant (Intercept)	2.22	2.37	2.32	2.20	2.20
SE	*0.07*	*0.08*	*0.05*	*0.05*	*0.05*
Time 2 (T2)	-0.28	-0.28	-0.22	-0.23	-0.22
SE	*0.05*	*0.05*	*0.05*	*0.05*	*0.06*
Girl		-0.29	-0.17	-0.12	-0.12
SE		*0.06*	*0.05*	*0.04*	*0.04*
Mothers' educational level			-0.07		
SE			*0.02*		
Initial reading level			-0.21	-0.13	-0.13
SE			*0.03*	*0.027*	*0.03*
Positive reading attitude			-0.43	-0.37	-0.37
SE			*0.03*	*0.03*	*0.03*
Negative MC. awareness				0.46	0.45
SE				*0.03*	*0.03*
Experimental conditions					
T2 * Cross age Peer Tutoring					0.02
SE					*0.08*
T2 * Same Age Peer Tutoring					-0.01
SE					*0.09*
T2 * Strategies Only					-0.12
SE					*0.11*
T2 * Free Reading					-0.04
SE					*0.12*
T2 * Control (= reference)					--
SE					--

(To be continued)

Table 4. Repeated measures multilevel analysis on negative reading attitude
(Continued)

	Model 1	Model 2	Model 3	Model 4	Model 5
Random Effects					
Class					
Cons / Cons	0.07	0.07	0.02	0.02	0.02
SE	*0.03*	*0.03*	*0.01*	*0.01*	*0.01*
T2 / Cons	-0.03	-0.03	-0.02	-0.01	-0.01
SE	*0.020*	*0.02*	*0.01*	*0.01*	*0.01*
T2 / T2	0.03	0.03	0.02	0.02	0.02
SE	*0.02*	*0.02*	*0.01*	*0.01*	*0.01*
Student					
Cons /cons	0.73	0.70	0.56	0.39	0.39
SE	*0.04*	*0.04*	*0.04*	*0.02*	*0.02*
T2 /cons	-0.37	-0.37	-0.35	-0.30	-0.30
SE	*0.03*	*0.03*	*0.03*	*0.02*	*0.02*
T2 / T2	0.65	0.65	0.64	0.63	0.63
SE	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*
Moth.ed.level /cons			-0.03		
SE			*0.01*		
Moth.ed.level / T2			-0.01		
SE			*0.01*		
Moth.edu.level/ Moth.edu.level			0.00		
SE			*0.01*		
Occasion					
Cons / Cons	0	0	0	0	0
SE	*0*	*0*	*0*	*0*	*0*
- 2log(lh)	2797.72	2773.73	2503.46	2335.59	2334.28
N (students)	623	623	623	623	623

Table 5. Repeated measures multilevel analyses on positive reading attitude

	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed Effects					
Constant (Intercept)	4.13	4.04	4.10	4.13	4.13
SE	*0.04*	*0.05*	*0.04*	*0.04*	*0.04*
Time2 (T2)	0.13	0.13	0.13	0.13	0.04
SE	*0.04*	*0.04*	*0.04*	*0.04*	*0.05*
Girl		0.16	0.13	0.11	0.11
SE		*0.05*	*0.04*	*0.04*	*0.04*
Mother's educational level			0.02		
SE			*0.02*		
Parents' att. children's literature			0.09	0.10	0.10
SE			*0.03*	*0.03*	*0.03*
Initial reading compreh. level			0.19	0.15	0.15
SE			*0.03*	*0.03*	*0.03*
Negative MC. awareness				-0.14	-0.14
SE				*0.03*	*0.03*
T2 * Cross Age Peer Tutoring					0.18
SE					*0.07*
T2 * Same Age Peer Tutoring					0.04
SE					*0.08*
T2 * Reading Strategies Only					0.30
SE					*0.10*
T2 * Free Reading					0.03
SE					*0.10*
T2 * Control (= reference)					--
SE					--

(To be continued)

Table 5. Repeated measures multilevel analyses on positive reading attitude (Continued)

	Model 1	Model 2	Model 3	Model 4	Model 5
Random Effects					
Class					
Cons /cons	0.02	0.02	0.01	0.01	0.01
SE	*0.01*	*0.01*	*0.01*	*0.01*	*0.02*
T2 /cons	0.01	-0.01	-0.00	-0.01	-0.01
SE	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*
T2 / T2	0.02	0.02	0.02	0.01	0.01
SE	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*
Student					
Cons /cons	0.46	0.45	0.39	0.38	0.38
SE	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*
T2 /cons	-0.21	-0.22	-0.22	-0.22	-0.22
SE	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*
T2 / T2	0.38	0.38	0.38	0.38	0.37
SE	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*
Initial reading level /cons			-0.04	-0.04	-0.04
SE			*0.01*	*0.01*	*0.01*
Initial reading level / T2			-0.01	-0.01	-0.01
SE			*0.02*	*0.02*	*0.02*
Init. read. lev. / Init. read. lev.			0.01	0.01	0.01
SE			*0.02*	*0.02*	*0.02*
Occasion					
Cons /cons	0	0	0	0	0
SE	*0*	*0*	*0*	*0*	*0*
- 2log(lh)	2200.26	2188.91	2104.57	2084.87	2071.85
N (students)	639	639	639	639	639

Appendix 2: Confirmatory factor analyses on all the subscales

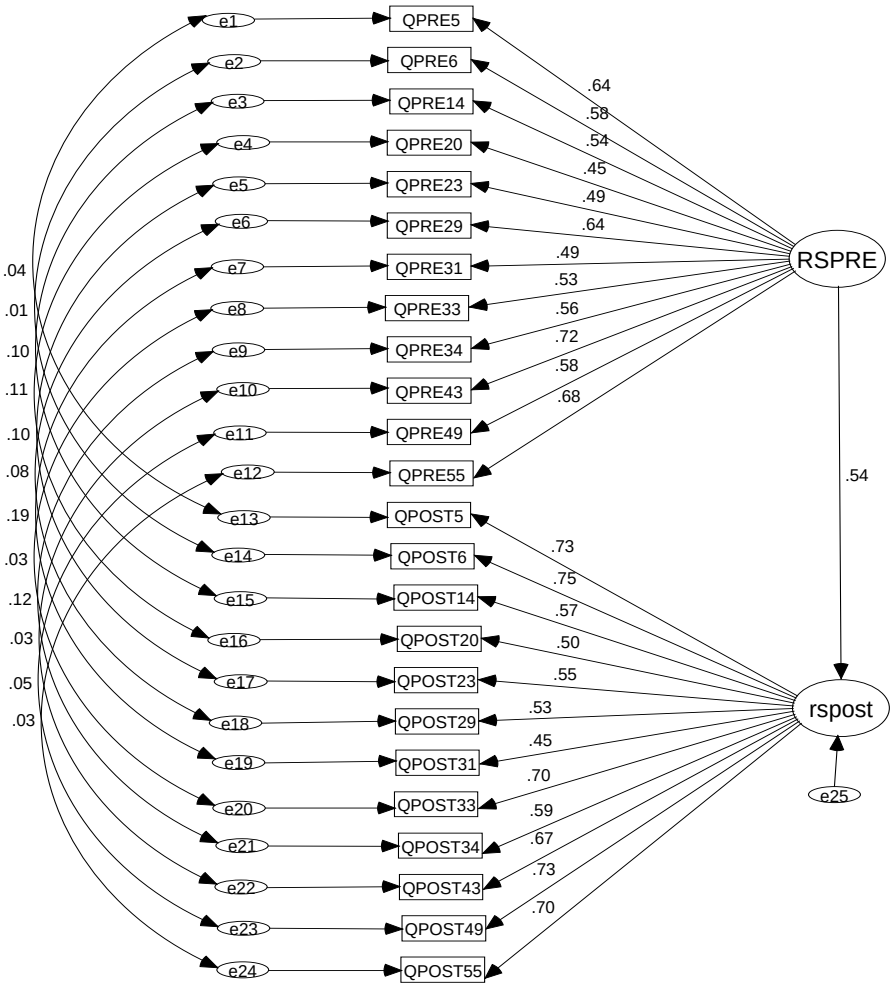


Figure 6. Confirmatory factor analysis of reading strategy use

RMSEA = .048 LO90 = .044 HI90 = .053

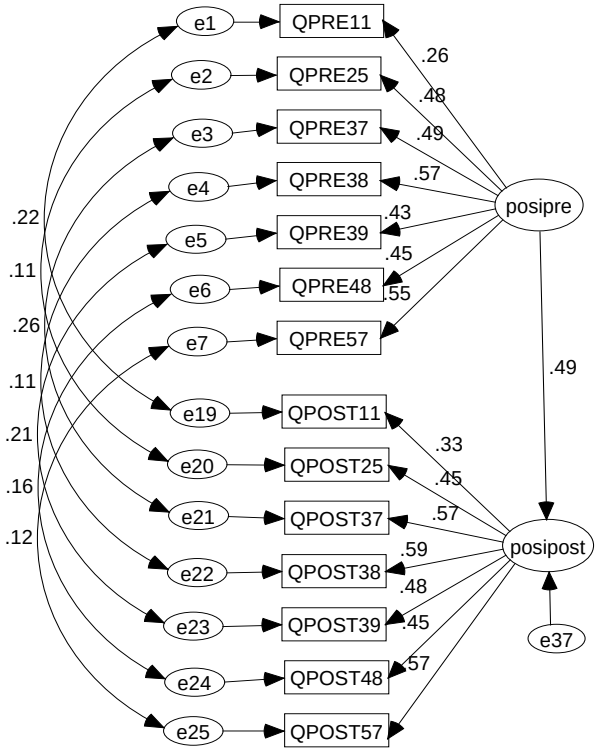


Figure 7. Confirmatory factor analysis of positive metacognitive awareness in reading

RMSEA = .047 LO = .038 HI90 = .056

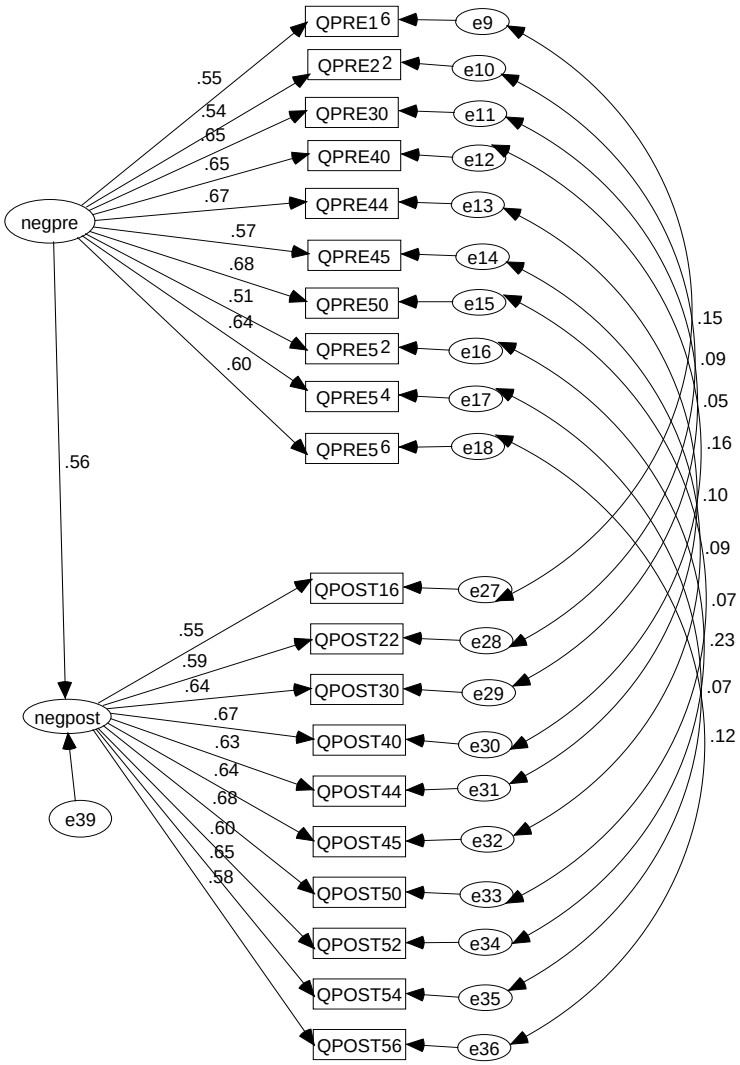


Figure 8. Confirmatory factor analysis of negative metacognitive awareness in reading

RMSEA = .053 LO90 = .047 HI90 = .059

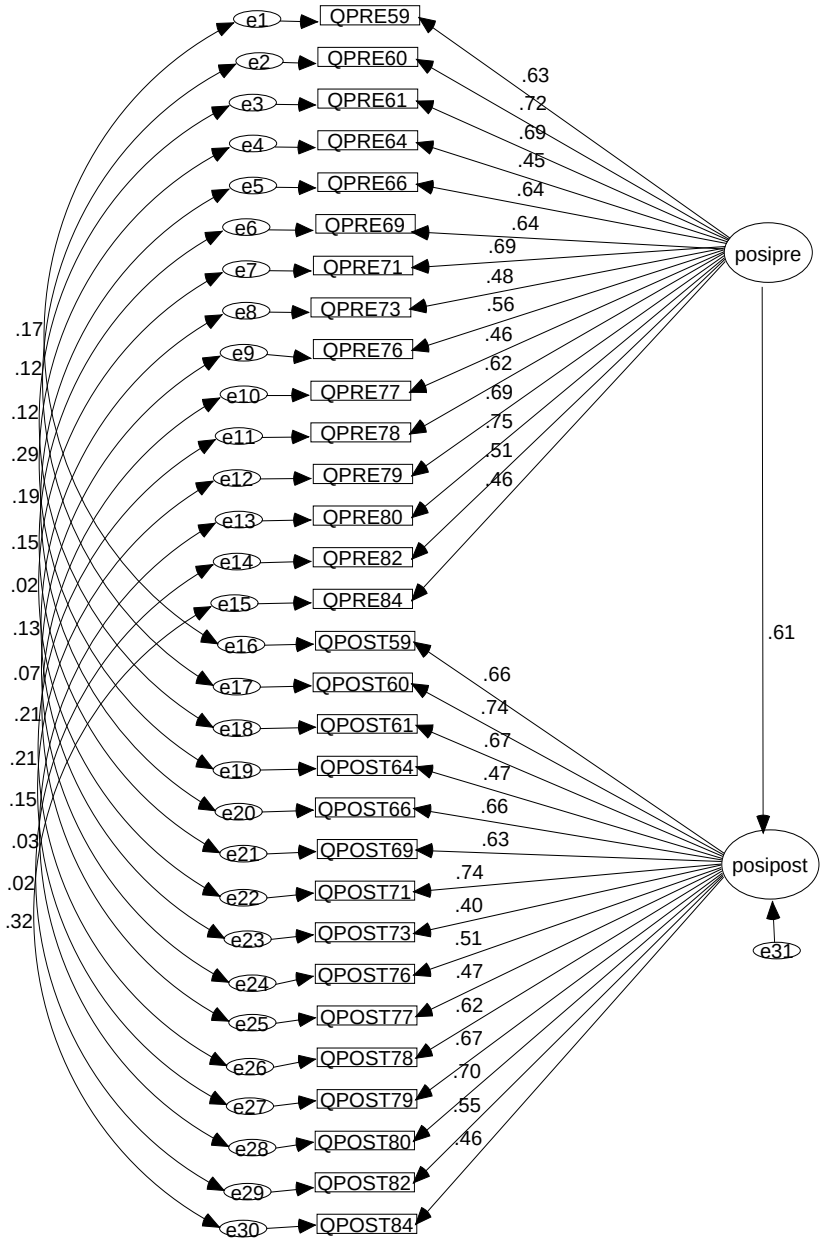


Figure 9. Confirmatory factor analysis of positive reading attitude

RMSEA = .051 LO90 = .047 HI90 = .054

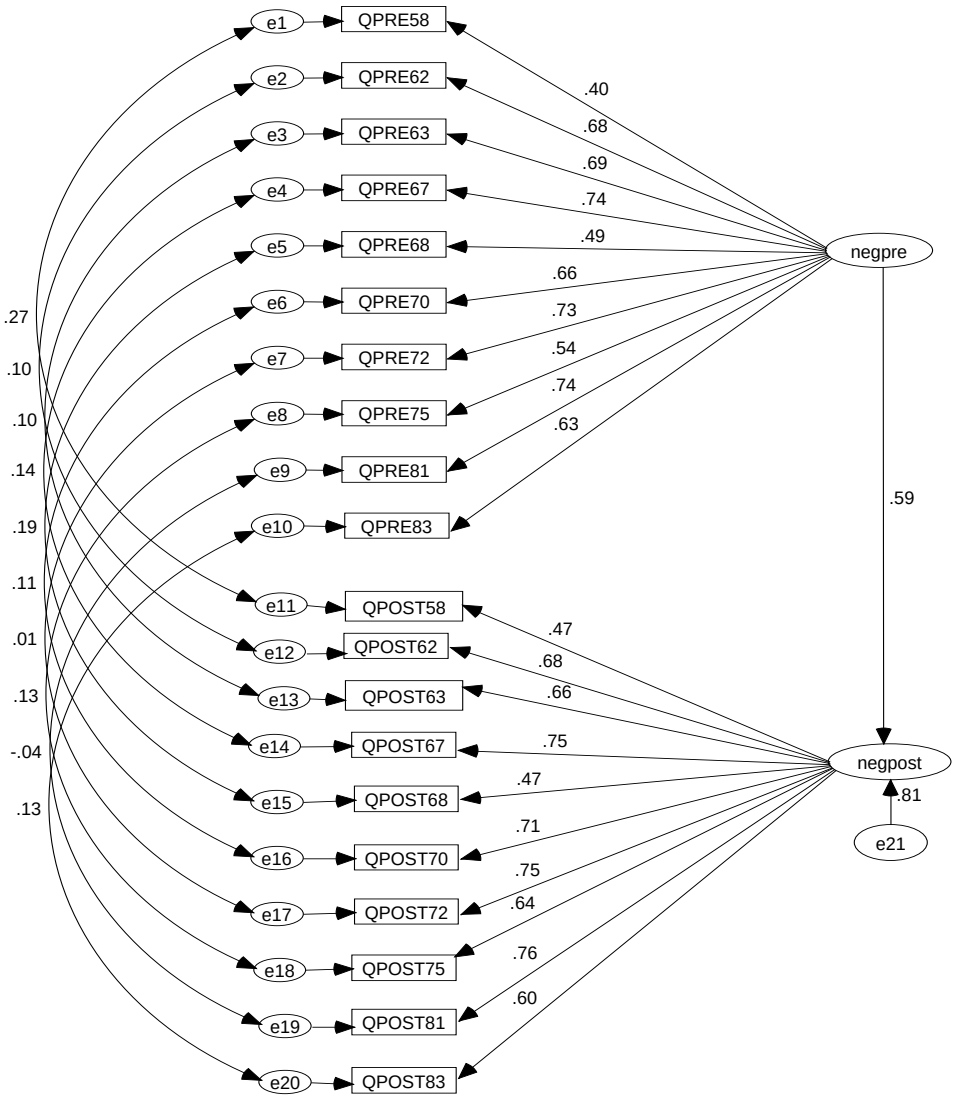


Figure 10. Confirmatory factor analysis of negative reading attitude

RMSEA = .040 LO90 = .033 HI90 = . 046

References

Aarnoutse, C. A. J. (1996). *Leesattitudeschaal* [*Reading attitude scale*]. Nijmegen (NL): Berkhout.

Almasi, J. F. (1996). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussion of literature. *Reading Research Quarterly*, 30, 314-351.

Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88, 18-37.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: a meta-analysis of findings. *American Educational Research Journal*, 19, 237-248.

Dole, J. A., Brown, K. J., & Trathen, W. (1996). The effects of strategy instruction on the comprehension performance of at-risk students. *Reading Research Quarterly*, 31, 62-88.

Du Toit, M. & Du Toit, S. (2001). *Interactive LISREL User's Guide*. Lincolnwood: Scientific Software International.

Fuchs, D., Fuchs, L. S., Mathes, P. G., Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34, 174-206.

Hartman, H. J. (2001). Developing students' metacognitive knowledge and skills. In H. J. Hartman (Ed.). *Metacognition in learning and instruction* (pp. 36-68). Dordrecht, The Netherlands : Kluwer Academic Publishers.

Li, J. (2003). U.S. and Chinese cultural beliefs about learning. *Journal of Educational Psychology*, 95, 2, 258-267.

Li, L. & Verhaeghe, J. P. (2005). Effects of explicit reading strategy teaching and peer tutoring on 5th graders' metacognition and reading attitudes. Paper presented at 86th Annual Meeting of American Educational Research Association (AERA), Montreal, Canada.

Li, L. & Verhaeghe, J. P. (2004). Fostering metacognition through Peer-tutoring in Chinese schools: interaction processes in cross-age and same-age reading dyads. Paper presented at the 85th Annual Meeting of American Educational Research Association (AERA), San Diego.

Marsh, H. W., Cairns, L., Relich, J., Barnes, J., & Debus, R. L. (1984). The relationship between dimensions of self-attribution and dimensions of self-concept. J*ournal of Educational Psychology*, 76(1), 3-32.

Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review*, 23, 59-80.

Mathes, P.G., Torgesen, J. K., & Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal*, 38, 371-410.

Nicholls, J. G. (1978). The development of the concept of effort and ability, perception of academic attainment and the understanding that difficult tasks require more ability. *Child Developmen*t, 49(3), 800-814.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.

Palincsar, A. S., David, Y. M., Winn, J. A., & Stevens, D. D. (1991). Examining the context of strategy instruction. *Remedial and Special Instruction*, 12, 43-53.

Pressley, M. & Allington, R. (1999). What should reading instructional research be the research of?, *Issues in Education,* 5, 1-35.

Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal*, 92, 513-555.

Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J., & Kurita, J. (1989). Strategies that improve children's memory and comprehension of text. *The Elementary School Journal,* 90, 3-32.

Raphael, T.E. (2000). Lessons from the Book Club Project. In B.M. Taylor, M.F. Graves, & P. Van Den Broeck (Eds.), *Reading for Meaning. Fostering Comprehension in the Middle Grades* (pp. 70-94). New York: Teachers Colleges Press.

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, Vol.*73, 4, 291-329.

Van Keer, H. (2002). *Reading strategies instruction and peer tutoring in primary school. A quasi-experimental study of the effects on students' reading comprehension and metacognition.* Proefschrift ingediend tot het behalen van de academische graad van Doctor in de Pedagogische Wetenschappen (Ghent University), Promotor: Prof. Dr. Jean Pierre Verhaeghe.

Van Keer, H. (2004). Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring. *British Journal of Educational Psychology*, 74(1), 37-70.

Veenman, M. V. J. & Van Hout-Wolters, B. H. A. M. (2002). Het meten van metacognitieve vaardigheden. In F. Daems, R. Rymenans & G. Rogiest (Eds.), *Onderwijsonderzoek in Nederland en Vlaanderen. Proceeding van de 29ste Onderwijsresearchdagen 2002 te Antwerpen* (pp. 102 – 103). Antwerpen: Universiteit Antwerpen.

Chapter 6

The Effects of Reading Strategy Instruction and Peer Tutoring on 2nd Graders' Use of Reading Strategies, Metacognitive Awareness and Reading Attitude*

1. Introduction

This chapter is a parallel chapter to the previous Chapter 5. It reports the effects of the same intervention in the second grade. The results on equivalent measures, i.e. students' reading strategy use, metacognitive awareness and reading attitude will be reported here. The theoretical grounds and experimental design are the same as in the 5th grade, and will be not repeated here. No significant intervention effects were found on any of the aforementioned measurements. To complete the picture of the whole study, this chapter will briefly describe how the instruments were constructed and then briefly present the results. At the end, a brief comparison with the Flemish study will be made.

Research hypotheses

Parallel to Chapter 5, this chapter investigates the effects of reading strategy instruction and peer tutoring practice on 2nd grade students' use of reading strategies, metacognitive awareness and reading attitude. The following hypotheses are stated:

* This chapter is not to be submitted for publication.

(1) Explicit instruction of reading strategies improves significantly 2nd grade students' use of reading strategies. Compared to students from control groups, and adjusting for various influences from student characteristics and their family background, a significant increase in. self-reported use of reading strategies will be observed for students in the experimental conditions

(2) Explicit instruction of reading strategies has a positive effect on 2nd grade students' reading attitude. Compared to those in the control condition, and adjusting for various influences from student characteristics and students' family background, a significant increase in their scores for positive reading attitude will be observed for students in the experimental conditions. Similarly, a significant decrease in their scores for negative reading attitude will be found for students in the experimental conditions.

(3) Explicit instruction of reading strategies has a positive effect on 2nd grade students' evolution in metacognitive awareness. For students in the experimental conditions significant increase in their scores for positive metacognitive awareness will be observed, compared to those in the control condition, having adjusted for various influences from students characteristics and their family background. Similarly, for the experimental groups a significant decrease in scores for negative metacognitive awareness will be observed.

(4) Peer-tutoring practice adds extra value to the experimental effects of reading strategies instruction. That is, for each of the aforementioned three aspects significant differences between the CA and RSA conditions on the one hand and the RS condition on the other hand will be observed, in favour of both peer tutoring conditions. In addition significant differences between the CA and the RSA conditions will be observed in favour of the CA condition.

2. Method

Sample

A total of about 900 2nd graders participated in this study, with about 40% of them in the control condition, 15% in the reading strategy condition, and about 23% in both the CA and the RSA condition. About half of the students are girls.

Design

The study can be characterized as a quasi-experimental intervention study, taking place in natural classroom settings, with whole class groups assigned to one of the three experimental conditions, using a pretest-posttest control group design. Data on students' strategy use, metacognitive awareness, and attitudes towards reading were collected prior to the start of the intervention and after finishing it.

Instrumentation

Measures for strategy use, metacognitive awareness and attitude towards reading are derived from data collected by means of a questionnaire containing 74 items with 5-point Likert scales. To avoid bias from the limited literacy level and reading ability of the 2nd graders, the questionnaires were read aloud by teachers, with thorough explanation on how to answer the questionnaire by taking one of five choices from "always" (accompanied by a very big circle) to "never" (accompanied by a very small circle). Whenever necessary, explanations were given about the statements in the questionnaire. Most items reflect some cognitive or metacognitive act that may occur during or shortly after reading. Students were asked to mark the alternatives that fit them best. To construct the scales, principal component analyses with Varimax rotation were conducted for pretest and posttest data, and for each of the variables (strategy use, reading attitude and metacognitive awareness separately. For each of the scales, very similar component structures were obtained for pretest and posttest data. Consequently, for pretest and posttest

the scales were constructed in the same way, using sum scores based on the same sets of items.

Use of reading strategies

The measure for use of reading strategies is based on items describing cognitive acts such as *"I stop to think: what happened first, then next?"* or *"I find [the meaning of] an unknown word from the context"*. Using a variety of wordings, the items reflect mainly the six reading strategies that were taught to the students in the program. Based on principle components analysis with Varimax rotation and reliability analyses, an a reliable scale was constructed, using 14 of the originally 27 items. Cronbach's alpha is 0.78 for the pretest data (N = 597) and 0.86 for the posttest data (N = 699). The scores on the scale for *use of reading strategies* range from 1 to 5, with a mean of 3.32 and SD = 0.74 for the pretest, and a mean of 3.76 with SD = 0.79 for the posttest.

Metacognitive awareness

The measure for *metacognitive awareness* focuses on students' awareness about themselves, the reading tasks and their performance in relation to reading tasks. Principle component analysis with Varimax rotation suggests a scale for *positive awareness* (based on 6 items such as: *"I think: I can tell this story"*), with a Cronbach's alpha coefficient of 0.72 for the pretest (N = 757) and 0.67 for the posttest (N = 806); and a scale for *negative awareness* (based on 10 items such as: *"I think: I don't understand, this is too difficult for me"* or *"I think: I can do this well only because I get help"*), with a Cronbach's alpha of 0.85 for the pretest data (N = 660) and 0.86 for the posttest data (N = 710) respectively.

The average scores for positive metacognitive awareness are normally distributed, ranging from 1 to 5 (M = 3.61, SD = 0.91) for the pretest and from 1.17 to 5.00 (M = 3.88, SD =0.78) for the posttest. The average scores for negative metacognitive awareness have comparable distributions, (M = 2.3, SD=0.93 for pretest and M = 2.38, SD=0.95 for posttest)

Positive and negative awareness are not correlated with each other (r = .03, p=0.382, N = 847 for the pretest and r =.062, p= .065, N = 895 for the posttest). From Fig. 1 and 2, it can be seen that a high incidence of positive metacognitive awareness can occur together with low, middle or high incidence of negative awareness. But a high incidence of negative awareness is more likely to occur with a high incidence of positive awareness and less likely to occur with a low incidence of positive awareness. This relationship shows that both scales cannot simply be subtracted one from the other in order to obtain a unidimensional metacognitive awareness scale.

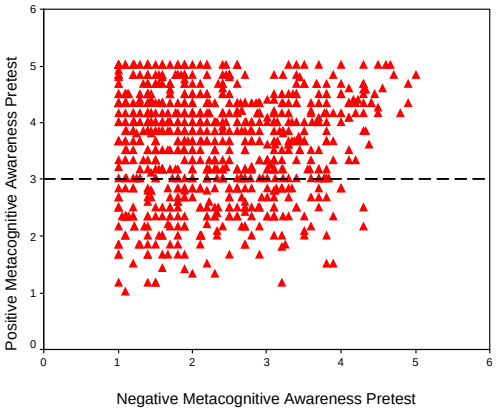


Fig.1. Scatterplot of negative vs. positive metacognitive awareness in pretest

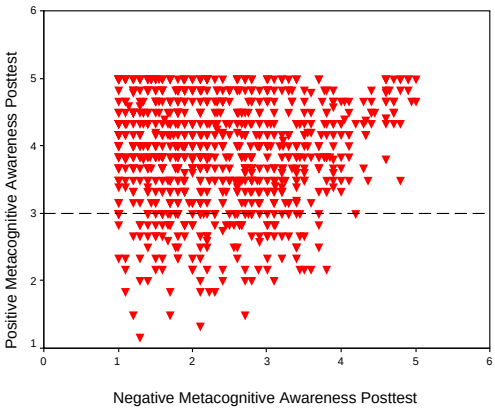


Fig. 2. Scatterplot of negative vs. positive metacognitive awareness in posttest

Reading attitude

To create a measure for reading attitude, 27 items based on Aarnoutse (1996) were presented to the students. The results of a principal component analysis suggest a positive subscale (10 items, alpha = .82, N = 674 for the pretest; alpha = .82, N = 744 for the posttest) and a negative subscale (9 items, alpha = .81, N = 633 for the pretest; alpha = .88, N = 744 for the posttest). For the positive subscale, average scores ranged from 1.00 to 5 (M = 4.11, SD = 0.79) for the pretest and from 1.00 to

5.00 (M = 4.25, SD = 0.73) for the posttest. The distributions for the negative subscale appears to mirror these characteristics, with average scores ranging from 1 to 5.00 (M = 2.38, SD = 0.95,) for the pretest and from 1 to 5.00 (M = 2.31, SD = 1.05) for the posttest. Both scales (average scores) correlate negatively with each other (r = -.23, p < .001, N = 846 for the pretest and r = -.29, p < .001, N = 884 for the posttest).

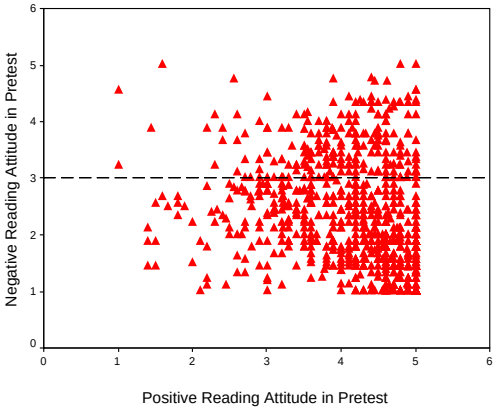


Fig. 3. Scatterplot of positive vs. negative reading attitude in pretest

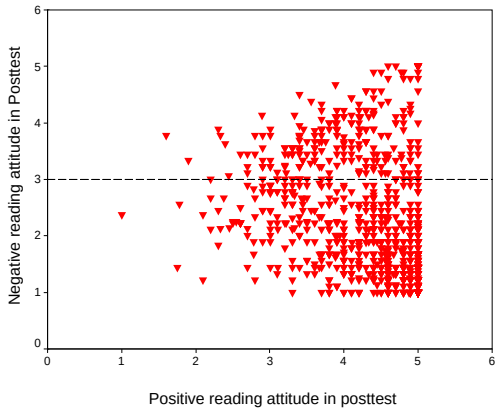


Fig. 4. Scatterplot of positive vs. negative reading attitude in posttest

Family background

Data on students' family background were collected by means of a parents' questionnaire. Parents were asked to provide information on their educational level, profession, family monthly income, parents' reading behaviors, reading interests and attitudes, early childhood rearing practices with regard to literacy education and children's actual reading behavior at home.

Analysis

We used the same multilevel with repeated measures approach to analyze the data in the 2nd grade as we did for the 5th grade data (see Chapter 5).

3. Results

Results in the Chinese study

Results for the five dependent variables (use of reading strategies, positive and negative metacognitive awareness, positive and negative reading attitude) are reported in Table 1 to Table 5 in the Appendix. No significant intervention effects were found on any of these five dependent variables. No significant differences were found between any of the four conditions.

There is in general a moderate correlation (r = .30) between the pretest and posttest results for most of the subscales. Use of reading strategies has a moderate correlation with metacognitive awareness and reading attitude (r = .30) in both pretest and posttest. From Table 1 to 5 in the Appendix, we can see that the mean differences between different groups on all the five subscales are rather small, but the standard errors are all relatively big. These five subscales probably didn't accurately measure what was intended to measure in the second grade students.

Comparison between the Chinese and the Flemish study

In the Chinese study, we only organized pretest and posttest questionnaires, just to evaluate the short-term intervention effects on the 2nd graders' use of reading strategy, metacognitive awareness and reading attitude. Our rational was that students' self-reported results on the three aspects right after the intervention could be attributed to the influence of the intervention., We assumed that, six months after the intervention, we could hardly trace the influence of intervention on any of the five dependent variables. Even if there would be any differences detected, we can hardly make any claim of the impact of the intervention on them. However, in the Flemish study the questionnaire was administered three times: at the pretest, the posttest and the retention test. The two Flemish studies showed a significant increase in the use of reading strategies in most of the experimental conditions at the posttest, and the CA condition even demonstrated a significant increase of

reading strategy use at retention test. In the 2nd study, a significant increase of positive self-related thoughts (metacognitive awareness) was also found.

4. Discussion

The results of the Flemish study (Van Keer, 2002) are more encouraging than the results of the Chinese study. The Flemish results fit the assumptions better than the Chinese results. The rather low correlations between pretest and posttest data in the Chinese study make us doubt about the reliability of the questionnaire for measuring students' use of reading strategy, metacognitive awareness and reading attitude. Compared to the 5th graders, Chinese 2nd graders don't seem to have clear and consistent concepts of their reading behaviors or attitude and awareness related to it. According to the Chinese language curriculum (Chinese Ministry of Education, 2001), the main task of the first year in Chinese primary school is to learn Pinyin, simple characters and words. The reading comprehension training in the first grade is not more than sentence comprehension. With such little experience in reading, most of students probably didn't have clear and stable ideas about their reading behaviors yet. Therefore, we are not sure about how well we have captured students' use of reading strategies and reading attitude in the pretest questionnaire. This may explain the low correlations between pretest and posttest.

Another reason might be related to the format the questionnaire, the five-point Likert scale. Although there are different sizes of circles accompanying the five alternatives to help students distinguish the magnitude of them. Although the questionnaire was filled in under the guidance of teachers, the cognitive tasks involved in responding to the items seemed to be quite challenging for the Chinese second graders.

5. Conclusions

The results didn't support any of the hypotheses we stated in the introduction. There are no clear indications that the explicit teaching of reading strategies nor the peer tutoring practice had benefits on 2nd grader' self-reported use of reading strategies, their metacognitive awareness or their reading attitude.

Appendix: Table 1. Multilevel analyses on 2nd graders' use of reading strategy

	Model 0	Model 1	Model 2	Model 3
Fixed effects				
Cons (Intercept)	3.303	3.262	3.132	3.198
SE	*0.055*	*0.078*	*0.088*	*0.071*
Posttest	0.454	0.381	0.379	0.413
SE	*0.055*	*0.078*	*0.078*	*0.079*
Nega. MC awareness				0.192
SE				*0.021*
Positive reading attitude				0.377
SE				*0.023*
City key school			0.312	0.251
SE			*0.099*	*0.079*
District key school			0.095	0.076
SE			*0.108*	*0.082*
Experimental conditions				
RS		0.017	0.060	0.018
SE		*0.132*	*0.128*	*0.103*
RSA		0.059	0.114	0.041
SE		*0.119*	*0.117*	*0.093*
CA		0.112	0.171	0.072
SE		*0.117*	*0.115*	*0.093*
SO * posttest		0.160	0.167	0.124
SE		*0.137*	*0.137*	*0.138*
RSA * posttest		0.166	0.169	0.165
SE		*0.123*	*0.123*	*0.123*
CA * posttest		0.045	0.049	0.001
SE		*0.121*	*0.121*	*0.122*
Random effects				
Class				
Cons / Cons	0.057	0.058	0.047	0.025
SE	*0.021*	*0.021*	*0.018*	*0.010*
Posttest / Cons	-0.034	-0.028	-0.030	-0.021
SE	*0.017*	*0.016*	*0.015*	*0.011*
Posttest / Posttest	0.046	0.038	0.038	0.036
SE	*0.021*	*0.018*	*0.018*	*0.018*
Student				
Cons / Cons	0.494	0.494	0.494	*0.355*
SE	*0.025*	*0.025*	*0.025*	*0.018*
Posttest / Cons	-0.332	-0.334	-0.334	-0.283
SE	*0.026*	*0.026*	*0.026*	*0.022*
Posttest / Posttest	0.778	0.778	0.777	0.778
SE	*0.041*	*0.041*	*0.041*	*0.041*
Occasion				
Cons / Cons	0	0	0	0
SE	*0*	*0*	*0*	*0*
-2log(lh)	3646.67	3641.07	3632.28	3192.26
N (cases)	1633	1633	1633	1562

Table 2. Multilevel analyses on 2nd graders' positive metacognitive awareness

	Model 0	Model1	Model 2
Fixed effects			
Cons (Intercept)	3.602	3.547	3.590
SE	*0.059*	*0.074*	*0.069*
Posttest	0.282	0.348	0.334
SE	*0.057*	*0.100*	*0.101*
Mothers' diploma		0.053	0.036
SE		*0.018*	*0.017*
Use of reading strategies pretest		0.496	0.373
SE		*0.030*	*0.032*
Positive reading attitude pretest		0.496	0.278
SE		*0.030*	*0.030*
Experimental conditions			
RS		0.103	0.049
SE		*0.124*	*0.118*
RSA		0.051	-0.030
SE		*0.111*	*0.105*
CA		0.009	-0.054
SE		*0.111*	*0.105*
SO * posttest		-0.080	-0.062
SE		*0.171*	*0.173*
RSA * posttest		-0.102	-0.093
SE		*0.154*	*0.155*
CA * posttest		-0.103	-0.067
SE		*0.153*	*0.155*
Random effects			
Class			
Cons / Cons	0.059	0.033	0.027
SE	*0.024*	*0.015*	*0.013*
Posttest / Cons	-0.049	-0.039	-0.037
SE	*0.021*	*0.019*	*0.018*
Posttest / Posttest	0.045	0.059	0.061
SE	*0.022*	*0.028*	*0.029*
Student			
Cons / Cons	0.773	0.536	0.492
SE	*0.038*	*0.030*	*0.027*
Posttest / Cons	-0.602	-0.474	-0.465
SE	*0.038*	*0.035*	*0.034*
Posttest / Posttest	1.038	1.026	1.025
SE	*0.053*	*0.058*	*0.058*
Occasion			
Cons / Cons	0	0	0
SE	*0*	*0*	*0*
-2log(lh)	4092.34	3027.18	2926.90
N (cases)	1668	1316	1308

Table 3. Multilevel analysis on 2nd graders' negative metacognitive awareness

	Model 0	Model1	Model 2
Fixed effects			
Cons (Intercept)	2.325	2.366	2.402
SE	*0.080*	*0.100*	*0.072*
Posttest	0.053	0.113	0.041
SE	*0.085*	*0.113*	*0.117*
Girl			-0.124
SE			*0.047*
Mothers' diploma			-0.077
SE			*.021*
Negative reading attitude pretest			0.405
SE			*0.027*
Use of reading strategies pretest			0.245
SE			*0.033*
Experimental conditions			
RS		-0.087	-0.064
SE		*0.166*	*0.115*
RSA		-0.224	-0.191
SE		*0.151*	*0.103*
CA		0.101	0.107
SE		*0.148*	*0.103*
SO * posttest		-0.002	0.035
SE		*0.188*	*0.192*
RSA * posttest		0.071	0.100
SE		*0.176*	*0.179*
CA * posttest		-0.332	-0.220
SE		*0.168*	*0.176*
Random effects			
Class			
Cons / Cons	0.128	0.102	0.023
SE	*0.044*	*0.036*	*0.012*
Posttest / Cons	-0.076	-0.054	-0.024
SE	*0.038*	*0.032*	*0.018*
Posttest / Posttest	0.139	0.124	0.120
SE	*0.049*	*0.045*	*0.045*
Student			
Cons / Cons	0.732	0.732	0.507
SE	*0.036*	*0.036*	*0.028*
Posttest / Cons	-0.455	-0.456	-0.368
SE	*0.034*	*0.034*	*0.031*
Posttest / Posttest	0.962	0.960	0.917
SE	*0.050*	*0.050*	*0.052*
Occasion			
Cons / Cons	0	0	0
SE	*0*	*0*	*0*
-2log(lh)	4250.210	4243.173	3045.79
N (cases)	1669	1669	1311

Table 4. multilevel analysis on 2nd graders' positive reading attitude

	Model 0	Model1	Model 2
Fixed effects			
Cons (Intercept)	4.103	4.004	3.927
SE	*0.048*	*0.071*	*0.063*
Posttest	0.133	0.147	0.108
SE	*0.046*	*0.069*	*0.074*
Girl			0.229
SE			*0.042*
Mothers' diploma			0.037
SE			*0.018*
Nega. MC awareness pretest			-0.199
SE			*0.026*
Use of reading strategies pretest			0.393
SE			*0.031*
Experimental conditions			
RS		0.121	0.105
SE		*0.125*	*0.102*
RSA		0.207	0.141
SE		*0.111*	*0.091*
CA		0.135	0.109
SE		*0.110*	*0.091*
SO * posttest		-0.035	-0.048
SE		*0.124*	*0.127*
RSA * posttest		-0.080	-0.053
SE		*0.110*	*0.114*
CA * posttest		0.051	0.059
SE		*0.109*	*0.114*
Random effects			
Class			
Cons / Cons	0.038	0.036	0.016
SE	*0.016*	*0.016*	*0.009*
Posttest / Cons	-0.021	-0.017	-0.015
SE	*0.013*	*0.012*	*0.022*
Posttest / Posttest	0.029	0.025	0.014
SE	*0.015*	*0.014*	
Student			
Cons / Cons	0.595	0.593	0.429
SE	*0.029*	*0.029*	*0.024*
Posttest / Cons	-0.397	-0.398	-0.321
SE	*0.027*	*0.027*	*0.026*
Posttest / Posttest	0.714	0.715	0.713
SE	*0.037*	*0.037*	*0.041*
Occasion			
Cons / Cons	0	0	0
SE	*0*	*0*	*0*
-2log(lh)	3671.99	3665.71	2689.69
N (cases)	1658	1658	1306

Table 5. Multilevel analysis on 2nd graders' negative reading attitude

	Model 0	Model1	Model 2	Model 3
Fixed effects				
Cons (Intercept)	2.401	2.494	2.556	2.552
SE	*0.081*	*0.105*	*0.069*	*0.080*
Posttest	-0.089	-0.006	-0.005	-0.007
SE	*0.073*	*0.104*	*0.110*	*0.109*
Girl			-0.161	-0.157
SE			*0.051*	*0.051*
Mothers' diploma			-0.084	-0.072
SE			*0.021*	*0.024*
Nega. MC awareness			0.477	0.478
SE			*0.031*	*0.031*
Use of RS at pretest			-0.117	-0.109
SE			*0.037*	*0.037*
City key school				-0.052
SE				*0.098*
District key school				0.174
SE				*0.084*
Experimental conditions				
RS		-0.046	-0.005	-0.063
SE		*0.174*	*0.110*	*0.102*
RSA		-0.278	-0.157	-0.188
SE		*0.158*	*0.098*	*0.095*
CA		-0.100	-0.140	-0.159
SE		*0.156*	*0.099*	*0.094*
SO * posttest		-0.067	-0.156	-0.154
SE		*0.180*	*0.187*	*0.185*
RSA * posttest		-0.060	-0.108	-0.095
SE		*0.163*	*0.171*	*0.170*
CA * posttest		-0.251	-0.257	-0.258
SE		*0.160*	*0.169*	*0.167*
Random effects				
Class				
Cons / Cons	0.131	0.118	0.015	0.006
SE	*0.045*	*0.041*	*0.010*	*0.007*
Posttest / Cons	-0.065	-0.060	-0.017	-0.005
SE	*0.033*	*0.030*	*0.015*	*0.012*
Posttest / Posttest	0.087	0.079	0.077	0.077
SE	*0.036*	*0.034*	*0.035*	*0.035*
Student				
Cons / Cons	0.769	0.768	0.570	*0.570*
SE	*0.038*	*0.038*	*0.032*	*0.032*
Posttest / Cons	-0.460	-0.462	-0.422	-0.422
SE	*0.038*	*0.038*	*0.036*	*0.036*
Posttest / Posttest	1.181	1.179	1.118	1.119
SE	*0.062*	*0.062*	*0.064*	*0.064*
Occasion Cons / Cons (SE)	0 *(0)*	0 *(0)*	0 *(0)*	0 *(0)*
-2log(lh)	4462.69	4453.61	3225.71	3222.00
N (cases)	1650	1650	1304	1304

References

Aarnoutse, C.A.J. (1996). *Leesattitudeschaal* [Reading attitude scale]. Nijmegen (NL): Berkhout.

Chinese Ministry of Education (2001). *Quan ri zhi jiu nian yi wu jiao yu yu wen ke cheng biao zhun* [*Full-time Nine-Year Compulsory Education Chinese Curriculum Standards*], Chinese Ministry of Education.

Van Keer, H. (2002). *Reading strategies instruction and peer tutoring in primary school. A quasi-experimental study of the effects on students' reading comprehension and metacognition.* Proefschrift ingediend tot het behalen van de academische graad van Doctor in de Pedagogische Wetenschappen (Ghent University), Promotor: Prof. Dr. Jean Pierre Verhaeghe.

General Discussion

1. Research background and theoretical scheme revisited

In this thesis, the findings from an intervention study in Chinese primary education are reported. The intervention program combined two lines of research in education – reading strategies instruction and peer tutoring. All students in the experimental conditions received direct instruction on reading strategies. In addition, these strategies were further practiced in cross-age (CA) or reciprocal same-age (RSA) dyads or by students individually (RS) under the teachers' help.

Our intervention program was based on an original Dutch program (Fukkink *et al*., 1997) and was also partly a replication of a similar study with Flemish primary school pupils (Van Keer, 2002, 2004; Van Keer & Verhaeghe, 2003, 2005). The main goals of our study, as is indicated in the left-hand side of the theoretical scheme, were to find out whether

(1) The positive effects of reading strategy instruction on students' reading comprehension growth that are frequently reported in western reading research literature, could also be found in Chinese primary school pupils;

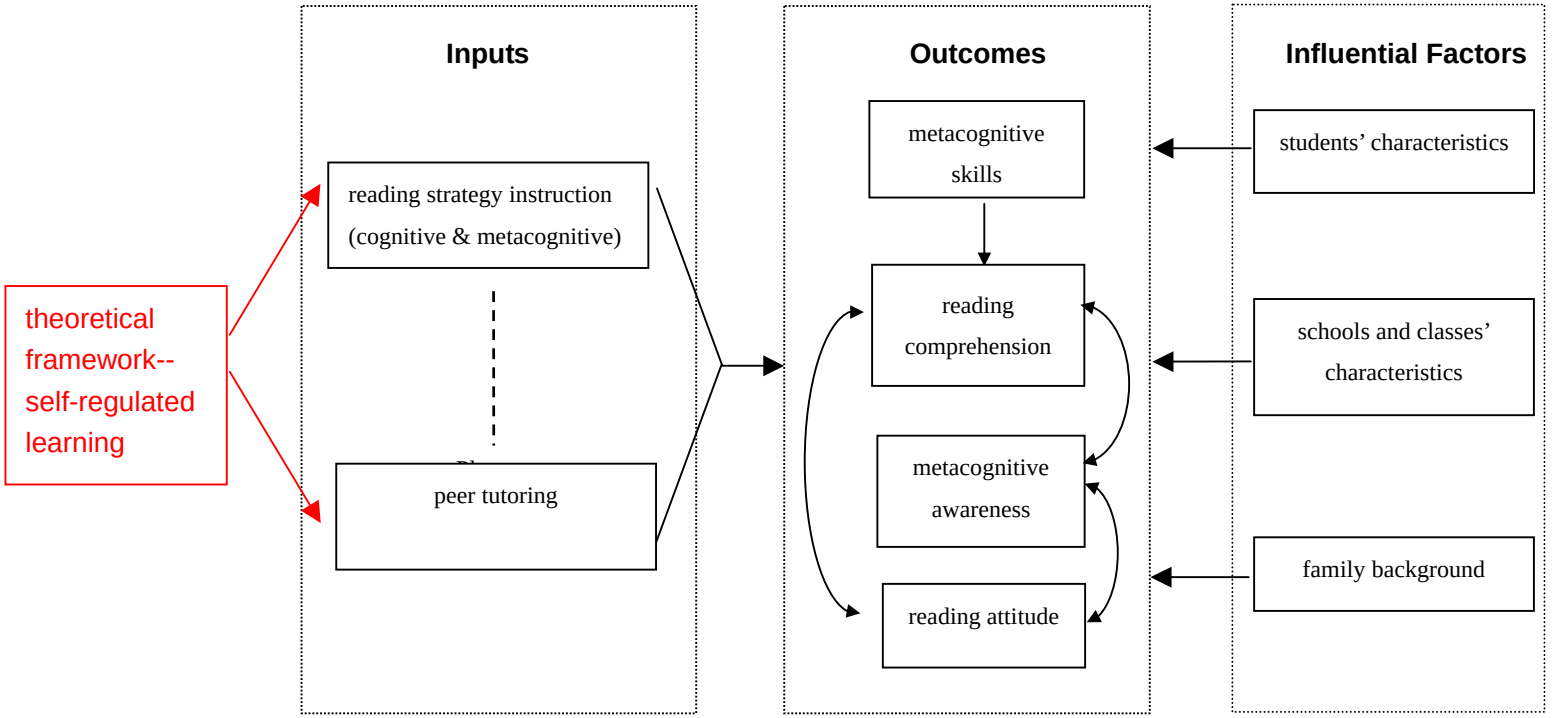
(2) The use of peer tutoring for practicing reading strategies would yield any extra value for Chinese students' reading comprehension.

When investigating these effects, the influential factors (as indicated on the right hand side of the scheme) across conditions were adjusted for.

We explored the treatment effects on students' reading comprehension growth in the short term and in the long run, both in 2nd and 5th grade. In addition to the two

central research topics, a number of other related questions were also explored: effects on students' use of reading strategies, reading attitude and metacognitive awareness. To gain a deeper insight in the processes that lead to the treatment effects, an in-depth investigation of interaction processes of peer tutoring dyads, and how these processes differ across different types of tutoring conditions, was also carried out. In these analyses we also adjusted for effects of students' initial reading level and gender.

Theoretical Scheme



2. Summary of significant findings

Table 1 and 2 report the main findings from our intervention study.

Table 1. Findings on reading comprehension, use of reading strategies, metacognitive awareness and reading attitude

		Effect Size			
	Measurements	RS	RSA	CA	Exp*
G5	reading comprehension				
	short-term effect	—	—	small	medium
	long-term effect	medium	medium	medium	large
	use of reading strategy	small	small	small	/
	reading attitude				
	positive	medium	—	small	/
	negative	—	—	—	/
	metacog. Awareness				
	Positive	—	—	small	/
	Negative	—	—	—	/
G2	reading comprehension				
	short-term effect	medium	marginal	medium	/
	long-term effect	—	—	medium	/
	use of reading strategy	—	—	—	/
	reading attitude				
	positive	—	—	—	/
	negative	—	—	—	/
	metacog. awareness				
	positive	—	—	—	/
	negative	—	—	—	/

Notes: * EXP, a merged experimental group with RS, CA, RSA groups;
— nonsignificant; / not investigated.

Table 2. Significant finding on CA-RSA differences on dyad interaction during reading process

	Effect Size
Reading activities	CA - RSA
strategic corrections by tutors	large
Non-strategic corrections by tutors	small
Overall reading strategy use by the tutor	large
Spontaneous use of reading strategies in tutor questions	large
Spontaneous use of reading strategies in tutor explanations	large
Overall frequency of “what”-level exchanges	large

The detailed results and more precise effect sizes have been reported in Chapters 2 to 6. In answering the research questions, the overall effects of the intervention could be summarized as follows:

(1) Reading strategy instruction is quite effective in improving students’ reading comprehension skills in both grades. All experimental conditions in the fifth grade demonstrated long-term gains, and all experimental conditions in the second grade had short-term gains in reading comprehension. In addition, reading strategy instruction has also significantly increased the (self-reported) use of reading strategies by fifth graders.

(2) When no distinction is made between the two types of peer tutoring, no significant extra value is added by peer tutoring practice. No significant differences were found between peer tutoring groups (CA and RSA together) and RS groups.

(3) CA students have benefited most from the intervention. In both grades the most consistent results (short term plus long term effects) for reading comprehension were found for the CA condition. In addition, for CA tutors, significant effects

were found with respect to the use of reading strategies, reading attitudes and positive metacognitive awareness.

(4) Comparing CA and RSA within-dyad interaction processes, CA dyads were significantly better than RSA dyads on a number of indicators of metacognitive activities, as is summarized in Table 2. This parallels well the findings that in 2nd grade, CA students are also significantly better than RSA dyads in keeping the intervention gains in reading comprehension for 6 months after the intervention.

(5) With regard to metacognitive awareness and reading attitudes, very few significant results were found. In Grade 5, students in CA showed significantly better results on positive awareness; for positive reading attitude both RS and CA showed better results. In Grade 2 no significant intervention effects were found at all in these dependent variables.

In addition to the aforementioned findings, there are also a few inconsistencies among the findings:

(a) There is some inconsistency between the observational study and the main study with regard to the differences between the two types of peer tutoring.

In the observation study, for a number of dyads, we found that CA tutors differed significantly from students in RSA dyads on a number of metacognitive activities, such as strategic monitoring, strategic questioning and explanations, and strategic information exchange (see Table 2). But when comparing CA and RSA self-report questionnaire data, we didn't find any significant differences between the two groups, neither in the use of reading strategies nor with respect to metacognitive awareness. In our study in the fifth grade, CA and RSA didn't differ in long-term reading comprehension gains, either. These inconsistencies could be partly due to the fact that when designing the observational study, we chose to observe only the strong and weak reading dyads, assuming to find significant differences between these high and low reading level dyads in strategic reading behaviors, and left the

middle level dyads unobserved. A reason for not having detected significant differences between CA and RSA dyads in the questionnaire data could be attributed to the limited power of data, as is discussed in this Chapter 5. However, in general, in both Grade 2 and Grade 5, CA students had relatively better results in the main study (Chapter 3 – 5). Therefore, there is a general agreement with the observational study.

(b) There is also an inconsistency between the findings for the growth in reading comprehension and the results for the self-reported use of reading strategies and metacognitive awareness in the second grade. As is summarized in Table 1, in the three conditions in the second grade there are significant short-term effects on reading comprehension growth, but we didn't find any parallel significant results for *reading strategy use* and *metacognitive awareness*. As is discussed earlier in Chapter 6, this is either because students have difficulties in matching what is described in the five-point Likert scale questionnaire with their general reading situation and experiences during reading, or – at least for these students – the questionnaire is not a reliable instrument to measure the use of reading strategies and metacognitive awareness.

3. Comparisons between Chinese and Flemish results

The similarities and dissimilarities between the Chinese and Flemish results can be summarized as follows:

(1) Reading strategy instruction

With regard to the intervention results in the 2nd grade, in the Chinese study, we have found significant short-term effects for all conditions — RS, CA, RSA — and significant long-term effects for the CA condition. CA is also significantly better than RSA in keeping the intervention gains in the long run. In the Flemish first study, results were less encouraging for the 2nd graders. There were no significant gains in any of the conditions in either posttest or retention test.

With regard to reading comprehension growth in the 5th grade, in the Chinese study, we have significant short-term effects in the CA condition and significant long-term effects in all three experimental conditions, although no significant differences were found among the three conditions on either short-term or long-term effects. In the Flemish study, short-term effects for RS and CA conditions were found; also significant long-term effects were found for CA condition; CA was significantly better than RSA on the long-term effects. It was on this latter result that we based the main hypothesis for our observational study to CA and RSA dyads. It is remarkable that this hypothesis was confirmed in our observational study, but no parallel finding was found in the fifth grade in our Chinese study on reading comprehension.

(2) Use of reading strategies, reading attitude and metacognitive awareness

With regard to the intervention effects on these three measures in the 2nd grade, no significant results were found in Chinese study. In contrast, the Flemish study was more encouraging. Significant intervention effects on the use of reading strategies were found in all three conditions. For the fifth grade, the Chinese-Flemish differences are reversed. In our study, in all three conditions we found significant intervention effects on use of reading strategies; significant intervention effects on positive reading attitude and positive metacognitive awareness in CA condition; significant intervention effects in positive reading attitude in RS condition. In the Flemish first study, only significant results were found on negative metacognitive awareness in RSA condition at retention test.

(3) The value of peer-tutoring practice

With regard to the hypothesis that peer-tutoring practice would add extra value on top of reading strategy instruction, neither the Flemish nor the Chinese study confirmed this hypothesis. Among the Chinese fifth graders, there is little difference between the experimental conditions and all of them have a medium effect size in the retention test. The reading strategy groups made comparably large

progress at the posttest as the two peer tutoring groups. In the retention test, the effect size of RS is even bigger (0.56) than for the CA (0.48) and RSA (0.44) conditions.

Among the Chinese second graders, CA has shown relatively better results than RS and RSA conditions, but there is no significant difference with any of the two groups either. Only between posttest and retention test, CA is marginally significantly better ($p = .054$) than RSA in keeping the experimental gains. In the Flemish study, for second grade no effects were found at all.

In reflecting on the implementation processes in the present Chinese study, we think that the value of peer tutoring practice might have been jeopardized by some practical restrictions, such as too large class size in combination with restricted physical space and amplification of the noise level by the dyads' discussions. In addition, although we have taken a great effort to enforce the treatment fidelity, as is described in Chapter 1, taking into account the scale of the intervention (more than 1,500 students in total), there was more danger for implementation infidelity compared to small scale interventions (Slavin, 1995).

With regard to the value of peer tutoring practice, another knowledge we have gained from both the Flemish and Chinese studies is that we should make a clear distinction between CA and RSA conditions, since these two types of peer tutoring differ considerably in their effects on a variety of indicators. These differences could be explained by information processing theory (e.g., Biemiller *et al*., 1998; Winne, 2001), and "shared limited capacity" theory in linguistics (Chall, 1983; Chen, Lau, Yung, 1993; Norman & Bobrow, 1975; Perfetti & Hogaboam, 1975; Rumelhart, 1978), as is discussed in Chapter 2.

In summary, we could say that the main differences in results between the Chinese and the Flemish studies regard the effect and value of cross-age tutoring, more specifically, who benefits most from cross-age peer tutoring experiences, the tutors

or the tutees. Based on the results from both studies, it is the Flemish tutors and Chinese tutees that have benefited most from the intervention.

When highlighting the differences in results of the intervention, more particularly with respect to the value of cross-age tutoring and its long-term effects on students' reading gains, some quite remarkable differences between Chinese and Flemish results became apparent. With regard to the long-term effect, it is rather remarkable that the Chinese 2nd graders in CA condition could keep their gains for at least six months after the conclusion of the intervention, while in the Flemish first study, there was neither short-term nor long-term effects. The significant short-term and long-term effects on the Chinese second grade tutees seem to fit very well with Vygotsky's *Proximal Development Zone* theory. Under the one-to-one guidance of elder tutors, Chinese tutees have made relatively more gains than 2nd graders in other conditions. However, this was not the case for the Flemish students. Various explanations were sought to address this difference. For example, this could be due to the average age difference between Chinese and Flemish students in the same grade. The difference in official regulations causes Chinese students on average to be about 3 months older than their Flemish counterparts. This age difference was confirmed by the actual data of the students involved in the studies. Since six to eight appears be a period of fast growth in metacognitive development (Corsale & Ornstein, 1980), three months difference in age might make a difference in the mastering of metacognitive skills. But would it be enough to explain the differences between Chinese and Flemish second grade students?

Another explanation for the difference in intervention effects in the second grade could be found in differences between Chinese and Flemish language. As is introduced in the general introduction of this thesis, the Chinese writing system is morpheme based and more complicated concepts are usually a combination of simpler morphemes that have a meaning of their own. In many cases the meaning of the more complex concept can be derived from its constituent morphemes (McBride-Change *et al*., 2003). This feature makes Chinese language more

analytic and transparent than Dutch. Besides, Chinese has little or almost no rules in grammar, words don't change their form and the precise meaning does not depend on the form the word. At the starting stage, students' main obstacle in reading is mainly lacking vocabulary. Fortunately, most Chinese characters consist of two parts: one part, called the phonetic radical, indicates more or less the pronunciation, and the other part, called semantic radical, refers to the meaning of the character (Shu & Anderson, 1997). Getting to know new characters from their morphological characteristics and try to use their phonetic and semantic radicals to remember its pronunciation and meaning is a common practice in language teaching. During our intervention, when teaching Strategy No.2 "enlarging vocabulary", we further strengthened this morphological awareness by encouraging students to guess (and then verify) the meaning of difficult words in reading materials. Probably the combined effects of the morphological characteristics of Chinese language, its morpheme based writing system and the explicit reinforcement from the intervention have fostered Chinese students' analytic and meaning reconstruction skills, which on turn have facilitated the Chinese reading process. For the reason mentioned above (see our reference to Vygotsky's *Zone of Proximal Development*) it is also not surprising that, compared to other second graders, more particularly CA tutees have benefited most from working in tutoring dyads.) It is very well conceivable that at a comparable young age, Flemish students cannot benefit that much yet from instruction in reading strategies and practice of metacognitive skills. With respect to the morphological structure of words (the ways words are formed) Dutch – as other Western languages – is a much more "abstract" language, less transparent for young children. Complex concept are very often expressed in long words with Latin and/or Greek roots; words take different forms (depending whether they are used as a verb, a noun, an adjective), for derived meanings words are formed with "abstract" prefixes and suffixes that by themselves have no meaning, etc. In daily life, in spoken language, Flemish children use these complex language forms intuitively. Because of the rather abstract level of these language features, systematic instruction in these

aspects is postponed to a later age. At a young age, Flemish children have little insight in the morphological structure of their language and therefore cannot benefit much from it to enhance their reading comprehension. So, for young Flemish students, there is not so much in the way a word is written that helps them to find the meaning of a word they do not know yet. Either they already know it, or they don't. Our second strategy is therefore of less use for them compared to the use their Chinese counterparts can make of it. Of course, they can use a dictionary or ask someone else for explanation – but these resources are normally not available when doing a test. The only other way to put this strategy into practice – derive the meaning from the context – requires insight at a much higher level (level of a paragraph or text) and therefore requires in fact the combined use of one of the other strategies as well (finding the meaning of a paragraph or text). In short, it could be argued that – because of its specific linguistic features – for a language as Dutch, reading strategies become only useful at a somewhat later age than for a language such as Chinese. That could explain some of the differences in intervention effects that were found between Flemish and Chinese second graders.

Another remarkable difference between the Chinese and the Flemish study regards the results in the 5th grade. We found long-term effects for all three conditions in Grade 5. CA is not significantly different from the RS or RSA condition and RS is relatively better in effect size, 0.56 in RS vs. 0.48 in CA. In addition to the explanations mentioned in Chapters 3 to 6, such as physical restrictions and big class size, these results seem, to some extend, reflect a difference in knowledge acquisition believes in the two cultures. Influenced by traditional believes, Chinese children, from very early age on, are told to show respect to teachers, to study hard to absorb knowledge and learn skills. Chinese students are assumed to respect teachers or other older persons in their role of knowledge transmitter. That is probably why, for a long time, no strong need or desire for collaborative or cooperative learning was felt among Chinese students and teachers. If at all, cooperative or collaborative learning in China is mainly practiced because of some

external demand from policy makers, who refer to such reasons as "preparation for the future life in modern society" (Yang, 2006; Fu, 2005). In the West, including Flanders, children and teachers are probably more influenced by a belief in the resourcefulness of knowledge co-construction. They believe that they can learn from interaction, and re- and co-construct knowledge from working with peers who don't necessary have more knowledge. Chinese CA tutors (and their teachers) probably tend to believe that the tutor's role is to teach tutees with the knowledge they have and to help them with reading; they probably hold less belief that they can benefit from this helping processes. Such a difference in knowledge acquisition believes might offer an answer to the differences between the Chinese and Flemish results in CA tutors' growth. On the other side, as is mentioned in Chapter 3, Chinese students in the 5th grade are probably influenced by Confucius' "self-perfection" ideas and strive a lot in learning the six reading strategies, since they were considered to be useful for them. It is possible that this strong individual motivation to perform well – that most probably is found equally much in all experimental conditions, overrules any effect of working as a CA tutor. That is probably why we could not detect any significant differences between the three conditions in reading comprehension growth. One could argue then that this Confucian striving for self-perfection would be equally true then for students in the control condition. That is true, but those students did not get reading strategies instruction, and hence were not offered any opportunity to demonstrate the very best of them by practicing those reading strategies.

4. Value of the study

The present Chinese study is a replication of a Dutch (Fukkink *et al*., 1997) and a Flemish intervention program (Van Keer & Verhaeghe, 2003, 2005). In the intervention, we included the same six reading strategies and the same two types of peer tutoring (CA and RSA) groups. Basically the same research design was adopted. On the practical implementation level, the Chinese teacher's manuals

were mainly based on the Flemish versions. For comparison reasons, we translated most of the instruments from Dutch into Chinese.

4.1. Theoretical contributions

(1) ***Reading strategy instruction.*** Our main contribution by carrying out this study in Chinese schools in Chinese reading is that we have shown that Chinese primary school pupils' reading comprehension can also be benefit from reading strategies instruction. Chinese language differs comprehensively from western languages, both in it basic structure and in its writing system. The processes involved in learning to read Chinese (for native Chinese speaking children) might therefore differ in some aspects from the processes involved in western children's learning to read in their own western language. At the beginning, it was not clear at all what the role of reading strategies would be for Chinese primary school children, compared to other aspects in learning to read. In our study, we have shown that reading strategy instruction can also improve Chinese students' reading comprehension achievement. Therefore, we can conclude that reading strategies are useful for reading in different languages.

(2) ***Peer tutoring.*** Although in our data we didn't strictly detect significant differences with other experimental conditions, our study confirmed that cross-age tutoring yield positive and consistent results. As was also indicated by the Flemish study, same-age tutoring is in general less favorable than cross-age tutoring for most of the aspects we have studied. The hypotheses we explored to explain the cross-cultural differences in effects of CA tutoring, may contribute to the understanding of how peer tutoring (and on a more general level: cooperative and collaborative learning) works, and how its effects probably are intertwined with culturally based beliefs about learning.

With respect to the processes involved in peer tutoring, a deeper insight was gained from the observational study (Chapter 2). By providing an in-depth analysis of the interaction processes within tutoring dyads, comparing CA and RSA dyads, this

observational study constitutes an important extension to Van Keer (2002)'s study. The observations and analysis of the interactions were carried out to find an explanation for the Flemish finding that the 5th grade tutors in the CA condition did generally better than RSA students and that CA tutors performed better in keeping what they have gained from the intervention for at least six months after its conclusion. As discussed above, we found a number of significant differences between CA and RSA dyads, all in favor of the CA dyads. These findings help to understand how cross-age tutoring settings provide better opportunities for developing metacognitive skills in tutors.

4.2. Strength in methodology

First, a major strength of our study is to be found in its ecological validity: This study is done in natural classes with their regular Chinese language teachers. To counter the potential threats involved in working with natural classes and regular teachers, we have taken great efforts to ensure treatment fidelity. Ecological validity is also embodied in the sampling (taking care of including students with a broad variety of social backgrounds and different types of schools) and in the analyses that took into account the influences from the broader environment and adjusted for the effect of these variables. Using multilevel analysis, the nested structure of the data was respected.

Second, there is the combination of qualitative and quantitative research approaches. Qualitative data collection and quantitative data analysis are integrated in the observational study. The software we used, proved to be a useful tool for making the transition from qualitative video-data into data that could be processed in a quantitative way. In the subsequent processing of the data, the rather uncommon way of using multilevel analysis enabled us to take into account the nested structure of the time unit based data (smaller time units within larger time units).

A third methodological strength of our study relates to the development of the reading comprehension test and the reliability of the measures derived from the questionnaire data.

We translated the Dutch reading comprehension tests used in the Flemish study into Chinese. To adapt to the Chinese situation, we replaced some texts (e.g., a text explaining a word game that doesn't fit Chinese language at all) with example texts from the same test. Considerable efforts were made to find good Chinese equivalents for specific idioms and expressions in the Dutch texts. Most important of all, we made sure that skills that are tested, such as finding the *meaning of words and expressions, recognizing main ideas, finding some details in the text, inferring the author's attitude and writing purpose, reasoning* and so on were tested equivalently in the Chinese version as in the Dutch version. Trials and a pilot study were carried out in the school year preceding the intervention study. In addition, extra efforts were taken to gather the necessary additional data for performing an IRT-calibration of the different tests we used. Using a three-parameter Item Response Theory model, we calibrated the tests used on three measurement occasions with a time span of one year and a half. As a result, the test scores for the different measurement occasions can be interpreted on one scale. This enabled us to model the growth curves of students' reading comprehension skills over time.

Processing the questionnaire data, we used exploratory factor analysis to construct highly coherent subscales and subsequently performed confirmatory factor analysis using structural equation modeling to check the reliability and overall quality of the measures derived from the questionnaire.

4.3. Practical significance

Based on our study, a number of recommendations for regular school practice can be made.

Generalizing reading strategy instruction to more schools. Both the Chinese and Flemish studies have shown the usefulness of reading strategy instruction for improving students' reading comprehension. We recommend generalizing this practice to more students in more classrooms. For use in China we have prepared very well written documents including teachers' manuals, students' exercise books, and teacher training materials. We have also accumulated experience in checking the implementation quality. All these could be useful for generalizing reading strategy instruction to more Chinese classrooms. For one step further, reading strategy instruction could be integrated into the daily Chinese reading instruction practice with careful planning. We also recommend interweaving the six reading strategy instruction with the content of the regular Chinese language curriculum and foster students' metacognitive awareness in using strategies to detect difficulties in reading and solve reading problems. We would further stress that at lower grades in primary school, teachers should make full use of the possibilities to develop students' awareness of the morphological characteristics of Chinese language and encourage students to derive clues for the meaning and pronunciation of new words from its radicals, to find the meaning of more complicated concepts from simpler morphemes; encourage and foster an analytical attitude in reading and reinforce the will to become independent readers and learners.

Recommendation of cross-age peer tutoring practice. We highly recommend CA peer tutoring as a cooperative learning method in regular classroom or extracurricular activities. We emphasize however that efficient use of CA peer tutoring requires that well-defined skills be practiced. Prior to the work in CA dyads, these skills need to be taught clearly and their practice in CA dyads needs good scaffolding. Also required is a careful preparation of the tutors with respect to the social skills they need to be effective tutors.

5. Limitations of the study and directions for further research

The research tested compressively the research hypotheses and answered the research questions. However, there are some limitations that cannot be overlooked.

(1) Measures on metacognition, metacognitive awareness, reading strategies, reading attitudes

Metacognition is a very crucial concept in the whole study; it is closely related to the effects of reading strategy instruction and the continuous growth in reading comprehension achievement. How to capture students' metacognitive processes remains a challenge in reading research. In the research field, there is no consensus on how to measure students' metacognition in reading. Some use structured interview, often off-line (Canney & Winograd, 1979; De Corte *et al*., 2001; Manion & Alexander, 1997). In addition, *close reading* or *error-detecting* tasks (Paris & Jacobs, 1984) were administered to test the level of students' metacognition in reading. All these methods have their limitations. For example, the structured interview method, either off-line or on-line, is often restricted by the language ability of the subjects. The measurement of the metacognitive level of younger students (e.g., below Grade 2) is often restricted by their ability to express themselves. And interview methods often mislead students to give desired answers (Jacobs & Paris, 1987). Besides, it takes a lot of time to interview and only relatively small numbers of students could be interviewed.

In our studies, we used observation (video analysis) and a self-report questionnaire to capture students' metacognitive level and processes. Observation was done in natural settings and it avoided disturbing students' reading and interaction processes. A self-report questionnaire was used to collect students' general metacognitive awareness. Both methods have their limitations.

A first limitation encountered in the video analysis relates to its restriction to a limited number of peer tutoring dyads and tutoring sessions. Videotaping and

analyzing the interactions in more dyads and sessions per dyad would have enabled us to perform other, also very interesting analyses for which now the statistical power of the dataset is insufficient. It was also impossible to capture individuals' reading activities and metacognitive processes in the RS, FR or CON groups using video recording. Therefore, by natural observation, no overall evaluation of the metacognitive levels of students from groups other than CA and RSA was possible. A third limitation is that based on the videotapes, we can only code the overt behaviors and conversations of the students in the dyads. We mostly had to *infer* what reading strategies they were using. We have no direct measure of their strategic reading behaviors. As there is a convention that metacognitive activities should be "conscious and deliberate" behaviors from the individuals (Hacker, Dunlosky & Graesser, 1998, p. 8), a threat involved in this "inference" coding is that the occurrence of metacognitive processes is overestimated. A lot of what looks "strategic" behaviors demonstrated by tutors or tutees might in fact be skillful reading behaviors operating at an unconscious level, therefore not being really strategic behaviors (Alexander, Graham, & Harris, 1998; Alexander, 2006; Nist & Holschuh, 2000). However, this argument only holds for some of our coding categories. For other categories, things are quite clear. When a tutor corrects a tutee, for example, it cannot be denied that the tutor was monitoring the tutee's reading. When a tutor asks a spontaneous question or gives a spontaneous explanation, in which the use of a reading strategy is apparent, there cannot be much doubt that he or she is actually regulating the tutee's reading behavior. In fact, the warning issued by Alexander (2006) and Nist and Holschuh (2000) would rather hold for inferences made with respect to the behaviors of a student who is reading. In our case, that is the tutee. There is less doubt that the interventions from the tutor are conscious and deliberate acts.

Using the self-report questionnaire we could collect information on the use of reading strategies and metacognitive awareness from all students in the different groups. But it turned out that there was almost no correlation between the

frequency of students' self-reported use of various reading strategies and their reading comprehension achievement, which puts the reliability and validity of the questionnaire in doubt. Moreover in Grade 2, between the subscales for *use of reading strategies*, *reading attitudes* and *metacognitive awareness*, there is only very low correlation (r = .3) between pretest and posttest. In contrast, correlations of .5 to .7 were found in the 5th grade. This low correlation in the second grade casts a lot of doubts on the reliability of these instruments for younger students. Therefore, in future research, more reliable and easily applicable measurements for capturing the use of reading strategies and metacognitive awareness are to be sought or developed.

(2) Measures of teachers' implementation quality and their general teaching performance.

Although great care has been taken to ensure treatment fidelity through frequent visits, use of checklists, regular meetings and discussion, and feedback, we observed some variation among teachers with regard to their teaching and coaching behaviors during the implementation of the intervention program. Many of the teachers' lessons (tutor preparation and introduction of reading strategies) were observed and videotaped and for each of those lessons I have graded teachers' practice on several aspects, such as how well they have understood the content, how well they had prepared for the teaching, how students participated in the class and how well students have understood the content after the introduction lessons and so on. But no independent coding of the videotaped lessons was done, due to time pressure and lack of availability of Chinese research assistants. However, in each school, directors in charge of Chinese teaching were interviewed about the teachers' performance in regular curriculum teaching, their age, professional ranks (in China, teachers are ranked according to their years of service and teaching quality), diplomas and highest diplomas to date. It is reasonable to assume that at least some of these variables could have played a role in the quality of the implementation and hence in the intervention effects. Including them in the

analysis could have refined the estimation of the intervention results. However, we didn't include these data in any of the analyses so far. Therefore, in a follow up analysis, we could try to adjust for the impact of these variables to gain a better estimation of the intervention effects.

(3) Limitations of the observational study

In the observational study, we used multilevel analysis to explore the differences between CA and RSA with respect to the interaction processes within the dyads. The videotapes were made throughout the duration of the intervention program. Principally, using a repeated measures design, it would have been possible to investigate evolutions in behavior and interactions over the course of the intervention. Although available, in the analysis we didn't include a variable indicating the specific period of the year or the reading strategy actually being practiced in the session. Instead, we just estimated the average differences between CA and RSA dyads irrespective of the time of the year or the reading strategy being practiced in that period. A weakness in our analysis method is that it would not reveal any evolution or fluctuations in the interaction processes. With our analysis, we just averaged these fluctuations out. It is well possible that the differences we found are even higher at the end of the intervention (or lower in case of a non-monotonous growth); it is also possible that we didn't detect some small, but significant real existing differences that were there at the end of the intervention program only. On the other hand, we wonder if the limited amount of dyads and observed sessions could give enough statistical power to model differences over time using a repeated measures approach.

(4) Directions for future research

Using the data-set of the present study, some more work could be done to address a number of the limitation discussed above, e.g. refining the analysis by including teachers' variables and data on the quality of their implementation practice; or refining the results of the observational study by including a time variable.

More generally speaking, more efforts are needed to find more reliable and consistent measurements to measure students' metacognition, using triangulation of methods, e.g. combining reading tasks with observation and interview.

With respect to its theoretical framework, the whole study is based on the theory of Self-regulated Learning (SRL), presenting an application in the field of reading, focusing on two elements – learning method and social environment. For future research, it would be meaningful to expand the research from reading to other subjects such as reading in foreign languages (e.g., English), mathematics, science, etc. The experience we have accumulated (e.g., with teacher training, upholding treatment fidelity, etc.) would be useful for such implementations in other fields; and the research methodology (e.g., interaction process analysis, intervention effect evaluation, etc.) could also be used in studying new research questions. In addition, some of the limitations in the present study could be addressed in those new studies in other fields. Research from various fields in self-regulated learning would enrich self-regulated learning theory and help to draw a clearer picture of how to cultivate self-regulated learners.

6. Final conclusions

The following conclusions can be drawn from our study:

(1) Reading strategy instruction in general has had significant positive effects in improving the level of Chinese students' reading comprehension. These gains have lasted for at least six months after concluding the intervention for the CA condition in both Grade 2 and Grade 5 and for the RSA and RS conditions in Grade 5.

(2) The analysis results have not supported that peer tutoring add extra value on top of reading strategies instruction, neither in the second nor in fifth grade. But in both grades the CA peer tutoring dyads have demonstrated the most consistent and sustainable progress, compared to the other conditions.

(3) In-depth analyses of the interactions within dyads have shown that CA dyads have significantly better performance than RSA in strategic reading behaviors, such as strategy based questioning and explanations by tutors, or the tutor's use of reading strategies when correcting the tutee's reading mistakes or addressing the tutee's reading difficulties.

References:

Alexander, P. A. (2006). *Psychology in Learning and Instruction* (pp. 146-166). Upper saddle river, NJ: Pearson.

Alexander, P. A., Graham, S., & Harris, K. (1998). A perspective on strategy research: Progress and prospects [Special issue]. *Educational Psychology Review*, 10, 129-154.

Beijing Educational Committee (2002). Bei Jing Shi Jiao Yu Wei Yuan Hui guan yu zhong xiao xue ru xue nian ling de gui ding.

Biemiller, A., Shany, M., Inglis, A., & Meichenbaum, D. (1998). Factors influencing children's acquisition and demonstration of self-regulation on academic tasks. In D.H, Schunk & B.J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp.203-224). New York: Guildford.

Canney, G., & Winograd, P. (1979). Schemata for reading and reading comprehension performance (Tech. Rep. No.120). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Chall, J. S. (1983). Building theories of reading ability: on the relation between individual differences I cognitive skills and reading comprehension. *Cognition*, 9, 73-113.

Chen, M. J., Lau L. L., & Yung Y. F.(1993). Development of component skills in reading Chinese, *International Journal of Psychology*, 28(4), 481-507.

Corsale, K., & Ornstein, P. A. (1980). Developmental changes in children's use of semantic information in recall. *Journal of Experimental Child Psychology*, 30, 231-245.

De Corte, E., Verschaffel, L., Van de Ven, A. (2001). Improving text comprehension strategies in upper primary school children: a design experiment. *British Journal of Educational Psychology, 74, 4, 531-560.*

Fu, R. (2005). Ke tang he zuo xue xi de wen ti yu shi zhi de fan si: Dui ke tang xiao zu tao lun jin xing de fan si [Problems and nature of cooperative learning in classroom: relection on small group discussion in classroom]. *Zhong Guo Ke Xue Jiao Yu* [*China Science Education*], 12, 97-98.

Fukkink, R., Van der Linden, S., Vosse, A., & Vaessen, K. (1997). *Stap Door! Handleiding* [*Step out! Manual*]. Utrecht, The Netherlands: Sardes.

Hacker, J, D., Dunlosky, J., Graesser, A. C. (1998). Metacognition in educational theory and practice. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Manion, V., & Alexander, J. M. (1997). The benefits of peer collaboration on strategy use, metacognitive causal attribution, and recall. *Journal of Experimental Child Psychology*, 67, 268-299.

McBride-Chang, C., Shu, H., Zhou, A., Wat C. P., & Wagner, R. K. (2003). Morphological awareness uniquely predicts young children's Chinese character recognition, *Journal of Educational Psychology*, 95, 4, 743-751.

Nist, S. L., & Holschuh, J. L. (2000). Comprehension strategies at the college level. In R. F. Flippo & D. C. Caverly (Eds.), *Handbook of college reading and study strategy research* (pp. 74-104). Mahwah, NJ: Lawrence Erlbaum.

Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processes, *Cognitive Psychology*, 7, 44-64.

Perfetti, C. A., & Hogaboam, T. (1975). Relationship between simple word decoding and reading comprehension skill, *Journal of Educational Psychology*, 67, 461-469.

Rumelhart, D. E. (1978). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shu, H., & Anderson, R. C. (1997). Role of radical awareness in the character and word acquisition of Chinese children. *Reading Research Quarterly*, 32, 78-98.

Slavin, R. (1995). Cooperative Learning and student achievement. In R. E. Slavin (2nd Eds.), *Cooperative learning: Theory, research, and practice* (pp. 14-49). Needham Heights, Massachusetts: Sallyn & Bacon.

Van Keer, H. (2002). *Reading strategies instruction and peer tutoring in primary school. A quasi-experimental study of the effects on students' reading comprehension and metacognition.* Proefschrift ingediend tot het behalen van de academische graad van Doctor in de Pedagogische Wetenschappen (Ghent University), Promotor: Prof. Dr. Jean Pierre Verhaeghe.

Van Keer, H. (2004). Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring. *British Journal of Educational Psychology*, 74, 37-70.

Van Keer, H., & Verhaeghe, J. P. (2003). Effects of explicit reading strategies instruction and peer tutoring in second and fifth grades. In E. De Corte, L. Verschafeel, N. Entwistle, & J. Van Merrienboer (Eds.), *Powerful learning environment: Unraveling basic components and dimensions* (pp. 211-228). Oxford: Elsevier.

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. *The Journal of Experimental Education*, 73(4), 291-329.

Winne, P. H. (2002). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman, H. S. Dale, (Eds), *Self-regulated learning and academic achievement: Theoretical perspectives.* Mahwah, New Jersey: Erlbaum.

Yang, Z. (2006). Ke tang jiao xue zhong xue yao zhe me duo “he zuo” ma ? Dui xin ke cheng zhong ke tang he zuo xue xi de si kao [Is so much ‘cooperation’ needed in classroom instruction? Thoughts about classroom copperative learning from the new curriculumn]. *Xiao Dai Jiao Yu Yan Jiu: Pu Jiao Yan Jiu* [*Research in Modern Education: Normal Education Research*], 71-72 & 90.