The background of the cover is a white surface with a fine, cracked texture. Scattered across this surface are various geometric shapes, primarily squares and rectangles, in vibrant colors: red, orange, and blue. Some shapes are solid, while others are partially obscured or layered. Small, dark, irregular shapes are also scattered throughout, adding to the abstract composition.

# Teaching reading strategies in classrooms

*Does it work?*

Mariska Okkinga



# Teaching reading strategies in classrooms

*Does it work?*

Mariska Okkinga

This research was partly funded by the Ministry of Education, Culture and Science (OCW), with an 'Onderwijsbewijs' grant, project 'Effecten van motiverende teksten en begrijpend leesinstructie op tekstbegrip in het vmbo', ODB10070

This research was partly funded by the Netherlands organization for Scientific Research (NWO), Netherlands Initiative for Educational Research (NRO), with a PROO Review Studies grant, project 405-15-715



This research was carried out in the context of the Interuniversity Centre for Educational Sciences (ICO)



All rights reserved. No part of this thesis may be reproduced, stored in any retrieval system, or transmitted, in any form or by any means without permission of the author.

<b>Printed by</b>	Ridderprint, The Netherlands
<b>Layout design</b>	Ridderprint, The Netherlands
<b>Cover design</b>	Bas Leerintveld
<b>Cover image</b>	Piet Mondriaan, Compositie in kleur B, 1917, Collectie Kröller-Müller Museum, Otterlo
<b>ISBN</b>	978-90-365-4448-1
<b>DOI</b>	10.3990/1.9789036544481

© 2017 **Mariska Okkinga**

Teaching reading strategies in classrooms

*Does it work?*

University of Twente, Enschede

# **TEACHING READING STRATEGIES IN CLASSROOMS DOES IT WORK?**

PROEFSCHRIFT

ter verkrijging van  
de graad van doctor aan de Universiteit Twente,  
op gezag van de rector magnificus,  
prof. dr. T.T.M. Palstra  
volgens besluit van het College voor Promoties  
in het openbaar te verdedigen  
op vrijdag 2 februari 2018 om 12.45 uur

door

Mariska Okkinga

geboren op 8 december 1983  
te Franekeradeel, Nederland

Dit proefschrift is goedgekeurd door:

Promotor: Prof. dr. P.J.C. Slegers

Co-promotor: Dr. A.J.S. van Gelderen

## **Promotiecommissie**

**Voorzitter:** Prof.dr. Th. A.J. Toonen

**Promotor:** Prof. dr. P.J.C. Slegers

**Co-promotor:** Dr. A.J.S. van Gelderen

**Overige leden:** Prof. dr. J.W.M. Kessels  
Prof. dr. P.C.J. Segers  
Prof. dr. C.M. de Glopper  
Prof. dr. H. van Keer  
Prof. dr. J.J.M. Schoonen



## CONTENTS

<b>Chapter 1</b>	General introduction	9
<b>Chapter 2</b>	Effects of reciprocal teaching on reading comprehension of low-achieving adolescents. The importance of specific teacher skills.	17
<b>Chapter 3</b>	Effectiveness of reciprocal teaching for reading comprehension: A two-year study in a whole-classroom setting with low-achieving adolescents.	41
	<i>Appendices Chapter 3</i>	78
<b>Chapter 4</b>	Does vocabulary knowledge matter in the effectiveness of teaching reading strategies? Differential responses from low-achieving adolescents on growth in reading comprehension.	83
	<i>Appendices Chapter 4</i>	110
<b>Chapter 5</b>	Effectiveness of reading-strategy interventions in classrooms: A meta-analysis.	115
	<i>Appendices Chapter 5</i>	140
<b>Chapter 6</b>	General discussion	159
	References	171
	Samenvatting in het Nederlands	183
	Dankwoord (Acknowledgements)	193
	About the Author   Publications and Presentations	199
	ICO Dissertation Series	205





# General introduction



## Introduction

Many low-achieving adolescents struggle with reading comprehension (e.g., Organisation for Economic Co-operation and Development [OECD], 2003; OECD, 2014). Since it is a fundamental skill in many school subjects, poor reading comprehension has serious implications for students' educational success and, consequently, for their later societal careers. Since the 1980's, reading programs have focused on teaching reading strategies in order to foster reading comprehension (Pressley & Afflerbach, 1995; Raphael, George, Weber, & Nies, 2009). However, results of research in teaching reading strategies to low-achieving students are mixed (Edmonds, et al., 2009; Fogarty et al., 2014; Simmons et al., 2014; Vaughn, 2013). The general aim of this dissertation is to gain further insight into the effects of instructing reading strategies on reading comprehension to students in whole-classroom settings. It does so, on the one hand by an experimental study of a specific reading strategy intervention, and on the other hand by a meta-analysis of 52 published studies reporting on reading strategy interventions conducted in whole-classroom contexts.

## Instructing Reading Strategies

Since the 1980's, and after Durkin's (1978) study demonstrating that comprehension instruction was virtually non-existent in elementary classrooms, research into reading comprehension instruction by means of the use of reading strategies, increased rapidly (Duke & Pearson, 2002). The underlying idea is that reading comprehension is a complex process in which the reader interacts with the text to construct a mental representation of the text, or a situation model (Kintsch, 1999 and 1998). Hence, if readers understand how to use reading (skills and) strategies as they read, their comprehension will be stimulated.

Reading comprehension is defined as a cognitive complex process involving several sets of knowledge and (meta) cognitive skills. Models of reading comprehension distinguish between lower and higher order skills that interact in the process of creating a mental representation of a text, or a situation model (Kintsch, 1998). Lower order skills include letter and word recognition (Perfetti, Landi, & Oakhill, 2005), while higher order skills refer to the ability to give meaning to words and sentences, make inferences, and make representations of paragraphs or a text as a whole (Aarnoutse & Van Leeuwe, 1988). To construct a coherent mental representation, the information in the text has to be integrated with the readers' background knowledge. It is known that vocabulary knowledge plays an important role in reading comprehension, as strong relationships are found between vocabulary knowledge and reading comprehension (Ouellette & Beers, 2010; Tilstra, McMaster, Van den Broek, Kendeou, & Rapp, 2009; Trapman, Van Gelderen, Van Steensel, Van Schooten, & Hulstijn, 2014; Van Gelderen et al., 2004; 2007; Van Steensel, Oostdam, Van Gelderen, & Van Schooten, 2016; Verhoeven & Van Leeuwe, 2008). Even within groups of low-achieving adolescents,

vocabulary knowledge explains reading comprehension of students of different ages to a substantial degree (Trapman et al., 2014; Trapman, Van Gelderen, Van Schooten, & Hulstijn, 2017; Van Steensel et al., 2016). Therefore, vocabulary knowledge is an important factor to take into account in research into low-achieving adolescents' reading comprehension.

A reading strategy is a mental tool a reader uses on purpose to monitor, repair, or foster comprehension (Afflerbach & Cho, 2009). The use of reading strategies is a deliberate and goal-directed attempt to construct meaning of text (Afflerbach, Pearson, & Paris, 2008), and as such, can refer to both metacognitive and cognitive strategies that aid the process of reading (Dole, Nokes, & Drits, 2009). Researchers have suggested many different strategies (Pressley & Afflerbach, 1995). They may involve an awareness of reading goals, the activation of relevant background knowledge, the allocation of attention to major content while ignoring irrelevant details, the evaluation of the validity of text content, comprehension monitoring, visualizing, summarizing, self-questioning and making and testing interpretations, predictions, and drawing conclusions (Duke, Pearson, Strachan, & Billman, 2011; Palincsar & Brown, 1984).

Reciprocal teaching is one of the most influential approaches for instructing reading strategies to foster reading comprehension (Palincsar & Brown, 1984; Palincsar, Brown, & Martin, 1987). Reciprocal teaching consists of a set of three principles: a) teaching comprehension-fostering reading strategies, b) expert modeling, scaffolding and fading; and c) students taking turns in practicing reading strategies and discussing with other students. The method includes the teacher explicitly modeling the use of reading strategies (Rosenhine & Meister, 1994) as well as scaffolding the application of reading strategies within groups of students working together. During this process, students become increasingly more capable of regulating their own reading process and the role of the teacher gradually fades. Many studies have demonstrated positive effects of this approach on reading comprehension (Kelly, Moore, & Tuck, 2011; Palincsar & Brown, 1984; Rosenshine & Meister, 1994; Spörer, Brunstein, & Kieschke, 2009).

### **Importance of Whole-Classroom Context**

Reciprocal teaching was originally designed by Palincsar and Brown (1984) for application in small group tutoring with researchers as tutors. In our study we question whether the context of small group tutoring can be transferred to whole-classroom teaching. According to several studies this transfer can be problematic (De Corte, Verschaffel, & Van de Ven, 2001; Edmonds, et al., 2009; Fogarty et al., 2014; McKeown, Beck, & Blake, 2009; Simmons et al., 2014; Vaughn, 2013). Implementation by regular teachers in whole-classroom settings may be far more difficult to accomplish than implementation by researchers in small tutoring sessions, as a few qualitative studies show. Teachers in whole-classroom settings face problems in the implementation of interventions directed at instructing reading strategies (Duffy, 1993;

Seymour & Osana, 2003; Hacker & Tenent, 2002). Teachers found the didactic principles of reciprocal teaching and the specific reading strategies that had to be taught hard to understand (Seymour & Osana, 2003), teachers found it hard to induce strategic thinking in students (Duffy, 1993), and students showed poor application of reading strategies and poor discourse skills while collaborating when teachers implemented reciprocal teaching in their classrooms (Hacker & Tenent, 2002). As a consequence, the teachers were hindered in changing from a teacher-centered to a student-centered approach. In addition, strategy instruction in regular classrooms has been criticized for failing to adequately focus students' attention on constructing coherent representations of text content (McKeown, Beck, & Blake, 2009). Therefore, this study examines whether the principles of reciprocal teaching to foster reading comprehension can be successfully implemented in whole-classroom settings with teachers as instructors. We believe that these whole-classroom settings are a more ecologically valid context for strategy instruction, given the financial and logistic situation in regular educational practice.

### **Context of the Experimental Study: 'Nieuwsbegrip' Program**

In 2010, the Dutch Ministry of Education, Culture and Science initiated funding for evidence-based research in the field of education. With this funding program 'Onderwijsbewijs', the aim was to fund research that could investigate whether existing educational approaches or programs work. Our experimental study was carried out in the context of this funding program.

More specifically, the dissertation initially reports on the results of an experimental study directed at testing the effects of a popular and widely used Dutch program for instructing reading strategies called 'Nieuwsbegrip' ("Newswise", CED Group, 2011) directed at low-achieving adolescent students. The program is based on the main principles of reciprocal teaching (strategy instruction, modeling and group work) and used for the training of five reading strategies (predicting, questioning, summarizing, inferencing, interpreting cohesive ties). Teachers had to model those reading strategies and to support their students working in small groups. Additionally, students were supposed to take turns in discussing the reading strategies that were used. The main goal of the experimental study was to find evidence whether the principles of reciprocal teaching used in 'Nieuwsbegrip' work in the context of whole-classroom instruction for low-achieving students.

'Nieuwsbegrip' is developed by the CED Group in Rotterdam and widely used in primary education (grades 1-6), with 80% of Dutch primary schools having a license for 'Nieuwsbegrip'. The program is increasingly used in secondary education (grades 7-9). Lessons are developed weekly by a team of developers at the CED Group. They are based on recent news texts (i.e., texts that had been issued the week before) about subjects close to students' everyday life (e.g., 'sugar in energy drinks', 'abdication of the Dutch queen', or

'20 years of text messaging'). The use of interesting texts aims to increase students' reading motivation (Guthrie & Wigfield, 2000; Schiefele, 1999). Each lesson contains a news text with a work sheet. The lessons can be downloaded by teachers from the program website ([www.nieuwsbegrip.nl](http://www.nieuwsbegrip.nl)) every week, starting Monday evening.

'Nieuwsbegrip' provides reading texts at 6 different reading levels differentiating between students from special education to students in pre-university education (until grade 10). In our experiment, we used texts at level 'B', which are recommended for first and second year prevocational students.

### **Design of the Experimental Study**

An experimental pretest-posttest design was used for a two-year intervention that was implemented in 20 prevocational classes (*vmbo basis-kader*) from 10 schools with a total of 369 students. At the class-level, students were randomly assigned to the experimental and control condition, which followed the business-as-usual reading comprehension program.

Students were tested twice a year for their reading comprehension. Additionally, tests were administered for vocabulary, non-verbal IQ and metacognitive knowledge. Teachers were trained in using the main principles of reciprocal teaching for the five reading strategies and subsequently coached in using these principles in their classrooms. In order to investigate the influence of implementation quality, lesson observations were done twice a year, in both the intervention and control classes. The results of these observations provided evidence whether the program was implemented according to the basic principles of reciprocal teaching and to what extent experimental and control teachers differed. After the intervention was completed, exit interviews with intervention teachers were held.

The role of implementation quality is central in the evaluation of the intervention in our study. First, results after one year of the intervention are reported, taking into account the moderating role of the realization of the main principles of reciprocal teaching in the experimental condition. Second, the results after two years are reported, while the moderating role of the main principles of reciprocal teaching is again analyzed. Third, we analyzed the moderating role of vocabulary knowledge in the intervention effects. We wanted to test the assumption that students with a low level of vocabulary knowledge benefit less from the reading strategy intervention compared to their classmates with a higher level of vocabulary knowledge.

### **Design of the Meta-analysis**

In order to examine the generalizability of findings of our experimental study considering the context of whole-classroom instruction of reading strategies, a meta-analysis of a large number of intervention studies was conducted. The main goal of this meta-analysis was to

find evidence for the overall effects of interventions directed at instructing reading strategies on reading comprehension in whole-classroom contexts. To establish the overall effects of instructing reading strategies, a search of the literature published in a period of more than a decade yielded a total of 52 studies, which comprised 125 experimental comparisons, for five different types of dependent variables: standardized tests for reading comprehension, researcher-developed reading comprehension tests, strategic ability, strategy knowledge and self-reported strategy-use.

In addition, moderation analyses were performed with the following characteristics: *intervention-* (e.g. different reading strategies, didactic principles, type of trainer), *student-* (grade and type of reader), and *study-design* (type of control group and type of design). These moderation analyses provide information about which of these characteristics may determine the magnitude of intervention effects found.

### **Outline of the Dissertation**

The general aim of this dissertation is to gain insight into the effects of interventions aimed at reading strategies to foster reading comprehension of students in whole-classroom settings. Results of the experimental study are reported in three different chapters. *Chapters 2 and 3* focus on the effects of the intervention and the moderating role of implementation quality. In *Chapter 2* the results after one year intervention are reported, while in *Chapter 3* the results after two years are reported. In *Chapter 4*, an additional analysis was carried out to investigate the moderating role of vocabulary knowledge in the intervention effects. In *Chapter 5*, findings from the meta-analysis are reported on the effects of interventions directed at reading strategies fostering students' reading comprehension in whole-classroom settings. Finally, in *Chapter 6*, an overview of the findings is presented and discussed. In addition, suggestions for further research and implications for practice are provided.

There is some overlap in the methods sections of *Chapter 2-4*, and in the literature overviews provided in the different chapters, as all of the chapters are written as independent journal papers. This means that each chapter can be read on its own.



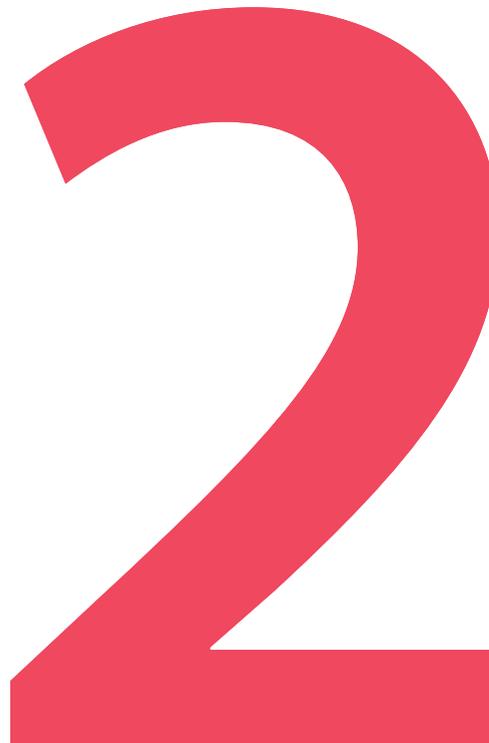




# Effects of reciprocal teaching on reading comprehension of low-achieving adolescents. The importance of specific teacher skills.



This chapter has been published as: Okkinga, M., Van Steensel, R, Van Gelderen, A.J.S., & Slegers, P.J.C. (2016). Effects of reciprocal teaching on reading comprehension of low-achieving adolescents. The importance of specific teacher skills. *Journal of Research in Reading*. doi: 10.1111/1467-9817.12082



**Abstract**

Low-achieving adolescents are known to have difficulties with reading comprehension. This article discusses how reciprocal teaching can improve low-achieving adolescents' reading comprehension in whole-classroom settings (as opposed to small-group settings) and to what extent intervention effects are dependent on teacher behavior. Over the course of one year, experimental teachers ( $n = 10$ ) were given extensive training and coaching aimed at using principles of reciprocal teaching, while control teachers ( $n = 10$ ) used their regular teaching method. Observations of teacher behavior were focused on instruction of reading strategies, modeling, and support of group work, and were performed in both experimental and control classes, comprising a total of 369 students (*mean age* = 13.01). Our study shows that reciprocal teaching contributed to adolescent low achievers' reading comprehension only when experimental teachers provided high-quality strategy instruction. In addition, results suggest that the quality of implementation of reciprocal teaching in whole-classroom settings should receive more research attention.

*Keywords:* reciprocal teaching, reading comprehension, reading strategies, low achieving students, teacher implementation

## Introduction

Many adolescent students, in particular low achieving ones, struggle with reading comprehension (e.g. Organisation for Economic Co-operation and Development [OECD], 2004; OECD, 2014). From several studies directed at adolescents, it is known that – in contrast to younger students – their reading comprehension is not so much dependent on efficient decoding of words, but much more by their vocabulary knowledge and their strategic skills in adapting their approach of the text to their reading goals (Trapman et al.; in press; Van Gelderen et al., 2004, 2007; Van Steensel, Oostdam, Van Gelderen, & Van Schooten, 2014). Therefore, reading *comprehension* instruction is regarded as an important part of the school curriculum. Since reading comprehension is a fundamental skill in many school subjects, difficulties can have serious implications for students' educational success and, consequently, for their later societal careers. Evidence-based reading comprehension programs that target low achieving adolescents are thus of vital importance. In this study, we analyze the effects of an intervention aimed at the improvement of reading comprehension based on principles of reciprocal teaching as introduced by Palincsar and Brown (1984). We examined its implementation in the everyday practice of Dutch language teachers, teaching Dutch low achieving adolescents (*mean age*=13.01) and we analyzed the association between instructional variation and intervention effects.

## Reciprocal Teaching

Reciprocal teaching (Palincsar & Brown, 1984) is a widely used method of instructing and guiding learners in reading comprehension. It consists of a set of three related instructional principles: a) teaching comprehension-fostering reading strategies, including predicting, question-generating, summarizing, and clarifying; b) expert modeling, scaffolding and fading; and c) students practicing and discussing reading strategies with other students, guided and coached by the teacher. Reciprocal teaching assumes a gradual shift of responsibility for the learning process from teacher to student, which includes the teacher explicitly modeling the use of reading strategies (Rosenhine & Meister, 1994) as well as scaffolding the application of reading strategies within the groups of students working together. It is assumed that by gradually fading teacher's support, students become increasingly more capable of regulating their own reading process. In this study, we consider reciprocal teaching as a method consisting of a set of several instructional principles, including direct instruction of reading strategies, teacher and student modeling, and group work.

### Effectiveness of Reciprocal Teaching

Many studies have confirmed the positive effects of reciprocal teaching (Rosenshine & Meister, 1994; Kelly, Moore, & Tuck, 2001; Spörer, Brunstein, & Kieschke, 2009). In a review by Rosenshine and Meister (1994), sixteen studies were analyzed. The authors found an overall positive effect on reading comprehension, with a median Cohen's effect size value ( $d = .32$ ) for standardized tests and a large effect size value ( $d = .88$ ) for researcher-developed tests. They also examined the effects of several moderator variables, of which two are particularly relevant for the current study: group size and type of interventionist (teacher or researcher). Regarding the former, they found contradictory results for studies where reciprocal teaching was applied in large groups ( $>18$ ), with two studies showing positive significant results, one study with mixed results, and one with non-significant results. Regarding the latter, they also found ambiguous results for teacher-led interventions, with two studies with positive significant results, three studies with mixed results and two studies with non-significant results. Thus, whether larger group size or teacher-led reciprocal teaching matter in finding positive results, is undecided.

In a more recent synthesis concerning reading interventions targeted at struggling readers between Grades 6 and 9 (Edmonds et al., 2009), seven studies focusing on reading comprehension were included. Most of these studies included some kind of instruction in reading strategies, with two of them using reciprocal teaching. The overall Cohen's effect size ( $d = 1.23$ ) on reading comprehension was very large. However, effects of possible moderators such as those reported by Rosenshine and Meister (1994) were not reported in this synthesis, which makes it difficult to draw definite conclusions about the question of whether reciprocal teaching is effective in whole-classroom settings with students' own teachers.

Reciprocal teaching was originally designed by Palincsar and Brown (1984) for small-group tutoring under the guidance of experts, in which small groups of students were taken out of the classroom (Palincsar, Brown, & Martin, 1987). In a whole-class setting, where 15-30 students are present, such extensive guidance as is provided in a small group might be quite difficult, if at all possible, as the teacher needs to pay attention to multiple groups of students within the classroom. Furthermore, small-group settings are often used in controlled experiments where the intervention is executed by the researchers instead of the students' own teachers. In comparison to researchers, who have extensive background knowledge about the theoretical basis of reciprocal teaching, the quality of implementation might be different for teachers, because they do not have the same background knowledge.

Studies in which teachers were followed during the implementation of reciprocal teaching or similar interventions suggest that the quality of implementation is indeed a serious problem (Duffy, 1993; Hacker & Tenent, 2002; Seymour & Osana, 2003). Duffy (1993) described the process of teachers becoming experts in reading strategies. Teachers were

followed during the implementation of a reading comprehension program, focusing on instructing reading strategies. During the study, the teachers were interviewed several times. A major conclusion from this study is that teachers realized that being able to model the use of strategies and explicitly relating strategy-use to text is not enough to induce strategic thinking in students that is useful for integrating process and content (Duffy, 1993).

Seymour and Osana (2003) found that teachers faced similar problems when they were trained in reciprocal teaching. In their study, two teachers were trained and observed during the implementation of reciprocal teaching. Interviews with the teachers revealed that their knowledge about reading strategies increased substantially during training, but their understanding of didactic principles was not developed optimally. Particularly, the teachers still did not fully understand what scaffolding entails at the end of the training.

These findings are corroborated by Hacker and Tenent (2002), who studied the application of reciprocal teaching in regular classrooms. They examined the way 17 teachers implemented reciprocal teaching and adapted the method to their own teaching practice over the course of three years. The researchers showed that teachers found it difficult to maintain the original format. First, they found that “student dialogues were hampered because of the students’ poor group discourse skills” as well as the poor application of reading strategies by the students, resulting in the observation “that there really was little for them to discuss” (Hacker & Tenent, 2002, p. 703). To deal with those problems, the teachers extended whole-class instruction of reading strategies to at least two months and they provided more scaffolding of strategy use in different kinds of contexts while at the same time providing scaffolding of the collaborative process. In other words, the teachers experienced difficulties in changing from a teacher-centered to a student-centered approach, which hampered the implementation of collaborative group work in discussing and practicing reading strategies. Second, Hacker and Tenent (2002) found that the students had difficulties with using all four reading strategies (predicting, questioning, summarizing and clarifying). Not all strategies were used, and the strategies that were used (summarizing and questioning) were “often being used inadequately” (p. 702). Students tended to ask superficial questions instead of making elaborations and reflections, and their strategy use could best be described as “mechanical” (p. 704).

The abovementioned studies into teachers’ implementation of reciprocal teaching give possible explanations of why previous experimental studies did not always support the success of reciprocal teaching in fostering reading comprehension. Whole-classroom application requires not only expert knowledge about the use of reading strategies on the part of the teachers, but also skills for regulating students’ collaborative process in different groups simultaneously.

## The Present Study

Our study aims to contribute to existing knowledge in two ways. First, we examined whether the principles of reciprocal teaching — originally developed for small-group tutoring (Palincsar & Brown, 1984; Palincsar, Brown, & Martin, 1987) — can be successfully used in whole-classroom settings in prevocational education to improve reading comprehension, in which the students' teachers are delivering the lessons (Woolley, 2011). Second, we analyzed whether intervention effects were moderated by the extent to which teachers were able to apply these principles. Issues of treatment fidelity have received little attention in reading intervention research (Edmonds et al., 2009; National Reading Panel, 2000; Rosenhine & Meister, 1994). Therefore, this study aims to add to the research base by analyzing moderation effects of specific treatment variables included in the principles of reciprocal teaching. This allows insight into the conditions under which the treatment will be effective in improving reading comprehension of low achieving adolescents in whole-classroom settings.

In this study, we will answer the following research questions:

- 1 Is reciprocal teaching provided by students' own teachers in whole-classroom settings effective in fostering reading comprehension of adolescent low achievers?
- 2 Does the quality of implementation of the three main principles of reciprocal teaching (strategy instruction, modeling and group work) moderate effects on reading comprehension?

## Method

### Sample Selection and Description

Our study focused on adolescent low achievers. Our operationalization of low achievement was based on educational track. The Netherlands have a tracked system of secondary education. After primary education, students are placed in one of three tracks—prevocational secondary education, senior general secondary education, pre-university education—on the basis of their scores on a general attainment test (Ministry of Education, Culture, & Science, 2006). Since students in prevocational education are generally characterized by poor reading skills (Dutch Education Inspectorate, 2008; Gille, Loijens, Noijons, & Zwitser, 2010), we selected our sample from schools offering this type of education.

We recruited schools in two ways. First, we contacted schools that had participated in a previous study on low achieving readers. Second, we contacted schools via a digital community of Dutch language teachers. Schools had to meet the following five criteria:

- Willingness to participate in a treatment study.
- They had (at least) two seventh grade classes.
- Each class had its own Dutch language teacher.

- The teachers were prepared to take part in the randomization procedure, implying that a) if their class was assigned to the treatment condition, they were prepared to take part in our training and coaching program and to weekly give our experimental lessons; and b) if their class was assigned to the control condition, they were prepared to not use our program nor discuss its contents with the colleague in the treatment condition.
- Control teachers were requested to use their regular language program during the language classes.

Ten different schools in different parts of the Netherlands were willing to participate. Within each school, two Dutch language teachers volunteered. Randomization was done at the class level within each school, resulting in a total of ten experimental and ten control classes, each with their teacher, divided over the ten schools. At the start of the study, these classes comprised 369 students, of which 189 were in the treatment condition (51%) and 180 in the control condition (49%). The students' mean age was 13.01 years ( $SD = 0.52$ ) at the start of the project. There was no statistically significant difference between the two conditions on this variable ( $t(366) = -1.27, p = .20$ ). There were relatively more girls in the sample ( $n = 200$ ; 54%) than boys ( $n = 169$ ; 46%), with relatively more girls than boys (59 vs 41%) in the treatment condition. The distribution in the control condition, however, was more equal (49 vs 51%). The difference in distribution between the two conditions was statistically significant ( $\chi^2(1) = 3.99, p = .046$ ).

More female than male teachers participated in the study ( $N = 15$  vs.  $N = 5$ ), with two male teachers in the treatment group and three males in the control group. The mean age of the teachers was 46.40 years ( $SD = 11.12$ ). On average they had 13.50 ( $SD = 13.73, \min = 1, \max = 38$ ) years of teaching experience in secondary education. No differences were found between the conditions on either variable, ( $t(14) = -.45, p = .66$ ) and ( $t(14) = .053, p = .96$ ), respectively.

## Design

We followed a pretest-posttest randomized controlled trial (Shadish, Cook, & Campbell, 2002). The design included one independent variable (treatment vs control) and one dependent variable (reading comprehension at posttest). We included four control variables: gender, reading comprehension at pretest, vocabulary knowledge at pretest, and IQ at pretest.

Gender was included, because generally girls are shown to have substantially greater reading skill than boys (e.g., Logan & Johnston, 2009). Vocabulary knowledge and IQ were included, as theoretical models suggest that reading comprehension draws heavily on both abilities (e.g., Just & Carpenter, 1976, 2004; LaBerge & Samuels, 1974; Rumelhart, 2004; Samuels, 2004), an assumption that is confirmed by much empirical evidence (e.g., Ouellette & Beers, 2010; Van Gelderen et al., 2004, 2007; Verhoeven & Van Leeuwe, 2008). We did not

include word recognition as a control in our analyses, because for adolescents in the age group of our study (age 13-16) it efficient word recognition is not related to their reading comprehension according to several studies (Trapman et al., in press; Van Gelderen et al., 2004, 2007; Van Steensel, Oostdam, Van Gelderen, & Van Schooten, 2014).

Finally, we included three moderator variables, covering the three didactic principles behind our treatment: direct instruction of reading strategies, teacher and student modeling, and group work.

### **Treatment**

Our intervention consisted of the training of teachers in the use of the three related instructional strategies of reciprocal teaching (Palincsar & Brown, 1984), that is:

- 1 Direct instruction of research-based reading strategies (see further). For each strategy, it was emphasized what the strategy entailed, how to use the strategy, when to use the strategy and why to use the strategy (Veenman, Van Hout-Wolters, & Afflerbach, 2006). Thus, teachers were required to give whole-class instruction about the different reading strategies, focusing on procedural knowledge.
- 2 Teacher and student modeling. Teachers were trained to model the use of reading strategies during plenary instruction by thinking aloud when reading text. They encouraged students to take over this role, both plenary and in small group sessions.
- 3 Group work. The primary objective of encouraging students to work in groups was to have them collaboratively apply reading strategies while thinking aloud during text reading. Teachers were given instructions on how to give feedback to the groups of students working together. For example, if a teacher noticed that the students were struggling with the application of a reading strategy, the teacher was instructed to model this strategy again and encourage and aid the students in doing this themselves.

Students received weekly lessons over a period of seven months within one school year. During the school year, the experimental teachers were trained and coached.

With respect to strategy instruction the intervention focused on five strategies that were shown to be related to reading comprehension in previous research (Dole, Duffy, Roehler, & Pearson, 1991; Palincsar & Brown, 1984; Pressley & Afflerbach, 1995; Van Silfhout, Evers-Vermeul, Mak & Sanders, 2014):

- 1 *Predicting*. On the basis of text features such as title, subheadings, and pictures, students are instructed to make predictions about text content before reading, and to check their predictions while reading.
- 2 *Summarizing*. Students are instructed to summarize sections of text, encouraging them to focus on main ideas and ignore irrelevant details as well as to check their understanding of the text so far.

- 3 *Self-questioning*. Students are instructed to generate questions about the text being read, helping them to focus on main ideas as well as to monitor understanding.
- 4 *Clarifying*. When confronted with a word or passage they do not understand, students are instructed to reread, read ahead, or, in the case of an unknown word, analyze it, and see whether its meaning can be inferred by looking at parts of the word.
- 5 *Interpreting cohesive ties*. Students are instructed to look for relationships between sentences or paragraphs that are connected, e.g. by using 'signal words' (different types of connectives).

The treatment was offered in the context of an existing program called "Nieuwsbegrip"<sup>®</sup>, developed by the CED Group in Rotterdam ("Comprehension of news", CED Group, 2011). Lessons were developed weekly by a team of developers at the CED Group. They were based on recent news texts (i.e., texts that had been issued the week before) about subjects close to students' everyday life (e.g., sugar in energy drinks, abdication of the Dutch queen, or 20 years of text messaging). The use of topical, interesting texts aimed to increase students' task motivation (Guthrie & Wigfield, 2000; Schiefele, 1999). Each lesson contained a news text with a work sheet. The lessons could be downloaded by teachers from the program website ([www.nieuwsbegrip.nl](http://www.nieuwsbegrip.nl)) every week, starting Monday evening.

Lessons were provided in sequences of six weeks. Each sequence consisted of six weekly lessons (approximately 45 minutes per lesson). In each of the first five lessons, the focus was on one reading strategy that was practiced in a central strategy assignment that was provided on a work sheet. In addition, students could work on other assignments (i.e., answering questions about the text) on the work sheet. In the final lesson of each sequence all strategies were practiced simultaneously. The idea behind this was that students have to be able to apply all strategies together during the reading process, selecting the right strategy at the right moment.

Each of the five strategies was trained several times during the year. This cyclical approach was assumed to result in the consolidation of strategy knowledge. Table 1 provides for each reading strategy an example of an assignment in which the focus is on the application of the reading strategy. Examples were generated from several work sheets that were used during the treatment.

**Table 1**

Examples of Strategy Assignments, Translated from Several Assignment Sheets from the Program “Nieuwsbegrip”

Strategy	Example
<b>Predicting</b>	This text has five subheadings. Write down for each subheading a) which thoughts it evokes and b) what you already know about the subject addressed in the subheading.
<b>Summarizing</b>	Read the text. Read paragraph by paragraph and underline in each paragraph the most important information. For each paragraph, write one or two sentences summarizing it. Use the words you underlined.
<b>Self-questioning</b>	Read the text. Note at least five questions that spring to mind while reading.
<b>Clarifying</b>	Search the text for difficult words. Try to uncover their meaning using these hints: a) reread the previous piece of text or read on, b) look at the illustrations in the text, c) look at the word: you might know part of the word, d) sometimes you have to use your own knowledge to figure out word meanings, or e) use a dictionary.
<b>Interpreting cohesive ties</b>	Read the text. Underline the signal words. Answer the questions, while noting the signal words: Which contrast is explained in lines 16-17? [signal word = however] Why are energy boosters unfit as sports drinks? [signal word = hence]

### Training and Coaching of Treatment Teachers

Treatment teachers took part in an extensive training and coaching program that was conducted by teacher trainers from the Rotterdam University of Applied Sciences, who had, in turn, been trained by the first three authors. In the first phase (October 2011-January 2012), teachers participated in three one-hour training sessions. In *Session 1*, they received general, practical information about the program (e.g., how to use the program website), theoretical information about the reading process and its components, and basic information about the three didactic principles behind the treatment (direct instruction of reading strategies, teacher and student modeling, and group work). In *Session 2*, in-depth information was provided about the nature, function, importance, and application of the five central reading strategies and on the way teachers could model the use of these strategies. Examples of modeling were provided by means of video clips and lesson protocols. In *Session 3* the focus was on group work and how, by means of scaffolded instruction, the use of reading strategies is transferred to the students. Attention was given to how the teacher can give feedback to groups of students and how his or her expert role is gradually faded. Two training sessions for the teachers occurred after the intervention had started, to give the teachers room to discuss their findings so far and to relate the content of the training sessions to their own practice.

Teachers were given a template for the lessons that would help them keeping focused on the reading strategies (See Table 2). The template was designed by the developers of the CED Group.

**Table 2**

Template for the Lessons that the Treatment Teachers used

Lesson structure	Activities
<b>Introduction</b>	<ul style="list-style-type: none"> <li>• Write the subject of the text and the central strategy of the lesson on the blackboard.</li> <li>• Introduce the subject and the central strategy with a whole-class approach and activate prior knowledge.</li> <li>• Write down questions students have about the text during orientation.</li> <li>• Read the first paragraph together and model the central strategy.</li> <li>• Invite a student to read the next paragraph while thinking aloud and applying the central strategy. Give support when necessary, that is, ask questions that stimulate the use of the reading strategy.</li> </ul>
<b>Processing</b>	<ul style="list-style-type: none"> <li>• Instruct the students to work together in groups of two or three. Let them work on the remainder of the work sheet.</li> <li>• Walk around to give the groups of students feedback. Focus on the central strategy and motivate the students to apply the strategy while thinking aloud. If necessary, model the strategy again.</li> </ul>
<b>Reflection</b>	<ul style="list-style-type: none"> <li>• Reflect with the students on the reading process as well as the content.</li> <li>• Together with the students, answer the questions they had before reading the text. Did reading the text answer those questions?</li> </ul>

*Note:* The template gives an overview of the activities the teacher should initiate to keep focused on the central strategy during the lesson. The template was designed by the developers of the “CED Group”.

In the second phase (February 2012–June 2012), teachers participated in three coaching sessions. A coaching session involved a classroom observation conducted by the trainer during an intervention lesson, followed by a feedback meeting of approximately 20 minutes on the same day. During the classroom observations, trainers used an observation scheme comparable to the one used by the researchers (see Classroom variables and treatment fidelity), directing the trainers’ attention and, consequently, their feedback to the central principles of the intervention (direct instruction of reading strategies, teacher and student modeling, and group work).

### Control Classes

Control classes were “business as usual”. Teachers in the control classes used the regular textbook for Dutch language that was used in their school. Among our schools, three different language textbooks were used. The textbooks and their teacher manuals were analyzed according to the three principles of instructional strategies in the treatment condition: instruction of reading strategies, modeling, and group work. Attention was given to reading strategies in all three textbooks. However, not all strategies that were covered

in the treatment condition were also covered in the control textbooks. Reading strategies that were often referred to were: predicting, clarifying, and attention to cohesive ties. Self-questioning did not occur and little attention was given to summarizing.

No attention was given to modeling by teachers or students in the teacher manuals of the control classes. Almost all of the assignments were individual and there were only a few instances where students were instructed to work together on an assignment.

## **Measures**

### ***Reading comprehension***

Reading comprehension was measured by means of the SALT-reading, a test that was validated for use among low achieving adolescents (Van Steensel, Oostdam, & Van Gelderen, 2013). The SALT-reading comprises eight tasks, each consisting of one or two texts and comprehension questions about those texts. The texts cover different genres (narrative, expository, argumentative, and instructive). They were selected from media students assumedly come across regularly in their daily lives: (school) books, newspapers, magazines, and official documents (such as regulations in a youth hostel). The eight tasks comprised a total of 59 test items, that were divided into three categories: items requiring students to retrieve relevant details from the text, items requiring students to make inferences on a local level (e.g., draw cause-effect relationships between sentences), and items requiring students to show their understanding of the macrostructure of the text (e.g., by inferring the main idea of the text or the intention of the author). The test consisted mainly of multiple choice questions but contained also five open-ended questions. Open-ended questions were coded by the researchers. The Cronbach's alpha coefficients of the pretest and posttest were .82 and .83, respectively.

### ***Vocabulary knowledge***

Vocabulary knowledge was assessed with a 73-item multiple-choice test, based on the receptive vocabulary test developed by Van Gelderen et al. (2003) and Van Gelderen et al. (2007). It measures the knowledge of nouns, verbs, adjectives, and adverbs belonging to the 23,000 words in a dictionary for junior high school students (see Hazenberg & Hulstijn, 1996 for details). Each item consists of a neutral carrier sentence with a bold-faced target word and four answer options, one of which represents a correct synonym. The Cronbach's alpha coefficient for the pretest was .86.

### ***IQ***

Intellectual ability was measured by administering the Raven Progressive Matrices, a non-verbal IQ test. The total test consists of 60 items, divided into 5 sets of 12 items. Each item represents a logical reasoning puzzle. The items become more difficult within a set and the

sets become increasingly difficult as well (Raven, Raven & Court, 1998). For students from the lowest tracks of prevocational education the last set was assumed to be too difficult and for this reason this set was omitted. The Cronbach's alpha coefficient was .82.

### ***Classroom variables and treatment fidelity***

To examine the moderator variables, we conducted classroom observations twice during the year. We devised an observation scheme for use both in the experimental and control conditions. Our aim was to examine a) whether the treatment teachers gave the lessons in the way we instructed them during the training and coaching program and b) whether the control teachers applied treatment principles, even though they were not trained by us. The scheme focused on three variables that were essential to the treatment: direct instruction of reading strategies, teacher and student modeling, and group work (Palincsar & Brown, 1984). We constructed these variables in the following manner, resulting in three four-point scales (0-3) to be used for further analysis:

- 1 Direct instruction of reading strategies. We distinguished four categories of behavior:
  - a Teachers provided no information on reading strategies (0 points).
  - b Teachers introduced the central strategy of the lesson (in the treatment condition) or any strategy (in the control condition), but provided no further explanation (1 point).
  - c Teachers introduced a strategy and explained about its nature, function, importance, and/or application (2 points).
  - d Teachers introduced a strategy, explained about its nature, function, importance, and/or application and interacted with the class about the strategy (3 points).
- 2 Teacher and student modeling. Here also, we distinguished four categories of behavior:
  - a Teachers did not use any modeling of strategy use (0 points).
  - b Teachers modeled strategy use (1 point).
  - c Teachers modeled strategy use and asked students to think aloud while using reading strategies, either individually (i.e., in front of the class) or in groups (2 points).
  - d Teachers modeled strategy use, asked students to think aloud, and provided them with feedback (3 points).
- 3 Group work, with four categories of behavior:
  - a Teachers did not have students work in groups (0 points).
  - b Teachers had students work in groups, but did not provide real feedback (1 point).
  - c Teachers had students work in groups and provided feedback, but not on collaboration, that is, teachers focused mainly on whether students had understood the assignment correctly, on whether their answers were correct, or on the meaning of unknown words (2 points).
  - d Teachers had students work in groups and provided feedback on collaboration (3 points).

The scales were constructed in such a way that a 3-point score would be the optimal

score for the purpose of the treatment. It should be noted that the scores within a scale were conditional: one could only score a 2 if both *b* and *c* were observed. This conditional approach proved to be appropriate in the classroom observations (i.e., we did not encounter a case in which *c* was observed, but *b* was not).

Before the start of the classroom observations, the observation scheme was piloted during two lessons, one in an experimental class and one in a control class. Two researchers filled out the observation scheme during the lessons, after which they compared their coding and discussed causes for any differences. If these discussions revealed that items were unclear or led to misinterpretation, the coding scheme was adjusted. Means were calculated over the two classroom observations.

Inter-rater reliability was calculated by means of observed agreement between two observers. In total, 16 from a total of 38 classroom observations were performed by two coders. Across these 16 observations, 94.22% observed agreement was obtained.

### **Procedure**

The reading comprehension pretest as well as the vocabulary and IQ tests were administered in the Fall of 2011, just before the start of the treatment, and the reading comprehension posttest was administered during May-June of 2012. All test administrations took place in classroom settings. The test sessions were introduced by a trained test leader. A familiar teacher was present to maintain order. Questions were answered by the test leaders following a standardized protocol. Students and teachers remained ignorant of test scores.

Classroom observations took place during January-February 2012 and during April-May 2012. During the classroom observations the researcher(s) sat at the back of the classroom to observe the teacher. In order to be able to check codings after the observation, the lessons were recorded using an audio-recorder carried by the teacher.

### **Attrition**

There was some attrition among teachers, but not due to a lack of motivation. One teacher in the treatment condition became terminally ill halfway during the school year. Because a replacement was only found after about two months, this class did not receive the treatment in this period. When a new teacher was found, she continued giving the treatment lessons and participated in our training program. Because of the replacement, we were not able to do classroom observations in this class. Therefore, we were not able to include this class in the analysis. A second teacher in the treatment condition became pregnant toward the end of the school year. During her leave, she was temporarily replaced by a new teacher, who continued giving the lessons and took part in the training. Finally, a teacher in the control condition found another job halfway during the school year; a new teacher immediately

replaced her. Both classes were included in our final analysis, because we succeeded in carrying out the planned classroom observations.<sup>1</sup>

There was some attrition among students, mainly because of transfers to different schools (7 students) and one student was ill for a long period of time. During the school year six new students entered the experimental and control classes.

## Analysis

Our sample had a hierarchical structure (students nested in classes, nested in schools). Because there was significant random variability at the class level, we performed multilevel analyses with the use of MLwiN 2.16 (Rasbash, Steele, Browne, & Goldstein, 2009). We tested whether a) the treatment had a significant positive effect on reading comprehension and b) whether the quality of teacher instruction moderated the effect of the treatment. Adding variables was done in the following order (Hox, 2010). First, all control variables were added (gender, reading comprehension at pretest, vocabulary at pretest, IQ at pretest), with the final three variables centered around the grand mean (Hox, 2010; Snijders & Bosker, 1999). Second, the moderator variables (strategy instruction, modeling, group work) were entered. It was not necessary to center the moderator variables around the grand mean, as zero was meaningful in the scoring of the classroom variables (See Classroom variables and treatment fidelity). Third, the independent variable (treatment vs control) was entered to answer the first research question. Finally, the interactions between the independent and moderator variables were entered to answer the second research question.

Of the 369 cases, 75 were incomplete due to missing values, either with missing values within a test or questionnaire or because students were not present at one of the test sessions due to illness (despite the fact that at each school at least one extra test session was organized). To prevent loss of information, single imputations using SPSS missing value analysis were performed for each variable at the item level, that is, missing items (as opposed to 'wrong' items) within a test or questionnaire were imputed. No missing values were imputed if the student was not present during the test session. As a result, 44 of the 75 cases (58.6%) with missing values could be included in the analyses (total  $N = 338$ ).

<sup>1</sup> We checked whether results were different when these classrooms were excluded from the analysis. This was not the case.

## Results

### Descriptive Statistics

Table 3 shows the mean student scores for all continuous variables, as well as correlations (posttest reading comprehension, pretest reading comprehension, IQ and vocabulary).

No significant differences are found between the treatment and the control condition. The highest correlation is found between posttest and pretest reading comprehension ( $r=.69, p<.01$ ).

**Table 3**

Comparison of Treatment and Control Students in terms of Reading Comprehension (posttest and pretest), Vocabulary, and IQ, as well as Correlations between the Variables

Variable	Treatment (n = 168)	Control (n = 170)	t-value	Correlations (N = 338)		
	Mean (SD)	Mean (SD)		2.	3.	4.
1. Posttest reading comprehension	37.37 (7.20)	36.22 (8.91)	1.30	.69*	.54*	.25*
2. Pretest reading comprehension	35.20 (7.24)	34.51 (8.53)	.80		.57*	.29*
3. Vocabulary	48.02 (8.28)	47.35 (9.17)	.71			.54*
4. IQ	35.95 (5.62)	35.19 (5.20)	1.28			

\* $p<.01$

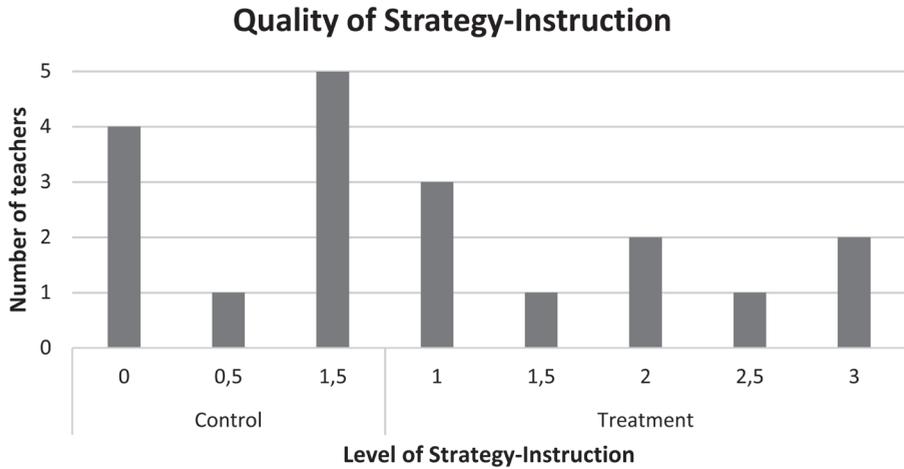
In Table 4, means and standard deviations are presented for the variables resulting from the classroom observations. As expected, the mean scores of the treatment group are higher than those of the control group, indicating that in the experimental classrooms modeling, strategy instruction and group work were more often observed than in the control classrooms. The difference between both groups is statistically significant on the .05 level for all variables, except for modeling. Given the small sample ( $N=19$ ), the non-significant difference in the case of modeling should not be given much weight. Because the scoring of the three variables is qualitatively different, the means and standard deviations presented in Table 4 cannot be compared one-on-one (See Classroom variables and treatment fidelity for an explanation of each variable).

Bar charts (See Figures 1-3) for each classroom variable show that only in a few instances treatment teachers scored maximally.

**Table 4**

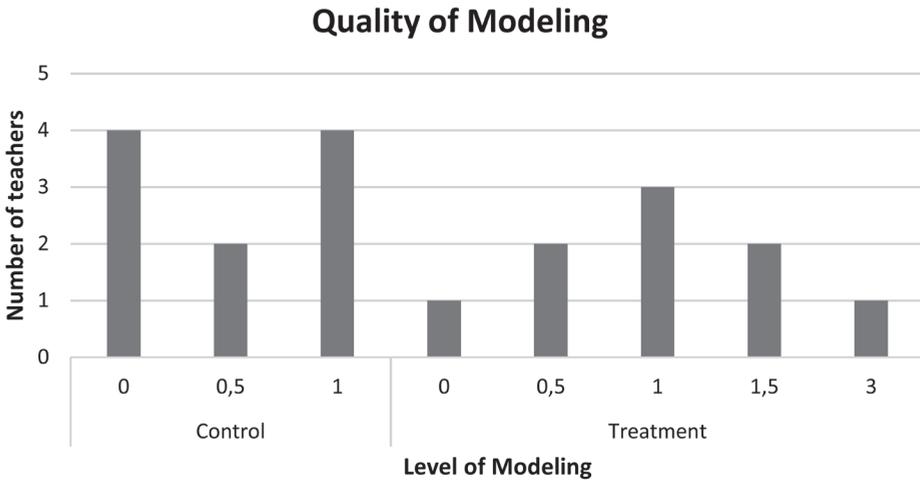
Comparison of Treatment and Control Teachers in terms of Classroom Observations: Strategy-Instruction, Modeling, and Group Work

Variable	Treatment (n = 9)	Control (n = 10)	t-value	p-value
	Mean(SD)	Mean(SD)		
Strategy-instruction	1.89(.82)	.80(.75)	-3.02	.008
Modeling	1.11(.86)	.50(.47)	-1.95	.068
Group work	1.94(1.21)	.50(.67)	-3.27	.005



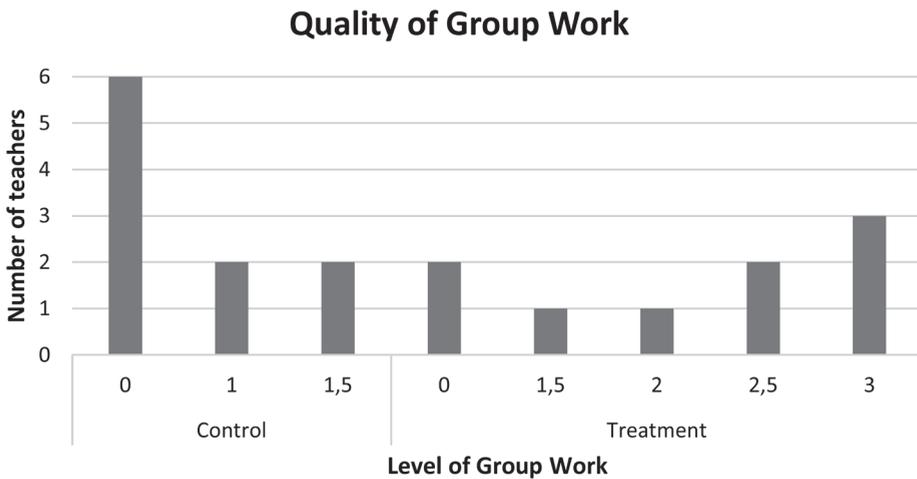
**Figure 1**

Bar chart for strategy-instruction, for both the control and treatment teachers. Scores are calculated as the mean over two classroom observations per teacher.



**Figure 2**

Bar chart for modeling, for both the control and treatment teachers. Scores are calculated as the mean over two classroom observations per teacher.



**Figure 3**

Bar chart for group work, for both the control and treatment teachers. Scores are calculated as the mean over two classroom observations per teacher.

## Multilevel Analyses

As a first step, we examined whether the data had a multilevel structure. A model with only a student level had an *I*GLS of 2373.309. A model with both a student and a class level had a significantly better fit (*I*GLS = 2321.862;  $\Delta$ *I*GLS = 51.447, *df* = 1,  $p < .001$ ). Adding a school level did not improve model fit (*I*GLS = 2320.221;  $\Delta$ *I*GLS = 1.641, *df* = 1,  $p > .05$ ). Therefore, in all further analyses a two-level structure was used. The empty two-level model is further referred to as Model 0.

Subsequently, the control variables were entered. Inclusion of these variables significantly increased model fit. As expected, both reading comprehension and vocabulary at pretest positively contributed to posttest reading comprehension. There was also an effect of gender: boys scored significantly lower on posttest reading comprehension than girls. The effect of IQ was non-significant, however: it appeared that pretest reading comprehension and vocabulary already accounted for the variance in IQ. Therefore, IQ was dropped from the model. The resulting model (Model 1, see Table 5) represented a significant increase in fit compared to Model 0 ( $\Delta$ *I*GLS = 210.156, *df* = 3,  $p < .001$ ).

In Model 2 (see Table 5), the moderator variables (strategy instruction, modeling, group work) were entered. This did not result in a significant increase in model fit ( $\Delta$ *I*GLS = 3.095, *df* = 3,  $p > .05$ ). This means that, overall, the quality of instruction did not influence students' reading comprehension.

Adding the treatment variable to the model (Model 3) did not result in a better fitting model either ( $\Delta$ *I*GLS = 0.458, *df* = 1,  $p > .05$ ). In other words, no main effect of the treatment on students' reading comprehension was found.

In the three subsequent models (Model 4a-c), we added the interactions between the independent and moderator variables (i.e., the interactions of treatment and each of the three observed instruction variables: strategy instruction, modeling and group work). Of these three models, only Model 4a resulted in a significant improvement of fit compared with Model 3 ( $\Delta$ *I*GLS = 5.033, *df* = 1,  $p < .05$ ), implying that the strategy instruction variable was a significant moderator of the treatment effect on reading comprehension ( $B = 3.183$ ,  $SE = 1.311$ , *df* = 14,  $p < .05$ ). In other words, elaborate strategy instruction had a positive effect on reading comprehension in the treatment classes, but not in the control classes. The size of this effect was considerable: the interaction effect was responsible for explaining an additional 37 per cent of class-level variance.

**Table 5**

Results of the Multilevel Analyses, Predicting Posttest Reading Comprehension by Gender, Pretest Reading Comprehension, Pretest Vocabulary (as Control Variables), Strategy-Instruction, Modeling, Group Work (as moderator variables), and Treatment (yes/no)

<b>N=338</b>	<b>Model 0</b>	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4a</b>	<b>Model 4b</b>	<b>Model 4c</b>
<b>Fixed effects</b>							
Intercept	36.409 (0.984)	37.431 (0.625)	37.398 (0.857)	37.875 (0.912)	36.621 (0.950)	37.460 (0.975)	38.637 (1.245)
Gender (Male = 1)		-1.442 (0.600)*	-1.364 (0.601)*	-1.368 (0.601)*	-1.363 (0.600)*	-1.355 (0.601)*	-1.370 (0.601)*
Reading comprehension pre		0.572 (0.048)***	0.570 (0.048)***	0.572 (0.048)***	0.572 (0.048)***	0.581 (0.049)***	0.573 (0.048)***
Vocabulary pre		0.167 (0.043)***	0.166 (0.043)***	0.166 (0.042)***	0.171 (0.042)***	0.167 (0.042)***	0.165 (0.042)***
Strategy instruction			1.144 (0.704)	1.327 (0.746)	-0.201 (0.898)	1.379 (0.724)	1.588 (0.787)
Modeling			-0.048 (0.883)	0.004 (0.876)	-0.171 (0.766)	-1.380 (1.616)	0.018 (0.857)
Group work			-0.252 (0.510)	-0.077 (0.565)	-0.618 (0.540)	-0.161 (0.554)	0.803 (1.137)
Treatment (treatment = 1)				-0.976 (1.434)	-0.210 (1.276)	-0.680 (1.418)	-1.650 (1.595)
Treatment * strategy instruction					3.183 (1.311)*		
Treatment * modeling						1.866 (1.854)	
Treatment * group work							-1.211 (1.371)
Level 1 random variance	50.831(4.025)	28.138(2.228)	28.129 (2.227)	28.129 (2.227)	28.160 (2.229)	28.150 (2.229)	28.139 (2.228)
Level 2 random variance	15.379(5.961)	4.267(1.920)	3.404 (1.638)	3.285(1.599)	2.068 (1.198)	2.979 (1.497)	3.063 (1.526)
Level 1 explained variance		45%	0%	0%	0%	0%	0%
Level 2 explained variance		72%	20%	3%	37%	9%	7%
Deviance (IGLS)	2321.862	2111.706	2108.611	2108.153	2103.120	2107.180	2107.394
Difference ( $\Delta$ IGLS)		210.156***	3.095	0.458	5.033*	0.973	0.759
Df		3	3	1	1	1	1

#  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

## Discussion

Our study set out to analyze how reciprocal teaching can improve low achieving adolescents' reading comprehension in whole-classroom settings and to what extent intervention effects are dependent on teacher behavior. Apart from analyzing the overall effects of the treatment in a whole-classroom setting (Research question 1), our aim was to examine whether effects were larger when teachers provided more elaborate instruction of reading strategies, engaged more in teacher modeling and promoted more student modeling, and when they supported more collaboration during group work (Research question 2). Answering our first research question, our study revealed no overall treatment effects: no significant differences were found between students in the treatment classes and the control classes on the reading comprehension posttest. Answering our second research question, we did find a moderator effect of instruction of reading strategies. This moderator effect implied that in the experimental condition more elaborate explanations of the nature, function, importance, and application of reading strategies positively contributed to students' reading comprehension. The effect was substantial: it explained an additional 37 per cent of the differences between classes after individual and class-level variables had been taken into account. In the control condition, there was no effect of strategy instruction. It thus seems that the frequent, systematic, and cyclical offering of reading strategies in our treatment set the stage for successful reading comprehension instruction.

Our results underscore the relevance of focusing on the quality of implementation of treatments in teacher-delivered classroom intervention studies. First of all, it is likely that there are important differences in implementation quality among teachers who are trained "on the job". As we have shown, neglecting such variation can result in overlooking meaningful effects. The effect of our reciprocal teaching intervention only appeared after taking the differences between teachers' application of strategy instruction into account.

Moreover, repeatedly measuring instructional behaviors essential to the treatment gives insight in the degree to which treatments such as these are successfully implemented by 'real teachers' (as opposed to researchers) and whether some elements are harder to apply than others. Our experience showed that even after a year of intensive training and coaching, application of the three instructional principles was less than optimal. Particularly, our observations showed that getting students to model reading strategies during group work was a challenge. Similar observations were made by Hacker and Tenent (2002) in an elaborate implementation study of reciprocal teaching: they showed that teachers found it particularly difficult to engage students in meaningful dialogues. Moreover, there was considerable variability among our experimental teachers in applying principles of reciprocal teaching: while some teachers fairly quickly succeeded in modeling reading strategies and having their students work in groups, others had more difficulties in incorporating these principles in their lessons. The latter seemed to be partly the result of

classroom management issues: in instances where students were unmotivated and showed oppositional behavior, teachers found it hard to gradually transfer control to students. These classroom management issues were non-existent in the original set-up of small groups of students under the guidance of a tutor (Palincsar & Brown, 1984; Palincsar, Brown, & Martin, 1987) and may explain why in previous research positive results were found of reciprocal teaching (Rosenshine & Meister, 1994; Kelly, Moore, & Tuck, 2001; Spörer, Brunstein, & Kieschke, 2009).

Interestingly, we found a moderator effect of instruction of reading strategies, but not of modeling or group work. There are at least two explanations for this observation. First, the difference between instruction of reading strategies and modeling or group work can be explained in terms of the extent of teacher versus student control. A higher score on the strategy instruction variable indicates more elaborate instruction by the teacher about the nature, function, importance, and/or application of reading strategies. This is the component of the treatment that is the most teacher-controlled and is also the most familiar, both for teachers and students (such instruction is commonly used in education, in every domain) and, thus, is probably easiest to implement. Also, it may be assumed that teachers have prior knowledge about reading strategies. Both direct instruction and knowledge of reading strategies are consistent with existing knowledge and practice of teachers, whereas modeling and group work are relatively unknown areas and therefore harder to master.

Second, modeling and group work are dependent on initiatives afforded to students: higher scores on the former imply that more modeling is being done by both teachers and students, higher scores on the latter imply more attention to group work. These components are dependent on teachers transferring control to their students and may not be part of many teachers' repertoire. For low achieving students in Dutch secondary education, it is quite uncommon that students work on tasks collaboratively in language arts lessons (De Milliano, 2013). Thus, both modeling (especially by students) and group work differ from regular classroom practice and require new skills from teachers, as they need to adapt their feedback to the level of the student and use techniques for motivating students to collaborate without direct teacher supervision.

Therefore, one explanation of the absence of moderator effects of modeling and group work is that these instructional strategies did not reach a certain 'threshold level' to become significant moderators (Simmons et al., 2014). The observational data of the classrooms seems to support this. The maximum score for modeling includes students being successfully encouraged to model reading strategy use themselves. However, only one treatment teacher managed to reach this stage. The same holds for group work: the maximum score for group work includes teachers changing the focus from correct responses to assignments to learning to apply reading strategies collaboratively. Only three treatment teachers reached this stage implying that most teachers did not attain this

level of practice. Our findings are supported to some extent by the outcomes of the study by Hacker and Tenent (2002) mentioned earlier. In this study a number of teachers were followed over the course of one or more years to examine how they implemented reciprocal teaching in their classrooms and to what extent they modified the method. The authors first of all observed that “the most pervasive problem that teachers faced with RT [Reciprocal Teaching] was getting students to learn and use the RT strategies in group dialogues” (2002: 712). In response, teachers tended to become more directive, providing more scaffolding in the form of whole-class instruction, teacher modeling, and direct guidance. This was particularly true in classes with many struggling readers.

It appears that the training and coaching offered in our study was sufficient for the application of one of the main elements of reciprocal teaching to bear fruit, namely strategy instruction. However, even for this component, not all trained teachers profited sufficiently to produce a significant difference between the experimental and the control condition. Some experimental classes were receiving significantly more strategy instruction than others resulting in a moderating effect of this instructional variable. For the two other main instructional components of reciprocal teaching, modeling and group work, however, we did not find significant moderating effects. Despite the fact that our teachers were provided with a quite extensive training and coaching program, we believe that even more training and coaching is needed for teachers to adapt new ways of teaching to such an extent that it enhances the learning process of their students, as compared with control students. This is in line with the findings of Hacker and Tenent (2002). In their research, teachers found it difficult to embrace new practices, such as letting students work together, and clung more tightly to practices that were known, such as direct instruction.

As for limitations, even though there was randomization at the class and the teacher level, students were not randomly distributed across the intervention and control group. For future research on reciprocal teaching, we recommend a more strict design with randomization at the student level. Secondly, more classroom observations could have provided more insight into the development of teachers’ implementation of the treatment, and consequently provide more valid conclusions. In this case, two observations were enough to find significant effects, but it would be a great addition in future research to show the developmental patterns of teachers in implementing an intervention.

Finally, in future research we recommend a study with teachers that are trained more extensively and more frequently than in this study. Teachers should probably be provided with more tools to be able to guide the students in their collaborative learning process. Coaching should be directed at increasing the quality of the dialogues among the students. Teachers in our study did not have many tools to facilitate the students in their collaborative group work. For teachers to become seasoned in new ways of teaching, they need to practice rigorously, up to the point where reciprocal teaching becomes routine; similar to the way we want students to become seasoned in the use of reading strategies.





# Effectiveness of reciprocal teaching for reading comprehension: A two-year study in a whole- classroom setting with low- achieving adolescents.



This chapter is submitted for publication as: Okkinga, M., Van Gelderen, A.J.S., Van Schooten, E., Van Steensel, R., & Slegers, P.J.C. (submitted). Effectiveness of reciprocal teaching for reading comprehension: A two-year study in a whole-classroom setting with low-achieving adolescents.



**Abstract**

Low-achieving adolescents are known to have difficulties with reading comprehension. This article discusses whether reciprocal teaching can improve low-achieving adolescents' reading comprehension in whole-classroom settings and to what extent treatment effects are dependent on teacher behavior. Over the course of two years, experimental teachers ( $n=10$ ) were given training and coaching aimed at using principles of reciprocal teaching, while control teachers ( $n=10$ ) used their regular teaching method. Observations of teacher implementation were focused on instruction of reading strategies, modeling, and support of group work, and were performed in both experimental and control classes, comprising a total of 369 students (grade 7). Our study shows that modeling positively moderated the effect of reciprocal teaching, but that, overall, there is no differential growth in adolescent low-achievers' reading comprehension. In addition, results suggest that the quality of implementation of reciprocal teaching in whole-classroom settings should receive more attention.

*Keywords:* reciprocal teaching, reading comprehension, reading strategies, low achieving students, teacher implementation

## Introduction

Reading comprehension is an essential skill for all students in their school careers. However, many students in secondary education, especially low-achieving students, struggle with reading comprehension (e.g., Organisation for Economic Co-operation and Development [OECD], 2003; OECD, 2014; Dutch Education Inspectorate, 2008; Kordes, Bolsinova, Limpens, & Stolwijk, 2013), resulting in difficulties with several school subjects. Not being able to comprehend texts can therefore have serious implications for students' educational success, and, consequently, for their later societal careers. Long-term, evidence-based reading comprehension programs that target low-achieving adolescents are thus of vital importance (Edmonds et al., 2009; Slavin, Cheung, Groff, & Lake, 2008; National Reading Panel, 2000). A well-known evidence-based method for teaching reading comprehension for low-achieving adolescents is called reciprocal teaching (Palincsar & Brown, 1984). This method was tested in numerous experiments in which researchers or other experts were instructing small groups of students (e.g. Palincsar & Brown, 1984; Palincsar, Brown, & Martin, 1987; Rosenshine & Meister, 1994; Spörer, Brunstein, & Kieschke, 2009). Evidence concerning the effectiveness of reciprocal teaching in whole-classroom settings in which students' regular teachers are responsible for delivering the intervention, however, is mixed. In this study, we investigated the effects of a two-year program for reading comprehension instruction with reciprocal teaching in whole-classroom settings of low-achieving adolescents in grades 7 and 8. From previous studies it appears that the quality of implementation of reading comprehension programs in whole-classroom settings is an important determinant of success (Chiu, 1998; De Boer, Donker, & Van der Werf, 2014; Swanson, Wanzek, Haring, Ciullo, & McCulley, 2011; Vaughn, 2013). Therefore, we analyzed the moderating effects implementation quality of reciprocal teaching on students' reading comprehension.

## Reciprocal Teaching

Reciprocal teaching (Palincsar & Brown, 1984) is a widely used method of instructing and guiding learners in reading comprehension. It consists of a set of three related instructional principles: a) teaching comprehension-fostering reading strategies, including predicting, question-generating, summarizing, and clarifying; b) expert modeling, scaffolding and fading; and c) students practicing and discussing reading strategies with other students, guided and coached by the teacher. Reciprocal teaching assumes a gradual shift of responsibility for the learning process from teacher to student, which includes the teacher explicitly modeling the use of reading strategies (Rosenhine & Meister, 1994) as well as scaffolding the application of reading strategies within the groups of students working together. It is assumed that by gradually fading teacher's support, students become increasingly more capable of regulating their own reading process.

Many studies have confirmed the positive effects of reciprocal teaching (Kelly, Moore, & Tuck, 2001; Rosenshine & Meister, 1994; Spörer, Brunstein, & Kieschke, 2009). In a review by Rosenshine and Meister (1994), sixteen experimental studies were analyzed. The authors found an overall positive effect on reading comprehension, with a median Cohen's effect size value ( $d = .32$ ) for standardized tests and a large Cohen's effect size value ( $d = .88$ ) for researcher-developed tests.

### **Instructors and Settings of Reciprocal Teaching**

Reciprocal teaching was originally designed by Palincsar and Brown (1987) for small-group tutoring under the guidance of experts, in which small groups of students were taken out of the classroom. From the literature, there are some indications that replacing experts by the students' regular teachers is not without problems. Most of the studies under investigation in the review by Rosenshine and Meister (1994) were small experimental studies in which students were taken out of the classroom and reciprocal teaching was delivered by researchers or research assistants. However, seven studies in the review were teacher-led and the effects on reading comprehension for those studies were ambiguous, with two studies with positive significant results, three studies with mixed results and two studies with non-significant results.

In other meta-analyses focused on the effects of reading comprehension interventions (Chiu, 1998) and on the effects of metacognitive instruction for several subjects among which reading comprehension (De Boer, Donker, & Van der Werf, 2014), lower effect sizes were obtained when regular teachers implemented the treatment compared to studies in which researchers implemented the treatment. De Boer, et al. (2014) suggest that there are several explanations for this result. It is possible that teachers are less motivated to invest in a high quality of implementation than researchers or they might suffer from less confidence. Thus, there is sufficient ground for investigating implementation quality of reciprocal teaching in whole-classroom settings and its moderating effects on students' reading comprehension.

Another important issue concerns the small group tutoring which is applied in studies into the effects of reciprocal teaching. In the original Palincsar and Brown set-up, small groups of students were taken out of the classroom and tutored by a researcher. In whole classroom settings, however such guidance of small groups may not be possible, because there is only one teacher overseeing multiple groups (Hacker & Tenent, 2002). In the present study, the focus is on both the whole-classroom setting and teachers as instructors.

### **Implementation Quality**

As we have seen that whole-classroom settings may influence treatment effects, it is

worthwhile to investigate implementation quality. In particular, it is of interest to study implementation quality when teachers are delivering the treatment in their own classroom, overseeing multiple groups of students' practicing reading strategies through reciprocal teaching.

Even though many researchers in the field of reading comprehension underscore the need to take into account implementation quality as moderator in the analysis of treatment effects on students outcomes, especially in whole-classroom settings (Andreassen & Bråten, 2011; Hulleman & Cordray, 2009; Larsen & Samdal, 2007; Swanson, Wanzek, Haring, Ciullo, & McCulley, 2011; Vaughn et al., 2013), these studies have not been carried out yet, to our knowledge. However, qualitative studies show that teachers in whole-classroom settings face problems in the implementation of reciprocal teaching or similar interventions (Duffy, 1993; Seymour & Osana, 2003; Hacker & Tenent, 2002). Results of those studies show that teachers find it hard to induce strategic thinking in students by modeling the use of strategies and explicitly relating strategy-use to text (Duffy, 1993). In addition, teachers found the didactic principles of reciprocal teaching and the specific reading strategies that had to be taught hard to understand (Seymour & Osana, 2003). Finally, Hacker and Tenent (2002) found that when teachers implemented reciprocal teaching in their classrooms, students showed poor application of reading strategies. To deal with that problem, teachers extended whole-class instruction of reading strategies and provided more scaffolding of strategy use (Hacker & Tenent, 2002). In addition, teachers found that students exhibited poor discourse skills while collaborating. As a consequence, the teachers were hindered in changing from a teacher-centered to a student-centered approach, hampering the implementation of collaborative group work in discussing and practicing reading strategies.

Quantitative studies in whole-classroom settings that focus on teaching reading strategies show that it is difficult to find positive effects on student's reading comprehension (De Corte, Verschaffel, & Van de Ven, 2001; Fogarty et al., 2014; McKeown, Beck, & Blake, 2009; Vaughn et al., 2013; Simmons et al., 2014). These interventions were based on reciprocal teaching or used one or more of its didactic principles (such as modeling or group work). Non-significant results on reading comprehension were obtained, which may be explained by problems with treatment adherence in whole-classroom settings (Simmons et al., 2014; De Corte, Verschaffel, & Van de Ven, 2001).

### **A Whole-Classroom Intervention in Grade 7**

The present study is a follow-up of a one-year intervention with low-achieving adolescents in grade 7 (Okkinga, Van Steensel, Van Gelderen, & Slegers, 2016). Effects of a reciprocal teaching treatment in whole-classroom settings and moderation effects of implementation quality were investigated. This experiment included training and coaching of experimental teachers. This study revealed no overall treatment effects on reading comprehension.

However, we did find an interaction between implementation quality of strategy-instruction and the treatment. This effect implied that in the experimental condition more elaborate explanations of the nature, function, importance, and application of reading strategies positively contributed to students' reading comprehension. The effect was substantial: it explained an additional 37 per cent of the differences between classes after individual and class-level variables had been taken into account.

A few conclusions can be drawn from these results. First, our results underscore the importance of including implementation quality in the analyses. Neglecting such variation can result in overlooking meaningful effects. Second, no moderation effects were found for two principles of implementation quality: modeling and group work. These two principles might be hard for teachers to master, which is also suggested by the observational data of teacher implementation in classrooms. Modeling requires that teachers model strategies in a fashion that is adaptive to the students' capacities (both reading skills and word knowledge) and is able to empathize with students' own thinking processes. Group work requires active participation from the students with the teachers transferring control to their students (Okkinga, Van Steensel, Van Gelderen, & Slegers, 2016). It is plausible that it requires more time to master the necessary skills underlying modeling and the guidance and supervision of group work effectively in a classroom with multiple small groups. Several authors point to the problem that mastering multi-component treatments fostering reading comprehension (such as the present one) is quite difficult for teachers (Scammacca, 2007; Scammacca, 2015; Roberts et al., 2013). Thus, it may be necessary for teachers to spend more time to become familiar with implementing both modeling and group work effectively.

### **The Present Study**

The present study presents final results of a two-year treatment, following the same cohort of students from grade 7 to 8. While in the first year the focus of coaching was on direct instruction of reading strategies, in the second year the focus was on coaching of teachers in their modeling behavior and their guidance of group work. The study investigates the role of implementation quality and its effects on students' reading comprehension.

Our present study adds to the research base by analyzing moderation effects of implementation quality of principles of reciprocal teaching in whole-classroom settings, in addition to treatment effects regarding students' reading comprehension. This allows insight into the conditions and necessary duration under which the treatment will be effective in whole-classroom settings with low-achieving adolescents. An explorative study of Chambers Cantrell et al. (2016) suggests that after a first year a second year of intervention directed at reading strategies of 6<sup>th</sup> and 9<sup>th</sup> grade low-achieving students did show promise for improvement in reading comprehension for those students that did not profit from a first year of intervention. In addition, they showed that teachers' quality of

implementation improved from year to year (in a course of four years). Therefore, combining implementation quality and students' reading comprehension in the analysis of effects of our two-year treatment is of interest.

In this study, we will answer the following research questions:

- 1 Is reciprocal teaching in the context of whole-classroom settings, over a period of two school years, effective in fostering reading comprehension of adolescent low achievers?
- 2 Does the quality of implementation of the three main principles of reciprocal teaching (strategy instruction, modeling and group work) moderate effects on reading comprehension?

## Method

### Design

A two-year longitudinal design with a randomized controlled trial was used in this study (Shadish, Cook, & Campbell, 2002). Randomization took place at the class level. At every participating school two classes, each with their own Dutch language teacher, took part in the study. Classes within each school were randomly assigned to either the control or treatment condition. The dependent variable, reading comprehension, was measured at four time points and was used as repeated measures.

We included five control variables on the student level. First, we included gender, because girls generally show greater reading skills than boys (Logan & Johnston, 2009; Schaffner, Philipp, & Schiefele, 2014). Additionally, we included non-verbal IQ, vocabulary knowledge and metacognitive knowledge, since theoretical models and empirical evidence suggest that reading comprehension draws heavily on these variables (Just & Carpenter, 1976, 2004; LaBerge & Samuels, 1974; Ouellette & Beers, 2010; Rumelhart, 2004; Samuels, 2004; Trapman, Van Gelderen, Van Steensel, Van Schooten, & Hulstijn, 2014; Van Gelderen et al., 2004; 2007; Verhoeven & Van Leeuwe, 2008). Finally, age was included as a control variable.

Additionally, we included two control variables on the class level: teacher replacement and cancelled classes. Six teachers (three treatment and three control teacher) were replaced during the study (see teacher replacements and attrition). For some schools, it was difficult to find replacements immediately. Therefore, we also included a class-level control variable "cancelled class" to account for the missed classes. This concerned two treatment classes in total. Those classes missed at least 6 weeks of Dutch language teaching before a replacement was found.

Finally, we included three moderator variables, covering the three didactic principles behind our treatment: whole-classroom instruction of reading strategies, teacher and student modeling, and group work.

### Sample Selection and Description

Our study focused on low achievers. Our operationalization of low achievement was based on educational track. The Netherlands have a tracked system of secondary education. After primary school, students are placed in one of three tracks: prevocational secondary education, senior general secondary education, or pre-university education. This decision is based on their scores on a general attainment test (directed at language, reading and mathematics) and their educational performance as assessed by their primary school teachers (Ministry of Education, Culture, & Science, 2006). Since students in prevocational education are generally characterized by poor reading skills (Dutch Education Inspectorate, 2008; Gille, Loijens, Noijons, & Zwitser, 2010), we selected our sample from schools offering this type of education.

We recruited schools in two ways. First, we contacted schools that had participated in a previous study on low achieving readers. Second, we contacted schools via a digital community of Dutch language teachers. Schools had to meet the following five criteria:

- Willingness to participate in a two-year treatment study.
- They had (at least) two seventh grade classes.
- Each class had its own Dutch language teacher.
- The teachers were prepared to take part in the randomization procedure, implying that
  - a) if their class was assigned to the treatment condition, they were prepared to take part in our training and coaching program and to weekly give our experimental lessons; and
  - b) if their class was assigned to the control condition, they were prepared to not use our program nor discuss its contents with the colleague in the treatment condition.
- Control teachers were requested to use their regular language program during the language classes.

Ten different schools in different parts of the Netherlands were willing to participate. Within each school, two teachers volunteered. Randomization was done at the class level within each school, resulting in a total of ten experimental and ten control classes, each with their teacher, divided over the ten schools. At the start of the study, these classes comprised 369 students, of which 189 were in the treatment condition (51%) and 180 in the control condition (49%). The students' mean age was 13.01 years ( $SD = 0.52$ ) at the start of the project. There was no statistically significant difference between the two conditions on this variable ( $t(366) = -1.27, p = .20$ ). There were relatively more girls in the sample ( $n = 200$ ; 54%) than boys ( $n = 169$ ; 46%), with relatively more girls than boys (59 vs 41%) in the treatment condition. The distribution in the control condition, however, was more equal (49 vs 51%). The difference in distribution between the two conditions was statistically significant ( $\chi^2(1) = 3.99, p = .046$ ).

More female than male teachers participated in the study ( $n = 15$  vs  $n = 5$ ), with two male teachers in the treatment group and three males in the control group. The mean age of the teachers was 46.40 years ( $SD = 11.12$ ). On average they had 13.50 ( $SD = 13.73$ ,  $min = 1$ ,  $max = 38$ ) years of teaching experience in secondary education. No differences were found between the conditions on either variable, ( $t(14) = -.45$ ,  $p = .66$ ) and ( $t(14) = .053$ ,  $p = .96$ ), respectively.

### Teacher Replacements and Attrition

During the two school years, six (3 control and 3 experimental teachers) of a total of twenty teachers were replaced during the study due to pregnancy, illness or a new job. Two of the three teachers in the control condition were immediately replaced. The third control teacher was replaced after the start of the second school year because of scheduling issues. It took the schools a few weeks to find a replacement for two treatment teachers. Finally, it was not possible to replace another treatment teacher in the second school year with a teacher who was willing to participate in the study. Therefore, this class dropped out ( $n=24$  students).

There was considerable attrition among the students. From a total of 369 students at the start of the project, 44 students changed schools, of which 19 students in the treatment condition and 25 in the control condition. Six students (5 from the treatment) switched classes within their school and three students were ill for a long period of time, of which two were in the treatment condition. Therefore, in total, 53 students dropped out of the study. The frequency distribution of these categories (students staying, changing schools, switching classes, and illness) across the treatment and control condition was statistically equal,  $\chi^2(3) = 4.78$ ,  $p = .19$ . Together with the dropped out class, a total of 292 students remained, of which 140 students were in the treatment condition and 152 students in the control condition.

### Treatment

Our treatment was based on the following three principles of reciprocal teaching (Palincsar & Brown, 1984), that is:

- 1 Whole-classroom instruction of reading strategies, focusing on procedural knowledge. This implies that for each strategy, it was emphasized what the strategy entailed, how to use the strategy, when to use the strategy and why to use the strategy (Veenman, Hout-Wolters, & Afflerbach, 2006).
- 2 Teacher and student modeling. Teachers were trained to model the use of reading strategies during plenary instruction by thinking aloud when reading text. They encouraged students to take over this role, both plenary and in small group sessions.
- 3 Group work. The primary objective of encouraging students to work in groups was to

have them collaboratively apply reading strategies while thinking aloud during text reading. Teachers were given instructions on how to give feedback to the groups of students working together. For example, if a teacher noticed that the students were struggling with the application of a reading strategy, the teacher was required to model this strategy again and encourage and aid the students in doing this themselves.

With respect to strategy instruction the treatment consisted of the following five strategies related to reading comprehension (Dole, Duffy, Roehler, & Pearson, 1991; Palincsar & Brown, 1984; Pressley & Afflerbach, 1995):

- 1 *Predicting*. On the basis of text features such as title, subheadings, and pictures, students are instructed to make predictions about text content before reading, and to check their predictions while reading.
- 2 *Summarizing*. Students are instructed to summarize sections of text, encouraging them to focus on main ideas and ignore irrelevant details as well as to check their understanding of the text so far.
- 3 *Self-questioning*. Students are instructed to generate questions about the text being read, helping them to focus on main ideas as well as to monitor understanding.
- 4 *Clarifying*. When confronted with a word or passage they do not understand, students are instructed to reread, read ahead, or, in the case of an unknown word, analyze it, and see whether its meaning can be inferred by looking at parts of the word.
- 5 *Interpreting cohesive ties*. Students are instructed to look for relationships between sentences or paragraphs that are connected, e.g. by using 'signal words' (different types of connectives).

Students received weekly lessons over a period of two school years, from October until June in the first year and from September until June in the second year. The treatment was offered in the context of an existing program called "Nieuwsbegrip"<sup>®</sup>, developed by an educational consultant organization ("Comprehension of news", CED Group, 2011). Lessons were developed weekly by a team of developers at the CED Group. They were based on recent news texts (i.e., texts that had been issued the week before) about subjects close to students' everyday life (e.g., sugar in energy drinks, abdication of the Dutch queen, or 20 years of text messaging). The use of topical, interesting texts aimed to increase students' task motivation (Guthrie & Wigfield, 2000; Schiefele, 1999). The lessons could be downloaded by teachers from the program website ([www.nieuwsbegrip.nl](http://www.nieuwsbegrip.nl)) every week, starting Monday evening.

Lessons were provided in sequences of six weeks. Each sequence consisted of six weekly lessons (approximately 45 minutes per lesson). In each of the first five lessons, the focus was on one reading strategy that was practiced in a central strategy assignment that was

provided on a work sheet. In addition, students could work on other assignments (i.e., answering questions about the text) on the work sheet.

Students practiced each of the five strategies several times during the year. This cyclical approach was assumed to result in the consolidation of strategy knowledge. In the final lesson of each sequence all strategies were practiced simultaneously. The idea behind this was that students have to be able to apply all strategies together during the reading process, selecting an appropriate strategy depending on their own needs. Appendix A provides examples of translated assignments from the program for each reading strategy.

### **Training and Coaching of Treatment Teachers**

Treatment teachers took part in an extensive training and coaching program that was provided by teacher trainers from the Rotterdam University of Applied Sciences, who had, in turn, been trained by three of the authors.

In the first phase (October 2011-January 2012), teachers participated in three one-hour training sessions. In *Session 1*, they received general, practical information about the program (e.g., how to use the program website), theoretical information about the reading process and its components, and basic information about the program's didactic principles (direct instruction of reading strategies, teacher and student modeling, and group work). In *Session 2*, in-depth information was provided about the nature, function, importance, and application of the five central strategies and on the way teachers could model the use of these strategies. Examples of modeling were provided by means of video clips and lesson protocols. In *Session 3*, the focus was on reciprocal teaching and how, by means of scaffolded instruction, the use of reading strategies is transferred to students. Attention was given to how the teacher can give feedback to groups of students and how his or her expert role is gradually faded.

Teachers were given a template for the lessons that would help them keeping focused on the reading strategies (See Table 1).

In the second phase (February 2012-May 2013), teachers participated in six coaching sessions; three coaching sessions during February-June 2012 and three coaching sessions during September 2012 – May 2013. A coaching session involved a classroom observation conducted by the trainer during a treatment lesson, followed by a feedback session of approximately twenty minutes on the same day. During the classroom observations, trainers used an observation scheme comparable to the one used by the researchers (see below). This scheme directed the trainers' attention to the three principles of the treatment (whole-classroom instruction of reading strategies, modeling, and group work). During the first year, coaching was mainly directed at instruction of the reading strategies and to a lesser extent to the teacher modeling those reading strategies. During the second year, the focus of coaching was on modeling and group work.

**Table 1**

Template for the Lessons that the Treatment Teachers used

Lesson structure	Activities
<b>Introduction</b>	<ul style="list-style-type: none"> <li>• Write the subject of the text and the central strategy of the lesson on the blackboard.</li> <li>• Introduce the subject and the central strategy with a whole-class approach and activate prior knowledge.</li> <li>• Write down questions students have about the text during orientation.</li> <li>• Read the first paragraph together and model the central strategy.</li> <li>• Invite a student to read the next paragraph while thinking aloud and applying the central strategy. Give support when necessary, that is, ask questions that stimulate the use of the reading strategy.</li> </ul>
<b>Processing</b>	<ul style="list-style-type: none"> <li>• Instruct the students to work together in groups of two or three. Let them work on the remainder of the work sheet.</li> <li>• Walk around to give the groups of students feedback. Focus on the central strategy and motivate the students to apply the strategy while thinking aloud. If necessary, model the strategy again.</li> </ul>
<b>Reflection</b>	<ul style="list-style-type: none"> <li>• Reflect with the students on the reading process as well as the content.</li> <li>• Together with the students, answer the questions they had before reading the text. Did reading the text answer those questions?</li> </ul>

*Note:* The template gives an overview of the activities the teacher should initiate to keep focused on the central strategy during the lesson. The template was designed by the developers of the “CED Group”.

### Control Classes

Control classes were “business as usual”. Teachers in the control classes used their regular textbook for Dutch language arts. Among our schools, three different language textbooks were used. The teacher manuals were analyzed to determine whether the three central principles of reciprocal teaching were present. No attention was given to modeling by teachers or students or group work in the teacher manuals of the control classes. Some reading strategies were mentioned in two teacher manuals (for example, “some assignment require activating prior knowledge”), but no guidelines were given for how to instruct reading strategies. The textbooks for students were analyzed for presence of the five reading strategies of the treatment program (predicting, summarizing, self-questioning, clarifying and interpreting cohesive ties). Attention was given to reading strategies in all three textbooks. However, not all strategies that were covered in the treatment condition were also covered in the control textbooks. Reading strategies that were often referred to were: predicting, clarifying, and attention to cohesive ties. This occurred in all three textbooks with similar frequency. Self-questioning did not occur in any of the textbooks. Summarizing

only occurred as a specific assignment after reading texts, but was not used as a reading strategy during reading. Almost all of the assignments were individual and there were only a few instances where students were instructed to work together on an assignment in all three textbooks.

## **Measures**

### ***Reading comprehension***

Reading comprehension was measured by means of the SALT-reading, a test that was validated for use among low-achieving adolescents (Van Steensel, Oostdam, & Van Gelderen, 2013). The SALT-reading comprises eight tasks, each consisting of one or two texts and comprehension questions about those texts. The texts cover different genres (narrative, expository, argumentative, and instructive). They were selected from media students assumedly come across regularly in their daily lives: (school) books, newspapers, magazines, and official documents (such as regulations in a youth hostel). The eight tasks comprised a total of 59 test items, that were divided into three categories: items requiring students to retrieve relevant details from the text, items requiring students to make inferences on a local level (e.g., draw cause-effect relationships between sentences), and items requiring students to show their understanding of the macrostructure of the text (e.g., by inferring the main idea of the text or the intention of the author). The test consisted mainly of multiple-choice questions but contained also five open-ended questions. The open-ended questions were coded by the researchers with the use of a coding scheme. The SALT-reading was administered at four time points. The Cronbach's alpha coefficients were .82, .83, .82, and .85 respectively, indicating sufficient reliability (Field, 2009).

### ***Vocabulary knowledge***

Vocabulary knowledge was assessed with a 73-item multiple-choice test, measuring the knowledge of nouns, verbs, adjectives, and adverbs belonging to the 23,000 words in a dictionary for junior high school students (see Hazenberg & Hulstijn, 1996, for details). Each item consists of a neutral carrier sentence with a bold-faced target word and four answer options, one of which represents a correct synonym. Vocabulary knowledge was administered two times and the average of both was used as a measure for vocabulary knowledge. The Cronbach's alpha coefficients were .86 and .85, respectively indicating sufficient reliability (Field, 2009).

### ***IQ***

Intellectual ability was measured by administering the Raven Progressive Matrices, a nonverbal IQ test. The total test consists of 60 items, divided into 5 sets of 12 items. Each

item represents a logical reasoning puzzle. The items become more difficult within a set and the sets become increasingly difficult as well (Raven, Raven & Court, 1998). For students from the lowest tracks of prevocational education the last set was assumed to be too difficult and for this reason this set was omitted. The Cronbach's alpha coefficient was sufficient: .82 (Field, 2009).

### ***Metacognitive knowledge***

Metacognitive knowledge was measured by a questionnaire consisting of 45 statements about text characteristics, reading and writing strategies (Trapman, Van Gelderen, Van Steensel, Van Schooten, & Hulstijn, 2014). It was based on the metacognitive knowledge test constructed by Van Gelderen et al. (2003) and Van Gelderen et al. (2007). Items consisted of correct or incorrect statements and students had to agree or disagree with each statement. An example of an incorrect statement is 'The order in which you present the information in your text is usually not relevant'. The Cronbach's alpha coefficient was .51. Although this indicates a rather low level of reliability (Field, 2009) we maintained the measure because in previous research it still predicted significant variance in reading comprehension (Trapman, Van Gelderen, Van Schooten & Hulstijn, submitted).

### ***Classroom variables and treatment fidelity***

To measure the moderator variables, we conducted classroom observations in both the experimental and control conditions twice each year, resulting in a total of four observations for each class. Our aim was to examine a) whether the treatment teachers provided the lessons in the way we instructed during the training and coaching program and b) whether the control teachers applied the three treatment principles, even though they were not trained to do so. Therefore, we devised an observation scheme focusing on the three main principles: whole-class teaching of reading strategies, teacher and student modeling, and group work. This was done in the following manner, resulting in three four-point scales (0-3) to be used for further analysis:

- 1 Whole-class teaching of reading strategies. We distinguished four categories:
  - a Teachers provided no information on reading strategies (0 points).
  - b Teachers introduced the central strategy of the lesson (in the treatment condition) or any strategy (in the control condition), but provided no further explanation (1 point).
  - c Teachers introduced a strategy and explained its nature, function, importance, and/or application (2 points).
  - d Teachers introduced a strategy, explained its nature, function, importance, and/or application and discussed this strategy with the class (3 points).

- 2 Teacher and student modeling. We distinguished four categories of behavior:
  - a Teachers did not use any modeling of strategy use (0 points).
  - b Teachers modeled strategy use (1 point).
  - c Teachers modeled strategy use and asked students to think aloud while using reading strategies, either individually (i.e., in front of the class) or in groups (2 points).
  - d Teachers modeled strategy use, asked students to think aloud, and provided them with feedback (3 points).
- 3 Group work. The following four categories were distinguished:
  - a Teachers did not order students to work in groups (0 points).
  - b Teachers ordered students to work in groups, but did not provide feedback (1 point).
  - c Teachers ordered students to work in groups, provided feedback, but focusing on students' understanding of the assignment, their answers to questions, or on unknown words (2 points).
  - d Teachers ordered students to work in groups and provided feedback on collaboration itself or collaboration directed to any of the previous issues (3 points).

The scales were constructed in such a way that a 3-point score would be the optimal score for the purpose of the treatment.

Before the start of the classroom observations, the observation scheme was piloted during two lessons, one in an experimental class and one in a control class. Two researchers filled out the observation scheme during the lessons, after which they compared their codes and discussed causes for any differences. The coding scheme was adjusted when needed.

The adjusted scheme was used for all observations. In order to be able to check codes after the observation, the lessons were recorded using an audio-recorder carried by the teacher. Means were calculated over the four classroom observations per class. Inter-rater reliability was calculated by means of observed agreement between two observers. In total, 30 from a total of 76 classroom observations were performed independently by two coders. Across these 30 observations, 93.89% observed agreement was obtained.

## Procedure

Tests were first administered in the fall of 2011, just before the start of the treatment. It concerned the SALT reading, vocabulary knowledge and non-verbal IQ. At the end of the first school year (May-June 2012), the SALT reading, vocabulary knowledge and metacognitive knowledge were administered. At the start of the second school year (September-October 2012), and at the end of the second school year (May-June 2013) the SALT-reading was administered. All test administrations took place in classroom settings. The test sessions were introduced by a trained test leader. A familiar teacher was present to maintain order.

Questions were answered by the test leaders following a standardized protocol.

Classroom observations took place during January-February 2012 and during April-May 2012 in the first school year. In the second school year, classroom observations were performed during October-November 2012 and April-May 2013. During the classroom observations the researcher(s) sat at the back of the classroom to observe the teacher.

After the treatment was completed, the first author held exit interviews with 7 of the 8 treatment teachers. The interviews were semi-structured and covered the following topics: how did teachers look back on the implementation of the treatment (e.g. did they encounter difficulties, if so, how did they solve these; what were advantages and disadvantages of the treatment; how did they perceive the training and coaching), how did their view of the principles of reciprocal teaching (whole-class instruction of reading strategies; modeling and group work) change during the course of the treatment, and did they see any changes for their students (e.g. in their views about reading or their learning outcomes). Interviews lasted 45-60 minutes. All interviews were recorded and transcribed.

### **Analyses**

After the collection of all data, the dataset contained data of 292 students (see teacher replacements and attrition). First, missing data within individual tests were imputed at the item level using the EM procedure from SPSS missing value analysis. The missing data in this procedure did never exceed 7% of the data matrix. If a student was not present during a test session, all the tests from that session were regarded as missing and these missing data were not imputed. This resulted in a loss of 54 additional students for the final analysis. Therefore, the final dataset contained a total of 238 students (110 students from the treatment condition and 128 students from the control condition).

Repeated measures multilevel analyses were performed to account for the hierarchical structure of the data (using MLwiN 2.16; Rasbash, Steele, Browne, & Goldstein, 2009). The time variable 'Occasion' (variance within students across times of measurement) was defined in months; with the first measurement of reading comprehension at month zero, and subsequent measurements at months 9, 12, and 22, respectively. These months correspond to the time points of the SALT-reading: September 2011, June 2012, September 2012, and June 2013. As there are four measurements of reading comprehension, the dataset is multiplied by four, resulting in 952 'cases'. Thus, growth is measured as a repeated measures.

We tested a) whether the treatment had a significant, positive effect on growth in reading comprehension by testing the interaction between treatment (yes or no) and occasion, and b) whether the classroom variables (, strategy-instruction, modeling and group work) moderated the treatment effect.

Adding predictors was done in the order Hox (2010) suggests. First, we tested whether adding a class or school level to the model significantly improved model fit. Levels

significantly improving model fit were added to the model. Second, we tested whether a model with random slopes both at the student or class level for the occasion variable improved model fit. The treatment variable is a class level variable, random slopes at class level indicate differences in growth between classes. If a treatment effect exists, we would expect significant model fit improvement by adding random slopes at class level to the occasion variable.

Third, we added the class level variables 'teacher replacement' and 'cancelled classes' to check whether we should include these variables as covariates. Fourth, we tested whether the student-level predictors gender, IQ, age, vocabulary knowledge, and metacognitive knowledge significantly improved model fit.

To answer the first research question, the treatment variable and the interaction between treatment and occasion were added to the model. A treatment effect implies a greater learning gain in the treatment group and thus a significant interaction effect between occasion and treatment<sup>2</sup>. To answer the second research question, we started with a model containing the significant predictors of the model resulting from the first research question. For each of the three moderator variables (strategy-instruction, group work and modeling), we checked separately whether adding the moderator variable and its interactions with the occasion and treatment variables significantly improved model fit. The interaction between occasion and moderator variable is indicative of an effect of the moderator variable on growth. The three way interaction (occasion\*moderator\*treatment) indicates a differential effect on growth of the moderator variable on students in the experimental and the control group.

Dichotomous independent class variables (teacher replacement and cancelled classes) and student variables (gender) are always scored 0 and 1. All continuous independent variables (IQ, age, vocabulary knowledge and metacognitive knowledge) are centered around their grand mean before adding them to the model (Hox, 2010). The number of levels needed in the analyses was tested by comparing nested models with one-sided Chi-square significance tests (Hox, 2010). Significance of predictors was tested both with Wald-tests (coefficient divided by the standard error) and by means of comparing nested models (with and without the predictors) with a Chi-square test<sup>3</sup>.

Regression coefficients for class-level variables were tested with number of classes as sample size ( $df = \text{number of classes} - \text{number of predictors} - 1$ ) (Hox, 2010).

<sup>2</sup> A significant main effect of the 'treatment variable' indicates a significant difference between treatment and control group on the dependent variable at the onset of the study, whereas the interaction between occasion and treatment indicates a difference in growth between treatment and control group on the repeatedly measured dependent variable (reading comprehension), which can be seen as the effect of the intervention.

<sup>3</sup> The difference in  $-2 \times \text{Loglikelihood}$  of nested models has a Chi-square distribution with a number of degrees of freedom equal to the difference in number of estimated parameters between both models.

## Results

### Descriptive Statistics

Table 2 shows the mean student scores for all student level variables (the pretest and the three reading comprehension, IQ, vocabulary, and metacognitive knowledge).

No significant differences were found (according to the t-tests) between the treatment and the control condition between any of the variables. This means that there were no significant differences between the control and experimental classes before the start of the treatment (time 1) on all student level variables including vocabulary knowledge, IQ, metacognitive knowledge, and reading comprehension. In addition, no significant differences were found between the control and experimental classes at all subsequent measurements for reading comprehension (time 2-4).

**Table 2**

Descriptives Student-Level Variables

Variable	Treatment (n=110)	Control (n=128)	t-value
	Mean (SD)	Mean (SD)	
Reading comprehension (time 1)	35.47(7.21)	34.67(8.38)	.79
Reading comprehension (time 2)	37.72(6.81)	36.72(8.69)	.97
Reading comprehension (time 3)	36.85(7.10)	36.93(8.60)	.08
Reading comprehension (time 4)	37.77(8.46)	39.28(8.53)	1.36
Vocabulary	49.66(6.80)	49.56(7.85)	.97
IQ	36.01(5.08)	35.36(5.24)	1.28
Metacognitive knowledge	26.26(4.19)	25.59(4.53)	1.19

\* $p < .05$

The development of level of instructional principles (strategy-instruction, modeling, and group work) was tested with three repeated measures ANOVA's<sup>4</sup>. For each of the instructional principles no main effects over time were found (strategy-instruction:  $F(3,48) = 1.84, p = .15$ ; modeling:  $F(3,48) = 2.77, p = .05$ ; and group work:  $F(3,48) = .73, p = .54$ ), nor were interaction effects between instructional principles and treatment found (strategy-instruction:  $F(3,48) = .78, p = .51$ ; modeling:  $F(3,48) = .06, p = .82$ ; and group work:  $F(3,48) = .95, p = .42$ ), suggesting that there was no systematic difference between the treatment and control teachers in growth of use of the principles in their lessons. Seeing that the observations

<sup>4</sup> The multilevel structure was tested but there was no significant variance at school level, therefore unilevel analyses were carried out.

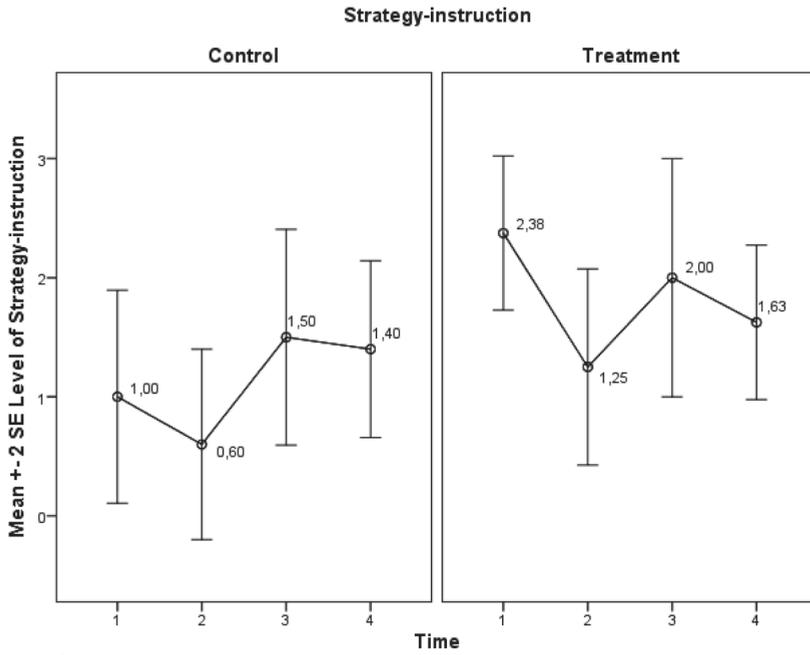
were based on 18 teachers in total, it is not surprising that no significant differences were found given the weak power of the analysis. To give a more precise impression of the development of the instructional principles over time in both the control and the treatment classes, Figures 1-3 are presented. It appears that strategy instruction was practiced more in the start of each academic year than at the end for both the control and the treatment teachers. Modeling was practiced more at the end of the first academic year and seems to decrease somewhat thereafter. Group work, however shows a slight increase in the second year of the treatment. Overall, the figures show that in the treatment group differences in application of each of the principles between the 8 teachers are quite large. Although the means are considerably higher than for the controls, that does not mean that each treatment teacher implements reciprocal teaching optimally. Ideally each of them should reach the maximum score, which is obviously not the case.

For each of the instructional principles, mean scores were calculated over the four time points. In Table 3, means and standard deviations are presented for the variables resulting from the classroom observations. As expected, the mean scores of the treatment group are higher than those of the control group, indicating that in the experimental classrooms modeling, strategy-instruction and group work were more often observed than in the control classrooms. The difference between both groups is statistically significant on the .05 level for all variables, except for strategy-instruction.

**Table 3**  
Descriptives Teacher-Level Variables

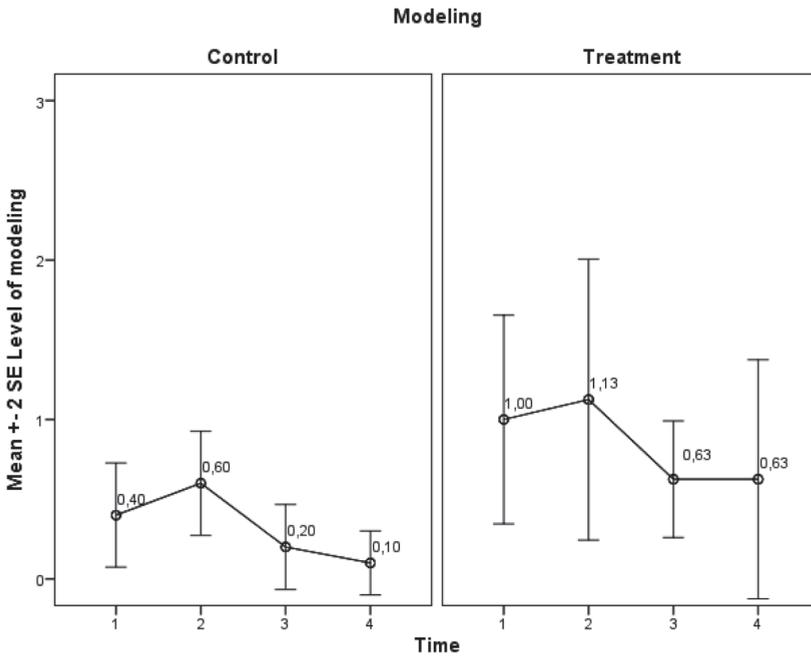
Variable	Treatment (n=8)	Control (n=10)	t-value	p-value
	Mean(SD)	Mean(SD)		
Strategy-instruction	1.81(.80)	1.13(.65)	2.02	.061
Modeling	.84(.65)	.33(.35)	2.15	.047
Group work	2.00(1.14)	.30(.33)	4.51	<.001

*Note.* Scoring between the three variables cannot be compared one-on-one. The meaning of the scoring (0-3) is different for each variable. See Classroom variables and treatment fidelity for an explanation of each variable.



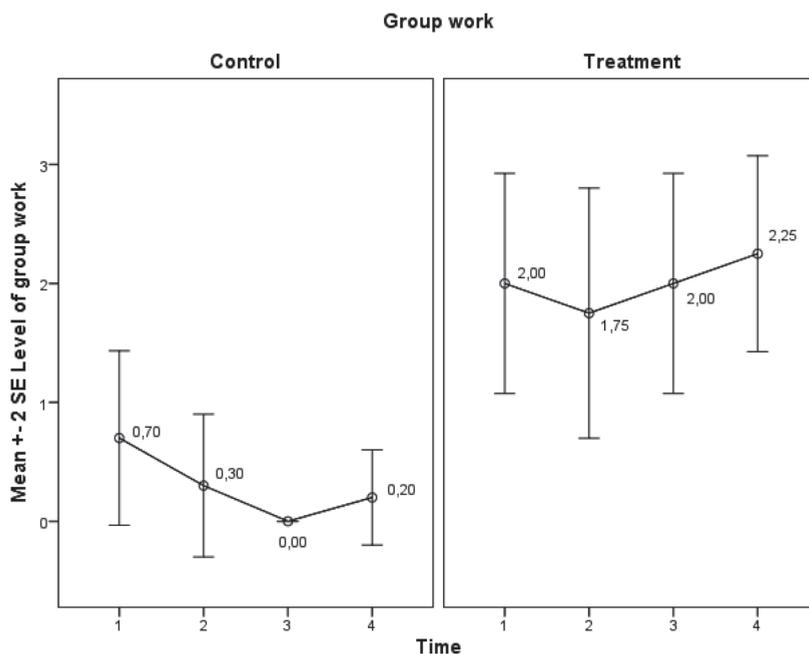
**Figure 1**

Error bars for strategy-instruction over time, for both the control ( $n=10$ ) and treatment ( $n=8$ ) teachers.



**Figure 2**

Error bars for modeling over time, for both the control ( $n=10$ ) and treatment ( $n=8$ ) teachers



**Figure 3**

Error bars for group work over time, for both the control ( $n=10$ ) and treatment ( $n=8$ ) teachers.

### Multilevel Analyses

As no significant random variance was found at the school level (see Appendix B), models with three levels were tested (occasion-, student-, and class level). Appendix B also shows that, random slopes were tested and found significant for both the class and student level. Next, the teacher-level control variables were entered in the model (see Appendix C). Both ‘teacher replacement’ and ‘cancelled classes’ did not significantly contribute to the model and were omitted from all further analyses. Subsequently, the student-level control variables were entered to control for differences between students at pretest. Inclusion of age and gender did not improve model fit (see Table 4), whereas vocabulary knowledge, metacognitive knowledge, and IQ did. Model 5 (see Table 4) is therefore the model referred to as model 0 in subsequent analyses.

In the next step, the interaction between occasion and treatment was entered. This effect was not significant implying there was no effect of the treatment on growth in reading comprehension (see Table 5, Model 2;  $\Delta IGLS = 2.131$ ,  $df = 1$ ,  $p > .05$ ).

Moderator effects of the teacher variables were tested subsequently. In Table 5, the results are depicted for the effect of strategy-instruction. We did not find a significant relationship between strategy-instruction and growth in reading comprehension: the

interaction between occasion and strategy-instruction was not significant (see Table 5, Model 5;  $\Delta IGLS = 2.165$ ,  $df = 1$ ,  $p > .05$ ). In addition, over time, no moderator effect of the level of strategy-instruction on the treatment effect was found (i.e. there was no significant interaction between strategy-instruction, occasion and treatment; see Table 5, Model 6;  $\Delta IGLS = .025$ ,  $df = 1$ ,  $p > .05$ ).

In Table 6, the results are reported for the effects of group work. No significant effect of group work on growth in reading comprehension (the interaction of group work and occasion) was found, nor was a moderator effect of group work found (the interaction of group work, occasion and treatment; see Table 6, Model 3;  $\Delta IGLS = .315$ ,  $df = 1$ ,  $p > .05$  and Model 4;  $\Delta IGLS = .007$ ,  $df = 1$ ,  $p > .05$ ).

Table 6 also shows the results for modeling. There was no significant effect of modeling on growth in reading comprehension (the interaction of modeling and occasion; see Table 6, model 7;  $\Delta IGLS = 1.162$ ,  $df = 1$ ,  $p > .05$ ;  $b = .083$ ,  $SE = .075$ ,  $p > .05$ ), but modeling did significantly moderate the effect of the treatment over time (the interaction of modeling, occasion and treatment. It appeared that in the treatment condition more elaborate modeling positively contributed to students' growth in reading comprehension (see Table 6, model 8;  $\Delta IGLS = 6.821$ ,  $df = 1$ ,  $p < .01$ ;  $b = .403$ ,  $SE = .141$ ,  $p < .05$ ). The moderator effect of modeling explains 13.69% of the variance at the class level and 43.75% of the variance in slopes at class level.





Table 4

Continued

Deviance testing									
-2*loglikelihood (deviance)	6054.507	6028.443	6022.513	5926.841	5928.564	5916.945	5916.923		
Difference between -2*loglikelihood		<b>26.064</b>	<b>5.930</b>	<b>95.692</b>	1.723 <sup>c</sup>	<b>11.619<sup>d</sup></b>	.022		
Difference df		1	1	1	(-)	1	1		
Compared to model		0	1	2	3 <sup>c</sup>	4 <sup>d</sup>	5		

<sup>a</sup> Variable is grand mean centered.

<sup>b</sup> =Model 6 from Appendix B.

<sup>c</sup> Model 4 is more parsimonious than model 3, therefore gender is omitted.

<sup>d</sup> Model 5 is more parsimonious than model 6, therefore Model 5 is used as model 0 in subsequent analyses.

**Bold =  $p < .05$ ; italicized =  $p < .01$ ; bold and italicized =  $p < .001$ .**

**Table 5** Multilevel Analyses with Reading Comprehension (repeatedly measured) as Dependent Variable to Establish Influence of Treatment (main effect) and Interaction between Strategy-Instruction, Occasion and Treatment, after Correcting for Control Variables (N=952 cases/238 student)

Model	model 1 <sup>b</sup>	model 2	model 3	model 4	model 5	model 6
<b>Fixed part</b>						
Intercept	<b>35.377</b> (.889)	<b>34.671</b> (.986)	<b>34.463</b> (.971)	<b>34.494</b> (1.005)	<b>34.174</b> (1.003)	<b>34.137</b> (1.030)
Occasion (in months)	<b>.155</b> (.040)	<b>.207</b> (.051)	<b>.207</b> (.051)	<b>.207</b> (.051)	<b>.232</b> (.050)	<b>.234</b> (.054)
IQ <sup>a</sup>	<b>.219</b> (.062)	<b>.219</b> (.062)	<b>.221</b> (.062)	<b>.221</b> (.062)	<b>.221</b> (.062)	<b>.221</b> (.062)
Vocabulary <sup>a</sup>	<b>.480</b> (.048)	<b>.479</b> (.048)	<b>.477</b> (.048)	<b>.477</b> (.048)	<b>.477</b> (.048)	<b>.477</b> (.048)
Metacognitive knowledge <sup>a</sup>	<b>.276</b> (.079)	<b>.277</b> (.079)	<b>.278</b> (.079)	<b>.278</b> (.079)	<b>.278</b> (.079)	<b>.278</b> (.079)
Treatment (1=treatment, 0=control)	-.710 (1.033)	.871(1.471)	1.322(1.491)	1.320(1.491)	2.006(1.524)	2.008(1.523)
Treatment*occasion		-.114(.076)	-.155(.076)	-.115(.076)	-.168(.079)	-.145(.169)
Strategy-instruction <sup>a</sup>			-.659(.742)	-.562(1.102)	-1.567(1.283)	-1.681(1.474)
Treatment*strategy-instruction				-.179(1.489)	-.173(1.486)	.036(1.995)
Occasion*strategy-instruction					.079(.052)	.088(.077)
Occasion*strategy-instruction*treatment						-.016(.104)

Table 5

Continued

<b>Random part (variances)</b>										
Class	8.157(3.425)	7.477(3.182)	6.612(2.909)	6.611(2.908)	6.126(2.732)	6.125(2.732)				
Class slope variance occasion	.022(.010)	.019(.009)	.019(.009)	.019(.009)	.015(.007)	.015(.007)				
Class covariance slope*intercept	-.346(.160)	-.298(.143)	-.270(.135)	-.270(.135)	-.231(.121)	-.231(.121)				
Student	13.081(2.663)	13.095(2.666)	13.098(2.665)	13.099(2.665)	13.101(2.666)	13.101(2.666)				
Student slope variance occasion	.018(.010)	.018(.010)	.018(.010)	.018(.010)	.018(.010)	.018(.010)				
Student covariance slope*intercept	.117(.123)	.116(.124)	.115(.123)	.115(.123)	.115(.124)	.115(.124)				
Occasion (rep. measures)	17.861(1.158)	17.861(1.158)	17.861(1.158)	17.861(1.158)	17.861(1.158)	17.861(1.158)				
<b>Deviance testing</b>										
-2*loglikelihood (deviance)	5916.529	5914.398	5913.722	5913.708	5911.543	5911.518				
Difference between	.016	2.131	.676	.014	2.165	.025				
-2*loglikelihood	1	1	1	1	1	1				
Difference df	0	1	2	3	4	5				
Compared to model										

<sup>a</sup> Variable is grand mean centered.<sup>b</sup> = Model 5 from Table 3.**Bold =  $p < .05$ ; italicized =  $p < .01$ ; bold and italicized =  $p < .001$**

**Table 6**

Multilevel Analyses with Reading Comprehension (repeatedly measured) as Dependent Variable to Establish Influence of Interaction between both Group Work, Occasion and Treatment, and Modeling, Occasion and Treatment, after Correcting for Control Variables ( $N=952$  cases/238 students)

Model	Model 0 <sup>b</sup>	Model 1	Model 2	Model 3
Fixed part				
Intercept	<b>34.671</b> (.986)	<b>34.242</b> (1.090)	<b>31.946</b> (1.859)	<b>31.667</b> (1.922)
Occasion (in months)	<b>.207</b> (.051)	<b>.207</b> (.051)	<b>.207</b> (.051)	<b>.228</b> (.063)
IQ <sup>a</sup>	<b>.219</b> (.062)	<b>.215</b> (.062)	<b>.219</b> (.062)	<b>.219</b> (.062)
Vocabulary <sup>a</sup>	<b>.479</b> (.048)	<b>.478</b> (.048)	<b>.482</b> (.048)	<b>.482</b> (.048)
Metacognitive knowledge <sup>a</sup>	<b>.277</b> (.079)	<b>.279</b> (.079)	<b>.289</b> (.079)	<b>.289</b> (.079)
Treatment (1=treatment, 0=control)	.871 (1.471)	1.917 (1.819)	3.839 (2.211)	4.462 (2.470)
Treatment*occasion	-.114 (.076)	-.115 (.076)	-.115 (.076)	-.161 (.112)
Group work <sup>a</sup>		-.565 (.665)	-3.599 (2.114)	-3.964 (2.211)
Treatment*group work			3.346(2.219)	3.336(2.218)
Occasion*group work				.028(.049)
Occasion*group work*treatment				
Modeling <sup>a</sup>				
Treatment*modeling				
Occasion*modeling				
Occasion*modeling*treatment				
Random part (variances)				
Class	7.477(3.182)	7.064(3.046)	6.602(2.896)	6.533(2.874)
Class slope variance occasion	.019(.009)	.019(.009)	.019(.009)	.018(.008)
Class covariance slope*intercept	-.298(.143)	-.288(.140)	-.288(.137)	-.282(.136)
Student	13.095(2.666)	13.089(2.665)	13.085(2.664)	13.082(2.664)
Student slope variance occasion	.018(.010)	.018(.010)	.018(.010)	.018(.010)
Student covariance slope*intercept	.116(.124)	.116(.123)	.115(.123)	.115(.123)
Occasion (rep. measures)	17.861(1.158)	17.861(1.158)	17.861(1.158)	17.861(1.158)
Deviance testing				
-2*loglikelihood (deviance)	5914.398	5913.699	5911.510	5911.195
Difference between -2*loglikelihood		.699	2.189	.315
Difference df		1	1	1
Compared to model		0	1	2

<sup>a</sup> Variable is grand mean centered.

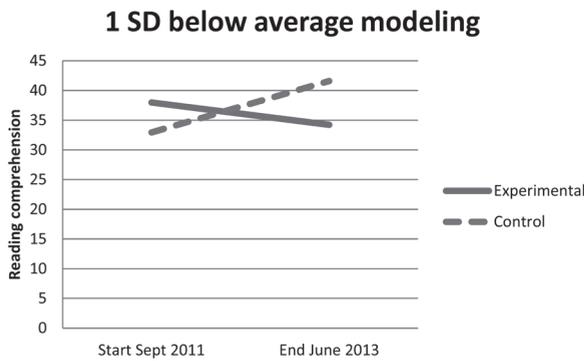
<sup>b</sup> = Model 2 from Table 4.

<sup>c</sup> This significant interaction effect indicates a 9 points higher score for the control group at the first measurement of reading comprehension for each point scored higher on modeling. Modeling was measured after the first measurement of reading comprehension (between the first and second measurement moment). For interpretation of the effect of modeling, the results of all main and interaction effects concerning modeling should be taken into account. This can therefore best be done by looking at Figures 5, 6 and 7.

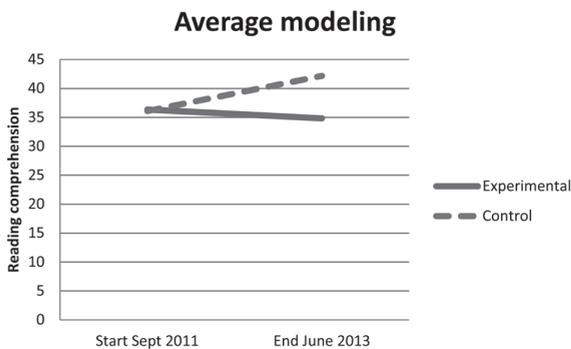
**Bold =  $p < .05$ ; italicized =  $p < .01$ ; bold and italicized =  $p < .001$ .**

Model 4	Model 5	Model 6	Model 7	Model 8
<b>31.799</b> (2.480)	<b>34.798</b> (1.033)	<b>35.578</b> (.962)	<b>35.381</b> (.968)	<b>36.099</b> (.965)
.218 (.134)	<b>.206</b> (.051)	<b>.206</b> (.051)	.178 (.055)	<b>.276</b> (.057)
<b>.219</b> (.062)	<b>.219</b> (.062)	<b>.218</b> (.062)	<b>.218</b> (.062)	<b>.218</b> (.062)
<b>.482</b> (.048)	<b>.480</b> (.048)	<b>.470</b> (.048)	<b>.470</b> (.048)	<b>.471</b> (.048)
<b>.289</b> (.079)	<b>.273</b> (.079)	<b>.260</b> (.079)	<b>.260</b> (.079)	<b>.261</b> (.079)
4.347 (2.826)	.601 (1.587)	.269 (1.369)	.691 (1.404)	.257 (1.347)
-.152 (.153)	-.114 (.076)	-.115 (.076)	-.156 (.082)	<b>-.345</b> (.094)
-3.792 (3.017)				
3.145(3.170)				
.015(.163)				
.014(.171)				
	.543 (1.077)	3.808(2.030)	2.983(2.168)	5.950(2.392)
		-4.909 (2.364)	-4.939 (2.365)	<b>-9.009</b> (2.789) <sup>c</sup>
			.083 (.075)	-.211 (.121)
				<b>.403</b> (.141)
6.527(2.872)	7.772(3.298)	4.800(2.297)	4.572(2.225)	3.946(2.015)
.018(.008)	.019(.009)	.018(.008)	.016(.008)	.009(.005)
-.282(.136)	-.309(.146)	-.204(.116)	-.182(.108)	-.115(.083)
13.083(2.664)	13.080(2.663)	13.029(2.658)	13.043(2.659)	13.001(2.655)
.018(.010)	.018(.010)	.018(.010)	.018(.010)	.018(.010)
.115(.123)	.116(.123)	.114(.123)	.113(.124)	.115(.123)
17.861(1.158)	17.861(1.158)	17.861(1.158)	17.861(1.158)	17.861(1.158)
5911.188	5914.165	5911.381	5910.219	5903.398
.007	.233	2.784	1.162	<b>6.821</b>
1	1	1	1	1
3	0	5	6	7

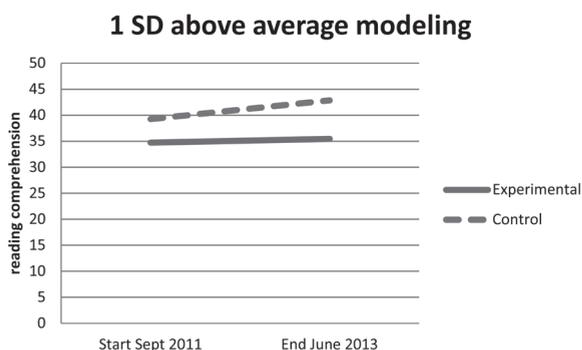
The interpretation of the moderating effect of modeling on growth in reading comprehension becomes clear when looking at regressions for different combinations of scores on the independent variables (Hox, 2010). For treatment we used two scores (0 and 1), for occasion we used the scores 0 and 3 for the first and the last time of measurement and for modeling we used three scores: one standard deviation below the mean, on the mean and one standard deviation above the mean. The resulting six regression lines are presented in Figures 4-6. These figures show that in cases where modeling is less elaborate or moderately elaborate, growth in treatment students' reading comprehension is less than in the control group. Only in the case of more elaborate modeling (Figure 6) the development in reading comprehension of the two groups appears to be similar.



**Figure 4**  
Regression lines of growth in reading comprehension for the experimental and control condition for 1 SD below average modeling.



**Figure 5**  
Regression lines of growth in reading comprehension for the experimental and control condition for average modeling.



**Figure 6**

Regression lines of growth in reading comprehension for the experimental and control condition for 1 SD above average modeling.

## Discussion

Our study set out to analyze how reciprocal teaching can improve low achieving adolescents' reading comprehension in whole-classroom settings and to what extent treatment effects are dependent on implementation quality. Apart from analyzing the overall effects of the treatment in a whole-classroom setting, our aim was to examine whether effects were larger when teachers provided more elaborate instruction of reading strategies, engaged more in teacher modeling, promoted more student modeling, and when they supported more collaboration during group work.

With regard to our first research question, our study revealed no overall treatment effects. No significant differences were found between students in the treatment classes compared to the control classes on the growth in reading comprehension. In this respect, our result is similar to what was found in several other studies analyzing effects of reading strategy instruction on reading comprehension in whole-classroom settings (Authors, in press; De Corte, Verschaffel, & Van de Ven, 2001; Fogarty et al., 2014; McKeown, Beck, & Blake, 2009; Simmons et al., 2014; Vaughn et al, 2013). Answering our second research question, we did not find significant moderation effects of strategy-instruction and group work on students' reading comprehension growth. Modeling, however, moderated the effect of the treatment over time.

### Observations of Implementation Quality

Implementation quality may have played an important role in preventing an overall positive effect of the treatment. First, our observational data of implementation quality showed that quality of strategy instruction, modeling and group work differed quite a lot between

treatment teachers. These data also showed that the average scores of these principles of reciprocal teaching, although higher than for the controls, were still not as high as could be wished for. Even on the last observations in the end of the second year, averages were only slightly higher – no significant differences were found – than in the end of the first year for strategy-instruction and group work, and even a little lower for modeling. These results are quite disappointing. It is plausible that a certain level of treatment-adherence should be met for the treatment to sort effect, as Simmons, Fogarty, Oslund et al. (2014) suggested. In their study, as in ours, the treatment involved multiple components. The authors argue that in such cases it is difficult to ensure that teachers implement each of the different components as intended (see also Roberts et al. 2013). It seems that even two years of training and coaching is not sufficient to improve implementation quality in a decisive way.

### **Main Treatment Effects**

There are at least two possible explanations for the lack of success of our treatment (research question 1). The first has to do with its length, the second with advantages of the business-as-usual control. The fact that we deal with a two-year treatment may not have worked out beneficially for our students. It is possible that a long-term intervention results in disappointing effects, because it is difficult to maintain experimental control ensuring high implementation quality over a longer period of time. This is similar to an effect noted by Roberts et al. (2013), who point out the negative effects of scale in several reading comprehension interventions. According to the authors, large scale interventions have a disadvantage in terms of internal validity and experimental control. Long-term interventions are also sensitive to such disadvantages. An example of how this might work was found in the exit interviews with teachers carried out after the treatment was completed. One of the main comments made was that teachers found it hard to maintain the motivation of the students in the second year of the treatment, and as a result found it hard to keep motivated themselves, as one teacher put it: “Two years of the same thing is a long time. The novelty effect is gone and they [students] want something different, something new.” This attitude of students and teachers might have resulted in a less than inspired way of bringing forward the classroom practices necessary for reciprocal teaching. Short-term interventions do not meet this problem of decrease in motivation and quality of implementation, as they appear as novel and interesting.

The second explanation for our disappointing result is the possibility that “business as usual” as practiced in the control group may have had some advantages over the treatment. For example, in the control group, teachers were free to select their own materials, adapt these materials or their teaching practices and had more opportunities to motivate their students by adapting their approach. The experimental treatment was quite rigid in the sense that it was prescribed what the teachers should teach, what materials and texts

they should use and how the pedagogical and didactic procedures should be realized (i.e. there were specific guidelines for strategy use, modeling and group work). This may have had negative side effects such as that our treatment teachers were less able to adapt their teaching to the needs of students in their classroom. In addition, it is possible that the experimental teachers felt less comfortable with such procedures as modeling and group work, which were not part of their normal classroom routines.

### **Moderation Effects**

Below, we will discuss the moderation effects for each of the three principles of reciprocal teaching (research question 2) and how they can be interpreted.

#### **Strategy-instruction**

We did not find a moderation effect for strategy-instruction on the basis of the two-year intervention, whereas there was a significant positive interaction for strategy-instruction on reading comprehension in the first year (Authors, in press). This difference can be explained by looking at the different roles that strategy instruction might have played in the first and second year of the intervention. The strategy component of the treatment was obviously more important in the first year, because students had to be familiarized at the start with the nature and function, of each strategy. In the second year of the treatment, however, it was not necessary to spend much attention to explaining the characteristics of the different reading strategies to the students, as they were already quite familiar with them. Students knew the characteristics of the different reading strategies, but now particularly needed to learn how to apply those strategies. Elaborate attention to teaching of the nature and function of the strategies in the second year can therefore be expected not to be more successful in fostering reading comprehension. Thus, it is not surprising that we found no overall moderation effect of strategy instruction after the two-year intervention.

#### **Modeling**

The moderation effect of modeling shows that growth in reading comprehension of the treatment and control students depends on teachers' attention to modeling (both by themselves and by students). It appeared that growth in reading comprehension in the case of less elaborate modeling (one standard deviation below average or on average) in the treatment group was less than in the control group. Only in the case of more elaborate modeling (one standard deviation above average) the growth in reading comprehension between the two groups appeared to be the same. In order to model reading strategies, a certain theatricality is needed. If the teacher is not comfortable with acting, this theatricality is difficult to master, as one teacher put it in our exit interviews: "It [modeling] didn't work,

the students thought it was strange when I tried to model". Another teacher pointed out that she found it difficult to teach the students how to model. In both cases, modeling was jeopardized. The impressions of these two teachers were representative for the majority of the treatment teachers. This can be seen in the mean scores for modeling in our observations, which were rather low. However, the significant moderation effect shows that a few of the treatment teachers nevertheless could make a difference by more elaborate modeling of reading strategies. Although the moderation did not result in more growth in reading comprehension for the treatment group, it shows promise for reciprocal teaching in the future. Given that even the best teachers in our study did not perform optimally in modeling, there is reason to expect that when modeling is practiced optimally, results on students' reading comprehension will be better than we have found.

### **Group work**

Regarding group work, we can conclude that the specific focus on coaching teachers in implementing this component of reciprocal teaching did result in a slight improvement in the quality of observed group work. Whereas in the first year teachers did not reach a higher average than 2 (meaning that group work was organized, but students received no teacher feedback on cooperation), in the second year the average score was higher (2.25 with a maximum of 3), meaning that feedback on collaboration was provided more frequently. Nevertheless, no positive effects of this improved group work were obtained in terms of improved reading comprehension, given that no significant moderation effect for group work was found.

In the original format of reciprocal teaching, small groups of students were taken out of the classroom. Under the guidance of a tutor, who had optimal control over the students' behavior, they practiced reading strategies while reading a text (Palincsar & Brown, 1984). In our treatment, the teachers were to manage up to five groups of students and provide guidance to all simultaneously. This means that compared to the original format, there is much less teacher control on collaboration and the quality of strategy use by the students in each group. The strength of reciprocal teaching may lie in the fact that there is enough time for the needs of each group of students. However, in a whole-classroom setting this time needs to be shared among multiple groups. This disadvantage of whole-classroom settings is supported by observations of Hacker and Tenent (2002). According to their in-depth analysis of teacher practice in reciprocal teaching, they concluded that group work was the most vulnerable component. The collaboration process between students was hampered because students did not practice reciprocal teaching in a productive way. Their discussions about the texts were rather superficial, and therefore did not reach a higher level of comprehension monitoring. In order to compensate for this problem teachers often returned to whole-classroom instruction, thereby jeopardizing one of the most important

aspects of reciprocal teaching: the fading of responsibility for the reading process to the students.

In addition, we have to consider the fact that our students were low-achievers and therefore may have needed much more guidance in group work than higher achieving students. Our students have to be supported in comprehension monitoring, because many of them are not used making inferences and practicing other types of deeper comprehension processes (Oakhill & Cain, 2007; Rapp, Van den Broek, McMaster et al., 2007; De Milliano, Van Gelderen, Slegers, 2014).

Moreover, it seems that the students needed more support for their collaboration process. From our own classroom observations of the treatment teachers we may conclude that the majority of them experienced serious problems with group work guidance, resulting in insufficient collaborative practice in reading strategies in the whole-classroom settings. In the exit interviews, teachers acknowledged that working in small groups is quite difficult in a whole-classroom setting. They found it hard to keep order and to keep the students motivated in the groups that the teacher was not supporting at that particular moment. Thus, one teacher asked herself: "Do I keep trying to work in groups, even if that is at the expense of learning outcomes?" Nevertheless, some of our teachers recognized the added value group work may have. In our exit interviews, a teacher, who was already proficient in applying group work before the start of the treatment, said: "If you mainly keep focused on whole-class instruction, it is difficult to get a glimpse of the reading process of the students. You keep repeating the reading strategies and hope that the students pick up what they need. For me, that is not enough. I want to exert more control on the [reading] process of the students." Another teacher added that a big advantage of group work is that all students take multiple turns, and that they can react immediately to each other. In a whole-classroom approach, students have to wait longer to take turns, and there is no time for every student to take a turn. But, "you need to be able to steer the group work in such a way that they [the groups of students] work effectively". These remarks about group work point to the fact that more intensive coaching may be needed to make this type of reciprocal teaching successful. Accordingly, many teachers mentioned that they would have appreciated more coaching of group work. Alternatively, it may be needed to include extra classroom assistance as group tutors.

### **Suggestions for Further Research and Implications for Practice**

The results of our study emphasize the importance of taking into account different aspects of quality of implementation as moderators in the analysis of treatment effects. Not taking into account quality of implementation may lead to overlooking meaningful effects, in particular in whole-classroom settings (Authors, in press; Hulleman & Cordray, 2009; Larsen & Samdal, 2007; Swanson, Wanzek, Haring, Ciullo, & McCulley, 2011; Vaughn et

al., 2013). Incorporating quality of implementation may also give clues to which treatment components contribute to the treatment effects. To our knowledge, our study is the first that systematically analyzes the moderating role of implementation quality of reading comprehension instruction with reciprocal teaching. We strongly recommend that future studies incorporate such moderation analyses in order to enhance our knowledge of conditions of successful application. In particular, it is of interest to find out what differences exist in successful implementation for different student populations (e.g. low ability vs high ability, younger vs older etc.).

In addition, our observations of classroom practices and exit interviews revealed large differences between treatment teachers in how they implemented reciprocal teaching. Therefore, it is recommended that prior insight into classroom practices of individual teachers are used to adapt the contents of training and coaching to their specific needs. For example, for teachers who have no prior experience with managing multiple groups in a whole-classroom setting may need support dividing their attention among the groups in an efficient and effective manner. Aspects of group work, such as group composition, group-size and ability grouping can then be discussed in detail and adapted to the needs of teachers and their classes. Such prior knowledge of individual teachers' classroom practices is useful in optimizing conditions for experimental research into reciprocal teaching, but it also may be useful for educational practice. Programs that use principles of reciprocal teaching in educational practice will certainly profit from such tailored training and coaching to the individual needs of teachers.

Finally, we need to acknowledge the fact that it is difficult to implement reciprocal teaching for low-achieving adolescents in whole-classroom settings. Even in our two-year treatment, treatment teachers did not succeed in an optimal implementation of reciprocal teaching. The main reason for this seems to be that the whole-classroom setting makes it difficult to attend to multiple groups of students at the same time and give them the guidance they need. However, given that the central objective of reciprocal teaching for reading comprehension instruction is to achieve more responsibility for students' comprehension processes, there is no doubt that the quality of group work should be a prime concern for educational practice. It is important that such quality can be guaranteed, so that students may experience that the use of reading strategies is not the goal but the means for using textual information for reaching their own goals. These goals may be strictly related to the school context (such as content area learning), but they are also relevant in a much wider context, such as their future professional careers.



## Appendix A

Examples of strategy assignments, translated from several assignment sheets from the program “Nieuwsbegrip”

Strategy	Example
<b>Predicting</b>	This text has five subheadings. Write down for each subheading a) which thoughts it evokes and b) what you already know about the subject addressed in the subheading.
<b>Summarizing</b>	Read the text. Read paragraph by paragraph and underline in each paragraph the most important information. For each paragraph, write one or two sentences summarizing it. Use the words you underlined.
<b>Self-questioning</b>	Read the text. Note at least five questions that spring to mind while reading.
<b>Clarifying</b>	Search the text for difficult words. Try to uncover their meaning using these hints: a) reread the previous piece of text or read on, b) look at the illustrations in the text, c) look at the word: you might know part of the word, d) sometimes you have to use your own knowledge to figure out word meanings, or e) use a dictionary.
<b>Interpreting cohesive ties</b>	Read the text. Underline the signal words. Answer the questions, while noting the signal words: <ul style="list-style-type: none"> <li>• Which contrast is explained in lines 16-17? [signal word = however]</li> <li>• Why are energy boosters unfit as sports drinks? [signal word = hence]</li> </ul>

## Appendix B

**Table B**

Multilevel Analyses with Reading Comprehension (repeatedly measured) as Dependent Variable to Establish Multilevel Structure of Data (N=952 cases/238 students)

Model	model 0	model 1	model 2	model 3	model 4	model 5	model 6 <sup>a</sup>
<b>Fixed part</b>							
Intercept	<b>35.232</b> (.499)	<b>34.661</b> (.965)	<b>34.385</b> (1.159)	<b>34.633</b> (.985)	<b>34.635</b> (.983)	<b>34.608</b> (1.050)	<b>34.593</b> (1.094)
Occasion (in months)	<b>.150</b> (.019)	<b>.150</b> (.019)	<b>.150</b> (.019)	<b>.150</b> (.021)	<b>.150</b> (.021)	<b>.152</b> (.038)	<b>.156</b> (.040)
<b>Random part (variances)</b>							
School			8.497 (6.123)				
Class		12.912 (5.315)	4.599 (3.730)	14.069 (5.651)	13.968 (5.636)	16.506 (6.575)	18.218 (7.174)
Class slope variance occasion						.019 (.009)	.022 (.010)
Class covariance slope*intercept						0 <sup>b</sup>	-.355 (.209)
Student	43.576 (4.486)	32.237 (3.586)	32.302 (3.592)	28.643 (3.485)	29.228 (4.139)	27.849 (4.020)	27.553 (3.994)
Student slope variance occasion				.036 (.009)	.038 (.011)	.019 (.010)	.018 (.010)
Student covariance slope*intercept				0 <sup>b</sup>	-.041 (.161)	.121 (.146)	.142 (.144)
Occasion (rep. measures)	21.050 (1.114)	21.050 (1.114)	21.050 (1.114)	17.949 (1.101)	17.861 (1.158)	17.861 (1.158)	17.861 (1.158)
<b>Deviance testing</b>							
-2*loglikelihood (deviance)	6132.566	6100.015	6097.977	6077.942	6077.883	6058.501	6054.507
Difference between -2*loglikelihood		<i>32.551</i>	2.038	<i>22.037</i>	.059	<i>19.382</i>	<b>3.994</b>
Difference df		1	1	1	1	1	1
Compared to model		0	1	1	3	4	5

<sup>a</sup>Model 6 shows that the multilevel structure of the data consists of three levels: Class, Student and Occasion. Random slopes are applied for both the class and student level, as slopes differ significantly for both classes and students

<sup>b</sup>= fixed at zero

**Bold**= $p < .05$ ; *italicized*= $p < .01$ ; **bold and italicized**= $p < .001$ .

## Appendix C

**Table C**

Multilevel Analyses with Reading Comprehension (repeatedly measured) as Dependent Variable to Verify Influence of Teacher Replacement (0=no, 1=yes) and Cancelled Classes (0=less than 6 weeks, 1=6 weeks or more) ( $N=952/238$ )

Model	model 0 <sup>a</sup>	model 1	model 2 <sup>b</sup>	model 3	model 4 <sup>c</sup>
<b>Fixed part</b>					
Intercept	34.593 (1.094)	35.292 (1.914)	34.940 (1.227)	34.961 (1.149)	34.666 (1.121)
Occasion (in months)	.156 (.040)	.157 (.040)	.178 (.044)	.157 (.040)	.173 (.038)
Teacher replacement (yes=1)		-3.160 (2.153)	-1.558 (2.619)		
Teacher replacement*occasion			-.101 (.094)		
Cancelled classes				-6.995 (3.997)	-1.392 (4.889)
Cancelled classes*occasion					-.333(.169)
<b>Random part (variances)</b>					
Class	18.218 (7.174)	18.306 (7.196)	17.829 (7.044)	19.654(7.650)	18.106 (7.136)
Class slope variance occasion	.022 (.010)	.022(.010)	.020 (.009)	.022 (.010)	.017 (.008)
Class covariance slope*intercept	-.355 (.209)	-.410 (.216)	-.380 (.206)	-.466 (.229)	-.378 (.198)
Student	27.553 (3.994)	27.542 (3.993)	27.548 (3.993)	27.559 (3.994)	27.555 (3.994)
Student slope variance occasion	.018 (.010)	.018 (.010)	.018 (.010)	.018 (.010)	.018 (.010)
Student covariance slope*intercept	.142 (.144)	.142 (.144)	.142 (.144)	.142 (.144)	.142 (.144)
Occasion (rep. measures)	17.861 (1.158)	17.861 (1.158)	17.861 (1.158)	17.861 (1.158)	17.861 (1.158)
<b>Deviance</b>					
-2*loglikelihood (deviance)	6054.507	6052.474	6051.360	6051.972	6048.400
Difference between -2*loglikelihood		2.033	1.114	2.535	3.572
Difference df		1	1	1	1
Compared to model		0	1	0	3

<sup>a</sup>=Model 6 from Appendix B.

<sup>b</sup>Model 2 shows that there is no significant differential growth for classes in which teacher replacement took place compared to classes in which this was not the case.

<sup>c</sup>Model 4 shows that cancelled classes did not account for significant differential growth in classes with more than 6 weeks of cancelled classes compared to less than 6 weeks of cancelled classes. Therefore, both teacher replacement and cancelled classes were omitted from further analyses.

**Bold**=  $p < .05$ , *Italicized* =  $p < .001$ .







# Does vocabulary knowledge matter in the effectiveness of teaching reading strategies? Differential responses from low-achieving adolescents on growth in reading comprehension



This chapter is submitted as: Okkinga, M., Van Gelderen, A.J.S., Van Schooten, E., Van Steensel, R., & Slegers, P.J.C. (submitted). Does vocabulary knowledge matter in the effectiveness of teaching reading strategies? Differential responses from low-achieving adolescents on growth in reading comprehension



**Abstract**

Low-achieving adolescents are known to have difficulties with reading comprehension. Previous research suggests that teaching reading strategies promotes reading comprehension in those students, but the results of those studies are mixed. Individual differences between students may explain these mixed results. This article discusses to what extent vocabulary knowledge influences the effect of a two-year intervention program focused on teaching reading strategies. It is hypothesized that the low-achieving students ( $N = 292$ ) with different levels of vocabulary knowledge respond differently to the treatment, as vocabulary knowledge is an important factor in reading comprehension. Results show that vocabulary knowledge moderates the effect of the treatment, suggesting that level of vocabulary knowledge matters in the effectiveness of the treatment. It is suggested that vocabulary knowledge is a prerequisite for the successful application of reading strategies. Especially in low-achievers with a lower vocabulary knowledge, cognitive overload may play a role when simultaneously applying newly learned reading strategies while trying to find out the meaning of multiple words which are needed for successful application of reading strategies.

*Keywords:* reciprocal teaching, reading comprehension, reading strategies, vocabulary knowledge, moderating effect

## Introduction

Many low-achieving adolescents struggle with reading comprehension (e.g., Organisation for Economic Co-operation and Development [OECD], 2003; OECD, 2014). Since it is a fundamental skill in many school subjects, poor reading comprehension has serious implications for students' educational success and, consequently, for their later societal careers. Since the 1980's, reading programs have focused on teaching reading strategies in order to foster reading comprehension (Pressley & Afflerbach, 1995; Raphael, George, Weber, & Nies, 2009). However, results of research in teaching reading strategies to low-achieving students are mixed (Edmonds, et al., 2009; Fogarty et al., 2014; Okkinga, Van Gelderen, Van Schooten, Van Steensel, & Slegers, submitted; Simmons et al., 2014; Vaughn, 2013). A possible explanation for the different results is that these studies do not take individual differences in students' skills into account. In the present study we analyzed whether low-achieving adolescents' vocabulary knowledge moderates the effect of instruction in reading strategies on their reading comprehension.

## Reading Comprehension and Vocabulary Knowledge

Reading comprehension is a complex process involving several sets of knowledge and skills. Most models of reading comprehension distinguish between lower and higher order skills that interact in the process of creating a mental representation of a text, or a situation model (Kintsch, 1998). To construct a coherent mental representation, the information in the text is integrated with the readers' background knowledge. Lower order skills include letter and word recognition (Perfetti, Landi, & Oakhill, 2005), while higher order skills refer to the ability to give meaning to words and sentences, make inferences, and make representations of paragraphs or a text as a whole (Aarnoutse & Van Leeuwe, 1988). Thus, vocabulary knowledge is a basic component of these higher order skills of reading comprehension.

It is safe to assume that if many words of a text are not properly understood, it becomes difficult to comprehend the text (Torgesen, 2000). This assumption is supported by ample empirical evidence in which a strong relationship between vocabulary knowledge and reading comprehension is found (Ouellette & Beers, 2010; Tilstra, McMaster, Van den Broek, Kendeou, & Rapp, 2009; Trapman, Van Gelderen, Van Steensel, Van Schooten, & Hulstijn, 2014; Van Gelderen et al., 2004; 2007; Van Steensel, Oostdam, Van Gelderen, & Van Schooten, 2016; Verhoeven & Van Leeuwe, 2008).

## Instructing Reading Strategies to Foster Reading Comprehension

A reading strategy is a mental tool a reader uses purposefully to monitor, repair, or bolster comprehension (Afflerbach & Cho, 2009). The use of reading strategies is a deliberate and goal-directed attempt to construct meaning of text (Afflerbach, Pearson, & Paris, 2008), and

as such, can refer to both metacognitive and cognitive strategies that aid the process of reading (Dole, Nokes, & Drits, 2009). Researchers have suggested many different strategies (Pressley & Afflerbach, 1995). Important strategies involve setting explicit reading goals, activating relevant background knowledge, allocating attention to major content while ignoring irrelevant details, the evaluating the validity of text content, comprehension monitoring, and making and testing interpretations, predictions, and drawing conclusions (Palincsar & Brown, 1984).

A widely used method of instructing, teaching, and guiding poorly reading adolescents in the use of reading strategies to foster reading comprehension is reciprocal teaching (Palincsar & Brown, 1984; Palincsar, Brown, & Martin, 1987). Reciprocal teaching consists of a set of three principles: a) teaching comprehension-fostering reading strategies b) expert modeling, scaffolding and fading; and c) students taking turns in practicing reading strategies and discussing with other students. The method includes the teacher explicitly modeling the use of reading strategies during the start of reciprocal teaching (Rosenhine & Meister, 1994) as well as scaffolding the application of reading strategies within the groups of students working together. During this process, students become increasingly more capable of regulating their own reading process and the role of the teacher gradually fades. Many studies have demonstrated positive effects of this approach (Kelly, Moore, & Tuck, 2011; Palincsar & Brown, 1984; Rosenshine & Meister, 1994; Spörer, Brunstein, & Kieschke, 2009). However, there are also indications that the approach of teaching reading strategies based on reciprocal teaching is not always successful in improving low-achievers' reading comprehension (Edmonds, et al., 2009; Fogarty et al., 2014; Okkinga, Van Gelderen, Van Schooten, Van Steensel, & Slegers, submitted; Simmons et al., 2014; Vaughn, 2013). An important factor that might explain these mixed results is the role of vocabulary knowledge in reading comprehension and strategy use.

Most interventions aimed at instructing reading strategies to foster reading comprehension in low achievers are based on research that is focused on characteristics of the reading process of good readers (Pressley & Afflerbach, 1995; Raphael et al., 2009). The idea is that when low-achievers imitate good readers' reading behavior, their reading problems will diminish. However, this line of reasoning might be problematic for low-achievers because they do not only differ from good readers in their use of reading strategies, but also in their level of vocabulary knowledge. Even within groups of low-achieving students vocabulary knowledge explains reading comprehension skill of students of different ages to a substantial degree (Trapman et al., 2014; Trapman, Van Gelderen, Van Schooten, & Hulstijn, 2017; Van Gelderen et al., 2004, 2007; Van Steensel et al., 2016). Low-achieving students with relatively little vocabulary knowledge may be less able to profit from strategies derived from the good readers' reading processes. In that case, these students are also less likely to profit from an intervention aimed at instructing reading

strategies than low-achievers scoring higher on vocabulary knowledge. For example, a strategy such as monitoring comprehension of sentences depends for a great part on the degree of accuracy of knowledge of the words used in these sentences. Low-achieving students that suffer from low vocabulary knowledge will therefore be at a disadvantage in practicing this type of monitoring, which is one of the most important strategies in the approach of reciprocal teaching practiced by Palincsar and Brown (1984).

This leads to the following hypothesis: low-achieving students differing in their level of vocabulary knowledge will not benefit similarly from an intervention aimed at instructing reading strategies to foster reading comprehension. In other words, low-achieving students may differ in vocabulary knowledge and these differences may influence students' response to a reciprocal teaching intervention directed at the improvement of reading strategies. Specifically, we hypothesize that students scoring low on vocabulary knowledge will benefit less from the intervention than students scoring high on vocabulary knowledge.

### **The Present Study**

In the present study, reciprocal teaching was used in a two-year intervention to improve reading comprehension of low-achieving adolescents. Previously, we found small effects of implementation quality of the intervention on growth in reading comprehension, but no overall effect on the experimental students' growth (Okkinga, et al., submitted). Nevertheless, it is possible that students' differing in vocabulary knowledge show different growth patterns dependent on the intervention. As discussed above, these differences in vocabulary knowledge can be quite decisive for reading comprehension and for the application of reading strategies and therefore may moderate the degree of growth in reading comprehension of both experimental and control students.

For that reason, we examine the moderating role of vocabulary knowledge in the effects of the intervention on growth in reading comprehension.

We will answer the following research question:

- 1 Is the effect of training reading strategies to foster reading comprehension in low-achieving students moderated by vocabulary knowledge?

## Method

### Design

A two-year longitudinal design with a cluster randomized controlled trial was used in this study (Shadish, Cook, & Campbell, 2002). Randomization took place at the class level. At every participating school two classes, each with their own Dutch language teacher, took part in the study. The dependent variable, reading comprehension, was measured at four time points. We included control variables on the student level (gender, IQ, language background, metacognitive knowledge and age). Finally, vocabulary knowledge was included as a moderator variable.

Gender was included as a control variable because girls generally show greater reading skills than boys (Logan & Johnston, 2009; Schaffner, Philipp, & Schiefele, 2016). Higher scores on IQ and metacognitive knowledge correlate with higher scores on reading comprehension and, therefore, also IQ and metacognitive knowledge were included as control variables (Just & Carpenter, 1976, 2004; LaBerge & Samuels, 1974; Rumelhart, 2004; Samuels, 2004; Ouellette & Beers, 2010; Van Gelderen et al., 2004, 2007; Verhoeven & Van Leeuwe, 2008; Trapman et al., 2014). Since empirical evidence suggests that factors contributing to reading comprehension are different for low-achieving students who mainly speak another language than Dutch at home compared to students who speak mainly Dutch at home (Trapman, 2015), language background was also included as a control variable. In addition, age was included as a control variable.

### Sample

Our study focused on low achievers. Our operationalization of low achievement was based on educational track. The Netherlands have a tracked system of secondary education. After primary school, students are placed in one of three tracks—prevocational secondary education, senior general secondary education, pre-university education— on the basis of their scores on a general attainment test and their educational performance as assessed by their primary school teachers (Ministry of Education, Culture, & Science, 2006). Since students in prevocational education are generally characterized by poor reading skills (Dutch Education Inspectorate, 2008; Gille, Loijens, Noijons, & Zwitser, 2010), we selected our sample from schools offering this type of education.

We recruited schools in two ways. First, we contacted schools that had participated in a previous study on low achieving readers. Second, we contacted schools via a digital community of Dutch language teachers. Schools had to meet the following five criteria:

- Willingness to participate in a treatment study.
- They had (at least) two seventh grade classes.
- Each class had its own Dutch language teacher.

- The teachers were prepared to take part in the randomization procedure, implying that a) if their class was assigned to the treatment condition, they were prepared to take part in our training and coaching program and to weekly give our experimental lessons; and b) if their class was assigned to the control condition, they were prepared to not use our program nor discuss its contents with the colleague in the treatment condition.
- Control teachers were requested to use their regular reading program during the language classes.

Ten different schools in different parts of the Netherlands were willing to participate. Within each school, two teachers volunteered. Randomization was done at the class level within each school, resulting in a total of ten experimental and ten control classes. At the start of the study, these classes comprised 369 students, of which 189 were in the treatment condition (51%) and 180 in the control condition (49%). The students' mean age was 13.01 years ( $SD = 0.52$ ) at the start of the project. The mean ages of the students in the two conditions did not differ significantly,  $t(366) = -1.27, p = .20$ . There were relatively more girls in the sample ( $n = 200$ ; 54%) than boys ( $n = 169$ ; 46%), with relatively more girls than boys (59 vs 41%) in the treatment condition. The distribution in the control condition, however, was more equal (49 vs 51%). The difference in distribution between the two conditions was statistically significant ( $\chi^2(1) = 3.99, p = .046$ ). Most of the students had a Dutch language background (84.2%), while a minority more often spoke another language than Dutch at home (15.8%). The distribution of language background across conditions did not differ statistically ( $\chi^2(2) = 0.024, p = .988$ ).

More female than male teachers participated in the study ( $n = 15$  vs.  $n = 5$ ). There were two male teachers in the treatment group and three in the control group. The mean age of the teachers was 46.40 years ( $SD = 11.12$ ). On average they had 13.50 ( $SD = 13.73, min = 1, max = 38$ ) years of teaching experience in secondary education. No significant differences were found between the conditions on age ( $t(14) = -.45, p = .66$ ) and years of experience ( $t(14) = .053, p = .96$ ).

### Attrition and Teacher Replacements

One experimental class dropped out of the study after one year, because the teacher was replaced by another teacher who refused to continue the cooperation. The data of this class (24 students) were therefore not included in this analysis. In two other experimental classes replacement of teachers took place. Both teachers got seriously ill, one after half a year, the other after the first year. They were replaced by other teachers who continued the cooperation, but it took the schools several weeks to find the replacement. In the control condition, three teachers were immediately replaced, due to illness, a new job or because the teacher had to teach other classes. One of these replacements occurred at the end of

the second year, one after one year, the third after half a year. In order to account for these replacements and missed classes, two control variables were added to our analysis: teacher replacement (0, 1) and cancelled classes (0, 1).

There was considerable attrition among the students. From a total of 369 students at the start of the project, 44 students changed schools, of which 19 students in the treatment condition and 25 in the control condition. Seven students ( $n_{\text{treatment}} = 4$ ;  $n_{\text{control}} = 3$ ) switched classes within their school. Three students were ill for a longer period of time during the study, of which two were in the treatment condition. The frequency distribution of these categories (students staying, changing schools, switching classes, and illness) across the treatment and control condition did not differ significantly,  $\chi^2(3) = 4.78, p = .19$ . In addition, some students ( $n = 48$ ) missed one (or more) of the four test sessions and were therefore excluded from the analyses. A total of 310 students were included in the analyses, of which 165 students in the treatment condition and 145 students in the control condition.

### Treatment

Our intervention consisted of the training of five strategies that were shown to be related to reading comprehension in previous research (Dole, Duffy, Roehler, & Pearson, 1991; Palincsar & Brown, 1984; Pressley & Afflerbach, 1995; Van Silfhout, Evers-Vermeul, Mak, & Sanders, 2014):

- 1 *Predicting*. On the basis of text features such as title, subheadings, and pictures, students are instructed to make predictions about text content before reading, and to check their predictions while reading.
- 2 *Summarizing*. Students are instructed to summarize sections of text, encouraging them to focus on main ideas and ignore irrelevant details as well as to check their understanding of the text so far.
- 3 *Self-questioning*. Students are instructed to generate questions about the text being read, helping them to focus on main ideas as well as to monitor understanding.
- 4 *Clarifying*. When confronted with a word or passage they do not understand, students are instructed to reread, read ahead, or, in the case of an unknown word, analyze it, and see whether its meaning can be inferred by looking at parts of the word.
- 5 *Interpreting cohesive ties*. Students are instructed to look for relationships between sentences or paragraphs that are connected, for instance, by using ‘signal words’ (different types of connectives).

These reading strategies were taught in the context of an existing program called “Nieuwsbegrip”®, developed by the CED Group in Rotterdam (“Newswise”, CED Group, 2011). Lessons were developed weekly by a team of developers at the CED Group and were based on recent news articles (i.e., texts that had been issued the week before) about subjects related to students’ everyday life (e.g., sugar in energy drinks, abdication of the Dutch queen, or 20 years of text messaging). The use of topical, interesting texts was assumed to increase

students' task motivation (Guthrie & Wigfield, 2000; Schiefele, 1999). The teachers could download the lessons from the program website ([www.nieuwsbegrip.nl](http://www.nieuwsbegrip.nl)) and were made available every Monday evening of the week.

Lessons were provided in sequences, each consisting of six weekly lessons (approximately 45 minutes per lesson). In each of the first five lessons, the focus was on one reading strategy that was practiced in an assignment provided on a work sheet. In addition, students could work on other assignments (i.e., answering questions about the text) on the work sheet.

Each of the five strategies was trained several times during the year. This cyclical approach was assumed to result in the consolidation of strategy use. In the final lesson of each sequence all strategies were practiced simultaneously. The idea behind this was that students have to be able to apply all strategies together during the reading process, selecting the right strategy at the right moment. Table 1 provides an example of an assignment for each reading strategy.

**Table 1**

Examples of Assignments for each Reading Strategy

Strategy	Example
Predicting	This text has five subheadings. Write down for each subheading a) which thoughts it evokes and b) what you already know about the subject addressed in the subheading.
Summarizing	Read the text. Read paragraph by paragraph and underline in each paragraph the most important information. For each paragraph, write one or two sentences summarizing it. Use the words you underlined.
Self-questioning	Read the text. Note at least five questions that spring to mind while reading.
Clarifying	Search the text for difficult words. Try to uncover their meaning using these hints: a) reread the previous piece of text or read on, b) look at the illustrations in the text, c) look at the word: you might know part of the word, d) sometimes you have to use your own knowledge to figure out word meanings, or e) use a dictionary.
Interpreting cohesive ties	Read the text. Underline the signal words. Answer the questions, while noting the signal words: <ul style="list-style-type: none"> <li>• Which contrast is explained in lines 16-17? [signal word = however]</li> <li>• Why are energy boosters unfit as sports drinks? [signal word = hence]</li> </ul>

The didactic approach underlying the program “Newswise” is reciprocal teaching (Palincsar & Brown, 1984). As mentioned earlier, reciprocal teaching is a widely used method of instructing and guiding learners in reading comprehension. It consists of a set of three related instructional principles: a) direct teaching of comprehension-fostering reading strategies, including predicting, question-generating, summarizing, and clarifying; b) expert modeling, scaffolding and fading; and c) students practicing and discussing reading strategies with other students, guided and coached by the teacher. Reciprocal teaching assumes a gradual shift of responsibility for the learning process from teacher to student,

which includes the teacher explicitly modeling the use of reading strategies (Rosenhine & Meister, 1994) as well as scaffolding the application of reading strategies within the groups of students working together. It is assumed that by gradually fading teacher support, students become increasingly more capable of regulating their own reading process.

Treatment teachers took part in an extensive training and coaching program that was conducted by teacher trainers from the Rotterdam University of Applied Sciences, who had, in turn, been trained by Authors 1, 2, and 4. In the first phase (October 2011-January 2012), teachers participated in three one-hour training sessions, covering the didactic principles of the programme. In the second phase (February 2012-May 2013), teachers participated in six coaching sessions; three coaching sessions during February-June 2012 and three coaching sessions during September 2012 – May 2013. A coaching session involved a classroom observation conducted by the trainer during an intervention lesson, followed by a feedback meeting of approximately twenty minutes on the same day (See for detailed information about treatment implementation and fidelity: Okkinga, et al., submitted).

### **Control Classes**

Control classes were “business as usual”. Teachers in the control classes used the regular textbook for Dutch language teaching that was used in their school. Among the schools in the study, three different language textbooks were used. The textbooks and their teacher manuals were analyzed according to the three principles of instructional strategies in the treatment condition: instruction of reading strategies, modeling, and group work. Attention was given to reading strategies in all three textbooks. However, not all strategies that were covered in the treatment condition were also covered in the control textbooks. Reading strategies that were referred to were: predicting, clarifying, and attention to cohesive ties. Self-questioning did not occur and little attention was given to summarizing.

No attention was given to modeling by teachers or students in the teacher manuals of the control classes. Almost all of the assignments were individual and there were only a few instances where students were instructed to work together on an assignment.

### **Treatment Fidelity**

We conducted classroom observations in both the experimental and control conditions twice each year, resulting in a total of four observations for each class. Our aim was to examine a) whether the treatment teachers provided the lessons in the way we instructed during the training and coaching program and b) whether the control teachers applied the three treatment principles, even though they were not trained to do so. We used four-point scales (0-3) to assess the treatment fidelity of the main three elements of the intervention: Whole-class teaching of reading strategies, teacher and student modeling, and group work (see for more information Okkinga et al., 2016). The four observations per class were

averaged to create one score for each element of treatment fidelity (Table 2). As can be seen in Table 2, the treatment classes scored higher on all three elements.

**Table 2**

Descriptives Treatment fidelity

Variable	Treatment (n = 8)	Control (n = 10)	t-value	p-value
	Mean (SD)	Mean (SD)		
Strategy-instruction	1.81 (.80)	1.13 (.65)	2.02	.061
Modeling	.84 (.65)	.33 (.35)	2.15	.047
Group work	2.00 (1.14)	.30 (.33)	4.51	<.001

Note. Scoring between the three variables cannot be compared one-on-one. The meaning of the scoring (0-3) is different for each variable. See Classroom variables and treatment fidelity for an explanation of each variable.

Inter-rater reliability was calculated by means of observed agreement between two observers. In total, 30 from a total of 76 classroom observations were performed independently by two coders. Across these 30 observations, 93.89% observed agreement was obtained.

## Measures

### *Reading comprehension*

Reading comprehension was measured by means of the SALT-reading, a test that was validated for use among low achieving adolescents (Van Steensel, Oostdam, & Van Gelderen, 2013). The SALT-reading comprises eight tasks, each consisting of one or two texts and comprehension questions about those texts. The texts cover different genres (narrative, expository, argumentative, and instructive) and were selected from media students come across regularly in their daily lives: (school) books, newspapers, magazines, and official documents (such as regulations in a youth hostel). The eight tasks comprised a total of 59 test items, that were divided into three categories: items requiring students to retrieve relevant details from the text, items requiring students to make inferences on a local level (e.g. cause-effect relationships between sentences), and items requiring students to show their understanding of the macro structure of the text (e.g., by inferring the main idea of the text or the intention of the author). The test consisted mainly of multiple choice questions but contained also five open-ended questions. The SALT-reading was administered at four time points (See Design). The Cronbach's alpha coefficients were .82, .83, .82, and .85 respectively.

***Vocabulary knowledge***

Vocabulary knowledge was assessed with a 73-item multiple-choice test, measuring the knowledge of nouns, verbs, adjectives, and adverbs belonging to the 23,000 words in a dictionary for junior high school students (see Hazenberg & Hulstijn, 1996, for details). Each item consists of a neutral carrier sentence with a bold-faced target word and four answer options, one of which represents a correct synonym. The vocabulary knowledge test was administered two times (at pretest and at the end of the first school year). The Cronbach's alpha coefficients were .86 and .85, respectively. The average of both test was used as a measure for vocabulary knowledge.

***Metacognitive knowledge***

Metacognitive knowledge was assessed with an adapted version of the metacognitive knowledge test directed at secondary students constructed by Van Gelderen et al. (2003) and Van Gelderen et al. (2007). The adaptation was directed at the population of low-achieving adolescents, which is more comparable to our sample (Trapman et al., 2014). Items consisted of correct or incorrect statements about text characteristics and reading and writing strategies. Students had to indicate whether they agree or disagree with each statement. An example of an incorrect statement about text characteristics is 'The order in which you present the information in your text is usually not relevant'. An example of a correct statement about text characteristics is 'Sometimes you need to know more than what is in the text to understand the text properly'. The test consisted of 45 items and was administered at the end of the first school year. The Cronbach's alpha coefficient was .51. Although this indicates a rather low level of reliability (Field, 2009) we maintained the measure because in previous research it still predicted significant variance in reading comprehension (Trapman et al., 2017).

***IQ***

Intellectual ability was measured by administering the Raven Progressive Matrices at pretest. The total test consists of 60 items, divided into 5 sets of 12 items. Each item represents a logical reasoning puzzle. The items become more difficult within a set and the sets become increasingly difficult as well (Raven, Raven & Court, 1998). For students from the lowest tracks of prevocational education the last set was assumed to be too difficult and for this reason this set was omitted. The Cronbach's alpha coefficient was .82.

***Language background***

Language background was assessed with a questionnaire asking students which language they predominantly speak with their mother. There were three options: 1) the student speaks

predominantly Dutch with his or her mother, 2) the student speaks predominantly a different language than Dutch with his or her mother, and 3) the student speaks as much Dutch as another language with his or her mother. Language background was operationalized as an ordinal variable with students who mainly spoke Dutch with their mothers coded as 0 and students who spoke a different language than Dutch with their mothers coded as 2, while students who spoke as much Dutch as a different language were coded as 1.

## Analyses

### Data imputation

After the collection of all data, the dataset contained data of 310 students who were present at all four test sessions. However, some students missed a few items on some of the tests (perhaps they forgot to go back to a difficult question). To prevent loss of students in the multilevel analyses, missing data were imputed at the item level (never exceeding 7% of the cases per item), using the EM procedure from SPSS missing value analysis.

### Procedure of multilevel analyses

Repeated measures multilevel regression analyses were performed to account for the hierarchical structure of the data (using MLwiN 2.16; Rasbash, Steele, Browne, & Goldstein, 2009). The time variable 'Occasion' (variance within students across times of measurement) was defined in months; with the first measurement of reading comprehension at month zero, and subsequent measurements at months 9, 12, and 22, respectively. These months correspond to the following time points: September 2011, June 2012, September 2012, and June 2013. Dichotomous independent class and student variables were always scored 0 and 1, all continuous independent variables (IQ, age, vocabulary and metacognitive knowledge) were centered around the grand mean before adding them to the model (Hox, 2010).

First, we tested whether adding a class or school variance level to the model significantly improved model fit. Levels significantly improving model fit were added to the model. Second, we tested whether a model with random slopes both at the student or class level improved model fit, as is recommended by Hox (2010). Results for these two steps are presented in Appendix A. Third, we added the class level variables 'teacher replacement' and 'cancelled classes' to check whether we should correct for these variables (see Appendix B). Fourth, we tested whether the student-level control variables gender, IQ, age, language spoken at home with mother, and metacognitive knowledge significantly improved model fit. In addition, vocabulary knowledge was included as predictor, because this was necessary for the later testing of its moderating effect (see Table 4). Fifth, we tested whether the treatment had a significant effect on growth in reading comprehension by testing

the interaction between treatment and occasion. Finally, we checked whether vocabulary moderated the treatment effect (see Table 5) by estimating the three way interaction between occasion, treatment and vocabulary knowledge.

### Testing of significance

The number of levels needed in the analyses was tested by comparing nested models with Chi-square significance tests (Hox, 2010). Significance of control variables and predictors were tested both with Wald-tests (coefficient divided by the standard error) and by means of comparing nested models (with and without the control variables or predictors) with a Chi-square test<sup>5</sup>. Regression coefficients for class level variables were tested with number of classes as sample size ( $df = \text{number of classes} - \text{number of predictors} - 1$ ) (Hox, 2010). Effect sizes were estimated by comparing the variances at different levels as well as the total variances of nested models.

## Results

### Descriptive Statistics

Table 3 shows the mean student scores for all student level variables (the four repeated measurements of reading comprehension and test scores for IQ, vocabulary knowledge, and metacognitive knowledge). The only student level variable showing a significant difference between treatment and control group (t-tests) is metacognitive knowledge.

**Table 3**

*Descriptives student-level variables*

Variable	Treatment ( $n = 165$ )	Control ( $n = 145$ )	t-value
	Mean (SD)	Mean (SD)	
Reading comprehension (time 1)	34.96 (6.98)	34.59 (8.50)	.41
Reading comprehension (time 2)	37.49 (6.92)	36.47 (8.81)	1.12
Reading comprehension (time 3)	36.45 (7.31)	36.94 (8.33)	.51
Reading comprehension (time 4)	37.35 (8.51)	39.24 (8.60)	1.75
Vocabulary knowledge	49.30 (7.42)	49.37 (7.69)	.08
IQ	35.96 (5.24)	35.09 (5.27)	1.45
Metacognitive knowledge	26.92 (4.36)	25.59 (4.46)	2.66*

\* $p = .008$

<sup>5</sup> The difference in  $-2 \times \text{Loglikelihood}$  (or Deviance) of nested models has a Chi-square distribution with a number of degrees of freedom equal to the difference in number of estimated parameters between both models.

## Multilevel Analyses

As no significant random intercept variance was found at the school level (see Appendix A), models with three levels were used (occasion-, student-, and class level). Appendix A also shows that random slopes for the occasion variable both at the class and student level significantly improved model fit.

Next, the teacher-level control variables were entered in the model (see Appendix B). Neither 'teacher replacement' nor 'cancelled classes' significantly improved model fit and both were therefore omitted from further analyses. Subsequently, the student-level control variables were entered to control for differences between students at pretest. Inclusion of age, gender, and language background did not improve model fit (see Table 4<sup>6</sup>), whereas vocabulary knowledge, metacognitive knowledge, and IQ did. Model 7 (see Table 4) is therefore the model referred to as model 0 in Table 5.

---

<sup>6</sup> In Table 4 the N of cases is determined by the sample N (310) times 4 (repeated measures) minus the missings on one or more of the repeated measures for reading comprehension (totalling 1129).

**Table 4**

*Multilevel Analyses with Reading Comprehension (repeatedly measured) as Dependent Variable to Establish Influence of Student-Level Variables (Language spoken at home with mother, IQ, Gender, Vocabulary Knowledge, Metacognitive Knowledge, and Age) (N=1129 cases/310 students)*

<b>Model:</b>	<b>model 0</b>	<b>model 1</b>	<b>model 2</b>
<b>Fixed part</b>			
Intercept	<b>34.684(.964)</b>	<b>34.761(.968)</b>	<b>33.846(1.039)</b>
Occasion (in months)	<b>.156(.036)</b>	<b>.156(.036)</b>	<b>1.56(.036)</b>
Language spoken at home with mother (contrast only Dutch vs Dutch half of the time)		.610(1.332)	
Language spoken at home with mother (contrast only Dutch vs mostly different language)		-1.390(1.303)	
Gender (male=0, female=1)			<b>1.517(.707)</b>
IQ			
Age			
Vocabulary knowledge			
Metacognitive knowledge			
<b>Random part (variances)</b>			
Class	15.851(5.879)	15.422(5.740)	15.809(5.859)
Class slope variance occasion	<b>.019(.008)</b>	<b>.019(.008)</b>	<b>.019(.008)</b>
Class covariance slope*intercept	-.260(.169)	-.260(.168)	-.254(.169)
Student	<b>26.826(3.437)</b>	<b>26.725(3.429)</b>	<b>26.496(3.411)</b>
Student slope variance occasion	<b>.018(.009)</b>	<b>.018(.009)</b>	<b>.018(.009)</b>
Student covariance slope*intercept	.191(.133)	.191(.133)	.172(.133)
Occasion (rep. Measures)	<b>17.560(1.065)</b>	<b>17.559(1.065)</b>	<b>17.562(1.065)</b>
<b>Variance explained</b>			
Class			.003
Class slope variance occasion			-
Student			.012
Student slope variance occasion			-
Occasion (rep. measures)			-
<b>Deviance testing</b>			
-2*Loglikelihood (deviance)	7179.080	7177.604	7174.540
Difference between -2*loglikelihood			<b>4.54</b>
Difference df			1
Compared to model		0	0

<sup>a</sup> Model 6 is more parsimonious than model 5

**Bold =  $p < .05$ ; italicized =  $p < .01$ ; bold and italicized =  $p < .001$**

model 3	model 4	model 5	model 6	model 7
<b>34.007(.978)</b>	<b>33.989(.983)</b>	<b>34.648(.738)</b>	<b>34.956(.038)</b>	<b>35.000(.659)</b>
<b>.156(.036)</b>	<b>.156(.036)</b>	<b>.155(.036)</b>	<b>.155(.036)</b>	<b>.156(.036)</b>
<b>1.319(.674)</b>	<b>1.348(.689)</b>	.552(.544)		
<b>.360(.064)</b>	<b>.359(.064)</b>	<b>.240(.052)</b>	<b>.242(.052)</b>	<b>.222(.051)</b>
	.128(.646)			
		<b>.523(.038)</b>	<b>.527(.038)</b>	<b>.473(.039)</b>
				<b>.287(.066)</b>
13.835(5.165)	13.865(5.174)	7.270(2.845)	7.243(2.837)	6.985(2.731)
<b>.019(.008)</b>	<b>.019(.008)</b>	<b>.019(.008)</b>	<b>.019(.008)</b>	<b>.019(.008)</b>
-.217(.157)	-.217(.157)	-.258(.128)	-.261(.129)	-.276(.129)
<b>23.301(3.155)</b>	<b>23.288(3.154)</b>	<b>12.673(2.309)</b>	<b>12.678(2.309)</b>	<b>11.374(2.206)</b>
<b>.017(.009)</b>	<b>.017(.009)</b>	<b>.018(.009)</b>	<b>.018(.009)</b>	<b>.018(.009)</b>
.161(.129)	.162(.129)	.112(.114)	.118(.114)	.134(.112)
<b>17.580(1.066)</b>	<b>17.581(1.066)</b>	<b>17.528(1.062)</b>	<b>17.527(1.062)</b>	<b>17.508(1.061)</b>
.125		.476		.036
-		-		-
.121		.456		.103
.056		-		-
-		.003		.001
7144.213	7144.174	6999.880	7000.898	6982.692
<b>30.327</b>	.039	<b>144.294</b>	<b>(-1.018)</b>	<b>18.206</b>
1	1	1	1	1
2	3	4	5a	6

To establish the effect of the treatment on growth in reading comprehension, in Table 5, we compared Model 1 (with effect on the pretest only) with Model 2. containing the interaction between occasion and treatment. This effect was not significant implying there was no effect of the treatment on growth in reading comprehension ( $\Delta IGLS = 1.827$ ,  $df = 1$ ,  $p > .05$ ).

Models 3 and 4 were conditional for testing the moderating effect of vocabulary in Model 5. The moderator effect of vocabulary was tested in Model 5 as the interaction between occasion, treatment and vocabulary (see Table 5). This interaction was significant ( $\Delta IGLS = 4.595$ ,  $df = 1$ ,  $p < .05$ ). Explained student-level variance on growth in reading comprehension equals 11.1%.



**Table 5**

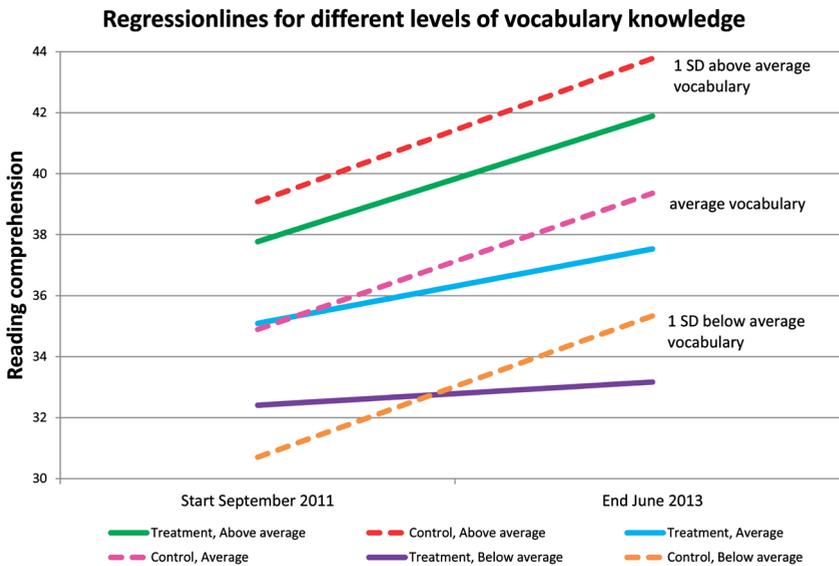
Multilevel Analyses with Reading Comprehension (repeatedly measured) as Dependent Variable to Establish Influence of Treatment over Time and Interaction between Vocabulary knowledge, Occasion and Treatment, after Correcting for Control Variables ( $N=1129$  cases/310 students)

Model	model 0	model 1	model 2
<b>Fixed part</b>			
Intercept	<b><i>35.000(.659)</i></b>	<b><i>35.509(.830)</i></b>	<b><i>34.872(.937)</i></b>
Occasion (in months)	<b><i>.156(.036)</i></b>	<b><i>.156(.036)</i></b>	<b><i>.204(.049)</i></b>
IQa	<b><i>.222(.051)</i></b>	<b><i>.223(.051)</i></b>	<b><i>.223(.051)</i></b>
Vocabulary knowledg <sub>ea</sub>	<b><i>.473(.039)</i></b>	<b><i>.473(.039)</i></b>	<b><i>.473(.039)</i></b>
Metacognitive knowledg <sub>ea</sub>	<b><i>.287(.066)</i></b>	<b><i>.293(.066)</i></b>	<b><i>.292(.066)</i></b>
Treatment (treatment = 1, control = 0)		-.998(.950)	.270(1.318)
Occasion * Treatment			-.096(.069)
Occasion * Vocabulary knowledg <sub>ea</sub>			
Treatment * Vocabulary knowledg <sub>ea</sub>			
Occasion * Treatment * Vocabulary knowledg <sub>ea</sub>			
<b>Random part (variances)</b>			
Class	<b><i>6.985(2.731)</i></b>	<b><i>7.422(2.866)</i></b>	<b><i>6.973(2.728)</i></b>
Class slope variance occasion	<b><i>.019(.008)</i></b>	<b><i>.019(.008)</i></b>	<b><i>.017(.008)</i></b>
Class covariance slope*intercept	-.276(.129)	-.301(.134)	-.267(.123)
Student	<b><i>11.374(2.206)</i></b>	<b><i>11.364(2.205)</i></b>	<b><i>11.371(2.206)</i></b>
Student slope variance occasion	<b><i>.018(.009)</i></b>	<b><i>.018(.009)</i></b>	<b><i>.018(.009)</i></b>
Student covariance slope*intercept	.134(.112)	.134(.112)	.134(.112)
Occasion (rep. measures)	<b><i>17.508(1.061)</i></b>	<b><i>17.508(1.061)</i></b>	<b><i>17.514(1.061)</i></b>
<b>Variance explained</b>			
Class			
Class slope variance occasion			
Student			
Student slope variance occasion			
Occasion (rep. measures)			
<b>Deviance testing</b>			
-2*loglikelihood (deviance)	6982.692	6981.709	6979.882
Difference between -2*loglikelihood	18.206	.983	1.827
Difference df	1	1	1
Compared to model		0	1

**Bold =  $p < .05$ ; italicized =  $p < .01$ ; bold and italicized =  $p < .001$**

model 3	model 4	model 5
<b>34.859(.949)</b>	<b>34.873(.947)</b>	<b>34.891(.949)</b>
<b>.205(.049)</b>	<b>.205(.049)</b>	<b>.203(.050)</b>
<b>.221(.051)</b>	<b>.223(.051)</b>	<b>.223(.051)</b>
<b>.445(.043)</b>	<b>.507(.057)</b>	<b>.547(.060)</b>
<b>.294(.066)</b>	<b>.289(.066)</b>	<b>.288(.066)</b>
.272(1.335)	.236(1.332)	.198(1.335)
-.096(.069)	-.095(.069)	-.092(.071)
.004(.003)	.004(.003)	-.001(.004)
	-.118(.072)	-.197(.081)
		<b>.011(.005)</b>
<b>7.205(2.800)</b>	<b>7.184(2.785)</b>	<b>7.228(2.797)</b>
<b>.017(.008)</b>	<b>.017(.008)</b>	<b>.018(.008)</b>
-.285(.127)	-.291(.127)	-.299(.130)
<b>11.309(2.199)</b>	<b>10.935(2.171)</b>	<b>10.901(2.166)</b>
<b>.018(.009)</b>	<b>.018(.009)</b>	<b>.016(.009)</b>
.140(.111)	.159(.110)	.167(.109)
<b>17.486(1.060)</b>	<b>17.491(1.060)</b>	<b>17.463(1.058)</b>
		-
		-
		.003
		.111
		.002
6977.604	6975.013	6970.418
1.178	2.591	<b>4.595</b>
1	1	1
2	3	4

The interpretation of the moderating effect of vocabulary knowledge on growth in reading comprehension becomes clear when looking at regressions for different combinations of scores on the independent variables (Hox, 2010). For treatment, we used two scores (0 and 1), for occasion we used the scores 0 and 3 for the first and the last time of measurement and for vocabulary we used three scores: one standard deviation below the mean, the mean, and one standard deviation above the mean. The resulting six regression lines are presented in Figure 1. The regression lines show that in cases where students score above average on vocabulary, there appears to be no difference between growth in reading comprehension for the control and treatment students. However, when students score below average on vocabulary knowledge, growth in reading comprehension appears to be different for control and treatment students; the control students outperform the treatment students over time. When looking at the regression lines for the treatment group only, it seems that the students scoring lower on vocabulary profit less than students scoring higher on vocabulary. In contrast, in the control group students high and low on vocabulary grow in equal measures on reading comprehension.



**Figure 1**

Regression lines for both treatment and control, split into 1 SD above average vocabulary, average vocabulary, and 1 SD above average vocabulary.

## Discussion

This study set out to analyze the moderating effect of vocabulary knowledge on low-achieving adolescents' growth in reading comprehension in a two-year intervention study. The intervention consisted of a program directed at the application of reading strategies. The control group followed lessons as usual at their schools. Our aim was to analyze whether students in the control and experimental conditions with different levels of vocabulary knowledge had different patterns of growth. Our study revealed a significant moderating effect of vocabulary knowledge. This means that vocabulary knowledge had a differential impact on students' growth in reading comprehension in the control and experimental groups. A closer look at the growth patterns in both groups showed that a lower vocabulary level resulted in less growth in reading comprehension in the experimental condition compared to the control condition, while there appears to be no difference in growth in reading comprehension between the two conditions for students with a higher vocabulary level. In addition, visualization of the growth patterns in both groups (Figure 1) suggests that the experimental condition had a discriminating effect for students with different levels of vocabulary, whereas the business-as-usual condition did not.

It is remarkable that students in the experimental condition were at a disadvantage compared to control students if their vocabulary knowledge was relatively low. An explanation for this result can be found in the different treatment of vocabulary in experimental and control classes. In the experimental condition students were stimulated to use strategies for clarification of unknown words in their texts. In contrast, in the control condition, it was frequently observed that teachers explained the meaning of difficult words to students prior to reading the texts. This can explain why students in the experimental condition met with more difficulties in compensating for their low vocabulary than the control students low in vocabulary. It seems plausible that, especially in deriving meaning from context, students with a low vocabulary are at disadvantage because they might not know the meaning of enough words necessary for using this clarification strategy. This is an indication that vocabulary knowledge is an important prerequisite for the application of reading strategies for low-achieving students as was hypothesized in this study.

Additionally, low vocabulary knowledge may lead to cognitive overload (Sweller, 1988) in the case of application of newly learned reading strategies that are not already part of the students' repertoire. Studies into reading comprehension components have suggested that limited working memory capacity may play a role in students' application of reading strategies (Van Gelderen et al., 2003; 2007). This may certainly be the case for low-achieving students with relatively low vocabulary knowledge, since they will have to deal with more meaning-related problems for building a situation model (Kintsch, 1988), while at the same time they have to apply new reading strategies, which was the case in the experimental condition. Students in the experimental condition had to direct their attention to their texts

in multiple ways in order to determine which strategy should be applied and how this strategy should be applied, while simultaneously keeping an eye on the meaning of multiple words which are needed for successful application of each strategy. These processes may interfere with each other, resulting in a sort of short circuit (Bossers, 1991; Taillefer, 1996). This is probably a greater risk for low vocabulary students in the experimental condition than in the control condition. In the control condition, this short circuit may not happen as easily, as these students apply reading strategies routinely (without much conscious attention).

From many studies it is known that vocabulary knowledge is a basic component of reading comprehension (Ouellette & Beers, 2010; Tilstra et al., 2009; Trapman et al., 2014; Van Gelderen et al., 2004, 2007; Van Steensel et al., 2016; Verhoeven & Van Leeuwe, 2008). The present study adds to this knowledge by suggesting that, in the case of low-achieving adolescents, vocabulary knowledge is not only involved in reading comprehension processes, but also in their application of reading strategies. Our reading strategy intervention, which was intended for facilitating low-achieving students' reading comprehension, may therefore have missed its target for those students that needed it the most, namely those with the lowest vocabulary. Low-achieving students with a higher level of vocabulary were not harmed (in the sense of showing similar progress compared to the control condition), but in comparison to the control condition the students with low vocabulary showed less progress in reading comprehension. This may also explain some of the mixed results of reading strategy interventions for low-achieving students mentioned in the introduction (Edmonds et al., 2009; Fogarty et al., 2014; Okkinga et al., submitted; Simmons et al., 2014; Vaughn et al., 2013).

### **Suggestions for Future Research**

Future studies into reading strategies interventions with low-achieving students may profit from the following suggestions. As vocabulary knowledge is an important factor in fostering reading comprehension but may also be important in the application of reading strategies for low-achieving students, future studies could investigate the optimal ways to include vocabulary instruction in the context of reading strategy interventions for this group. This means that low-achieving students should not only receive instruction in applying specific reading strategies, but also be supported in their use of vocabulary knowledge needed for successful application of reading strategies. The original approach used by Palincsar and Brown (1984) to reciprocal teaching can be used to this end. In this format tutors support struggling readers in small group settings by steering group conversations directed at the application of reading strategies. The approach aims at producing a natural dialogue between the group members and the tutor, while the tutor supports all students' attempts at text comprehension. In such a small group tutoring approach, a specific vocabulary support condition allows tutors to converse intensely with the students, focusing specifically

on their problems with understanding meanings of words encountered in the text. Studying the effects of such additional vocabulary support could shed light on the question whether low-achieving students profit more from reading strategy instruction when their deficits in vocabulary are simultaneously being repaired.

Apart from vocabulary knowledge, other student characteristics can be considered as moderator variables. For example, the level of self-regulatory skills may influence how students respond to an intervention targeted at the use of reading strategies. Self-regulatory skills refer to planning, executing, and controlling behavior while performing a task (Boekaerts & Simons, 1993). In the case of reading comprehension, self-regulatory skills are important, for example, in knowing which reading strategy to apply or monitoring one's comprehension during reading. There is evidence that self-regulatory skills applied in reading tasks may determine low-achieving students' reading comprehension (De Milliano, Van Gelderen & Slegers, 2016). Thus, the application of reading strategies (in the context of an intervention) may also be dependent on the level of self-regulatory skills of students.

In general, the fact that moderator variables such as vocabulary knowledge and self-regulatory skills may play an important role in the success of reading strategy interventions makes it critical to understand how the intervention is affected by students' skills, strengthening the need for investigating new moderator effects. This may especially be the case in whole-classroom settings, as instruction in whole-classroom settings is likely to be targeted at the average level of students, whereas instruction in small groups, as originally proposed by Palincsar and Brown (1984), is likely to be more aligned to the individual needs of the students in the small group.

### **Implications for Educational Practice**

It is difficult for teachers to implement reading strategy training with reciprocal teaching for low-achieving adolescents in whole-classroom settings successfully (Okkinga et al., submitted). Our study confirms this difficulty, because it shows no overall advantage of the intervention in growth of reading comprehension. An important explanation given for the lack of success of the whole-classroom approach is that teachers cannot attend to multiple groups of students simultaneously and give the students the individual guidance they need. The present study deepens this explanation, as it shows that individual differences in vocabulary knowledge impact on the effects of the experimental intervention for low-achieving adolescents. This suggests that other approaches are needed to facilitate reading comprehension for this group of students with low vocabulary knowledge. In addition, the results of the control group show that it is possible to make students with low vocabulary knowledge profit similarly to their classmates with higher vocabulary knowledge.

It is hard for teachers to be aware of the problems that individual students with low vocabulary knowledge may encounter. Therefore, it is advisable for teachers that are

involved in teaching low-achieving adolescents to identify students that are better served with additional vocabulary support. For this group, the focus on application of reading strategies can lead to frustration.

In the original Palincsar and Brown set-up (1984) with small groups of students, the tutor can support the group process from beginning to end and adapt instruction and guidance to the individual students within that group. In such small-group settings, it is much easier for a tutor to tell whether students need more help in clarifying the meaning of words than in a whole-classroom setting. Although it is not possible for a single teacher to support several small groups in the classroom sufficiently to monitor their dialogues related to reading, it is possible to support at least one such group that may need the support the most, while other students are working independently. This type of classroom organization may allow teachers to give the weakest readers the attention in vocabulary support they need.



## Appendix A

**Table A**

Multilevel Analyses with Reading Comprehension (repeatedly measured) as Dependent Variable to Establish Multilevel Structure of Data ( $N=1129$  cases/310 students)

Model	model 0	model 1	model 2
<b>Fixed part</b>			
Intercept	<b>35.048 (.435)</b>	<b>34.690 (.886)</b>	<b>34.683 (.974)</b>
Occasion (in months)	<b>.151 (.017)</b>	<b>.151 (.017)</b>	<b>.151 (.017)</b>
<b>Random part (variances)</b>			
School			3.353 (4.846)
Class		<b>12.512 (4.728)</b>	9.078 (5.185)
Class slope variance occasion			
Class covariance slope*intercept			
Student	<b>42.958 (3.935)</b>	<b>31.489 (3.118)</b>	<b>31.499 (3.118)</b>
Student slope variance occasion			
Student covariance slope*intercept			
Occasion (rep. measures)	<b>20.475 (1.012)</b>	<b>20.471 (1.011)</b>	<b>20.472 (1.011)</b>
<b>Deviance testing</b>			
-2*loglikelihood (deviance)	7274.617	7227.612	7227.097
Difference between -2*loglikelihood		<b>47.005</b>	.515
Difference df		1	1
Compared to model		0	1

\* A priori fixed coefficient

<sup>a</sup> Although adding the covariance of slope\*intercept is not significantly different from 0 (model 5 to model 6), since the slope variance is significant we have to include the covariance of the slope\*intercept in subsequent models. Therefore, model 6 is the “zero model” in Table B.

**Bold =  $p < .05$ ; italicized =  $p < .01$ ; bold and italicized =  $p < .001$**

model 3	model 4	model 5	model 6 <sup>a</sup>
<b>34.689 (.894)</b>	<b>34.689 (.894)</b>	<b>34.678 (.937)</b>	<b>34.684 (.964)</b>
<b>.152 (.020)</b>	<b>.152 (.020)</b>	<b>.155 (.035)</b>	<b>.156 (.036)</b>
13.142 (4.929)	13.185 (4.947)	14.793 (5.508)	15.851 (5.879)
		<b>.018 (.008)</b>	<b>.019 (.008)</b>
		0*	-.260 (.169)
<b>28.421 (3.034)</b>	<b>28.065 (3.535)</b>	<b>27.010 (3.452)</b>	<b>26.826 (3.437)</b>
<b>.036 (.009)</b>	<b>.035 (.010)</b>	<b>.018 (.009)</b>	<b>.018 (.009)</b>
0*	.027 (.146)	.175 (.134)	.191 (.133)
<b>17.525 (1.007)</b>	<b>17.581 (1.068)</b>	<b>17.557 (1.065)</b>	<b>17.560 (1.065)</b>
7202.546	7202.514	7182.110	7179.080
<b>25.066</b>	.032	<b>20.404</b>	3.030
1	1	1	1
1	3	4	5

## Appendix B

**Table B**

Multilevel Analyses with Reading Comprehension (repeatedly measured) as Dependent Variable to Verify Influence of Teacher Replacement (0=no, 1=yes) and Cancelled Classes (0=less than 6 weeks, 1=6 weeks or more) ( $N=1129/310$ )

Model	model 0	model 1	model 2
<b>Fixed part</b>			
Intercept	<b>34.684(.964)</b>	<b>35.444(1.055)</b>	<b>35.006</b>
Occasion (in months)	<b>.156(.036)</b>	<b>.157(.036)</b>	<b>.156(.036)</b>
Teacher replacement (yes = 1, no = 0)		<b>-3.023(1.889)</b>	
Cancelled classes (> 6 weeks)			-2.117(2.398)
Treatment (treatment = 1, control = 0)			
Occasion*Treatment			
<b>Random part (variances)</b>			
Class	<b>15.851(5.879)</b>	<b>15.094(5.638)</b>	<b>15.852(5.883)</b>
Class slope variance occasion	<b>.019(.008)</b>	<b>.019(.008)</b>	<b>.019(.008)</b>
Class covariance slope*intercept	-.260(.169)	-.297(.170)	-.282(.172)
Student	<b>26.826(3.437)</b>	<b>26.827(3.438)</b>	<b>26.819(3.347)</b>
Student slope variance occasion	<b>.018(.009)</b>	<b>.018(.009)</b>	<b>.018(.009)</b>
Student covariance slope*intercept	.191(.133)	.191(.133)	.191(.133)
Occasion (rep. measures)	<b>17.560(1.065)</b>	<b>17.565(1.065)</b>	<b>17.565(1.065)</b>
<b>Deviance testing</b>			
-2*loglikelihood (deviance)	7179.080	7176.735	7178.338
Difference between -2*loglikelihood		2.345	.742
Difference df		1	1
Compared to model		0	0

\* a priori fixed coefficient, \*\* total variance summed (without covariance slope\*intercept)

Note: no explained variances since all effects are non-significant

**Bold =  $p < .05$ ; italicized =  $p < .01$ ; bold and italicized =  $p < .001$**

model 3	model 4
<b>34.967(1.307)</b>	<b>34.376(1.366)</b>
<b>.156(.036)</b>	<b>.204(.048)</b>
-561(1.732)	.617(1.924)
	-.096(.068)
<b>16.14(5.972)</b>	<b>15.756(5.848)</b>
<b>.019(.008)</b>	<b>.016(.007)</b>
-.274(.172)	-.244(.16)
<b>26.82(3.437)</b>	<b>26.818(3.437)</b>
<b>.018(.009)</b>	<b>.018(.009)</b>
.191(.133)	.191(.133)
<b>17.561(1.065)</b>	<b>17.567(1.066)</b>
7178.985	7177.106
.095	1.879
1	1
0	3





# Effectiveness of reading-strategy interventions in classrooms: A meta-analysis



This chapter is submitted as: Okkinga, M., Van Steensel, R., Van Gelderen, A.J.S., Van Schooten, E., Slegers, P.J.C., & Arends, L.R. (submitted). Effectiveness of reading-strategy interventions in classrooms: A meta-analysis.



## Abstract

Research has demonstrated that in controlled experiments in which small groups are being tutored by researchers, reading strategy instruction is highly effective in fostering reading comprehension (Palincsar & Brown, 1984). It is unclear, however, whether reading-strategy interventions are equally effective in whole-classroom situations in which the teacher is the sole instructor for the whole class. This meta-analysis focuses on the effects of reading strategy interventions in whole-classroom settings.

Results of studies on the effectiveness of reading strategy interventions in whole-classroom settings were summarized ( $N_{\text{studies}} = 52, K = 125$ ) to determine the overall effect on reading comprehension and strategic ability. In addition, moderator effects of intervention, study, and student characteristics were explored. The analysis demonstrated a very small effect on reading comprehension (Cohen's  $d = .186$ ) for standardized tests and a medium effect (Cohen's  $d = .431$ ) on researcher-developed reading comprehension tests. A medium overall effect was found for strategic ability (Cohen's  $d = .786$ ). Interventions in which 'setting reading goals' was part of the reading-strategy package, effects tended to be larger. In addition, effects were larger for interventions in which the trainer was the researcher as opposed to teachers and effect sizes tended to be larger for studies conducted in grades 6-8. Implications of these findings for future research and educational practice are discussed.

*Keywords:* meta-analysis, reading strategy, intervention, reading comprehension

## Introduction

Many students struggle with reading comprehension (e.g. Organisation for Economic Co-operation and Development [OECD], 2014). Since reading comprehension is a fundamental skill in all school subjects, problems with this skill have serious implications for students' educational success and, consequently, for their later societal careers. From the literature it is known that students who are struggling readers have problems reading strategically (Paris, Lipson, & Wixson, 1983). Good readers monitor their understanding of the text, while making use of different reading strategies such as predicting, activating prior knowledge, summarizing during reading, question generating and clarifying (e.g. Palincsar & Brown, 1984). Therefore, interventions aimed at fostering reading comprehension in low achievers are often based (or focused) on this type of reading strategies (Pressley & Afflerbach, 1995).

Many studies have demonstrated positive effects of reading strategy interventions on reading comprehension and previous meta-analyses established that the effects of these interventions are quite large (e.g. Rosenshine & Meister, 1994; Sencibaugh, 2007; Swanson, 1999). However, many studies have been conducted in controlled settings in which experimenters are instructors (as opposed to regular teachers) and in which instruction is given to small groups of students (as opposed to regular classrooms in which multiple groups of students work simultaneously). Therefore, it is unclear whether reading strategies interventions are as effective in whole-classroom settings as it is in more controlled settings (Droop, Van Elsäcker, Voeten & Verhoeven, 2016). This is an important lack in the current research base, considering that reading comprehension strategies have found their way into curriculum materials in the last decades.

This meta-analysis is carried out to provide more insight into the effects of reading strategy interventions on reading comprehension in whole classroom settings. In addition, it explores moderator effects of intervention, study-design, and student characteristics.

## Teaching Reading Strategies and Didactic Principles

Since the 1980's, and after Durkin's study (1978) demonstrating that comprehension instruction was virtually non-existent in elementary classrooms, research into reading comprehension instruction by means of the use of reading strategies, increased rapidly (Duke & Pearson, 2002). The underlying idea is that reading comprehension is a complex process in which the reader interacts with the text to construct a mental representation of the text, or a situation model (Kintsch, 1988; 1998). Hence, if readers understand how they can use comprehension skills as they read, their comprehension will be stimulated.

A reading strategy is a mental tool a reader uses on purpose to monitor, repair, or bolster comprehension (Afflerbach & Cho, 2009). The use of reading strategies is a deliberate and goal-directed attempt to construct meaning of text (Afflerbach, Pearson, & Paris, 2008), and as such, can refer to both metacognitive and cognitive strategies that aid the

process of reading (Dole, Nokes, & Drits, 2009). Researchers have suggested many different strategies (Pressley & Afflerbach, 1995). They may involve an awareness of reading goals, the activation of relevant background knowledge, the allocation of attention to major content while ignoring irrelevant details, the evaluation of the validity of text content, comprehension monitoring, visualizing, summarizing, self-questioning and making and testing interpretations, predictions, and drawing conclusions (Duke, Pearson, Strachan, & Billman, 2011; Palincsar & Brown, 1984).

There is a variety of approaches directed at instructing reading strategies to foster reading comprehension relevant to our study. For example, one of the approaches is reciprocal teaching (Palincsar & Brown, 1984; Palincsar, Brown, & Martin, 1987). Reciprocal teaching consists of a set of three principles: a) teaching comprehension-fostering reading strategies, b) expert modeling, scaffolding and fading; and c) students taking turns in practicing reading strategies and discussing with other students. Another approach is called Collaborative Strategic Reading in which students in groups have more differentiated roles (leader, clunk expert, gist pro) than in reciprocal teaching and there is more attention to whole-class instruction (Vaughn et al. 2013). In addition, there is the approach called Concept Oriented Reading in which more emphasis is placed on motivational engagement support provided by the teacher. For example, by providing choice in reading materials to increase students' intrinsic motivation, emphasizing importance of reading, competence support and stimulating collaboration (Guthrie & Klauda, 2014). Other approaches emphasize self-regulatory strategies for example Mason, Davison, Hammer et al. (2013) and Jitendra, Hoppes, and Xin (2000), while Durukan (2011) integrates reading strategy instruction with writing strategies. Although these approaches differ, they have also important similarities. The most important similarities are the use of whole-classroom strategy instruction, the modeling of strategies and students working in small groups.

Interventions using reading strategies according to the above-mentioned approaches appear to be not always successful in improving reading comprehension (De Corte, Verschaffel, & Van de Ven, 2001; Edmonds, et al., 2009; Fogarty et al., 2014; McKeown, Beck, & Blake, 2009; Simmons et al., 2014; Vaughn, 2013). The complexity of the didactic principles of the combination of strategy instruction, modeling and guided group work that is used in these interventions may explain why it can be difficult to achieve improvements in reading comprehension in whole-classroom situations. In such situations it is hard to maintain implementation quality of the intervention, given that one instructor (teacher or researcher) has to supervise several groups of students simultaneously, as opposed to small-group tutoring. This explanation is supported by a few qualitative studies that show that teachers in whole-classroom settings face problems in the implementation of interventions using the principles of strategy instruction, modeling and group work (Duffy, 1993; Seymour & Osana, 2003; Hacker & Tenent, 2002). Teachers found it hard to induce strategic thinking in

students (Duffy, 1993). In addition, students showed poor application of reading strategies and poor discourse skills while collaborating (Hacker & Tenent, 2002), in which case students become too distracted to form coherent representations of text content (McKeown, Beck, & Blake, 2009). As a consequence, the teachers were hindered in changing from a teacher-centered to a student-centered approach.

### **Effects of Reading-Strategy Interventions: Findings from Previous Reviews and Meta-Analyses**

The report of the National Reading Panel (2000) identified 16 types of interventions directed at reading comprehension, of which six were regarded to be effective. Five of these types can be defined as reading strategy interventions: 1) comprehension monitoring, 2) graphic and semantic organizers, 3) generating questions, 4) summarizing and 5) multiple strategy instruction.

In the past decades, several systematic meta-analyses on the effects of several types of interventions for fostering reading comprehension (including reading strategies) have been conducted (e.g. Berkeley, Scruggs, & Mastropieri, 2010; Edmonds, et al. 2009; Slavin, et al., 2009; Scammacca, Roberts, Vaughn, & Stuebing, 2015; Swanson, 1999). Most of those meta-analyses are directed at a specific group of students, for example students with learning disabilities (Berkeley, Scruggs, & Mastropieri, 2010; Swanson, 1999), adolescent struggling readers (Edmonds, et al. 2009; Slavin, et al., 2008) or elementary students (Slavin et al. 2009) and include a wide variety of interventions aimed at fostering reading comprehension. In these meta-analyses, interventions focusing specifically on reading strategies yield mixed results. Effect sizes for reading strategy interventions range from large to very small: Edmonds (2009), for instance, established an average effect of  $d = 1.23$ , whereas Slavin et al. (2009) found an overall effect size of  $d = 0.21$ .

Meta-analyses that focus specifically on the effects of reading strategies interventions are from Rosenshine and Meister (1994), Chiu (1998) and more recently from Sencibaugh (2007). In the review by Rosenshine and Meister (1994), sixteen experimental studies of reciprocal teaching, conducted between 1984 and 1992, were summarized. The authors found an overall positive effect on reading comprehension; with a median Cohen's effect size value ( $d = .32$ ) for standardized tests and a large Cohen's effect size value ( $d = .88$ ) for researcher-developed tests.

Chiu (1998) synthesized studies that incorporate metacognitive interventions to foster reading comprehension, which involved reading strategies such as self-questioning, summarizing or inferencing. He analyzed 43 studies, ranging from second grade to college, which were conducted between 1978 and 1995. The overall effect size was larger for researcher-developed tests ( $d = .61$ ) than for standardized tests ( $d = .24$ ) and effect sizes were larger when researchers delivered instruction compared to teachers.

Sencibaugh (2007) focused on students with learning disabilities and analyzed 15 studies, conducted between 1985 and 2005, testing the effect of reading strategy interventions on reading comprehension. He distinguished interventions with a focus on 'auditory-language dependent' strategies (such as summarizing, self-questioning, inferencing) and 'visually-dependent' strategies (such as semantic organizers or visual attention therapy). The overall effect of the former was 1.18 and on the latter 0.94. However, no separate effect sizes were reported for researcher-developed tests ( $n = 10$ ) and standardized tests ( $n = 5$ ).

Taken together, all of them have made relevant contributions to the growing insight into effective reading-strategy interventions. However, the following important question remains. It is unclear how the above findings from meta-analyses relate to the context of whole-classroom instruction. In the overall effect sizes presented in previous meta-analyses, it does not become clear whether the strategy interventions are successful in such whole-classroom contexts. We believe that this is a significant omission in the research literature, because the teaching of reading strategies has become a standard part of the reading curriculum in primary and secondary education. Hence, a focused meta-analysis on effectiveness of reading strategy instruction in whole-classroom settings is needed to shed light on whether such teaching is fruitful in regular educational practice.

Additionally, previous meta-analyses give rise to two questions. First is the finding that regular teachers are less successful in delivering strategy instruction than researchers. This can be concluded from the moderation analyses reported in Chiu, (1998) and Scammacca et al., (2015). A possible explanation for this finding is provided by Seymour and Osana (2003) stating on the basis of teacher interviews that teachers found the strategies to be taught hard to understand (for example the distinction between strategies such as questioning and clarifying). In addition, Duffy (1993) reported that teachers were not familiar with the definitions of strategies to be taught.

Second is that in previous meta-analyses, effect sizes were significantly larger for researcher-developed than for standardized tests (Chiu, 1998; De Boer, Donker, & Van der Werf, 2014; Rosenshine & Meister, 1994; Swanson, 1999). This issue is relevant for educational practice, especially if the analysis focuses on whole classrooms. In that context it is important to decide whether the teaching of reading strategies should be directed at performance on standardized reading tests or that this teaching is intended for more specific reading objectives (that can be achieved by the use of reading strategies).

### **The Present Study**

The goal of the study is to estimate the effects of reading strategy interventions in whole-classroom settings on students' reading comprehension and strategic abilities. In addition, we explore the moderating effects of intervention-, student-, and study design characteristics.

In terms of intervention characteristics, we focus on type of reading strategies instructed, type of trainer (regular teacher vs experimenter) and type of didactic principle used (modeling, group work and scaffolding). As explained above these variables may influence the effect-size of strategy interventions. In addition, moderating effects of type of educational context (language or content area classes) are analyzed, because these contexts are quite different in nature for the application of reading strategies (general reading comprehension vs knowledge acquisition) (Guthrie & Davis, 2003; De Milliano, 2013). Finally, moderation effects of duration of the intervention are included, to find confirmation for the findings of Scammacca et al. (2015) that shorter interventions generally result in larger effect sizes.

As reading strategy interventions focused on different student populations, we also analyzed moderation effects of grade (3-12) and different types of readers (typical or low-achieving), as research showed that strategy interventions are most effective for low-achieving readers (Edmonds et al., 2009).

Finally, we took into account several study-design characteristics that are important in determining the validity of the studies (Cooper, Hedges, & Valentine, 2009). We focused on design of the experiment, as it is documented that studies in which a quasi-experimental design is used show larger effect sizes compared to studies with a randomized design (Lipsey, 2003). In addition, the type of control condition (business-as-usual vs controlled control group) was used as a moderator. Controlled control groups refer to conditions in which the control students were given a different intervention (for example a vocabulary intervention) or one of the components of the reading strategy intervention. It is plausible that differences between experimental conditions and controlled control groups are smaller than differences between experimental conditions and business-as-usual control groups.

The following research questions are addressed:

- 1 What are the effects of reading strategy interventions in whole classroom settings on students' reading comprehension measured by standardized and researcher-developed tests?
- 2 Which intervention-, student-, and study design characteristics moderate the effects of reading strategy interventions?

## Method

### Inclusion Criteria

We chose the year 2000 as a starting point for our literature search, because from this year onwards the instruction of reading strategies started to become more and more mainstream in education (Pressley, 2002). Criteria for inclusion of studies in this meta-analysis were as follows:

1. The participants were in grades 3-12.
2. The study measured the effects of reading strategy-interventions on students' reading comprehension skills in regular classroom settings.
3. The dependent measure(s) included quantitative measures of reading comprehension.
4. The study compared an experimental group, participating in the intervention, to a control group that did not.
5. The article was written in English, but the study could have taken place in any country.
6. Pretest data were available.
7. The information provided should be sufficient for calculating effect sizes.

The following exclusion criteria were utilized:

1. The study's treatment focuses on strategies that pertain to decoding and morphemic strategies.
2. The treatment takes place in foreign language classes (e.g. Chinese students learning English).
3. The treatment is implemented outside the classroom (e.g. one-on-one tutoring or remedial teaching).
4. The treatment takes place during a summer school.
5. The treatment is lab-based (e.g. experiments in which students are given individual instruction with a computer).
6. The treatment is a curriculum-wide program in which it does not become clear which specific reading strategies are taught and how this was done (e.g. Success-for-All)
7. The student population of the study consists mainly of students with a developmental disorder (e.g. autism or ADHD), students are deaf or hearing impaired, or students suffer from aphasia.
8. The study design is a single-subject or single-case research design.

### Literature Search

Two databases were accessed; ERIC and PsycINFO. Search queries consisted of synonyms of 'reading comprehension', 'intervention', 'strategy-instruction', and 'children'. See Appendix A for the search syntax. Articles had to be peer-reviewed and written in English. Based on the

criteria 31 articles were included. The initial search was carried out in April 2012. An update on 11 May 2015 resulted in 16 additional articles. Snowballing resulted in 5 more articles. An overview of the database search and selection is presented in Appendix B.

### **Coding Procedures**

Based on Cooper, Hedges, and Valentine (2009), we devised a coding scheme containing both theoretical and statistical elements. This scheme was piloted and refined until the first two authors reached agreement on all topics. At the start of the second round of literature search, two coders joined the team. They had completed training on how to use the coding scheme and had reached a high level of reliability.

Interrater reliability was measured at the level of the decision to include articles to be coded. From the first batch of the literature search (until 2012), twenty articles were randomly selected. The two new coders (who were oblivious to the articles thus far included in the meta-analysis) independently checked the twenty articles and, based on the aforementioned criteria, decided which of those should be included in the meta-analysis and which should not. Overall interrater reliability between all coders was calculated as percentage agreement, which reached 86%.

Regular meetings between the coders were held to discuss particular issues or concerns and to collaboratively decide in cases of doubt how to interpret aforementioned criteria when coding and other coding problems, such as the definition of types of reading strategies.

The coding scheme included the following five elements: intervention characteristics, student characteristics, study design characteristics, and measurement characteristics and statistics.

### **Intervention characteristics**

Intervention characteristics pertain to who implemented the treatment (researcher/teacher/ 'other'), the educational context in which the study took place (language classes/content area classes/ 'other'), whether the treatment included scaffolding, modeling and/or group work to teach reading strategies, and which reading strategies were taught.

To guide the coding of reading strategies, we took the seminal work of Pressley and Afflerbach (1995) as a starting point in identifying and sorting the many reading strategies that are described and reported in experimental studies. These authors analysed 40 think-aloud studies and reported strategies that were executed by good readers as they go through a text.

**Table 1**

List of coded reading strategies

<b>Reading strategy</b>	<b>Description</b>
<i>Before reading</i>	
Predicting	Make predictions about text content before reading, based on text features such as title, subheadings, and pictures
Activating prior knowledge	Doing a mental search of what the reader already knows about the text
Setting reading goals	Defining what the reader wants to achieve by reading the text
<i>During reading</i>	
Questioning	Ask questions to oneself about important aspects of the text to monitor understanding
Paraphrasing	Restating the meaning of a small passage
Summarizing (during)	Providing a short account of a main idea in a paragraph, for the purpose of checking understanding of the text so far
Inferencing	Relating information in the text to prior knowledge
Underlining important information	Deciding what is important and highlight/underline this information.
Use of graphic organizers / visual representation	Making illustrations that depict relationships among the key concepts in a text
Using text structure	Identify the global structure of a text, and using signal words for local structure
Using mental imagery	Forming mental images of the text to promote deeper understanding
Explicit monitoring strategies	Focusing on specific ways of improving text comprehension: clarifying word meanings, setting boundaries for monitoring comprehension, error detection and fix-up strategies (e.g. rereading)
<i>After reading</i>	
Summarizing (after)	Stating the main ideas of the text to check on understanding
Memorizing	Recall the main ideas for later use

Reading strategies included goal directed activities that occur before a text is read as it is read, and after the reading of the text is completed. For example, setting reading goals is a reading strategy utilized before one starts reading, while inferencing is used during reading. Summarizing has been observed both during and after reading, and therefore is listed both under 'during reading' and 'after reading'. In our analysis we included all strategies that were explicitly directed at the comprehension of text on the word-, sentence- or whole-text level. Strategies that were directed at focusing students' attention to specific ways of improving their text comprehension (e.g. clarifying word meanings, setting boundaries for monitoring comprehension or error detection), were taken together in a strategy called

'explicit monitoring strategies'. Examples of such comprehension directed strategies are: clarifying the meaning of a word, error detection and fix-up strategies such as rereading. Strategies only used for decoding words, such as spelling, phonemic analysis and phonics were excluded.

Table 1 lists the strategies that we coded with the accompanying studies. For each reading strategy, we coded whether the reading strategy was taught (1), or not (0).

### **Student characteristics**

With regard to student characteristics, we wanted to be able to differentiate among student populations. Therefore, we coded the grade(s) in which the study took place and type of student (typical students/learning disabled/low-achieving or struggling readers).

### **Study design characteristics**

As for study design characteristics, we coded the design of the study (randomized, quasi-experimental, matched, and 'other'), and whether a delayed posttest was administered.

### **Outcome measures**

With respect to the outcome measures, we coded reading comprehension measured with standardized and researcher developed tests. Furthermore, as many studies also reported on measures for strategic ability (i.e. the quality of application of reading strategies), strategy knowledge and/or self-reported strategy use, we also included these in our analysis. We distinguished immediate posttests and delayed posttests in our analysis. Thus, we ran analyses for five outcome measures (i.e. reading comprehension standardized, reading comprehension researcher-developed, strategic ability, strategy knowledge, self-report strategy-use).

In cases in which multiple outcome measures (for example, three measures of strategic ability) were reported for one of our outcome measures, we decided to include one of those based on the following decision tree:

1. If both subtests and a total score of the instrument were analyzed and reported, we used the total score in our analysis (this was often the case for reading comprehension).
2. If different instruments were reported and it was not possible to include a total score, we chose an instrument that measured summarizing or main idea identification (this was often the case for measures of strategic ability).
3. If 1) and 2) were not available, we chose the first measure that was described in the study.

To determine the effect size (Cohen's *d*), we coded group size (*n* experimental and *n* control), whether and which covariates were used in the specific statistical comparison by

the authors, pretest- and posttest values (means and standard deviations) and the type of statistic the effect size is based on (t-value, F value, regression weight) with accompanying degrees of freedom and p-values.

### **Method of Analysis**

In a number of studies multiple experiments were described, or multiple samples were researched, resulting in more than one experimental comparison within one study. For that reason, the unit of analysis was “experimental comparison”. As described above, we distinguished five outcome measures, and also distinguished between immediate and delayed posttests. Thus, for one experimental comparison it was possible to have 10 outcome measures, which were analyzed separately.

To calculate the average effect sizes (Cohen's *d*) (Cohen, 1988) for our outcome measures, a random effects model was used. To take into account differences in sample size between the comparisons, the effect sizes are weighted based on the variances within the samples and the between-study variation.

To analyze whether the variance in effect sizes can be attributed to differences in intervention-, student-, and study design characteristics, moderation analyses were carried out using mixed effects models for categorical moderators (for example, type of design or reading strategy yes/no). The *Q*-statistic was calculated to analyze between-group differences for the categorical moderators (Lipsey & Wilson, 2001).

Publication bias was tested by applying Duval and Tweedie's trim-and-fill method (Duval & Tweedie, 2000). A random effects model was used to estimate if there were any interventions missing in the meta-analysis.

All analyses were performed by a statistician (Author 5), with a registered version of the Comprehensive Meta-Analysis software package (version 3; Biostat, Englewood, NJ).

## **Results**

### **Descriptives**

A total of 52 articles met our eligibility criteria. Within those 52 articles, 125 eligible effects were found. Of those 125 effects, 89 effects measured reading comprehension. Sixty-one effects concerned reading comprehension standardized as dependent variable, of which 9 were delayed posttests. For reading comprehension researcher developed tests, 28 effects were found, of which 4 were delayed posttests. For strategic ability 22 effects were found, of which 5 were delayed posttests. For strategy knowledge a total of 8 effects were found, of which 3 were delayed posttests. Finally, for self-reported strategy-use 6 (immediate) effects were found.

Mean duration of the interventions was 47.11 hours, with a standard deviation of 55.01. The range of duration in hours was 6-233. The interventions were spread over 17.47 weeks on average with a standard deviation of 11.37. In Appendix C an overview of the key characteristics is given of all experimental comparisons in the meta-analysis.

### **Main effects**

First, main overall effect sizes for our outcome measures were analyzed. In Table 2, an overview of the effect sizes per outcome measure is displayed. All overall effects were positive, but not all were significantly different from zero. For reading comprehension, the effect sizes for both immediate standardized measures (Cohen's  $d = 0.186$ ) and delayed measures (Cohen's  $d = 0.167$ ) can be considered trivial (Cohen, 1988), whereas the effect size for researcher developed measures was significant but small for immediate tests (Cohen's  $d = 0.431$ ), and large for delayed researcher-developed tests (Cohen's  $d = .947$ ). The difference between the effect sizes for standardized and researcher developed tests of reading comprehension is also significant,  $Q(1) = 10.599$ ,  $p = .001$ ; with effects on researcher-developed tests being larger.

In terms of reading strategy related outcome measures, the immediate measures of strategic ability and self-reported strategy use were significantly different from zero with respectively medium and small effect sizes (Cohen's  $d = 0.786$  and Cohen's  $d = 0.358$ ), while strategy knowledge was not significantly different from zero. For the delayed reading strategy measures, effect sizes were small and trivial, and, probably due to the small number of comparisons, not significantly different from zero.

Heterogeneity analyses (Table 2) and forest plots (Appendix D) show large and significant variation in effect sizes across the studies, which justifies moderation analyses for our three dependent variables of interest (immediate measures of reading comprehension standardized, researcher-developed tests, and strategic ability).

**Table 2**

Results of meta-analyses of overall effect sizes

Measure	Effect size analysis				Heterogeneity analysis		
	K	Cohen's <i>d</i> (SE)	LB; UB	<i>p</i>	<i>Q</i> ( <i>df</i> ), <i>p</i>	<i>I</i> <sup>2</sup>	<i>τ</i> <sup>2</sup> (SE)
Immediate posttest							
Reading comprehension standardized	52	<b>0.186</b> (0.027)	0.132; 0.240	<.001	<i>Q</i> (51) = 108.12, <i>p</i> <.001	52.83	.015 (.007)
Reading comprehension researcher developed	24	<b>0.431</b> (0.070)	0.294; 0.569	<.001	<i>Q</i> (23) = 155.20, <i>p</i> <.001	85.18	.077 (.046)
Strategic ability	17	<b>0.786</b> (0.147)	0.498; 1.074	<.001	<i>Q</i> (16) = 326.48, <i>p</i> <.001	95.10	.317 (.237)
Strategy knowledge	5	0.366 (0.070)	0.228; 0.503	.104	<i>Q</i> (4) = 3.45, <i>p</i> <.486	0.00	
Strategy use self-report	6	<i>0.358</i> (0.139)	0.085; 0.630	.010	<i>Q</i> (5)=20.47, <i>p</i> = .001	75.57	.084 (.074)
Delayed posttest							
Reading comprehension standardized	9	<i>0.167</i> (0.058)	0.053; 0.281	.004			
Reading comprehension researcher developed	4	<b>0.947</b> (0.149)	0.654; 1.240	<.001			
Strategic ability	5	0.382 (0.088)	-0.051; 0.550	.104			
Strategy knowledge	3	0.117 (0.091)	-0.061; 0.295	.199			
Strategy use self-report	0	-					

Note. *k* = number of samples; **bold** *p* <.001; *italicized* *p* <.01; LB; UB = Lower bound; Upper bound 95% confidence intervals.

### Moderator effects

We chose to analyze only moderator effects on 'immediate' outcome measures, as only a few studies included delayed measures. For the same reason, we did not analyze moderator effects for strategy knowledge (*n* = 5) and strategy-use self-report (*n* = 6). All significant moderators of the three outcome measures (reading comprehension standardized, reading comprehension researcher-developed, and strategic ability) are presented in Table 3.

### Intervention characteristics

As a first step, moderation analyses were performed on the three overarching categories of reading strategies ('before reading', 'during reading' and 'after reading'). None of those

moderation analyses were significant, apart from 'before reading' for strategic ability (see Table 3), which was negative.

In the next step, the separate reading strategies were examined. Only 'setting reading goals' appeared to have a significant contribution to the overall effect size on all three outcome variables (Table 3). Interestingly, this contribution was positive for both reading comprehension outcome measures, but negative for strategic ability. In other words, when 'setting reading goals' was part of the intervention, the overall effect size of strategic ability was lower for the intervention than for the control. Furthermore, 'underline important information' was a significant contributor to strategic ability.

In addition, a number of other reading strategies showed negative contributions to different outcome measures. For reading comprehension standardized, this was the case for 'mental imagery' and 'memorizing'. For researcher developed tests, this was the case for 'comprehension directed strategies' and the use of 'graphic organizers / visual representation'. Lastly, for strategic ability, 'predicting' and 'prior knowledge' were negative contributors to the intervention.

As for other intervention characteristics, 'modeling' (applying reading strategies while thinking aloud by the teacher) significant contributor for measures of strategic ability, but not for the reading comprehension measures. The educational context in which the intervention took place mattered only in the case of strategic ability. A higher overall effect size was obtained in language arts classes compared to content area classes. The type of trainer of the intervention only mattered in the case of researcher developed tests with the effect size for researchers as trainers being larger than the effect size of teachers as trainers.

We did not find significant contributions of the following intervention characteristics: scaffolding, group work, questioning, summarizing (during reading), inferencing, text structure, paraphrasing, hinge words, summarizing (after reading), and duration in weeks, number of sessions, total hours of the intervention (time per session\*number of sessions) and intensity of the intervention (total hours / number of weeks).

**Table 3**

Significant moderators for reading comprehension standardized, reading comprehension researcher-developed, and strategic ability.

<b>Moderator</b>	<b>Dependent variable</b>	<b>k / k (/ / k etc)</b>
<i>Intervention characteristics</i>		
Before reading (yes / no) *	SA	9 / 8
Predicting (yes / no)*	SA	9 / 8
Prior knowledge (yes / no)*	SA	1 / 16
Setting reading goals (yes / no)*	SA	1 / 16
Setting reading goals (yes / no)	RC-St	3 / 49
Setting reading goals (yes / no)	RC-RD	1 / 23
Comprehension directed strategies (yes / no)*	RC-RD	10 / 14
Graphic organizers/ visual representation (yes / no)*	RC-RD	8 / 16
Underline important information (yes / no)	SA	2 / 15
Mental imagery (yes / no)*	RC-St	6 / 46
Memorizing (yes / no)*	RC-St	6 / 46
Modeling (yes / no)	SA	15 / 2
Trainer (teacher / researcher)	RC-RD	19 / 5
Context (language arts / content area)	SA	16 / 1
<i>Student characteristics</i>		
Reader type (low achievers / typical development)	SA	3 / 12
Grade (1 = grades 3-5, 2 = grades 6-8, 3 = grades 9-12)	RC-RD	10 / 12 / 2
<i>Study design characteristics</i>		
Control type (business as usual / controlled control group)	RC-RD	18 / 6

*Note.* RC-St = reading comprehension standardized test, RC-RD = reading comprehension researcher-developed test, SA = strategic ability; k = number of samples, \* = negative effect of moderator

Cohen's <i>d</i> ( <i>SE</i> ) category 1 / Cohen's <i>d</i> ( <i>SE</i> )category 2 / etc.	<i>Q</i> , <i>p</i>
0.430 (0.078) / 1.169 (0.238)	$Q(1) = 8.698, p = .003$
0.430 (0.078) / 1.169 (0.238)	$Q(1) = 8.698, p = .003$
0.293 (0.149) / 0.820 (0.153)	$Q(1) = 6.081, p = .014$
0.339 (0.158) / 0.816 (0.153)	$Q(1) = 4.701, p = .030$
0.493 (0.111) / 0.171 (0.027)	$Q(1) = 7.868, p = .005$
1.300 (0.312) / 0.404 (0.069)	$Q(1) = 7.837, p = .005$
0.271 (0.101) / 0.568 (0.098)	$Q(1) = 4.437, p = .035$
0.262 (0.094) / 0.525 (0.094)	$Q(1) = 3.911, p = .048$
1.759 (0.524) / 0.670 (0.153)	$Q(1) = 3.984, p = .046$
0.048 (0.056) / 0.209 (0.030)	$Q(1) = 6.350, p = .012$
0.112 (0.027) / 0.219 (0.033)	$Q(1) = 6.165, p = .013$
0.851 (0.168) / 0.431 (0.126)	$Q(1) = 3.976, p = .046$
0.337 (0.069) / 0.962 (0.281)	$Q(1) = 4.663, p = .031$
0.820 (0.153) / 0.293 (0.149)	$Q(1) = 6.081, p = .014$
1.115 (0.189) / 0.612 (0.170)	$Q(1) = 3.922, p = .048$
0.387 (0.065) / 0.618 (0.154) / 0.091 (0.237)	$Q(2) = 6.323, p = .042$
0.552 (0.080) / 0.022 (0.101)	$Q(1) = 17.135, p < .001$

### Student characteristics

When looking at student characteristics (Table 3), reader type mattered only for strategic ability measures, with a higher overall effect size for low-achievers compared to typically developing students. Grade was a significant contributor for researcher developed tests, with the largest overall effect size for students in grades 6-8.

### Study-design characteristics

When looking at study design characteristics as moderators (Table 3), the type of control condition mattered only for researcher-developed tests of reading comprehension. A higher overall effect size was observed when the control condition was a business-as-usual compared to controlled control groups. Design (randomized experiment vs quasi-experiment) did not influence effect sizes on the three outcome measures.

### Publication bias

We also tested for publication bias by applying the trim-and-fill method (Duval & Tweedie, 2000). The funnel plot for reading comprehension standardized (immediate) showed evidence of asymmetry (see Appendix E). The addition of the 'missing' studies imputed using the Duval and Tweedie's trim-and-fill method (Duval & Tweedie, 2000) shifted the effect size from Cohen's  $d = 0.186$  to Cohen's  $d = 0.115$  (0.093 ; 0.208), but still significant. Egger's test (Egger, Smith, Schneider, & Minder, 1997) confirmed the presence of publication bias,  $t(50) = 2.087, p = .042$ . The opposite was found for researcher-developed tests (immediate), with evidence of asymmetry on the right side of the funnel plot (see Appendix E). With the Duval and Tweedie's trim-and-fill method (Duval & Tweedie, 2000) the effect size shifted from Cohen's  $d = 0.431$  to Cohen's  $d = 0.522$  (0.382 ; 0.662), but Egger's regression intercept (Egger, Smith, Schneider, & Minder, 1997) did not confirm this,  $t(22) = 1.489, p = .151$ , suggesting a weak indication for publication bias. No indications of publication bias were found for strategic ability (immediate) with a symmetric funnel plot (See Appendix E) and an Egger's regression intercept of  $t(15) = 0.507, p = .620$ .

### Discussion

This study set out to summarize the overall effects of interventions aimed at instructing reading strategies on both reading comprehension and strategic ability in whole-classroom settings. In addition, it was determined whether intervention-, student-, and study design characteristics influenced the effects on reading comprehension and strategic ability. To establish the overall effects of reading strategies interventions, a search of literature published from 2000 onwards yielded a total of 51 studies, which comprised 220 effect sizes.

We found a significant, but very small, effect on reading comprehension (Cohen's  $d = .186$ ) for standardized tests and a small significant effect on researcher-developed reading comprehension tests (Cohen's  $d = .431$ ). Effects were significantly larger when researcher-developed tests were used compared to standardized tests. A significant medium overall effect was found for strategic ability tests (Cohen's  $d = .786$ ) and a significant, but small effect for self-reported strategy use (Cohen's  $d = .358$ ). For delayed tests, we found a significant, but very small effect on standardized reading comprehension tests (Cohen's  $d = 0.167$ ) and a large effect for researcher-developed reading comprehension tests (Cohen's  $d = 0.947$ ). No significant effects were found for delayed tests of strategic ability and knowledge.

Moderation effects were analyzed for intervention characteristics (type of strategies instructed, type of trainer, didactic principles, educational context and duration). From all types of strategies discriminated in this study, only 'setting reading goals' was found to positively moderate the effects on reading comprehension tests (both standardized and researcher-developed). For some strategies even a negative effect was found. For standardized tests of reading comprehension, this was the case for 'mental imagery' and 'memorizing'. For researcher developed tests, this was the case for 'explicit monitoring strategies' and the use of 'graphic organizers / visual representation'. Lastly, for strategic ability, 'predicting', 'setting reading goals' and 'prior knowledge' were negative contributors to the effects of interventions. Intervention characteristics which positively influenced effect sizes were 'modeling' (strategic ability), the educational context of language arts classes (strategic ability), and researchers as type of trainers (researcher-developed tests).

For student and study design characteristics, larger effect sizes were obtained for low-achievers compared to typically developing students (strategic ability), students in grades 6-8 (researcher developed tests), and when control classes were business-as-usual (researcher-developed tests). It should be noted that all moderation analyses are exploratory in nature and are not based on explicit manipulation of experimental variables.

### **Main Effects of Reading-Strategy Interventions**

Our findings concur with the findings of the meta-analyses of Rosenhine and Meister (1994) and Chiu (1998). They also found quite small effects (respectively  $d = .32$  and  $d = .24$ ) of reading strategy interventions on standardized tests. The finding that effects on standardized tests are hard to accomplish with reading strategy interventions has already been recorded in several instances (Paris, Cross & Lipson, 1984). The transfer of instructed strategies to standardized reading tests is probably difficult to accomplish, because it may require quite different strategies than were provided in the instruction, such as how to handle multiple choice questions and how to interpret typical reading comprehension questions that are posed in standardized tests with limited time for thinking and using strategies. In sum, results of our study call into question whether the skills needed for achieving higher scores on

standardized measures for reading comprehension can be improved by teaching students how to apply a limited set of reading strategies. On top of that, it shows that there is a systematic difference in the skills required for standardized tests and researcher-developed tests, justifying that both measures are treated separately.

On the other hand, we found a stronger effect size for researcher-developed reading comprehension tests, suggesting that such tests are better suited to capture the learning effects of strategy instruction on reading comprehension than standardized tests. Assuming that researcher developed tests for reading comprehension measure significant aspects of students' comprehension of texts, this shows that reading strategies taught in whole classrooms may be a valuable addition to students' reading development. In addition, our finding that for delayed tests there was a large effect on researcher-developed measures of reading comprehension ( $d = .947$ ) gives substantial support to the usefulness of instruction in reading strategies. This is an important finding because it shows that the reading strategies taught in interventions are quite durable and students keep on using them effectively also long after the intervention has stopped.

Furthermore, we found a quite large effect of reading strategy interventions on strategic ability. Tests for strategic ability require applying the learned strategies, which are qualitatively scored. This finding is of relevance because it shows that application of strategies is improved by strategy training. Furthermore, we found a small effect on students' self-reports of strategy use, indicating that students' awareness of the type of strategies that are taught can be increased by the reading strategy interventions. However, surprisingly, we did not find significant effect sizes for strategy knowledge. This can be due to low reliability scores of the tests used, as was reported for half of the experimental comparisons (Cronbach's  $\alpha$  range = .33 - .60; Souvignier & Mohkleserami, 2006).

### **Moderation of Intervention Characteristics**

Our analysis of moderation effects of types of strategies taught showed that only a few reading strategies did affect the effect sizes, while some of these strategies had even negative effects. In most of the studies included in the meta-analysis several strategies are taught in combination, only in a few cases did interventions focus on one reading strategy only (DiCecco, 2002; Jitendra, 2000; Lubliner & Smetana, 2005; Miller et al., 2011; Ng, Bartlett, Chester, & Kersland, 2013; Redford, Thiede, Wiley, & Griffin, 2012). This means that the overall effect sizes for reading comprehension and strategic ability have to be attributed to the teaching of different combinations of reading strategies instead of one of the types of strategies discriminated. As a consequence, the results of our moderation analysis for types of reading strategies should be interpreted in terms of whether it matters if one or other type of strategy is part of the package offered. In most cases, our findings show that this is not the case: it does not seem to matter if, for example, inferencing is part of the package of

reading strategies taught. In some cases (see Table 3), inclusion of some types of strategies in the package (e.g. explicit monitoring strategies, or the use of graphic organizers) results even in negative effects on reading comprehension in comparison to studies that do not contain these types. That, however, does not mean that such types used in isolation would result in negative effects. It only means that packages of strategies without them result in higher effect sizes.

An exception to the above is the positive effects for one of the 15 types of trained strategies on the measures for reading comprehension. This strategy, 'setting reading goals', requires students to critically reflect on their reading goals before reading. The finding of this type of strategy to be effective is based on only 3 experimental comparisons for standardized tests and one experimental comparison for researcher developed tests. Therefore, the finding must be treated with caution. However, it is interesting that this type of strategy sticks out as an important of reading strategy instruction, because goal setting may be one of the main determinants of a successful task approach, especially when it is important to select relevant information from texts (Rouet & Britt, 2011). We also found that 'setting reading goals' had a negative effect on measures of strategic ability. However, there was only one experimental comparison for the combination of 'setting reading goals' and strategic ability, thus the result cannot be generalized.

For strategic ability, we also found that 'underline important information' and the didactic principle of 'modeling' were positive contributors. Modeling refers to trainers (and eventually students) thinking aloud while reading and thereby exposing their cognitive process of comprehending texts and making clear how to use reading strategies when reading. It appears that this is a useful practice in comparison to approaches that do not contain modeling, because it increases effects on strategic ability. 'Underline important information' is a reading strategy that was instructed in four studies (Guthrie et al., 2004; Ponce, López, & Mayer, 2012; Souvignier & Mokhlesgerami, 2006; Sung, Chang, & Huang, 2008). In these studies, strategic ability was measured by underlining important parts of a text, which is closely aligned to the strategy 'Underlining important information'. This may explain the effect found on strategic ability.

As expected, effect sizes for reading comprehension (researcher-developed) were smaller for interventions in which instruction was given by the teacher compared to interventions in which researchers provided instruction, an outcome which is in line with earlier findings (Chiu, 1998; Scammacca, 2015). Thus, researchers seem to be better able to deliver the interventions in whole-classroom settings than teachers. This supports the results of qualitative studies in which teachers were followed implementing reading strategy instruction in their classrooms (Duffy, 1993; Seymour & Osana, 2003; Hacker & Tenet, 2002). They found that teachers in whole-classroom settings face problems in the implementation of interventions directed at instructing reading strategies. For example,

teachers found the didactic principles of reciprocal teaching and the specific reading strategies that had to be taught hard to understand (Seymour & Osana, 2003), teachers found it hard to induce strategic thinking in students (Duffy, 1993), and students showed poor application of reading strategies and poor discourse skills while collaborating when teachers implemented reciprocal teaching in their classrooms (Hacker & Tenet, 2002).

An interesting finding is the fact that duration of the intervention did not influence the overall effect size. Thus, it did not matter how long interventions lasted for the effect size. It should be noted that there were very few long interventions (> 1 year), which may have limited this effect to become significant.

### **Moderation Effects of Study-Design and Student Characteristics**

As expected, studies in which the control group comprised a business-as-usual control group, effect sizes tended to be larger compared to controlled control groups for researchers developed reading comprehension tests. Controlled control groups refer to conditions in which the control students were given a different intervention (for example a vocabulary intervention) or one of the components of the reading strategy intervention. This finding seems logical given that the difference between a controlled-control group and the intervention group is most likely smaller than between business-as-usual control groups and intervention groups.

In terms of student characteristics, it seems that students in grades 6-8 profited the most from reading strategy interventions to improve reading comprehension, measured by researcher developed tests. A possible explanation is that middle grades students are increasingly required to learn from texts in school and therefore learning reading strategies is of more direct use to them. In addition, low-achievers' strategic ability improved more from strategy instruction than was the case for typically developing students. This finding suggests that low-achievers profit more from strategy instruction in strategic ability tasks than typical students. On the other hand, we found that both types of readers profit equally from reading strategy instruction on measures of reading comprehension.

### **Limitations**

We chose to only include published articles as a way to ensure research quality, but this means that we did not take into account unpublished research reports. Generally, effects estimated in published work tend to be higher than in unpublished reports, which could affect the results of the meta-analysis. For that reason, we used the trim-and-fill method (Duval & Tweedie, 2000) to estimate whether our meta-analysis was subject to publication bias. For reading comprehension standardized tests the effect size decreased significantly (from  $d = .186$  to  $d = .115$ ), while for researcher developed tests there was a significant increase

(from  $d = .431$  to  $d = .522$ ). No indications of publication bias were found for strategic ability. Thus, there was a small publication bias for our reading comprehension outcome measures, but this does not change our main conclusions for these measures.

A second major limitation was that we could not include implementation quality as moderator. Implementation quality refers to the degree in which interventions are carried out as intended, and especially in whole-classroom contexts in which teachers are often the trainers as opposed to researcher, this moderator is of importance. Many of the studies that we analyzed above do not give enough information about the quality of implementation to include this variable in our analysis, such as the way that teachers and trainers worked according to protocol, the way that students responded to instruction and the quality of training and coaching of teachers. It is quite interesting for future research to find out whether interventions with high implementation quality succeed in improving students' reading comprehension skills more than interventions with lower implementation quality, as it is known that especially when studies are performed outside strong controlled settings, implementation quality is of importance in finding effects (Hulleman & Cordray, 2009).

### **Suggestions for Future Research**

There were a few studies that were potentially interesting, but we could not include those because of inadequate or missing data to calculate the effect size. We therefore call for more rigorous descriptions of statistical data in future research. Both journal editors and researchers should take care in registration of statistical results and should take into account that studies might be used in a meta-analysis in the future. For example, presenting pretest data is very helpful in determining a proper effect size. Also presenting student characteristics (e.g. gender or age) of each condition, as opposed to characteristics that apply to the whole student sample, is helpful to accurately synthesize data. For intervention studies in particular, it is interesting to be able to include moderators pertaining to intervention duration (e.g. intervention duration in terms of total hours and number of weeks) and more elaborate descriptions of what the intervention specifically entailed are helpful in determining whether a study is eligible for coding (e.g. training procedures for the trainers who delivered the intervention, what strategies were taught, what kind of tasks were the students required to do in the lessons, what didactic principles were underlying the intervention).

A second suggestion for future research pertains to the mediating effect of strategy-use by the students. The only feature that was experimentally manipulated in the studies included in the meta-analysis was the presence or absence of an intervention aimed at instructing reading strategies. The underlying assumption in the studies is that the strategies taught are also used by the students, but we cannot be sure that was always the case (Donker et al., 2014). It would be interesting to investigate in future research how strategy-use by students mediates the level of reading comprehension.

### Implications for Educational Practice

Our meta-analysis gives rise to some important implications for the use of interventions aimed at fostering reading strategies in whole-classroom contexts. In the first place, the question whether reading strategy interventions can be effective for improving students' reading comprehension can be answered in the affirmative. Although the overall effect size of strategy interventions on researcher-developed tests appears to be small, the effect on the delayed measures is stronger (large) and shows that strategy instruction can be considered to be quite durable in achieving effects on reading comprehension. The researcher-developed tests used in our analyzed interventions are presumably directed at reading comprehension skills that are needed for specific types of comprehension problems, which may arise in textbooks used in school (e.g. in content area teaching). For that reason, the results of our meta-analysis give support to the teaching of reading strategies for the purpose of improving such textbook reading in school (in contrast to improving achievement on standardized reading tests).

As the studies used a package of multiple strategy instruction to foster students' reading comprehension, our findings do support the recommendations provided by the National Reading Panel (2000). The panel concludes that teaching of multiple reading strategies "leads to increased learning of the strategies, to specific transfer of learning, to increased memory and understanding of new passages, and, in some cases, to general improvements in comprehension" (p. 4-52). In addition, our study suggests that the strategy called 'setting reading goals' is a promising one for adding to such packages of multiple strategies, since it has shown to have a positive moderating effect on reading comprehension. This type of strategy requires that students think beforehand what the purpose of reading is, for example, knowing something, answering questions or enjoying a story (e.g. Aaron et al., 2008). This means that teachers should make their students think and be aware of *their* purpose for reading and ask questions that elicit such awareness. Setting reading goals is an important type of strategy in task-oriented reading, which is often asked for in content-area classes. In task-oriented reading, students need to select relevant parts of the text, specific for the task at hand, they need to know when to skip information, and they need to know when to search for information (Vidal-Abarca, Mañá, & Gil, 2010). To do this successfully, knowing the goal(s) of reading is necessary. Thus, especially for content-area teachers, asking their students to think about their reading goals might be a valuable addition to learning from texts.

Our study has made it clear that reading strategy instruction seems to be especially effective in the middle (grades 6-8), and earlier grades (3-5), as substantial effects on researcher-developed measures of reading comprehension are found ( $d = .618$  and  $d = .387$ ). This implies that instructing reading strategies in whole-classroom settings should begin from the earliest grades to be effective. In contrast, for the group of older students

(grades 9-12) the effects on reading comprehension is negligible. Although, the results found for the latter were derived from only two experimental studies, it remains doubtful whether strategy instruction for this oldest group in our analysis is fruitful.

Lastly, we need reflect on the result that regular teachers are not as proficient as researchers in successfully implementing an intervention aimed at instructing reading strategies. This finding calls for more emphasis in teacher education on the teaching of reading strategies to foster reading comprehension. As the National Reading Panel (2000) recommends, such instruction should be extensive, especially on *how* teachers should teach those strategies.

## Conclusion

Summarizing our findings, we conclude that reading-strategy interventions in whole classroom settings can be beneficial, especially for students in grades 3-8. Larger effects were found for researcher-developed tests of reading comprehension, compared to standardized tests, which shows that researcher-developed tests are more sensitive to the specific reading strategies learned by students. In addition, the finding that delayed (researcher developed) tests demonstrated a large effect on reading comprehension is certainly an important support for the case that strategy interventions may be quite durable in achieving effects on reading comprehension in whole-classroom settings.

## Appendix A

### Search syntax for ERIC and psycINFO

#### #1 Reading Comprehension

reading comprehension/ OR reading skills/ OR reading/ OR reading ability/ OR reading achievement/ OR reading comprehension/ OR reading development/ OR reading education/ OR reading improvement/ OR reading skills/ OR reading.ti,ab,id.

#### #2 Intervention

intervention/ OR school based intervention/ OR intervention.ti,ab,id. OR educational programs/ OR literacy programs/ OR educational program evaluation/ OR program effectiveness/ OR program evaluation/ OR reading programs/ OR improvement programs/ OR program\*.ti,ab,id. OR training\*.ti,ab,id. OR experiment.ti,ab,id. OR instructional effectiveness/

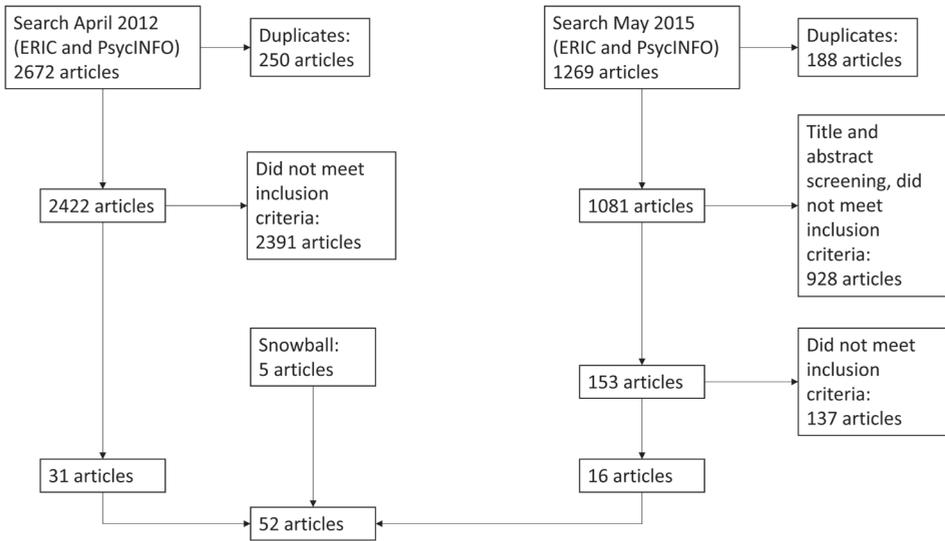
#### #3 Strategy-instruction

strategies/ OR reading strategies/ OR learning strategies/ OR strateg\*.ti,ab,id. OR reading instruction/ OR metacognition/ OR metacogn\*.ti,ab,id. OR self regulated learning/ OR self-regul\*.ti,ab,id. OR question generat\*.ti,ab,id. OR questioning/ OR questioning techniques/ OR summar\*.ti,ab,id. OR predict\*.ti,ab,id. OR comprehension monitor\*.ti,ab,id. OR inferenc\*.ti,ab,id. OR teaching methods/ OR teaching/

#### #4 Children

(school age 6 12 yrs OR adolescence 13 17 yrs).ag. OR (early childhood education OR elementary education OR elementary secondary education or grade 1 or grade 2 or grade 3 or grade 4 or grade 5 or grade 6 or grade 7 OR grade 8 or grade 9 or grade 10 or grade 11 or grade 12 or intermediate grades or junior high schools or middle schools or primary education or secondary education).el. OR grade 1/ OR grade 2/ OR grade 3/ OR grade 4/ OR grade 5/ OR grade 6/ OR grade 7/ OR grade 8/ OR grade 9/ OR grade 10/ OR grade 11/ OR grade 12/ OR childhood development/ or child development/ OR child psychology/ OR child behavior/ OR adolescent development/ or children/ OR child\*.ti,ab,id. or puberty/ or puberty.ti,ab,id. or teen\*.ti,ab,id. or young\*.ti,ab,id. or youth\*.ti,ab,id. or preadolescents/ OR adolesc\*.ti,ab,id. or girl\*.ti,ab,id. or boy\*.ti,ab,id. or students/ or elementary school students/ or primary education/ OR primary school students/ or middle school students/ or junior high school students/ or high school students/ or school\*.ti,ab,id.

## Appendix B



**Figure B.**

Database search and selection diagram.

## Appendix C

**Table C**

Key characteristics of the studies included in the meta-analysis

Author, year of publication	Dependent variable	Post-test	d	SE	n exp	n control	Grades
Aaron, 2008	RC-St	I	0,458	0,181	46	97	3-5
Alfassi, 2009-a	RC-St	I	0,054	0,187	61	54	7
Alfassi, 2009-b	RC-RD	I	0,237	0,188	61	54	7
Allor, 2010	RC-St	I	-0,035	0,263	34	25	3-4
Andreassen, 2011-a	RC-St	I	0,007	0,153	84	86	5
Andreassen, 2011-b	RC-RD	I	0,440	0,155	84	86	5
Andreassen, 2011-c	SA	I	0,293	0,149	90	91	5
Berkeley 2011-a	RC-RD	I	0,465	0,120	156	132	6-7
Berkeley 2011-c	SR	I	0,789	0,117	177	142	6-7
Calhoon, 2005	RC-St	I	0,624	0,333	18	20	6-8
Cantrell, 2010-a	RC-St	I	0,218	0,116	171	131	6
Cantrell, 2010-b	SR	I	0,232	0,160	94	67	6
Cantrell, 2010-e	RC-St	I	0,076	0,107	194	159	9
Cantrell, 2010-f	SR	I	0,232	0,160	94	67	9
Chambers-Cantrell, 2016-a	RC-St	I	0,151	0,060	605	530	6
Chambers-Cantrell, 2016-b	RC-St	I	0,096	0,060	593	535	9
Denton, 2008	RC-St	I	0,030	0,325	20	18	6-8
DiCecco, 2002-a	RC-RD	I	-0,490	0,435	10	12	6-8
Durukan, 2011-a	RC-RD	I	1,532	0,340	24	21	7
Durukan, 2011-b	RC-RD	D	1,315	0,329	24	21	7
Elbro, 2013-a	RC-St	I	0,438	0,137	151	85	6
Elbro, 2013-c	SA	I	0,674	0,139	151	85	6
Faggella-Luby, 2011-a	RC-St	I	0,232	0,374	12	18	5
Faggella-Luby, 2011-b	SA	I	0,602	0,381	12	18	5
Faggella-Luby, 2011-d	RC-St	I	1,055	0,375	14	19	6
Fang, 2010	RC-St	I	0,504	0,136	140	93	6
Fogarty, 2014-a	RC-St	I	0,062	0,068	411	448	6-8

Reader type	Context	Trainer	Hrs	Weeks	Reading strategies	Didactic principles	Study design
LD	Lang	Researcher	80	20	PK, Goal, Quest, SumD, CDS	Sc/Mod/GW	matched
TD	Lang	Teacher	24	16	Pred, Quest, SumD, CDS	Sc/Mod/GW	random
TD	Lang	Teacher	24	16	Pred, Quest, SumD, CDS	Sc/Mod/GW	random
LD	Lang	Teacher	225	60	Pred, SumD, Infer, GO, TS	GW	random
TD	Cont	Teacher	67,5	18	Pred, PK, Quest, SumD, CDS	Sc/Mod/GW	quasi
TD	Cont	Teacher	67,5	18	Pred, PK, Quest, SumD, CDS	Sc/Mod/GW	quasi
TD	Cont	Teacher	67,5	18	Pred, PK, Quest, SumD, CDS	Sc/Mod/GW	quasi
TD	Cont	Teacher	1	1	PK, Quest, SumD, CDS, GO	no	random
TD	Cont	Teacher	1	1	PK, Quest, SumD, CDS, GO	no	random
LD	Lang	Teacher	34	31	Pred, SumD	Sc/Mod/GW	random
LA	Lang	Teacher	108,5	40	Quest, SumD, CDS, Im	Sc/Mod/GW	matched
LA	Lang	Teacher	108,5	40	Quest, SumD, CDS, Im	Sc/Mod/GW	matched
LA	Lang	Teacher	135,83	40	Quest, SumD, CDS, Im	Sc/Mod/GW	matched
LA	Lang	Teacher	135,83	40	Quest, SumD, CDS, Im	Sc/Mod/GW	matched
LA	Lang	Teacher	233,33	44	Quest, Pph, Memo	no	random
LA	Lang	Teacher	223,33	44	Quest, Pph, Memo	no	random
LA	Lang	Teacher	29	13	Quest, SumD	no	random
LD	Lang	Teacher	13,33	4	GO	no	random
TD	Lang	Researcher	10	5	Pred, Quest, SumD	Sc/GW	random
TD	Lang	Researcher	10	5	Pred, Quest, SumD	Sc/GW	random
TD	Lang	Teacher	6	6	Infer, GO	Mod	random
TD	Lang	Teacher	6	6	Infer, GO	Mod	random
LA	Lang	Researcher	22,5	18	Quest, SumD, GO	Sc/Mod	random
LA	Lang	Researcher	22,5	18	Quest, SumD, GO	Sc/Mod	random
LA	Lang	Researcher	22,5	18	Quest, SumD, GO	Sc/Mod	random
TD	Cont	Teacher	6,42	22	Pred, Quest, SumD, GO, TS	Sc/Mod	random
TD	Lang	Teacher	30	12	Pred, Quest, SumD, CDS, GO	Sc/GW	random

Table C

Continued

Author, year of publication	Dependent variable	Post-test	d	SE	n exp	n control	Grades
Fogarty, 2014-b	RC-RD	I	0,049	0,068	411	448	6-8
Graves, 2011a-a	RC-St	I	0,101	0,278	28	24	6
Graves, 2011a-b	RC-St	I	0,061	0,310	27	17	6
Graves, 2011b	RC-St	I	-0,042	0,289	25	23	6
Guthrie, 2004-a	RC-RD	I	0,435	0,132	257	76	3
Guthrie, 2009-a	RC-St	I	0,529	0,213	53	40	5
Guthrie, 2009-b	RC-St	I	0,963	0,278	41	22	5
Guthrie, 2014-i	RC-RD	I	0,130	0,085	269	288	7
Jitendra, 2000-a	RC-RD	I	1,675	0,406	18	15	6-8
Jitendra, 2000-b	RC-RD	D	0,606	0,357	18	15	6-8
Kim, 2006-a	RC-St	I	0,456	0,348	16	18	6-8
Kim, 2006-b	SA	I	0,951	0,362	16	18	6-8
Klingner, 2004	RC-St	I	0,155	0,138	113	98	3-5
Lau, 2007-a	SA	I	1,272	0,270	21	63	7
Lau, 2007-h	RC-RD	I	0,296	0,253	21	63	7
Lederer, 2000-a	RC-RD	I	0,347	0,179	62	65	4-6
Lee, 2014	RC-St	I	0,383	0,197	53	52	3
Lovett, 2012	RC-St	I	0,371	0,139	197	71	9
Lubliner, 2005	RC-RD	I	0,324	0,216	77	30	5
Lucariello, 2012-a	RC-St	I	-0,242	0,299	23	22	3
Lucariello, 2012-b	SA	I	1,426	0,334	23	22	3
Lundberg, 2013	RC-RD	I	-0,441	0,320	20	20	6-12
Mason, 2013-a	RC-St	I	0,795	0,311	29	18	4
Mason, 2013-b	RC-St	D	0,543	0,305	29	18	4
McCown, 2014-e	SR	I	-0,034	0,207	58	39	5
McCown, 2014-a	RC-St	I	0,714	0,213	58	39	5
McKeown, 2009-a	RC-RD	I	-0,109	0,231	37	38	5

Reader type	Context	Trainer	Hrs	Weeks	Reading strategies	Didactic principles	Study design
TD	Lang	Teacher	30	12	Pred, Quest, SumD, CDS, GO	Sc/GW	random
LA	Lang	Researcher	30	10	Quest, SumD	Mod/GW	matched
LA	Lang	Researcher	30	10	Quest, SumD	Mod/GW	matched
LA	Lang	Researcher	30	10	Quest, SumD	Mod/GW	random
TD	Cont	Teacher	90	12	PK, Quest, SumD, Ull, GO	Sc/Mod/GW	quasi
HA	Lang	Teacher	90	12	SumD, Infer, CDS	Sc/Mod/GW	matched
LA	Lang	Teacher	116	12	SumD, Infer, CDS	Sc/Mod/GW	matched
TD	Cont	Teacher	16,7	4	SumD, Infer, GO	Sc/Mod/GW	random
LD	Lang	Researcher	5,33	2	SumD	Mod	random
LD	Lang	Researcher	5,33	2	SumD	Mod	random
LD	Lang	Teacher	16,66	11	Pred, SumD, CDS, SumA	Mod/GW	quasi
LD	Lang	Teacher	16,66	11	Pred, SumD, CDS, SumA	Mod/GW	quasi
Mixed	Lang	Teacher	53,33	40	Pred, SumD, CDS, SumA	Mod/GW	matched
LA	Lang	Researcher	18,66	6	SumD, Infer, CDS, TS	Mod/GW	quasi
LA	Lang	Researcher	18,66	6	SumD, Infer, CDS, TS	Mod/GW	quasi
TD	Cont	Researcher	12	4	Pred, Quest, SumD, CDS	Sc/GW	quasi
TD	Lang	Teacher	10,66	8	Pred, SumD	GW	random
RD	Lang	Teacher	65	20	Pred, PK, Quest, SumD, CDS	Sc/Mod/GW	quasi
LA	Lang	Teacher	18	12	SumD	Sc/Mod	quasi
TD	Lang	Researcher	10,6	4	Infer, Im	Mod	random
TD	Lang	Researcher	10,6	4	Infer, Im	Mod	random
LD	Lang	Teacher	8	8	Pred, Quest, SumD, CDS	Sc/Mod/GW	quasi
LA	Lang	Researcher	10	9	Pred, Goal, SumD, Infer, CDS, SumA	Sc/Mod	random
LA	Lang	Researcher	10	9	Pred, Goal, SumD, Infer, CDS, SumA	Sc/Mod	random
TD	Cont	Researcher		12	Pred, SumD, CDS, SumA	GW	quasi
TD	Cont	Researcher		12	Pred, SumD, CDS, SumA	GW	quasi
TD	Lang	Teacher	7,25		Pred, Quest, SumD, Infer, CDS	no	random

Table C

Continued

Author, year of publication	Dependent variable	Post-test	d	SE	n exp	n control	Grades
McKeown, 2009-d	SA	I	0,194	0,232	37	38	5
McKeown, 2009-e	SK	I	0,422	0,232	39	37	5
Miller, 2011-a	RC-RD	I	-0,011	0,325	20	18	3-8
Miller, 2011-g	RC-RD	D	0,644	0,333	20	18	3-8
Ng, 2013-a	RC-St	I	0,789	0,330	22	18	5
Ng, 2013-b	RC-St	I	0,642	0,323	23	18	5
Ng, 2013-c	RC-RD	I	0,508	0,309	27	18	5
Orbea, 2010-a	RC-St	I	0,284	0,131	118	117	3-4
Orbea, 2010-b	RC-St	I	0,170	0,135	115	107	3-4
Ponce, 2012	RC-St	I	0,414	0,116	742	83	4
Proctor, 2011	RC-St	I	-0,013	0,129	129	111	5
Radcliffe, 2008-a	RC-RD	I	1,300	0,312	23	27	6
Radcliffe, 2008-b	SR	I	0,741	0,293	23	27	6
Redford, 2012-a	RC-RD	I	0,001	0,239	39	32	7
Reis, 2008-a	RC-St	I	-0,017	0,099	251	173	3-5
Reis, 2008-b	RC-St	I	0,367	0,181	56	70	3-5
Reis, 2011-a	RC-St	I	-0,110	0,127	148	106	3-5
Reis, 2011-b	RC-St	I	-0,030	0,109	175	161	3-5
Reis, 2011-c	RC-St	I	0,110	0,136	108	109	3-5
Reis, 2011-d	RC-St	I	-0,010	0,135	136	91	3-5
Reis, 2011-e	RC-St	I	0,270	0,160	82	76	3-5
Rojas-Drummond, 2014	RC-RD	I	1,917	0,221	60	60	5
Schünemann, 2013-a	RC-St	I	0,109	0,155	127	62	5
Schünemann, 2013-b	RC-St	D	-0,038	0,155	127	62	5
Schünemann, 2013-c	RC-St	I	0,441	0,159	117	62	5
Schünemann, 2013-d	RC-St	D	0,058	0,157	117	62	5
Schünemann, 2013-e	SA	I	0,347	0,156	127	62	5
Schünemann, 2013-f	SA	D	0,142	0,155	127	62	5

Reader type	Context	Trainer	Hrs	Weeks	Reading strategies	Didactic principles	Study design
TD	Lang	Teacher	7,25		Pred, Quest, SumD, Infer, CDS	no	random
TD	Lang	Teacher	7,25		Pred, Quest, SumD, Infer, CDS	no	random
LD	Lang	Teacher	9	3	SumD	Sc/Mod	random
LD	Lang	Teacher	9	3	SumD	Sc/Mod	random
LA	Lang	Teacher	6	6	TS	no	quasi
LA	Lang	Teacher	6	6	TS	no	quasi
LA	Lang	Teacher	6	6	TS	no	quasi
TD	Lang	Teacher	8	8	Pred, SumD, Infer, CDS, TS	Sc	quasi
TD	Lang	Teacher	8	8	Pred, SumD, Infer, CDS, TS	Sc	quasi
TD	Lang	Teacher	45	25	Quest, SumD, UII, TS, Pph	no	quasi
TD	Lang	Teacher	26,67	16	Pred, Quest, SumD, CDS	no	random
TD	Cont	Teacher	999	5	Pred, Goal, Quest, SumD, Infer, GO	Sc/Mod	quasi
TD	Cont	Teacher	999	5	Pred, Goal, Quest, SumD, Infer, GO	Sc/Mod	quasi
TD	Lang	Teacher	0,875	1	GO	no	random
TD	Lang	Teacher	70	14	Pred, Infer, CDS	Sc/Mod/GW	random
TD	Lang	Teacher	70	14	Pred, Infer, CDS	Sc/Mod/GW	random
TD	Lang	Teacher	110	22	Pred, Infer, CDS	Sc/Mod/GW	random
TD	Lang	Teacher	110	22	Pred, Infer, CDS	Sc/Mod/GW	random
TD	Lang	Teacher	110	22	Pred, Infer, CDS	Sc/Mod/GW	random
TD	Lang	Teacher	110	22	Pred, Infer, CDS	Sc/Mod/GW	random
TD	Lang	Teacher	110	22	Pred, Infer, CDS	Sc/Mod/GW	random
TD	Lang	Teacher	27	28	SumD, Infer, TS	Sc/GW	quasi
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	quasi
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	quasi
TD	Lang	Researcher	10,5	7	Pred, Goal, Quest, SumD, CDS	Sc/Mod/GW	quasi
TD	Lang	Researcher	10,5	7	Pred, Goal, Quest, SumD, CDS	Sc/Mod/GW	quasi
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	quasi
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	quasi

Table C

Continued

Author, year of publication	Dependent variable	Post-test	d	SE	n exp	n control	Grades
Schünemann, 2013-g	SA	I	0,339	0,158	117	62	5
Schünemann, 2013-h	SA	D	0,485	0,159	117	62	5
Simmons, 2010	RC-St	I	0,021	0,086	324	232	4
Simmons, 2014-a	RC-St	I	0,034	0,066	489	432	7-10
Simmons, 2014-b	RC-RD	I	0,069	0,072	409	368	7-10
Souvignier, 2006-a	RC-St	I	-0,022	0,150	95	84	5
Souvignier, 2006-b	RC-St	D	0,258	0,150	95	84	5
Souvignier, 2006-c	SK	I	0,384	0,151	95	84	5
Souvignier, 2006-d	SK	D	0,297	0,151	94	84	5
Souvignier, 2006-g	RC-St	I	-0,046	0,137	146	84	5
Souvignier, 2006-h	RC-St	D	0,105	0,137	146	84	5
Souvignier, 2006-i	SK	I	0,555	0,139	146	84	5
Souvignier, 2006-j	SK	D	-0,009	0,137	146	84	5
Souvignier, 2006-m	RC-St	I	-0,036	0,152	89	84	5
Souvignier, 2006-n	RC-St	D	0,378	0,153	89	84	5
Souvignier, 2006-o	SK	I	0,225	0,153	89	84	5
Souvignier, 2006-p	SK	D	0,084	0,152	89	84	5
Sporer, 2009a-a	RC-RD	I	0,428	0,148	95	91	7
Sporer, 2009a-b	RC-St	I	0,372	0,148	95	91	7
Sporer, 2009a-c	SA	I	0,228	0,147	95	91	7
Sporer, 2009a-e	SK	I	0,244	0,147	95	91	7
Sporer, 2009a-g	SR	I	0,221	0,147	95	91	7
Sporer, 2009b-aa	RC-St	D	0,037	0,178	60	66	3-6
Sporer, 2009b-b	SA	I	0,536	0,201	42	66	3-6

Reader type	Context	Trainer	Hrs	Weeks	Reading strategies	Didactic principles	Study design
TD	Lang	Researcher	10,5	7	Pred, Goal, Quest, SumD, CDS	Sc/Mod/GW	quasi
TD	Lang	Researcher	10,5	7	Pred, Goal, Quest, SumD, CDS	Sc/Mod/GW	quasi
TD	Cont	Teacher	27	18	Quest, SumD	Sc/Mod/GW	random
TD	Lang	Teacher	40,5	18	Pred, Quest, SumD, Infer, CDS, GO, SumA	Sc/Mod/GW	random
TD	Lang	Teacher	40,5	18	Pred, Quest, SumD, Infer, CDS, GO, SumA	Sc/Mod/GW	random
TD	Lang	Teacher	15	20	PK, Quest, SumD, CDS, Ull, Im, Memo	Sc/Mod/GW	quasi
TD	Lang	Teacher	15	20	PK, Quest, SumD, CDS, Ull, Im, Memo	Sc/Mod/GW	quasi
TD	Lang	Teacher	15	20	PK, Quest, SumD, CDS, Ull, Im, Memo	Sc/Mod/GW	quasi
TD	Lang	Teacher	15	20	PK, Quest, SumD, CDS, Ull, Im, Memo	Sc/Mod/GW	quasi
TD	Lang	Teacher	15	20	PK, Quest, SumD, CDS, Ull, Im, Memo	Sc/Mod/GW	quasi
TD	Lang	Teacher	15	20	PK, Quest, SumD, CDS, Ull, Im, Memo	Sc/Mod/GW	quasi
TD	Lang	Teacher	15	20	PK, Quest, SumD, CDS, Ull, Im, Memo	Sc/Mod/GW	quasi
TD	Lang	Teacher	15	20	PK, Quest, SumD, CDS, Ull, Im, Memo	Sc/Mod/GW	quasi
TD	Lang	Teacher	15	20	PK, Quest, SumD, CDS, Ull, Im, Memo	Sc/Mod/GW	quasi
TD	Lang	Teacher	15	20	PK, Quest, SumD, CDS, Ull, Im, Memo	Sc/Mod/GW	quasi
TD	Lang	Teacher	10,5	9	Pred, SumD, SumA	Sc/Mod/GW	random
TD	Lang	Teacher	10,5	9	Pred, SumD, SumA	Sc/Mod/GW	random
TD	Lang	Teacher	10,5	9	Pred, SumD, SumA	Sc/Mod/GW	random
TD	Lang	Teacher	10,5	9	Pred, SumD, SumA	Sc/Mod/GW	random
TD	Lang	Teacher	10,5	9	Pred, SumD, SumA	Sc/Mod/GW	random
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	random
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	random

Table C

Continued

Author, year of publication	Dependent variable	Post-test	<i>d</i>	SE	<i>n</i> exp	<i>n</i> control	Grades
Sporer, 2009b-cc	SA	D	0,230	0,198	42	66	3-6
Sporer, 2009b-e	RC-RD	I	1,219	0,214	42	66	3-6
Sporer, 2009b-g	SA	I	0,879	0,187	60	66	3-6
Sporer, 2009b-gg	RC-St	D	0,058	0,197	42	66	3-6
Sporer, 2009b-l	SA	I	0,538	0,201	42	66	3-6
Sporer, 2009b-q	SA	D	0,544	0,201	42	66	3-6
Sporer, 2009b-t	RC-RD	D	1,043	0,210	42	66	3-6
Sporer, 2009b-u	RC-St	D	0,403	0,199	42	66	3-6
Sporer, 2009b-w	SA	D	0,558	0,182	60	66	3-6
Sung, 2008-j	SA	I	1,305	0,272	31	35	6
Sung, 2008-m	RC-St	I	0,468	0,254	34	30	6
Sung, 2008-u	SA	I	1,247	0,274	34	30	6
Wijekumar, 2012-a	RC-St	I	0,255	0,040	1280	1300	4
Wijekumar, 2012-c	SA	I	1,473	0,048	1100	1080	4
Wijekumar, 2012-d	RC-RD	I	0,312	0,043	1100	1080	4
Wijekumar, 2014-a	RC-St	I	0,134	0,040	1258	1227	5
Wijekumar, 2014-c	SA	I	0,499	0,044	1156	1004	5
Wijekumar, 2014-d	RC-RD	I	0,368	0,043	1161	1012	5

Note: RC-RD = reading comprehension researcher-developed, RC-St = reading comprehension standardized, SA = strategic ability, SK = strategy knowledge, SR = strategy self-report, I = immediate posttest, D = delayed posttest, *d* = Cohen's *d*, SE = standard error of *d*, LD = learning disabled, TD = typical development, LA = low-achievers, HA = high achievers, RD = reading disabled, Lang = language arts classes, Cont = content area classes, Hrs = total duration of intervention in hours, Weeks = total duration of intervention in number of weeks, Pred = predicting; PK = prior knowledge; Goal = goal setting; Quest = questioning; SumD = summarizing during reading, Infer = inferencing, CDS = comprehension directed strategies; Ull = underline important information, GO = graphic organizers / visual representation, TS = text structure, Im = imagery, Pph = paraphrasing, HW = hinge words, SumA = summarizing after reading, Memo = memorizing, Sc= scaffolding, Mod = modeling, GW = group work; BUA = business-as-usual; Control = controlled control group; Random = randomized experiment; Quasi = quasi-experiment.

Reader type	Context	Trainer	Hrs	Weeks	Reading strategies	Didactic principles	Study design
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod	random
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	random
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	random
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod	random
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod	random
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	random
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	random
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	random
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	random
TD	Lang	Researcher	10,5	7	Pred, Quest, SumD, CDS	Sc/Mod/GW	random
HA	Lang	Researcher	18,33	11	Quest, SumD, Infer, CDS, Ull, GO	Sc/Mod	random
LA	Lang	Researcher	18,33	11	Quest, SumD, Infer, CDS, Ull, GO	Sc/Mod	random
LA	Lang	Researcher	18,33	11	Quest, SumD, Infer, CDS, Ull, GO	Sc/Mod	random
TD	Lang	Teacher	17,27	28	SumD, TS, HW	Sc/Mod	random
TD	Lang	Teacher	17,27	28	SumD, TS, HW, Memo	Sc/Mod	random
TD	Lang	Teacher	17,27	28	SumD, TS, HW, Memo	Sc/Mod	random
TD	Lang	Teacher	24,375	26	SumD, TS, HW, Memo	Sc/Mod	random
TD	Lang	Teacher	24,375	26	SumD, TS, HW, Memo	no	random
TD	Lang	Teacher	24,375	26	SumD, TS, HW, Memo	no	random

## Appendix D

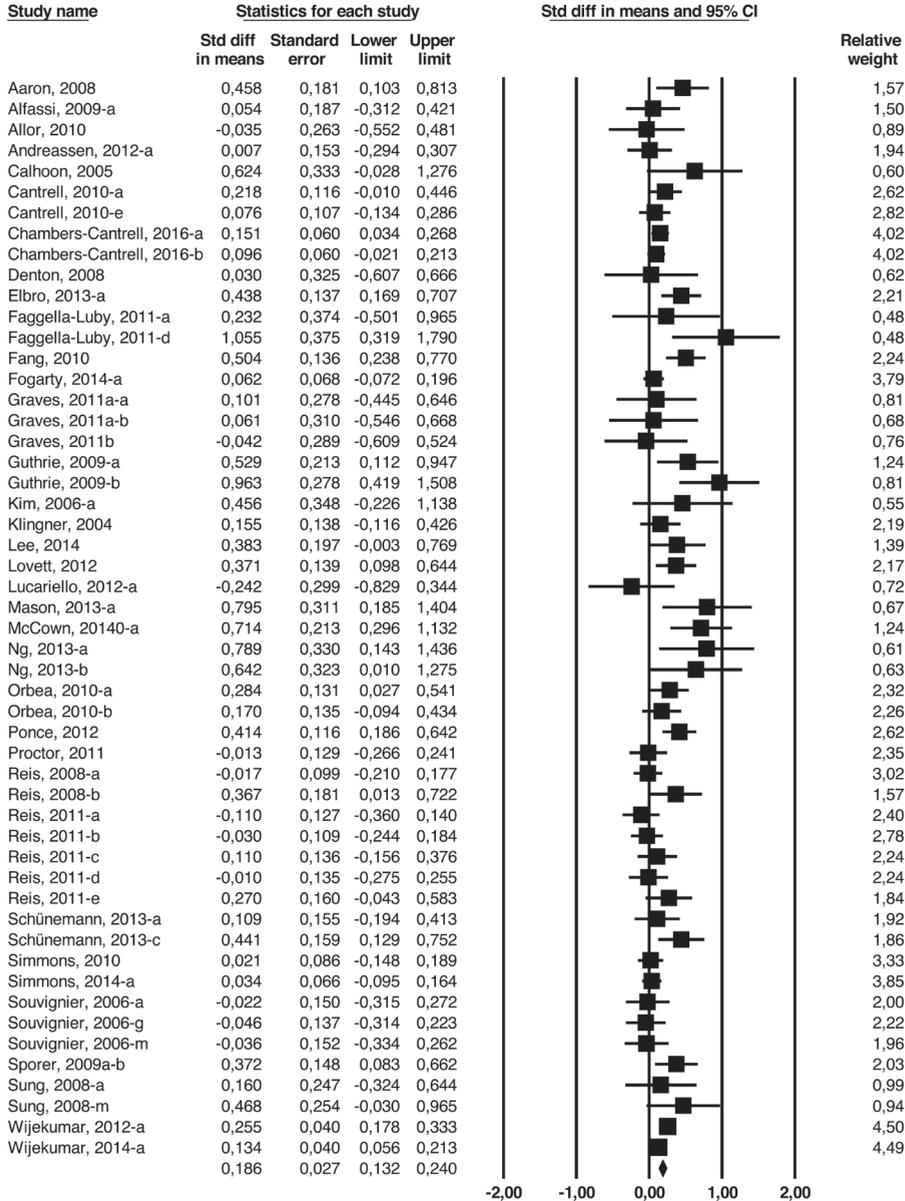
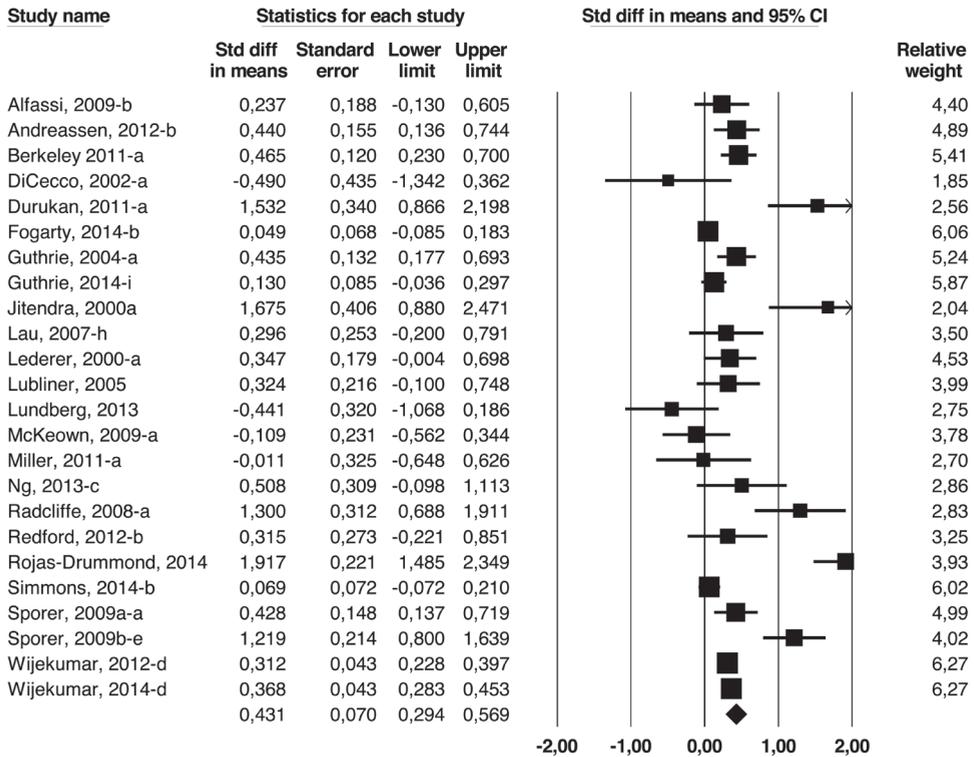


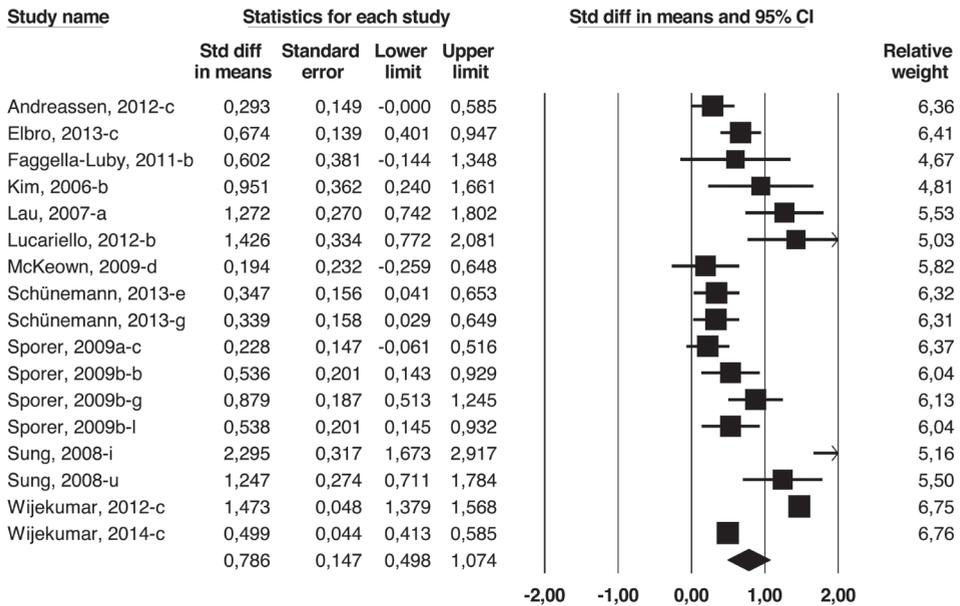
Figure D1

Forest plot for reading comprehension standardized (immediate). Forest plot showing standardized mean differences (SMDs) and 95% confidence intervals (CIs) of individual studies and a summary effect (represented by a diamond).

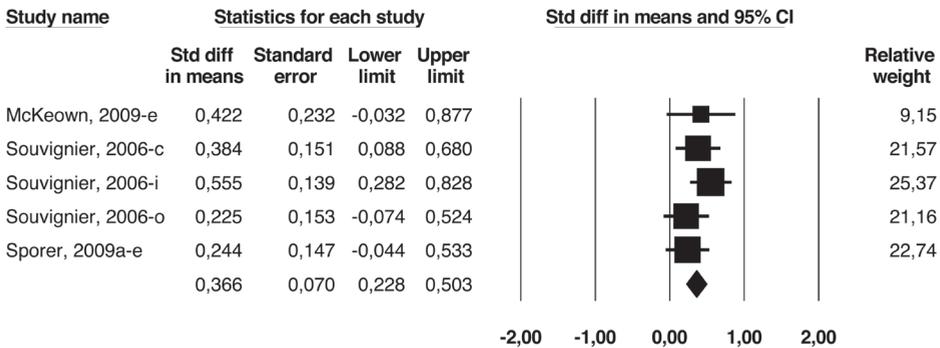


**Figure D2**

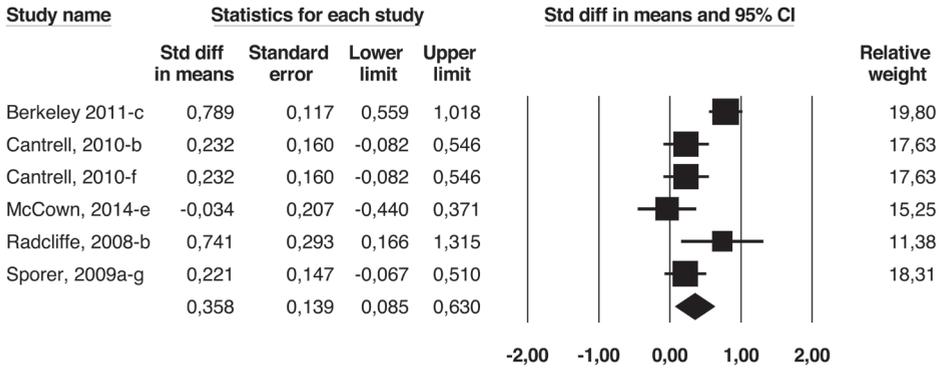
Forest plot for researcher-developed reading comprehension (immediate). Forest plot showing standardized mean differences (SMDs) and 95% confidence intervals (CIs) of individual studies and a summary effect (represented by a diamond).



**Figure D3**  
Forest plot for strategic ability (immediate). Forest plot showing standardized mean differences (SMDs) and 95% confidence intervals (CIs) of individual studies and a summary effect (represented by a diamond).



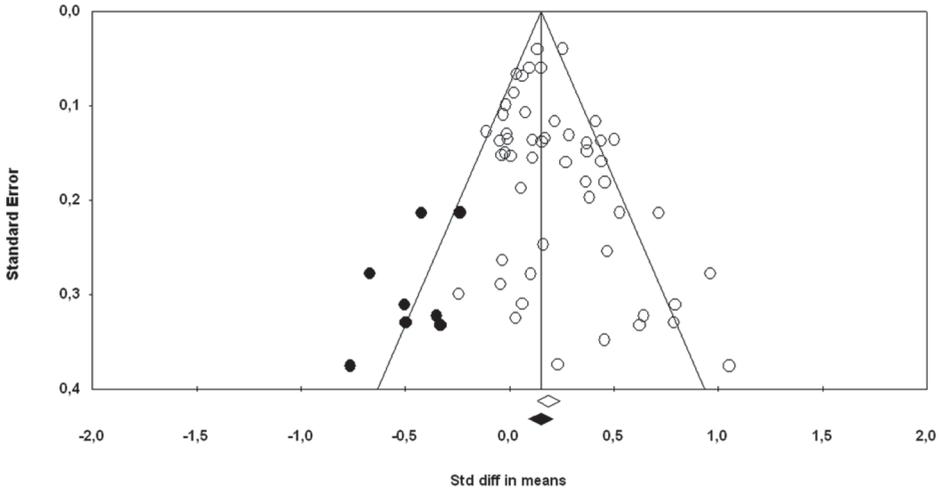
**Figure D4**  
Forest plot for strategy knowledge (immediate). Forest plot showing standardized mean differences (SMDs) and 95% confidence intervals (CIs) of individual studies and a summary effect (represented by a diamond).



**Figure D5**

Forest plot for strategy-use self-report (immediate). Forest plot showing standardized mean differences (SMDs) and 95% confidence intervals (CIs) of individual studies and a summary effect (represented by a diamond).

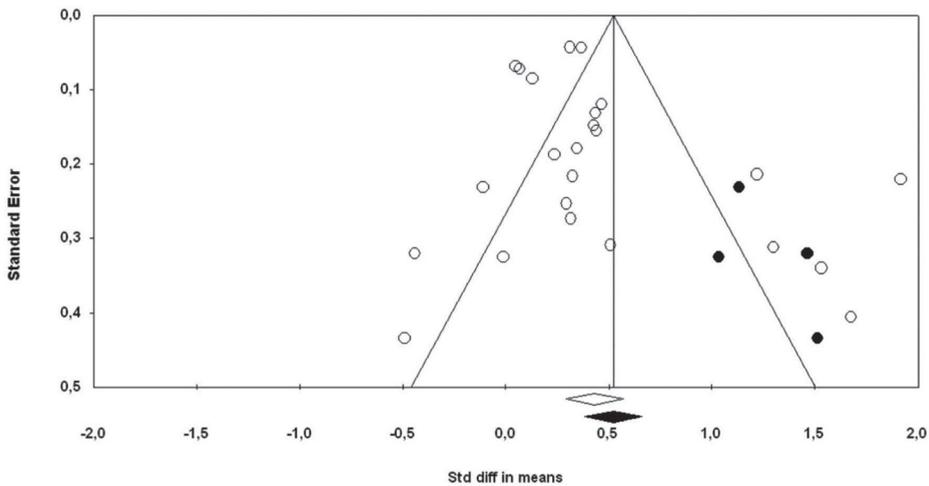
## Appendix E



**Figure E1**

Funnel plot of standard error by effect size for reading comprehension standardized (immediate).

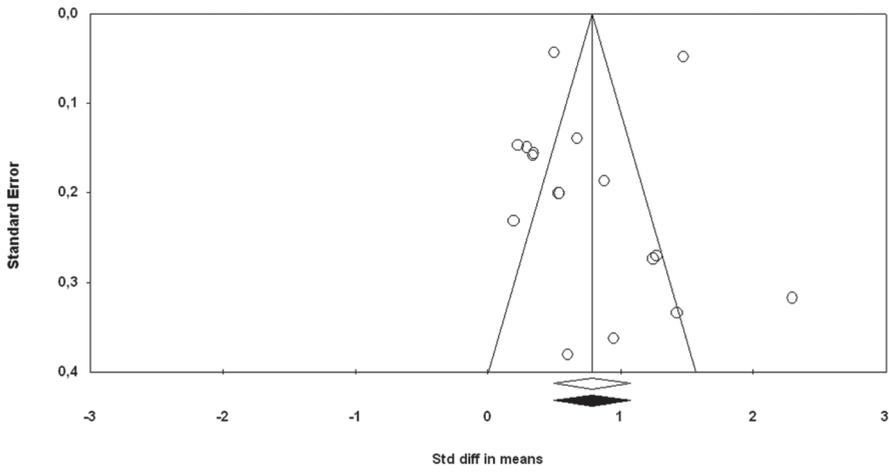
*Note.* The observed effects are represented by an open circle; imputed effects are represented by a filled circle. The open diamond at the bottom represents our mean effect size, the filled diamond represents the mean effect size based on the total number of effects, including imputed effects.



**Figure E2**

Funnel plot of standard error by effect size for researcher developed reading comprehension (immediate).

*Note.* The observed effects are represented by an open circle; imputed effects are represented by a filled circle. The open diamond at the bottom represents our mean effect size, the filled diamond represents the mean effect size based on the total number of effects, including imputed effects.



**Figure E3**

Funnel plot of standard error by effect size for strategic ability (immediate).

*Note.* The observed effects are represented by an open circle; imputed effects are represented by a filled circle. The open diamond at the bottom represents our mean effect size, the filled diamond represents the mean effect size based on the total number of effects, including imputed effects





General discussion



## Introduction

This thesis aimed to gain insight into the improvement of low-achieving adolescents reading comprehension through the instruction of reading strategies in whole-classroom settings. In the first place, a two-year experimental study among prevocational students was conducted in which regular teachers were trained and coached to instruct reading strategies and guide their students in the application of reading strategies through reciprocal teaching.

An important feature of the experimental study is the setting in which the treatment took place; we conducted research in ecologically valid settings, namely in whole classrooms with the students' regular teachers. Given that instruction of reading strategies in such settings is demanding and difficult for teachers (Duffy, 1993; Hacker & Tenent, 2002; Seymour & Osana, 2003), implementation quality of the intervention is an important issue to attend to (Chapter 2 and Chapter 3). A second important feature of the study is the longitudinal design – we followed the teachers and students for two school years. The combination of these features makes this study unique. To our knowledge, this is the first long-term experimental study conducted in a whole-classroom setting, which incorporates implementation quality as a moderator of the treatment effects.

The treatment was based on an existing program '*Nieuwsbegrip*' ('Newswise'), which is meant to be used for multiple school years, and as such is expected to have continued effects on students' reading comprehension skills. The program is based on principles of reciprocal teaching, which were originally developed for small group tutoring (Palincsar & Brown, 1984), but transfers these principles to whole-classroom settings.

The effect of instructing reading strategies on growth in reading comprehension was investigated after one year (*Chapter 2*) and two years of treatment (*Chapter 3*). In both chapters the moderating effect of implementation quality of the treatment was analyzed. In addition, the moderating role students' vocabulary knowledge plays in the overall treatment effect on reading comprehension was analyzed (*Chapter 4*). Finally, a meta-analysis (*Chapter 5*) was conducted in which the overall effectiveness of instructing reading strategies in whole classroom settings on reading comprehension was examined. In addition, the moderating effects of student-, intervention-, and study design characteristics were examined.

In this chapter, the main findings of the experimental study and the meta-analysis are discussed, followed by a discussion of these findings. The chapter ends with suggestions for future research into instructing reading strategies in whole-classroom settings and implications for educational practice.

## Main Findings

From the experimental study it appeared that no main effects of the intervention were found, after either one or two years. Furthermore, we have learned that the quality of implementation is an important factor moderating the effectiveness of the reading strategy intervention. Despite the absence of a main effect of the treatment, during the first year of the treatment (*Chapter 2*), elaborate strategy-instruction moderated the effect on reading comprehension. When strategy-instruction was more elaborate (i.e. more elaborate explanations of the nature, function, importance and application of reading strategies), this positively impacted students' reading comprehension in the experimental condition but not in the control condition. This effect of strategy-instruction was not found after two years of treatment. Instead, elaborate *modeling* of strategies influenced the effect on growth in reading comprehension (*Chapter 3*). Growth in reading comprehension in the experimental condition depended on elaborate modeling (both modeling by teachers and by their students). When modeling was more elaborate, no differences in growth in reading comprehension were found between the experimental and control condition. On the other hand, it appeared that in the case of less elaborate modeling growth in reading comprehension in the experimental group was less than in the control group.

The experimental study additionally revealed that students with different levels of vocabulary knowledge responded differently to the treatment, while this was not the case for the control condition (*Chapter 4*). When students in the experimental condition scored lower on vocabulary knowledge, their growth in reading comprehension was smaller than students in the control condition with lower scores on vocabulary knowledge. For students with higher scores on vocabulary knowledge, there appears to be no difference in the effect on their growth in reading comprehension between the two conditions. Thus, one might say that the treatment had a discriminating effect on the basis of students' vocabulary knowledge, whereas the business-as-usual control worked similar for students with different vocabulary levels.

The meta-analysis showed that interventions focused on instructing reading strategies in whole-classroom settings can be beneficial to promote reading comprehension. This effect was significantly higher for researcher-developed tests (Cohen's  $d = .431$ ) compared to standardized tests of reading comprehension (Cohen's  $d = .186$ ). A medium overall effect was found for strategic ability (Cohen's  $d = .786$ ). For self-reported strategy use a small but significant effect was found (Cohen's  $d = .358$ ). For delayed tests, we found a very small but significant effect for standardized reading comprehension tests (Cohen's  $d = 0.167$ ) and a medium effect for researcher developed reading comprehension tests (Cohen's  $d = 0.521$ ). No significant effects were found for delayed tests of strategic ability and knowledge.

Moderation effects were analyzed for intervention characteristics (type of strategies instructed, type of trainer, didactic principles, educational context and duration), for student

characteristics (low- achievers vs typical development and grade) and study design (type of control group and randomized vs quasi-experimental) For types of strategies, however, it was not possible to isolate the moderating effects of specific strategies, because in almost all studies involved, strategies were taught in combination with other strategies. Therefore, the moderating effects of strategy types were always calculated as an additional effect in comparison to studies that did not use that specific type of strategy. From all types of strategies discriminated in our study, only 'setting reading goals' was found to positively moderate the effects on reading comprehension tests (both standardized and researcher-developed, in comparison to studies that did not use 'setting reading goals'. As expected, larger effect sizes were obtained for studies in which researchers were the instructor compared to the teachers. For didactic principles the only significant effect was that modeling was beneficial for strategic ability. In the context of language arts classes larger effects were found on strategic ability in comparison to content area classes. There were no significant moderation effects found for the duration of the experimental treatments.

For low-achieving students, reading strategy interventions resulted in larger effects on strategic ability, than for typically developing students. In addition we found that for students in grades 6-8 effects were larger on researcher developed reading comprehension tests than for students in lower or higher grades. For the study-design characteristics we only found a significant moderating effect for the type of control group: in studies that used business as usual control groups larger effects were found for researcher developed reading comprehension tests in comparison to studies that used controlled control groups (in which a different intervention was implemented e.g. for keeping time on task similar between experimental and control conditions).

### **Difficulties in Finding Intervention Effects of Reading-Strategy Instruction in Whole-Classroom Contexts**

Overall, it appeared that positive results for reading strategies interventions aimed at promoting reading comprehension in whole-classroom settings were difficult to find. The experimental study showed no main effects of strategy instruction after two years of treatment, while the meta-analysis showed a very small effect for standardized tests of reading comprehension. These results contrast with findings from previous experimental studies. Several explanations for the lack of main positive effects in our experimental study are discussed.

First, teachers may have difficulties to adapt principles of reciprocal teaching successfully to a whole-classroom setting. In the original approach of reciprocal teaching, small groups of students were taken out of the classroom. Under the guidance of a tutor, who had optimal control over the students' behavior, they practiced reading strategies while reading a text (Palincsar & Brown, 1984). In our treatment, the teachers were to manage

up to five groups of students and provide guidance to all simultaneously. This means that, compared to the original approach, the teachers had much less control on collaboration of and the quality of strategy use by the students in each group in their classroom. Thus, in a whole-classroom setting the available time needs to be shared among multiple groups. This makes group work – one of the principles of reciprocal teaching – far more difficult to realize as also demonstrated by observations of Hacker and Tenent (2002). According to their in-depth analysis of teacher practice in reciprocal teaching, they concluded that group work was the most vulnerable component. The collaboration process between students was hampered because students did not practice reciprocal teaching in a productive way. Their discussions about the texts were rather superficial, and therefore did not reach a higher level of comprehension monitoring. In order to compensate for this problem teachers often returned to whole-classroom instruction, thereby jeopardizing one of the most important aspects of reciprocal teaching: the fading of responsibility for the reading process to the students. In addition, we have to consider the fact that our students were low-achievers and therefore may need much more guidance in group work than higher achieving students. Moreover, our students may need more support in comprehension monitoring as studies have shown that low-achievers are not used to making inferences and practicing other types of deeper comprehension processes (Oakhill & Cain, 2007; Rapp, Van den Broek, McMaster et al., 2007; De Milliano, Van Gelderen, Slegers, 2014). Our own classroom observations of the treatment teachers do, suggest that the majority of them indeed experienced serious problems with group work guidance, resulting in insufficient collaborative practice in reading strategies in the whole-classroom settings. Many of our teachers, therefore mentioned that they would have appreciated more coaching of group work.

Secondly, the duration of the experimental study may not have worked beneficially for promoting students' reading comprehension. Several factors may have played a role: loss of motivation, lack of challenging texts, the repeated use of the same type of texts, the rigid use of the same strategies in a certain order, and dropout issues over the course of two school years. Several teachers indicated in the exit interviews that they found it hard to keep the students motivated for such a long time. There are several reasons why motivation might have decreased in the students, as a result of the long duration of the intervention. The difficulty of the texts ('B' level according the CED-group) did not change over the course of two years, which may have caused a lack of challenge for the students in the long run. According to the norms of the CED-group, the text difficulty was appropriate for our prevocational students. Although we stimulated teachers to work with texts of the next level of difficulty ('C' level) at the end of the first school year, they reported that using more difficult and longer texts did not work well in their classrooms. They indicated that it took too long to read the texts and there were too many difficult words in the texts for the students to work independently as a group. In hindsight, we should have tried to

convince the teachers to work with the more difficult texts in the second year, providing more challenging tasks for the students. An advantage of increasing the difficulty of texts is that students are supported in using reading strategies in more complex texts in a flexible manner.

Another issue, which may be related to student motivation in the long run, is that the program is quite rigid in several aspects. The repeated use of the same type of texts in the program 'Nieuwsbegrip' may have become boring to the students, even though every week a newly written text was used (based on a recent news item). Furthermore, the focus in each lesson was always on one of the reading strategies (the 'central strategy' for that lesson), whether or not this was functional for comprehension. This approach runs the risk of reading strategies becoming an end in itself, instead of a means for comprehension. This may also have decreased students' motivation for practicing reading strategies in the long run. Also, the sequencing of activities in each lesson was the same. The fact that the same activities had to be carried out over and over again, may have harmed students' motivation in the end.

Another reason why the long duration of our experimental study may have limited the success of the treatment is related to the number of teachers that dropped out during the two years of the experimental study. Even though we tried our best to replace those teachers and train them, this may have affected the implementation quality and, in turn, the growth in reading comprehension of the students. This is especially the case for those schools that had trouble finding replacements, resulting in skipped lessons.

Finally, it is possible that a long-term intervention results in disappointing effects, because it is difficult to maintain experimental control ensuring high implementation quality over a longer period of time. According to Roberts et al. (2013) large-scale interventions (including long-term interventions) have a disadvantage in terms of internal validity and experimental control and are sensitive to loss of experimental control resulting in lower implementation quality. We did not find indications that duration of the intervention influenced effect sizes in our meta-analysis. However, this can be due to the fact that there were almost no interventions that lasted longer than a year.

Third, the low level of vocabulary knowledge in the target group of prevocational students may also explain the lack of overall positive treatment effects. As our findings clearly showed, low-achieving students with the lowest levels of vocabulary knowledge were at a disadvantage in the experimental condition. Moreover, even the students with a highest level of vocabulary knowledge in the experimental condition did not outperform control students with the highest level of vocabulary. These findings show that the treatment did not adjust sufficiently to the low vocabulary knowledge in the target group of prevocational students. For example, students with a low vocabulary might not know the meaning of enough words in a text necessary for using the reading strategy 'clarification' effectively.

Thus, vocabulary knowledge is an important prerequisite for the successful application of reading strategies for low-achieving students.

Lastly, effects of instruction in reading strategies appear to be much harder to find for standardized tests than for researcher-developed tests which are designed to measure effects on near transfer (Chiu, 1998; Paris, Cross & Lipson, 1984; Rosenshine & Meister, 1994). The meta-analysis showed a significant difference between standardized and researcher-developed tests and this difference became even larger for delayed tests of reading comprehension. In our experimental study, we measured growth in reading comprehension using a reading comprehension test that was designed for the measurement of general reading comprehension development of prevocational students over a period of three years (Van Steensel, Oostdam, & Van Gelderen, 2013). This test has much in common with standardized measures, as it was designed for the measurement of general reading comprehension development of prevocational students over a period of three years (Van Steensel, Oostdam, & Van Gelderen, 2013). Therefore, the test did not mimic the type of reading comprehension questions that were used in the program nor did it focus on the application of reading strategies taught. This may explain why we did not find an overall positive effect of our treatment. If we had used researcher-developed tests, directed at the application of taught reading strategies in similar news-related texts as included in the intervention, would probably have captured differences between treatment and control students.

### **One or Two Years of Treatment: Different Results**

The longitudinal design of the experimental study showed that in the first year treatment effects were moderated by elaborate strategy instruction, while for the whole intervention the degree of elaborate modeling was a significant moderator of treatment effects. On the other hand, in the first year elaborate strategy instruction had a different meaning compared to the role of elaborate modeling over the whole duration of the treatment. After one year of treatment, strategy-instruction was a significant moderator, with more elaborate strategy-instruction resulting in higher scores on reading comprehension in the experimental condition compared to the control condition. After two years of treatment, modeling was a significant moderator. However, this time, the moderation indicated that experimental students who received more elaborate modeling did not differ in growth in reading comprehension from the control students. Moreover, experimental students who received less elaborate modeling were at a disadvantage compared to the control condition.

The difference between the type of significant moderators may be explained by a change in the focus of the treatment. In the first year, students needed to learn the meaning of the reading strategies and teachers focused therefore on *direct instruction* of the reading

strategies. In the second year of the treatment, however, the focus was on *how* the reading strategies should be applied. Therefore, coaching of teachers in the second year focused more on modeling.

The fact that the moderator effect of modeling did not indicate that students in the experimental condition profited from more elaborate modeling might be explained by the dropout of teachers. Especially in the transition between first and second year of treatment, some treatment teachers needed to be replaced. Although replaced teachers were trained and coached as well, they missed the experience of the previous year. Even teachers who remained for the whole duration of the treatment reported in the exit-interviews that they found it difficult to apply the principles of modeling the reading strategies themselves and to stimulate their students to do so as well. Therefore, the replacing teachers, who were newly trained, probably experienced even more difficulties with the principles of modeling. Inspection of the observational data of modeling seem to confirm this assumption. Two replacing teachers indeed scored lower on modeling in the second year compared to the remaining teachers. In fact, both replacing teachers did not model at all. This may explain why elaborate modeling did not result in an advantage for the experimental students after two years.

## **Suggestions for Future Research**

### **Implementation quality**

The findings of our experimental study clearly showed that implementation quality is an important factor for the success of reading strategies interventions in whole-classroom settings. Unfortunately, in the meta-analysis we could not investigate the influence of implementation quality as moderator since most studies did not incorporate implementation quality in their analysis. In future research, studies into the effectiveness of a reading strategies interventions for promoting student reading comprehension should include measures of implementation quality as a moderator in their analysis as this deepens our insight in which interventions contribute the most in enhancing reading comprehension of students. This is especially important in studies in whole-classroom settings with regular teachers (Hulleman & Cordray, 2009; Larsen & Samdal, 2007; Swanson, Wanzek, Haring, Ciullo, & McCulley, 2011; Vaughn et al., 2013).

### **Attention for collaboration process**

Our findings suggest that the majority of teachers experienced serious problems with group work guidance. This may have resulted in insufficient collaborative practice in reading strategies in the groups of low achieving students. As low-achieving students need more support for collaborative learning to benefit from reading strategy training, more research

on how practicing and discussing reading strategies with other students can be enhanced in a whole-classroom setting, is needed. For instance, an ICT-tool that supports the students in their collaborative group work by providing scaffolds that help students regulate their collaborative process may induce more meaningful dialogues between the students about the application of reading strategies. Findings from studies into computer-supported collaborative learning, in which students are working together remotely, can be used for designing and evaluating an ICT-tool that can stimulate students to collaboratively apply reading strategies within a whole-classroom setting (e.g. Karakostas & Demetriadis, 2011).

### **Optimizing interventions aimed at instructing reading strategies**

As vocabulary knowledge is an important factor in fostering reading comprehension but may also be important in the application of reading strategies for low-achieving students, future studies should investigate the optimal ways to include vocabulary instruction in the context of reading strategy interventions for this group. This means that low-achieving students should not only receive instruction in applying specific reading strategies, but also be supported in vocabulary knowledge needed for successful application of these reading strategies. Apart from vocabulary knowledge, self-regulatory skills may also affect how students respond to an intervention targeted at the use of reading strategies. Self-regulatory skills refer to planning, executing, and controlling behavior while performing a task (Boekaerts & Simons, 1993). In the case of reading comprehension, self-regulatory skills are important, for example, in knowing which reading strategy to apply and how to apply the chosen strategy. There is evidence that self-regulatory skills applied in reading tasks may determine low-achieving students' reading comprehension (De Milliano, Van Gelderen & Slegers, 2016).

Therefore, in future studies, the role of self-regulatory skills in the application of reading strategies by low-achieving students may be investigated. More specifically in an intervention study it could be examined whether self-regulatory skills for the application of reading strategies of low-achieving students can be fostered for example with guidance from a tutor. Previous research by Schünemann, Spörer and Brunstein (2013) gives support to the assumption that students may indeed profit from such combination of self-regulation skills and reading strategy instruction.

The experimental study was conducted in the context of language arts classes. Nevertheless, content area teachers from several of our intervention schools noted that their students had great difficulty transferring the reading strategies taught in language arts classes to their subjects (e.g. social sciences, history, or geography). Furthermore, these content area teachers mentioned that their students had great trouble learning from texts. As research has shown that support for language difficulties is scarce in social studies classes (De Milliano, 2013), future research should focus on how students can transfer

the knowledge and skills they have learned from reading strategy intervention programs effectively to different subject areas. This is especially of interest since reading strategies are seen as tools that can be used in any reading context, not only as tools specifically suited for language arts classes. To this end, ways of collaboration between language arts teachers and content area teachers should be investigated in order to determine which of them are suited for realizing the targeted transfer of reading strategies from one context to the other.

### **Implications and Recommendations for Educational Practice**

Several implications for educational practice can be drawn from our study. First, we need reflect on the result that regular teachers are not as proficient as researchers in successfully implementing an intervention aimed at instructing reading strategies. This finding has implication(s) for teacher education programs. Teacher training institutes should pay more attention to *how* teachers should teach reading strategies aimed at fostering reading comprehension (National Reading Panel, 2000). In order for teachers to know how strategies should be taught, it seems necessary that teachers have a deeper understanding of the reading comprehension process. Thus, not only knowledge of the strategies is important, but also how those strategies fit into the teachers own experience in using reading strategies for comprehension. This deeper understanding of the processes involved in reading comprehension should start during teacher education. For example, teachers could be stimulated to think aloud while reading a text and reflect on what reading strategies they use and whether the use of those strategies is helpful or not in understanding.

In the original approach of reciprocal teaching as developed by Palincsar and Brown (1984), with small groups of students guided by a tutor the tutor supports the group process and adapts instruction and guidance to the individual students within that group. In whole-classroom settings, with one teacher overseeing multiple groups of students simultaneously, this approach is far more difficult or even hardly possible. Nevertheless, a few recommendations can be made in this respect. In the first place, teachers may spend effort in assessing which students need the tutoring approach the most. Teachers may observe and identify what their students are capable of and what kind of reading problems students encounter. Furthermore, teachers may sit with one group of students while other groups work independently, and rotate the group of students for intensive support. Within the intensive support group, the teacher may let the students think aloud while reading to observe their reading process. Such an approach does pose strong demands on effective classroom management as teachers need to actively engage the students in learning who are working independently within their group, without the constant supervision of the teacher.

As reading strategies are tools for supporting students to better comprehend texts, teachers should take into account that reading strategies are a means to an end and never an end in itself. Teachers need to instruct students how to use reading strategies in a flexible manner and to become aware of the degree to which they understand the text in order to select the strategy that might help them the most at that time.

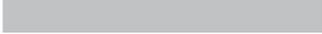
Next, teachers need to motivate students by providing challenging tasks (De Milliano, 2013), by providing variation in type of texts, by letting students select their own texts, and by instructing students how to read the texts they find interesting (Guthrie & Klauda, 2014) in order to enhance students' engagement for reading at school.

Finally, teachers need to be aware that instructing reading strategies is not a quick fix to increase students' performance on standardized tests. Nevertheless, instructing reading strategies is useful as part of reading comprehension instruction. The results of the meta-analysis show that effects on reading comprehension are sustained, on researcher-developed tests, even after the intervention has ended. Together with the indications that implementation quality in whole-classroom settings can be improved, this gives rise to optimism for the perspective of enhancing low-achieving students' reading comprehension by means of instructing reading strategies.

## Conclusion

Overall, reading-strategy programs can be beneficial in whole-classroom settings for low-achieving adolescents to improve reading comprehension. But, attention for implementation quality is crucial to create the best circumstances in which such programs can thrive. For schools to implement reading-strategy programs such as 'Nieuwsbegrip' it is beneficial to invest in teacher training, specifically focused on knowledge of *the nature and characteristics* of reading comprehension processes, *how to diagnose* the reading problems of their students, and *how to instruct* reading strategies and guide group work in whole-classroom settings.





## **REFERENCES**

**SAMENVATTING IN HET NEDERLANDS  
DANKWOORD (ACKNOWLEDGEMENTS)  
ABOUT THE AUTHOR | PUBLICATIONS  
AND PRESENTATIONS  
ICO DISSERTATION SERIES**



## References

\* Included in the meta-analysis

- \*Aaron, P.G., Joshi, R.M., Gooden, R., & Bentum, K.E. (2008). Diagnosis and treatment of reading disabilities based on the component model of reading: An alternative to the discrepancy model of LD. *Journal of Learning Disabilities*, 41, 67-84. doi:10.1177/0022219407310838
- Aarnoutse, C.A.J. & Van Leeuwe, J.F.J. (1988). Het belang van technisch lezen, woordenschat en ruimtelijke intelligentie voor begrijpend lezen [Importance of decoding, vocabulary and spatial intelligence for reading comprehension]. *Pedagogische Studiën [Pedagogical Studies]* 65, 49-59.
- Afflerbach, P., & Cho, B. (2009). Identifying and describing constructively responsive comprehension strategies in new and traditional forms of reading. In: S.E. Israel & G.G. Duffy (Eds.). *Handbook of research on reading comprehension*. New York, NY: Routledge.
- Afflerbach, P., Pearson P., & Paris, S.G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, 61(5), 364-373. doi: 10.1598/RT.61.5.1
- \*Alfassi, M. (2009). The efficacy of a dialogic learning environment in fostering literacy. *Reading Psychology*, 30(6), 539-25; 563.
- \*Allor, J. H., Mathes, P. G., Roberts, J. K., Cheatham, J. P., & Champlin, T. M. (2010). Comprehensive reading instruction for students with intellectual disabilities: Findings from the first three years of a longitudinal study. *Psychology in the Schools*, 47(5), 445-466.
- \*Andreassen, R., & Bråten, I. (2011). Implementation and effects of explicit reading comprehension instruction in fifth-grade classrooms. *Learning and Instruction*, 21(4), 520-537. doi: <https://doi.org/10.1016/j.learninstruc.2010.08.003>
- \*Berkeley, S. & Riccomini, P.J. (2011). QRAC-the-Code: A Comprehension Monitoring Strategy for Middle School Social Studies Textbooks. *Journal of Learning Disabilities*, 46(2) 154–165. doi: 10.1177/0022219411409412
- Berkeley, S., Scruggs, T. E., & Mastropieri, M. A. (2010). Reading comprehension instruction for students with learning disabilities, 1995—2006: A Meta-analysis. *Remedial and Special Education*, 31(6), 423-436.
- Boekaerts, M. & Simons, P.R.J. (1993). *Leren en instructie, Psychologie van de leerling en het leerproces*. [Learning and Instruction. Psychology of the student and the learning process] Assen: Dekker & van de Vegt.
- Bossers, B. (1991). On thresholds, ceilings and short-circuits: The relation between L1 reading, L2 reading and L2 knowledge. *AILA Review*, 8, 45-60.
- \*Calhoun, M.B. (2005). Effects of a peer-mediated phonological skill and reading comprehension program on reading skill acquisition for middle school students with reading disabilities. *Journal of Learning Disabilities*, 38(5), 424 - 433.
- \*Cantrell, S. C., Almasi, J. F., Carter, J. C., Rintamaa, M., & Madden, A. (2010). The impact of a strategy-based intervention on the comprehension and strategy use of struggling adolescent readers. *Journal of Educational Psychology*, 102(2), 257-280.
- CED-Groep. (2011). *Algemene Handleiding Nieuwsbegrip [Manual "Newswise"]*.
- \*Chambers Cantrell, S., Almasi, J.F., Rintamaa, M., & Carter, J.C. (2016). Supplemental reading strategy instruction for adolescents: A randomized trial and follow-up study. *The Journal of Educational Research*, 109(1), 7-26. doi: 10.1080/00220671.2014.917258

- Chiu, C. W. T. (1998, April). Synthesizing metacognitive interventions: What training characteristics can improve reading performance? Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Cohen, J. (1988). *Statistical power analysis for the social sciences*. Hillsdale, NJ: Erlbaum.
- Cooper, L., Hedges, H., & Valentine, J. (Eds.). (2009). *The handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation.
- De Boer, H., Donker, A.S., & Van der Werf, M.P.C. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review of Educational Research*, 84(4), 509-545. doi: 10.3102/0034654314540006
- 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*, 71, 531-559. doi: 10.1348/00070990115866
- De Milliano, I., Gelderen, A. van; Slegers, P. (2014). Types and sequences of self-regulated reading of low-achieving adolescents in relation to reading task achievement. *Journal of Research in Reading*. doi:10.1111\_1467-9817.12037
- De Milliano, I.I.C.M. (2013). *Literacy development of low-achieving adolescents: The role of engagement in academic reading and writing (Dissertation)*. Amsterdam: University of Amsterdam.
- \*Denton, C. A., Wexler, J., Vaughn, S., & Bryan, D. (2008). Intervention provided to linguistically diverse middle school students with severe reading difficulties. *Learning Disabilities Research & Practice*, 23(2), 79-89.
- \*DiCecco, V. M., & Gleason, M. M. (2002). Using graphic organizers to attain relational knowledge from expository text. *Journal of Learning Disabilities*, 35, 306-320
- Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research*, 61(2), 239-264. doi: 10.3102/00346543061002239
- Dole, J. A., Nokes, J. D., & Drits, D. (2009). Cognitive strategy instruction. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 347- 372). New York: Routledge.
- Donker, A. S., De Boer, H., Kostons, D., van Ewijk, C. D., & Van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, 11, 1-26.
- Droop, M., Van Elsäcker, W., Voeten, M.J.M., & Verhoeven, L. (2016). Long-Term Effects of Strategic Reading Instruction in the Intermediate Elementary Grades. *Journal of Research on Educational Effectiveness*, 9(1), 77-102. doi: 10.1080/19345747.2015.1065528
- Duffy, G. G. (1993). Teachers' progress toward becoming expert strategy teachers. *The Elementary School Journal*, 94(2), 109-120. doi: 10.1086/461754
- Duke N. K., Pearson P. D., Strachan S. L., Billman A. K. (2011). Essential elements of fostering and teaching reading comprehension. In Samuels S. J., Farstrup A. E. (Eds.), *What research has to say about reading instruction* (4th ed., pp. 51-93). Newark, DE: International Reading Association.
- Duke, N. K., & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In A. E. Farstrup & S. J. Samuels (Eds.). *What research has to say about reading instruction* (3rd ed., pp. 205-242). Newark, DE: International Reading Association.
- Durkin, D. (1978-1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 15, 481-553.

- \*Durukan, E. (2011). Effects of cooperative integrated reading and composition (CIRC) technique on reading-writing skills. *Educational Research and Reviews*, 6(1), 102-8; 109.
- Dutch Education Inspectorate (2008). *Basisvaardigheden taal in het voortgezet onderwijs: Resultaten van een Inspectieonderzoek naar taalvaardigheid in de onderbouw van het vmbo en praktijkonderwijs* [Basic language skills in secondary education: Results of an inspectorate study into language skills in the first two years of prevocational secondary education and practical training]. Utrecht: Dutch Education Inspectorate.
- Duval, S. and Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455-463. doi:10.1111/j.0006-341X.2000.00455.x
- Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A Synthesis of Reading Interventions and Effects on Reading Comprehension Outcomes for Older Struggling Readers. *Review of Educational Research*, 79(1), 262-300. doi: 10.3102/0034654308325998
- Egger, M., Smith, D.S., Schneider, M., Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629-634.
- \*Elbro, C. & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading*, 17(6), 435-452. doi: 10.1080/10888438.2013.774005
- \*Faggella-Luby, M., & Wardwell, M. (2011). RTI in a middle school: Findings and practical implications of a tier 2 reading comprehension study. *Learning Disability Quarterly*, 34(1), 35-15; 49.
- \*Fang, Z., & Wei, Y. (2010). Improving middle school students' science literacy through reading infusion. *Journal of Educational Research*, 103(4), 262-12; 273.
- Field, A. P. (2009). *Discovering statistics using SPSS: and sex and drugs and rock 'n' roll* (third edition). London: Sage publications.
- \*Fogarty, M., Oslund, E., Simmons, D., Davis, J., Simmons, L., Anderson, L., Clemens, N., Roberts, G. (2014). Examining the effectiveness of a multicomponent reading comprehension intervention in middle schools: A focus on treatment fidelity. *Educational Psychology Review*, 26(3), 425-449. doi: 10.1007/s10648-014-9270-6
- Gille, E., Loijens C., Noijons, J., & Zwitser, R. (2010). *Resultaten PISA-2009, Praktische kennis en vaardigheden van 15-jarigen* [PISA Results 2009, Practical knowledge and skills of 15-years old students]. Arnhem: Cito.
- \*Graves, A. W., Brandon, R., Duesbery, L., McIntosh, A., & Pyle, N. B. (2011a). The effects of tier 2 literacy instruction in sixth grade: Toward the development of a response-to-intervention model in middle school. *Learning Disability Quarterly*, 34(1), 73-14; 86.
- \*Graves, A. W., Duesbery, L., Pyle, N. B., Brandon, R. R., & McIntosh, A. S. (2011b). Two studies of tier II literacy development: Throwing sixth graders a lifeline. *Elementary School Journal*, 111(4), 641-21; 661.
- Guthrie, J. T., & Davis, M. H. (2003). Motivating struggling readers in middle school through an engagement model of classroom practice. *Reading & Writing Quarterly*, 19(1), 59-85.
- \*Guthrie, J. T., & Klauda, S. L. (2014). Effects of classroom practices on reading comprehension, engagement, and motivations for adolescents. *Reading Research Quarterly*, 49(4), 387-30; 416.
- \*Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., ... & Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, 96(3), 403.

- Guthrie, J.T., & Wigfield, A. (2000). Engagement and motivation in reading. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Volume III* (p. 403-422). New York: Erlbaum.
- \*Guthrie, J.T., McRae, A., Coddington, C.S., Lutz Klauda, S., Wigfield, A., & Barbosa, P. (2009). Impacts of comprehensive reading instruction on diverse outcomes of low- and high-achieving readers. *Journal of Learning Disabilities, 42*(3), 195-214. doi:10.1177/0022219408331039
- Hacker, D. J. & Tenent, A. (2002). Implementing reciprocal teaching in the classroom: Overcoming obstacles and making modifications. *Journal of Educational Psychology, 94*, 699-718. doi: 10.1037//0022-0663.94.4.699
- Hazenberg, S., & Hulstijn, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics, 17*, 145-163. doi: 10.1093/applin/17.2.145
- Hazenberg, S., & Hulstijn, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics, 17*, 145-163. doi: 10.1093/applin/17.2.145
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. Second Edition. New York: Routledge.
- Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: the role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*(1), 88-110. doi:10.1080/19345740802539325.
- \*Jitendra, A., Hoppes, M. K., & Xin, Y. (2000). Enhancing main idea comprehension for students with learning problems: The role of a summarization strategy and self- monitoring instruction. *The Journal of Special Education, 34*, 127-139.
- Johnston, P. (1984). Assessment in reading. In P.D. Pearson, R. Barr, M.L. Kamil & P. Mosenthal (Eds.) *Handbook of reading research* (pp 147-182). New York: Longman.
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441-480. doi: 10.1016/0010-0285(76)90015-3
- Just, M. A., & Carpenter, P. A. (2004). A theory of reading: From eye fixations to comprehension. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading*. Fifth edition (pp. 1182-1218). Newark, DE: International Reading Association.
- Karakostas, A., & Demetriadis, S. (2011). Enhancing collaborative learning through dynamic forms of support: the impact of an adaptive domain-specific support strategy. *Journal of Computer Assisted Learning, 27*(3), 243-258. doi: 10.1111/j.1365-2729.2010.00388.x
- Kelly, M., Moore, D.W., & Tuck, B.F. (2001). Reciprocal teaching in a regular primary school classroom. *Journal of Educational Research, 88*, 53-61. doi: 10.1080/00220671.1994.9944834
- \*Kim, A., Vaughn, S., Klingner, J. K., Woodruff, A. L., Reutebuch, C. K., & Kouzekanani, K. (2006). Improving the reading comprehension of middle school students with disabilities through computer-assisted collaborative strategic reading. *Remedial and Special Education, 27*(4), 235-249. doi: 0.1177/07419325060270040401
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182. doi:/10.1037/0033-295X.95.2.163
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- \*Klingner, J.K, Vaughn, S., Arguelles, M.E., Tejero Hughes, M., & Ahwee Leftwich, S. (2004). Collaborative strategic reading: "Real-World" lessons from classroom teachers. *Remedial and Special Education, 25*, 291-302. doi:10.1177/07419325040250050301

- Kordes, J., Bolsinova, M., Limpens, G., & Stolwijk, R. (2013). PISA resultaten 2012: Praktische kennis en vaardigheden van 15-jarigen. Nederlandse uitkomsten van het Programme for International Student Assessment (PISA) op het gebied van leesvaardigheid, wiskunde en natuurwetenschappen in het jaar 2012 [PISA results 2012: skills and knowledge of 15-years olds. Results of Dutch students in PISA in regard to reading, mathematics and science in the year 2012]. Arnhem, the Netherlands: Cito.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323. doi: 10.1016/0010-0285(74)90015-2
- Larsen, T, Samdal, O. (2007). Implementing Second Step: Balancing Fidelity and Program Adaptation. *Journal Of Educational And Psychological Consultation*, 17(1), 1-29.
- \*Lau, K. & Chan, D.W. (2007). The effects of cognitive strategy instruction on Chinese reading comprehension among Hong Kong low achieving students. *Reading & Writing*, 20, 833-857. doi:10.1007/s11145-006-9047-5
- \*Lederer, J.M. (2000). Reciprocal teaching of social studies in inclusive elementary classrooms. *Journal of Learning Disabilities* 33(1), 91-106.
- \*Lee, Y. (2014). Promise for Enhancing Children's Reading Attitudes Through Peer Reading: A Mixed Method Approach. *The Journal of Educational Research*, 107(6), 482-492. doi: 10.1080/00220671.2013.836469
- Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *Annals of the American Academy of Political and Social Science*, 587, 69-81. doi: 10.1177/000271620225079
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage publications.
- Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where the differences lie. *Journal of Research in Reading*, 32, 199-214. doi: 10.1111/j.1467-9817.2008.01389.x
- \*Lovett, M.W., Lacerenza, L., De Palma, M., & Frijters, J.C. (2012). Evaluating the efficacy of remediation for struggling readers in high school. *Journal of Learning Disabilities*, 45, 151-169. doi:10.1177/0022219410371678
- \*Lubliner, S., & Smetana, L. (2005). The effects of comprehensive vocabulary instruction on title I students' metacognitive word-learning skills and reading comprehension. *Journal of Literacy Research*, 37, 163-200. doi: 10.1207/s15548430jlr3702\_3
- \*Lucariello, J. M., Butler, A. G., & Tine, M. T. (2012). Meet the "reading rangers": Curriculum for teaching comprehension strategies to urban third graders. *Penn GSE Perspectives on Urban Education*, 9(2), 12.
- \*Lundberg, I. & Reichenberg, M. (2013). Developing Reading Comprehension Among Students With Mild Intellectual Disabilities: An Intervention Study. *Scandinavian Journal of Educational Research*, 57(1), 89-100, doi: 10.1080/00313831.2011.623179
- \*Mason, L.H., Davison, M.D., Hammer, C.S. et al. (2013). Knowledge, writing, and language outcomes for a reading comprehension and writing intervention. *Reading & Writing*, 26, 1133-1158. doi:10.1007/s11145-012-9409-0
- \*McCown, M., & Thomason, G. (2014). Informational text comprehension: Its challenges and how collaborative strategic reading can help. *Reading Improvement*, 51(2), 237-253.
- \*McKeown, M. G., Beck, I. L., & Blake, R. G. K. (2009). Rethinking reading comprehension instruction: a comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, 44, 218-253. doi;

- \*Miller, C. A., Darch, C. B., Flores, M. M., Shippen, M. E., & Hinton, V. (2011). Main idea identification with students with mild intellectual disabilities and specific learning disabilities: A comparison of explicit and basal instructional approaches. *Journal of Direct Instruction*, 11, 15-29.
- Milliano, I. de, Gelderen, A. van; Slegers, P. (2016). Types and sequences of self-regulated reading of low-achieving adolescents in relation to reading task achievement. *Journal of Research in Reading*, 39(2), 229-252. doi:10.1111\_1467-9817.12037
- Ministry of Education, Culture, & Science (2006). *The education system in the Netherlands 2006*. The Hague: Ministry of Education, Culture, & Science/Dutch Eurydice Unit.
- National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.
- \*Ng, C. C., Bartlett, B., Chester, I., & Kersland, S. (May 2013). Improving reading performance for economically disadvantaged students: Combining strategy instruction and motivational support. *Reading Psychology*, 34(3), 257-300.
- Oakhill, J. & Cain, K. (2007). Issues of causality in children's reading comprehension. In D.S.McNamara (Ed.), *Reading comprehension strategies: Theories, interventions and technologies*. (p. 47-72). New York: NY: Erlbaum.
- Okkinga, M., Van Gelderen, A.J.S, Van Schooten, E., Van Steensel, R & Slegers, P.J.C. (submitted). Effectiveness of reciprocal teaching for reading comprehension: A two-year study in a whole-classroom setting with low-achieving adolescents.
- Okkinga, M., Van Steensel, R., Van Gelderen, A.J.S., & Slegers, P.C.J. (2016). Effects of reciprocal teaching on reading comprehension of low-achieving adolescents. The importance of specific teacher skills. *Journal of Research in Reading*.
- \*Orbea, J. M. M., & Villabeitia, E. M. (2010). The teaching of reading comprehension and metacomprehension strategies. A program implemented by teaching staff. *Anales De Psicología*, 26(1), 112-122.
- Organisation for Economic Co-operation and Development [OECD] (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.
- Organisation for Economic Co-operation and Development [OECD] (2014). "Profile of student performance in reading", in *PISA 2012 Results: What students know and can do (Volume I, Revised edition, February 2014): Student performance in mathematics, reading and science*, OECD Publishing.
- Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: how oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*, 23, 189-208. doi: 10.1007/s11145-008-9159-1
- Palincsar, A. S., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175. doi: 10.1207/s1532690xci0102\_1
- Palincsar, A. S., Brown, A., & Martin, S.M. (1987). Peer interaction in reading comprehension instruction. *Educational Psychologist*, 22, 231-253. Doi: 10.1080/00461520.1987.9653051
- Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8, 293-316.
- 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 (6), 1239-1252.

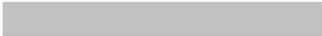
- Perfetti, C.A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In: M.J. Snowling & C. Hulme (Eds.). *The science of reading: A handbook*. Oxford: Blackwell. doi: 10.1002/9780470757642.ch13
- \*Ponce, H.C., López, M.J., Mayer, R.E. (2012). Instructional effectiveness of a computer-supported program for teaching reading comprehension strategies. *Computers & Education*, 59, 1170-1183. doi: 10.1016/j.compedu.2012.05.013
- Pressley, M. (2002). Metacognition and self-regulated learning. In A.E. Farstrup & S.J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 291-309). Newark, DE: International Reading Association.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.
- \*Proctor, C. P., Dalton, B., Uccelli, P., Biancarosa, G., Mo, E., Snow, C., et al. (2011). Improving comprehension online: Effects of deep vocabulary instruction with bilingual and monolingual fifth graders. *Reading and Writing*, 24(5), 517-544.
- \*Radcliffe, R., Caverly, D., Hand, J., & Franke, D. (2008). Improving reading in a middle school science classroom. *Journal of Adolescent and Adult Literacy*, 51(5), 398-11; 408.
- Raphael, T.E., George, M., Weber, C.M., & Nies, A. (2009). Approaches to teaching reading comprehension. In: S.E. Israel & G.G. Duffy (Eds.). *Handbook of research on reading comprehension*. New York, NY: Routledge.
- Rapp, D.N., Van den Broek, P., McMaster, K.L., Kendeou, P. & Espin, C.A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, 11(4), 289–312. doi: 10.1080/10888430701530417
- Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2009). *A user's guide to MLwiN*. Version 2.10. Bristol: University of Bristol, Centre for Multilevel Modelling.
- Raven, J., Raven, J.C., & Court, J.H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Section 1: General Overview. San Antonio, TX: Harcourt Assessment.
- \*Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction*, 22(4), 262-9; 270.
- \*Reis, S.M., Eckert, R.D., McCoach, D.B., Jacobs, J.K., & Coyne, M. (2008) Using enrichment reading practices to increase reading fluency, comprehension, and attitudes. *The Journal of Educational Research*, 101(5), 299-315. doi: 10.3200/JOER.101.5.299-315
- \*Reis, S.M., McCoach, D.B., Little, C.A., Muller, L.M., & Kaniskan, R.B. (2011). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal*, 48, 462-501. doi:10.3102/0002831210382891
- Roberts, G., Fletcher, J.M., Stuebing, K.K., Barth, A.E., & Vaughn, S. (2013). Treatment effects for adolescent struggling readers: An application of moderated mediation. *Learning and Individual Differences*, 23 (1), 10-21. doi: <http://dx.doi.org/10.1016/j.lindif.2012.09.008>
- \*Rojas-Drummond, S., Mazón, N., Littleton, K., & Vélez, M. (2014). Developing reading comprehension through collaborative learning. *Journal of Research in Reading*, 37(2), 138–158. doi: 10.1111/j.1467-9817.2011.01526.x
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64, 4, 479-530. doi: 10.3102/0034654306400447
- Rouet, J.F. & Britt, M.A. (2011). Relevance processes in multiple document comprehension. In: M.T. McCrudden, J.P. Magliano, & Schraw, G. (Eds.), *Text relevance and learning from text* (pp 19-52). USA: Information Age Publishing Inc.

- Rumelhart, D. E. (2004). Toward an interactive model of reading. In R. B. Ruddell, & N. J. Unrau (Eds.), *Theoretical models and processes of reading*. Fifth edition. Newark, DE: International Reading Association.
- Samuels, S. J. (2004). Toward a theory of automatic information processing in reading, revisited. In R. B. Ruddell, & N. J. Unrau (Eds.), *Theoretical models and processes of reading*. Fifth edition (p. 1127–1148). Newark, DE: International Reading Association.
- Scammacca, N., Roberts, G., Vaughn, S., & Stuebing, K.K. (2015). A meta-analysis of interventions for struggling readers in grades 4-12:1980-2011. *Journal of Learning Disabilities*, 48(4), 369-390. doi: 10.1177/0022219413504995
- Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. Portsmouth, NH: Center on Instruction.
- Schaffner, E., Philipp, M., and Schiefele, U. (2016) Reciprocal effects between intrinsic reading motivation and reading competence? A cross-lagged panel model for academic track and nonacademic track students. *Journal of Research in Reading*, 39, 19–36. doi: 10.1111/1467-9817.12027
- Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading*, 3(3), 257–279. doi: 10.1207/s1532799xssr0303\_4
- \*Schünemann, N., Spörer, N., Brunstein, J.C. (2013). Integrating self-regulation in whole-class reciprocal teaching: A moderator–mediator analysis of incremental effects on fifth graders’ reading comprehension. *Contemporary Educational Psychology*, 38, 289–305. Doi: 10.1016/j.cedpsych.2013.06.002
- Sencibaugh, J.M. (2007). Meta-analysis of reading comprehension interventions for students with learning disabilities: Strategies and implications. *Reading Improvement*, 44(10), 6-22.
- Seymour, J.R. & Osana, H.P. (2003). Reciprocal teaching procedures and principles: two teachers’ developing understanding. *Teaching and Teacher Education*, 19, 325-344. doi: 10.1016/S0742-051X(03)00018-0
- Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Silfhout, G., Evers-Vermeul, J., Mak, W. M., & Sanders, T. J. M. (2014). Connectives and layout as processing signals: How textual features affect students’ processing and text representation. *Journal of Educational Psychology*, 4, 1036-1048. doi: 10.1037/a0036293
- \*Simmons, D., Fogarty, M., Oslund, E.L., Simmons, L., Hairrell, A., Davis, J., Anderson, L., Clemens, N, Vaughn, S., Roberts, G., Stillman, S., & Fall, A. (2014). Integrating content knowledge-building and student-regulated comprehension practices in secondary English arts classes. *Journal of Research on Educational Effectiveness*, 7, 309-330. doi: 10.1080/19345747.2013.836766
- \*Simmons, D., Hairrell, A., Edmonds, M., Vaughn, S., Larsen, R., Willson, V., et al. (2010). A comparison of multiple-strategy methods: Effects on fourth-grade students’ general and content-specific reading comprehension and vocabulary development. *Journal of Research on Educational Effectiveness*, 3(2), 121-36; 156.
- Slavin, R. E., Cheung, A., Groff, C., & Lake, C. (2008). *Effective reading programs for middle and high schools: A best-evidence synthesis*. *Reading Research Quarterly*, 43(3), 290-322
- Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). *Effective reading programs for the elementary grades: A best-evidence synthesis*. *Review of Educational Research*, 79(4), 1391-1466.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

- \*Souvignier, E. & Mokhlesgerami, J. (2006). Using self-regulation as a framework for implementing strategy-instruction to foster reading comprehension. *Learning & Instruction*, 16, 57–71. doi:10.1016/j.learninstruc.2005.12.006
- \*Spörer, N., & Brunstein, J. C. (2009a). Fostering the reading comprehension of secondary school students through peer-assisted learning: Effects on strategy knowledge, strategy use, and task performance. *Contemporary Educational Psychology*, 34(4), 289–297.
- \*Spörer, N., Brunstein, J.C., & Kieschke, U. (2009). Improving students' reading comprehension skills: Effects of strategy instruction and reciprocal teaching. *Learning and Instruction*, 19, 272–286. doi: 10.1016/j.learninstruc.2008.05.003
- \*Sung, Y., Chang, K., & Huang, J. (2008). Improving children's reading comprehension and use of strategies through computer-based strategy training. *Computers in Human Behavior*, 24(4), 1552–1571.
- Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2011). Intervention fidelity in special and general education research journals. *The Journal of Special Education*, 47(1), 3–13. doi: 10.1177/002246691141951.
- Swanson, H. (1999). Reading Research for Students with LD: A meta-analysis of intervention outcomes. *Journal of learning disabilities*, 32(6), 504–532. doi: 10.1177/002221949903200605
- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12, 257–285. doi:10.1207/s15516709cog1202\_4
- Taillefer, G. F. (1996). L2 Reading ability: Further insight into the short-circuit hypothesis. *The Modern Language Journal*, 80, 461–477. doi:10.1111/j.1540-4781.1996.tb05466.x
- Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P. & Rapp, D. (2009). Simple but complex: components of the simple view of reading across grade levels. *Journal of research in reading*, 32(4), 383–401. doi: 10.1111/j.1467-9817.2009.01401.x
- Torgesen, J.K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, 15(10), 55–64. doi: 10.1207/SLDRP1501\_6
- Trapman, M., Gelderen, A. van, Schooten, E. van & Hulstijn, J. (2017). Reading comprehension level and development in native and language minority adolescent low achievers: Roles of linguistic and metacognitive knowledge and fluency. *Reading and Writing Quarterly*, 33 (3), 239–257. DOI: 10.1080/10573569.2016.1183541
- Trapman, M.J.W. (2015). Reading and writing development of low-achieving adolescents. The roles of linguistic knowledge, fluency, and metacognitive knowledge. Amsterdam: University of Amsterdam.
- Trapman, M.J.W., Van Gelderen, A., Van Steensel, R., Van Schooten, E. and Hulstijn, J. (2014). Linguistic knowledge, fluency and meta-cognitive knowledge as components of reading comprehension in adolescent low achievers: differences between monolinguals and bilinguals. *Journal of Research in Reading*, 37, 3–21. doi: 10.1111/j.1467-9817.2012.01539.x
- Van Gelderen, A., Schoonen, R., De Glopper, K., Hulstijn, J., Simis, A. Snellings, P., & Stevenson, M. (2003). Roles of linguistic knowledge, metacognitive knowledge and processing speed in L3, L2 and L1 reading comprehension; a structural equation modelling approach. *International Journal of Bilingualism*, 7(1), 7–25. doi: 10.1177/13670069030070010201
- Van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P. & Stevenson, M. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: a componential analysis, *Journal of Educational Psychology* 96(1), 19–30. doi: 10.1037/0022-0663.96.1.19

- Van Gelderen, A., Schoonen, R., De Glopper, K., Hulstijn, J., Snellings, P., Simis, A. et al. (2003). Roles of linguistic knowledge, metacognitive knowledge and processing speed in L3, L2 and L1 reading comprehension: A structural equation modeling approach. *International Journal of Bilingualism*, 7(1), 7–25. doi:10.1177/13670069030070010201
- Van Gelderen, A., Schoonen, R., Stoel, R. D., De Glopper, K., & Hulstijn, J. (2007). Development of adolescent reading comprehension in Language 1 and Language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology*, 99(3), 477-491. doi: 10.1037/0022-0663.99.3.477
- Van Silfhout, G., Evers-Vermeul, J., Mak, W. M., & Sanders, T.J.M. (2014). Connectives and layout as processing signals: How textual features affect students' processing and text representation. *Journal of Educational Psychology*, 106(4), p. 1036-1048. doi: 10.1037/a0036293
- Van Steensel, R., Oostdam, R., & Van Gelderen, A. (2013). Assessing reading comprehension in adolescent low achievers: subskills identification and task specificity. *Language Testing*, 30(1), 3-21. doi: 10.1177/0265532212440950
- Van Steensel, R., Oostdam, R., Van Gelderen, A., & Van Schooten, E. (2016). The role of word decoding, vocabulary knowledge and meta-cognitive knowledge in monolingual and bilingual low-achieving adolescents' reading comprehension. *Journal of Research in Reading*, 39, 312–329. doi: 10.1111/1467-9817.12042
- Vaughn, S., Roberts, G., Klingner, J. K., Swanson, E. A., Boardman, A., Stillman-Spisak, S. J., et al. (2013). Collaborative strategic reading: Findings from experienced implementers. *Journal of Research on Educational Effectiveness*, 6, 137–163. doi: 10.1080/19345747.2012.741661
- Veenman, M.V.J., Van Hout-Wolters, B.H.A.M. & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition & Learning*, 1(1), 3-14. doi: 10.1007/s11409-006-6893-0
- Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, 22, 407-423. doi: 10.1002/acp.1414
- Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual differences for self-regulating task-oriented reading activities. *Journal of Educational Psychology*, 102(4), 817. doi: http://dx.doi.org/10.1037/a0020062
- \*Wijekumar, K.K., Meyer, B.J.F., & Lei, P. (2012). Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development*, 60, 987–1013. doi: 0.1007/s11423-012-9263-4
- \*Wijekumar, K.K., Meyer, B.J.F., Pui-Wa Lei, Yu-Chu Lin, Johnson, L.A., Spielvogel, J.A., Shurmatz, K.M., Ray, M., & Cook, M. (2014). Multisite randomized controlled trial examining intelligent tutoring of structure strategy for fifth-grade readers. *Journal of Research on Educational Effectiveness*, 7(4), 331-357. doi: 10.1080/19345747.2013.853333
- Woolley, G. (2011). *Reading comprehension: Assisting children with learning difficulties*. Dordrecht, The Netherlands: Springer International.





REFERENCES

**SAMENVATTING IN HET NEDERLANDS**

DANKWOORD (ACKNOWLEDGEMENTS)

ABOUT THE AUTHOR | PUBLICATIONS

AND PRESENTATIONS

ICO DISSERTATION SERIES



## Samenvatting in het Nederlands

### Achtergrond

Veel leerlingen op de middelbare school hebben moeite met begrijpend lezen (bijvoorbeeld Organisation for Economic Co-operation and Development ([OECD], 2003; OECD, 2014). Ook in Nederland zijn veel jongeren onvoldoende in staat teksten te lezen op het vereiste niveau. Met name vmbo-leerlingen hebben moeite om teksten te begrijpen. Zij hebben, een gemiddelde leerachterstand van twee jaar op het gebied van begrijpend lezen ten opzichte van leerlingen die instromen op de havo en vwo, (Inspectie, 2006). Dit gegeven is zorgelijk: hoe kunnen wij deze leerlingen helpen om beter te leren begrijpend lezen?

Uit onderzoek is bekend dat leerlingen die zwak zijn in begrijpend lezen, vaak problemen hebben met strategisch lezen. Goede strategische lezers controleren hun begrip van de tekst. Zij doen dit onder andere door te voorspellen waar een tekst over gaat en deze voorspellingen tijdens het lezen te controleren en waar nodig aan te passen. Daarnaast activeren zij hun voorkennis rondom het onderwerp van de tekst en integreren deze telkens met nieuw opgedane kennis uit de tekst (Paris, Lipson, & Wixson, 1983). Het aanleren van dergelijke leesstrategieën aan vmbo-leerlingen zou dus het begrijpend lezen kunnen bevorderen.

Sinds de jaren 80 van de vorige eeuw is er een toenemende belangstelling voor leesprogramma's die dergelijke leesstrategieën aanleren. Het onderliggende idee is dat begrijpend lezen een complex proces is waarbij de lezer actief de tekst probeert te representeren: het zogenoemde 'situatie model' (Kintsch, 1999; 1998). Om dit te laten slagen, is het van belang dat lezers de nieuw opgedane informatie uit de tekst kunnen relateren aan hun eigen voorkennis. Ook besteden moderne programma's aandacht aan het bevorderen van de woordenschat van leerlingen bij het lezen. Woordenschat is een belangrijke voorspeller van begrijpend lezen, ook in groepen van zwakke lezers (e.g. Trapman et al., 2014; Van Steensel et al., 2016). In onderzoek naar begrijpend lezen is het dus belangrijk om rekening te houden met woordenschat.

Een leesstrategie is een mentale tool die de lezer gebruikt om het leesproces te controleren te sturen of te bevorderen (Afflerbach & Cho, 2009). Het gebruik van leesstrategieën is een doelgerichte poging om een tekst te begrijpen (Afflerbach, Pearson, & Paris, 2008; Dole, Nokes, & Drits, 2009). Voorbeelden van leesstrategieën zijn: leesdoelen specificeren, de activatie van relevante voorkennis, visualiseren, samenvatten, vragen stellen over de tekst en voorspellen waar een tekst over gaat (Duke, Pearson, Strachan, & Billman, 2011; Palincsar & Brown, 1984).

Een van de bekendste methoden voor het aanleren van leesstrategieën is 'reciprocal teaching' (vertaald als: 'rolwisselend leren'). 'Reciprocal teaching' bestaat uit een set van drie principes: (a) directe instructie van verschillende leesstrategieën, b) docenten modelen het

gebruik van leesstrategieën door hardop te denken tijdens het lezen van een tekst, waarna deze rol wordt overgedragen op leerlingen; (c) leerlingen worden aangemoedigd om in groepjes de betekenis van de tekst te achterhalen (Palincsar & Brown, 1984; Rosenshine & Meister, 1994). Veel studies hebben positieve effecten laten zien van deze methode (e.g. Kelly, Moore, & Tuck, 2011; Palincsar & Brown, 1984, Rosenshine & Meister, 1994, Spörer, Brunstein, & Kiescke, 2009).

Toch zijn er ook indicaties dat ‘reciprocal teaching’ niet in alle gevallen even effectief werkt. Zo werden de eerste studies naar ‘reciprocal teaching’ verricht in kleine groepjes leerlingen die uit de klas werden gehaald waarbij zij ondersteund en begeleid werden door een van de onderzoekers. Implementatie van de methode door reguliere docenten in reguliere klassen is veel lastiger te realiseren dan in zulke kleine gecontroleerde contexten. In verschillende kwalitatieve studies wordt gedemonstreerd dat docenten in reguliere klassen problemen ervaren bij de implementatie van interventies gericht op het aanleren van leesstrategieën (Duffy, 1993; Seymour & Osana, 2003; Hacker & Tenent, 2002). Zo vonden docenten het lastig om de didactische principes van ‘reciprocal teaching’ toe te passen en om leerlingen aan te zetten tot strategisch lezen. Docenten vertelden verder dat hun leerlingen grote moeilijkheden ondervonden bij het samenwerkend leren, waarbij leerlingen gevraagd werd om per toerbeurt hardop denkend te lezen leesstrategieën toe te passen. Meer recent werd strategie-instructie in reguliere klassen bekritiseerd omdat te veel aandacht voor het uitvoeren van leesstrategieën kan verhinderen dat leerlingen hun aandacht richten op het construeren van een coherent situatie model van de tekst (McKeown, Beck, & Blake, 2009).

Het doel van deze dissertatie is om meer inzicht te krijgen in hoe het aanleren van leesstrategieën aan vmbo-leerlingen in reguliere klassen, het begrijpend lezen kan bevorderen. Enerzijds is er een experiment uitgevoerd in het vmbo waarbij het effect van strategie-instructie op begrijpend lezen is gemeten. Anderzijds is een meta-analyse uitgevoerd, gericht op recente effectstudies naar strategie-instructie die plaats vonden in reguliere klassen, waarbij de gevonden effecten op het begrijpend lezen zijn samengevat.

### **Context van het experiment: ‘Nieuwsbegrip’**

De hoofdstukken 2, 3 en 4 van deze dissertatie doen verslag van de resultaten van een tweejarige experimentele studie die de effecten van een veel gebruikt leesprogramma, genaamd ‘Nieuwsbegrip’, meet in de context van reguliere vmbo klassen (basis/kader).

De methode Nieuwsbegrip is gebaseerd op de principes van ‘reciprocal teaching’. Het aanleren van vijf leesstrategieën staat centraal (voorspellen, samenvatten, vragen stellen, ophelderen van onduidelijkheden en verbanden leggen), waarbij het voordoen van de strategieën door de docent en het hardop denkend oefenen van de leesstrategieën door leerlingen in kleine groepjes belangrijke aspecten zijn. Nieuwsbegrip is ontwikkeld door

de CED-Groep in Rotterdam en wordt vooral gebruikt in basisscholen. Ongeveer 80% van de basisscholen heeft een licentie van Nieuwsbegrip. Het programma wordt echter ook vaak ingezet op middelbare scholen in de onderbouw. Elke week wordt één nieuwe les ontwikkeld door een ontwikkelteam, waarbij de inhoud van de teksten altijd een recent nieuwsitem belichten. In elke les staat één van de strategieën centraal. Nieuwsbegrip biedt teksten en lessen aan op zes verschillende leesniveaus, variërend van leerlingen in het speciaal onderwijs, tot leerlingen in havo/vwo leerjaar 3/4. In het onderhavige experiment is gebruikt gemaakt van teksten op niveau 'B', dit zijn teksten geschikt voor vmbo-leerlingen in leerjaar 1-2.

Tien vmbo-scholen, elk met twee vergelijkbare brugklassen (een interventie- en een controleklas), hebben meegedaan aan deze studie ( $N_{\text{leerlingen}}=382$ ). De klassen ( $N=20$ ) zijn samen met hun docent, aselect toegewezen aan een van beide condities. Tijdens de interventie werden de docenten intensief getraind en gecoacht om te leren werken met de methode. Controledocenten werkten op hun eigen manier met de reguliere methode Nederlands. Tijdens het onderzoek is er vier keer een begrijpend leestoets afgenomen: een voormeting aan het begin van het eerste schooljaar, een eerste nameting aan het einde van dat jaar, een retentiemeting na de zomervakantie en een laatste nameting aan het einde van het tweede schooljaar. Daarnaast zijn ook woordenschat, metacognitieve kennis en IQ gemeten zodat hiervoor gecontroleerd kon worden in de analyses. Twee keer per jaar vonden klasobservaties plaats waarin de implementatiekwaliteit werd gemeten op drie gebieden (strategie-instructie, 'modeling', groepswork). In de controleklassen werden vergelijkbare observaties gedaan waarin gemeten werd in hoeverre de controledocenten vergelijkbare instructie aanboden, ook al waren zij niet getraind om te werken met de interventie. Na afloop van de interventie vonden exitinterviews plaats met de Nieuwsbegrip docenten die twee jaar hebben meegewerkt aan het onderzoek.

### Meta-analyse

De resultaten van de experimentele studie hebben geleid tot het uitvoeren van de meta-analyse. Een meta-analyse vat de statistische resultaten samen van verschillende empirische studies die een vergelijkbare onderzoeksvraag beantwoorden. Het doel van deze meta-analyse was om de effecten van strategie-instructie interventies samen te vatten. Bovendien is nagegaan of effecten verklaard worden door kenmerken van de interventie (zoals duur van de interventie), door kenmerken van de steekproef (bijvoorbeeld leeftijd van de leerlingen) en door kenmerken van studie-opzet (zoals het wel of niet randomiseren van de steekproef).

Er is een literatuursearch gedaan met gepubliceerde artikelen vanaf het jaar 2000. Uiteindelijk werden 52 studies geïdentificeerd voor inclusie, met een totaal van 125 experimentele vergelijkingen die meegenomen werden, verdeeld over 5 typen afhankelijke

variabelen: gestandaardiseerde toetsen van begrijpend lezen, door onderzoekers ontwikkelde toetsen van begrijpend lezen, vaardigheid in het toepassen van leesstrategieën, kennis van leesstrategieën en zelf-gerapporteerd strategiegebruik. Daarnaast werd onderscheid gemaakt tussen toetsen die direct waren afgenomen na afloop van de interventie en toetsen die na enige tijd na afloop van de interventie werden afgenomen. In de analyses is ook gekeken naar modererende effecten van leerling-, interventie- en studie kenmerken.

### **Belangrijkste bevindingen**

De hoofdstukken 2 en 3 beschrijven het effect van het programma Nieuwsbegrip op begrijpend lezen, na één en twee jaar interventie. In beide hoofdstukken is implementatiekwaliteit als modererende variabele meegenomen in de bepaling van het hoofdeffect. Er is geen hoofdeffect gevonden van Nieuwsbegrip op begrijpend lezen in beide interventie-jaren; de experimentele groep groeide niet meer in begrijpend lezen dan de controlegroep. Wel was er in het eerste jaar (Hoofdstuk 2) een significant moderatoreffect van strategie-instructie. Als de strategie-instructie door de docent uitgebreider was, had dit een positief effect op begrijpend lezen van de leerlingen in de interventie conditie, maar niet in de controle conditie. Ditzelfde moderator effect was niet meer aanwezig na twee jaar interventie. 'Voordoelen van leesstrategieën had na twee jaar wel een significant moderatoreffect (Hoofdstuk 3). Dit effect betekent dat de groei van begrijpend lezen in de interventie conditie afhankelijk was van hoe goed er voorgedaan werd (zowel door docenten als leerlingen). Als er beter dan gemiddeld voorgedaan werd, waren er geen verschillen tussen de interventie- en controle conditie in groei in begrijpend lezen. Als er slechter dan gemiddeld voorgedaan werd, leek de interventie conditie in het nadeel te zijn, aangezien groei in begrijpend lezen dan achterliep op de groei in de controle conditie.

In hoofdstuk 4 is gekeken naar de invloed van woordenschat op de interventie en de groei in begrijpend lezen. Uit deze studie bleek dat, afhankelijk van het niveau van woordenschat, de leerlingen in de interventie een andere groei in begrijpend lezen doormaken dan leerlingen in de controle conditie. Als leerlingen in de interventie conditie een lage woordenschat hadden, was hun groei in begrijpend lezen lager dan dat van vergelijkbare leerlingen in de controle conditie. Als leerlingen in de interventie conditie juist een hoge woordenschat hadden, was er geen verschil met leerlingen met een hoge woordenschat in de controle conditie.

De meta-analyse beschreven in Hoofdstuk 5 laat zien dat interventies gericht op het aanleren van leesstrategieën in klassensituaties bevorderlijk kunnen werken om het niveau van begrijpend lezen te verhogen. Effecten waren gemiddeld significant hoger voor toetsen die ontwikkeld waren door onderzoekers (Cohen's  $d = 0.431$ , een klein effect) dan voor gestandaardiseerde toetsen van begrijpend lezen (Cohen's  $d = 0.186$ , een erg klein

effect). Een matige effectgrootte werd gevonden voor vaardigheid in het toepassen van leesstrategieën (Cohen's  $d = 0.786$ ), terwijl voor zelf-gerapporteerd strategiegebruik een klein effect werd gevonden (Cohen's  $d = 0.358$ ). Voor de effecten op langere termijn vonden we een zeer klein effect voor gestandaardiseerde toetsen (Cohen's  $d = 0.167$ ), terwijl er een groot effect was voor door onderzoekers ontwikkelde toetsen van begrijpend lezen (Cohen's  $d = 0.947$ ).

We vonden verschillende moderatoreffecten op begrijpend lezen. Zo werden er grotere effecten gevonden voor studies waarin de onderzoekers de instructie aan leerlingen gaven, dan voor studies waarin docenten de instructie gaven. Leesstrategie-interventies sorteerden grotere effecten voor relatief zwakke leerlingen ('low-achieving') dan voor gemiddelde leerlingen. Daarnaast vonden we grotere effecten voor interventies die plaats vonden in de bovenbouw lagere school en onderbouw middelbare school ('grades 6-8') dan voor jongere en oudere leerlingen. Er is ook gekeken naar de verschillende leesstrategieën die aangeboden werden in de interventies. De moderatoreffecten hiervan zijn echter lastig te interpreteren, aangezien de meeste interventies werken met een pakketje van aangeboden leesstrategieën. Het is dus niet mogelijk om het effect van één leesstrategie te bepalen. Wel kan bepaald worden wat de extra waarde is van een bepaalde strategie, bovenop het pakketje. Van alle leesstrategieën die we gedefinieerd hebben, heeft alleen het 'bepalen van leesdoelen' een extra positief effect op begrijpend lezen.

## Conclusies en discussie

De resultaten uit dit proefschrift laten zien dat het lastig is om positieve effecten op begrijpend lezen van interventies gericht op het aanleren van leesstrategieën in reguliere klassen te vinden. We vonden geen hoofdeffecten van 'Nieuwsbegrip' na twee jaar interventie en we vonden een heel klein effect op gestandaardiseerde toetsen van begrijpend lezen in de meta-analyse. Verklaringen hiervoor kunnen gezocht worden in de complexiteit van de klassensituatie, de lange duur van de interventie, het lage woordenschat niveau van de leerlingen, en de keuze voor een gestandaardiseerde toets van begrijpend lezen.

In de originele 'reciprocal teaching' methode werden groepjes leerlingen uit de klas gehaald waarna zij onder begeleiding van een tutor aan de slag gingen met het lezen van teksten. Dit geeft de tutor een grote mate van controle op het gedrag van de leerlingen en ook een beter zicht op welke begeleiding de leerlingen nodig hebben. In de klassensituatie, zoals gebruikt in het experiment, werden docenten geacht 4-5 groepjes leerlingen in het oog te houden en te begeleiden. Dit betekent dat de docenten veel minder controle hadden op de samenwerking binnen de groepjes, maar ook veel minder tijd om per groepje te besteden aan begeleiding. Dit betekent dat het groepswerk, zoals bedoeld in de originele methode, veel lastiger te realiseren is in reguliere klassen, waardoor positieve resultaten

op begrijpend lezen moeilijker te behalen zijn. De observaties bevestigden dit beeld: veel docenten uit de experimentele conditie ervaarden moeilijkheden in het begeleiden van de groepjes en docenten gaven aan in de exitinterviews dat zij meer coaching op het gebied van groepswork op prijs hadden gesteld.

De lange duur van de interventie, twee schooljaren lang, heeft mogelijk ook een rol gespeeld in de gevonden resultaten. Zo vonden de docenten het moeilijk om de leerlingen voor zo'n lange termijn gemotiveerd te houden. Verschillende factoren kunnen hier een rol hebben gespeeld. Zo werd de moeilijkheidsgraad van de teksten niet aangepast gedurende de interventie, wat de motivatie van de leerlingen niet ten goede kwam. Ondanks dat we docenten gevraagd hebben om met het volgende niveau teksten te werken in de klas (waarbij ze een overgang zouden maken van type 'B' naar type 'C' teksten), rapporteerden de docenten dat de C-teksten te lang waren en te veel moeilijke woorden te bevatten om te kunnen gebruiken in de lessen. Een andere factor die mee kan hebben gespeeld in de motivatie, is dat het programma 'Nieuwsbegrip' tamelijk rigide is. Het type teksten was telkens hetzelfde ('nieuws'), de cyclus van leesstrategieën die aan bod kwamen bleef hetzelfde (in elke les was de focus op één van de vijf leesstrategieën) en de activiteiten binnen een les werden ook telkens herhaald op dezelfde manier. Deze herhaling kan negatieve gevolgen hebben gehad voor de motivatie van de leerlingen.

Een ander probleem met de lange duur van de interventie heeft te maken met uitval van docenten. Uitval van docenten vond grotendeels plaats in de overgang van het eerste jaar naar het tweede jaar en had uiteenlopende redenen (ziekte, zwangerschap, verandering van baan). Ondanks dat we geprobeerd hebben om deze docenten te vervangen en de vervangers te trainen, heeft dit mogelijk effect gehad op de kwaliteit van de implementatie, en op de groei in begrijpend lezen van de leerlingen. Dit is in het bijzonder het geval voor scholen die moeite hadden om vervangende docenten te vinden waardoor leerlingen veel lessen hebben gemist.

Een andere verklaring is te vinden in het lage woordenschatniveau van de leerlingen. Zoals in hoofdstuk 4 is aangetoond, waren de leerlingen met een onder gemiddeld niveau van woordenschat in de interventie conditie in het nadeel ten opzichte van vergelijkbare leerlingen uit de controle conditie. En de leerlingen in de interventie conditie met een bovengemiddelde score op woordenschat, scoorden niet hoger op begrijpend lezen dan vergelijkbare leerlingen uit de controle conditie. Dit zou kunnen betekenen dat de interventie niet voldoende was afgestemd op het niveau van woordenschat van de vmbo-leerlingen (basis-kader) die meededen aan de studie. Het zou bijvoorbeeld kunnen dat leerlingen met de minste woordenschat niet voldoende woorden kenden uit de teksten om de strategie 'ophelderend van onduidelijkheden' effectief te kunnen toepassen. Het lijkt er dus op dat woordenschat een belangrijke voorwaarde is voor het succesvol toepassen van leesstrategieën door vmbo-leerlingen.

Een laatste verklaring voor de gevonden effecten kan gezocht worden in de keuze voor een gestandaardiseerde toets van begrijpend lezen. Uit de meta-analyse komt duidelijk naar voren dat effecten van strategie-instructie interventies op gestandaardiseerde toetsen van begrijpend lezen veel kleiner zijn dan op door onderzoekers ontwikkelde toetsen. Dit verschil werd zelfs groter voor toetsen die pas een tijd na de interventie werden afgenomen. In het experiment werd gebruik gemaakt van een algemene toets voor begrijpend lezen, speciaal ontwikkeld voor leerlingen in het vmbo, de SALSA-toets. Deze toets is vergelijkbaar met gestandaardiseerde toetsen, omdat hij zich niet specifiek richt op de toepassing van geleerde strategieën en bovendien verschilt in de typen teksten die aangeboden worden in vergelijking met de teksten uit Nieuwsbegrip.

### **Aanbevelingen voor de onderwijspraktijk**

De uitkomsten uit dit proefschrift hebben ook implicaties voor het onderwijs. Ten eerste zijn er aanbevelingen voor lerarenopleidingen. Een belangrijke uitkomst is dat docenten meer moeite hebben met het succesvol implementeren van een interventie gericht op het aanleren van leesstrategieën dan onderzoekers. Dit betekent dat er meer aandacht zou moeten zijn binnen het programma voor leraren in opleiding *hoe* docenten leesstrategieën zouden moeten aanleren (National Reading Panel, 2000). Hierbij zou naast didactische uitleg, ook aandacht moeten zijn voor het proces van begrijpend lezen en hoe leesstrategieën kunnen ingrijpen op dit proces.

Hoewel de uitkomsten van dit proefschrift laten zien dat het lastig is om in een reguliere klas ‘reciprocal teaching’ van de grond te krijgen, kunnen er toch aanbevelingen worden gedaan hoe docenten dit kunnen aanpakken. In de eerste plaats is het belangrijk dat docenten de leerlingen kunnen identificeren die het meeste baat hebben bij begeleiding van hun leesproces. Docenten zouden bijvoorbeeld bij één groepje leerlingen kunnen zitten terwijl de rest individueel aan het werk is, en dit groepje intensief kunnen begeleiden. Om meer zicht te hebben op het begrijpend leesproces van de leerlingen, zou de docent de leerlingen in het groepje kunnen vragen om hardop te denken tijdens het lezen.

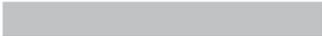
Aangezien het gebruik van leesstrategieën een manier is om teksten beter te begrijpen, is het belangrijk dat docenten zich ervan bewust zijn dat dergelijke strategieën een middel, en niet het doel vormen. Het flexibel kunnen inzetten van de verschillende leesstrategieën is daarom van belang. Dit betekent dat docenten hun leerlingen zouden moeten helpen de strategie te kiezen waar de leerling op dat moment het meeste baat bij heeft om de tekst beter te begrijpen.

Tenslotte is het belangrijk dat docenten zich ervan bewust zijn dat het aanleren van leesstrategieën niet een snelle oplossing is om scores op gestandaardiseerde toetsen te verhogen. Desondanks kan het aanleren van leesstrategieën een nuttige bijdrage leveren aan begrijpend leesonderwijs. De resultaten van de meta-analyse laten namelijk zien dat

effecten op begrijpend lezen blijvend zijn, en zelfs groter worden na verloop van tijd bij niet-gestandaardiseerde toetsen.

Concluderend zijn de resultaten van dit proefschrift positief voor het onderwijs in leesstrategieën. Enerzijds zijn er duidelijke indicaties hoe de implementatiekwaliteit van dit type onderwijs kan worden verbeterd. Anderzijds zijn er duidelijke indicaties dat leesstrategie-onderwijs leidt tot beter tekstbegrip. Dit is een bemoedigend gegeven voor de toekomst van begrijpend leesonderwijs voor vmbo (basis/kader) leerlingen.





REFERENCES

SAMENVATTING IN HET NEDERLANDS

**DANKWOORD (ACKNOWLEDGEMENTS)**

ABOUT THE AUTHOR | PUBLICATIONS

AND PRESENTATIONS

ICO DISSERTATION SERIES



## Dankwoord

Toen ik begon met studeren aan de Universiteit van Humanistiek had ik nooit kunnen bedenken dat ik later onderzoeker zou worden. Onderzoek doen was in mijn ogen saai en lastig. Lastig is het inderdaad af en toe gebleken, maar saai nooit! Pas toen ik tijdens mijn studie psychologie in Leiden in aanraking kwam met statistiek, begon onderzoek te leven voor mij. De docenten die daaraan hebben bijgedragen, hebben mij langzaam maar zeker op het pad gezet dat promoveren heet. Ik noem hier dan ook graag hun namen: Rien van der Leeden voor zijn altijd inspirerende en duidelijke colleges, Kees Verduin voor zijn coördinatorschap van de SPSS-student-assistenten, Marcel Veenman als fijne scriptiebegeleider en David Whitebread and Deborah Pino-Pasternak for welcoming me to Cambridge University as a research intern. Your enthusiasm and curiosity for doing research have inspired me to become a researcher myself.

Ik was dan ook zo blij dat ik, ondanks mijn te late binnenkomst op de sollicitatie, als promovenda werd aangenomen op het project **Begrijpend Lees**experiment (BELEX).

## Begeleidersteam

Amos van Gelderen, als co-promotor en begeleider was jij van onschatbare waarde. Door jou heb ik geleerd om sneller hulp te vragen en het hele promotietraject als een *team effort* te zien. Jouw schrijfkwaliteiten bewonder ik zeer en ik ben je dankbaar voor alle sessies samen achter de pc gezeten om alinea's om te gooien en zinnen bij te schaven, zodat elk artikel leest als een Perzisch tapijt. Je bent een geweldige mentor! Dank je wel hiervoor.

Peter Slegers, als promotor was jouw sterkste kant mij vragen te stellen. Vragen over de wetenschap, vragen over waarom ik bepaalde keuzes genomen heb, en vragen over wat ik nu precies bedoelde. Jij zette me letterlijk aan het denken. Dank je wel voor al jouw vragen waarmee je als geen ander de zwakke plekken in een ogenschijnlijk logisch verhaal weet aan te wijzen.

Roel van Steensel, begonnen als postdoc op het BELEX-project, nu bijzonder hoogleraar aan de VU; jij hebt me wegwijs gemaakt in het uitvoeren van grotere onderzoeksprojecten. Jouw planmatige aanpak, jouw geduld en consciëntieusheid zijn een groot voorbeeld voor mij. De periode dat we samen een kamer deelden op het Kohnstamm was naast leerzaam ook erg plezierig. Dank je wel.

Erik van Schooten, als statisticus verbonden aan het project, heb jij mij verder de weg gewezen in de wondere wereld van MLwiN. Ik kijk met veel plezier terug op onze 'statistiekdagen', en naast het opschonen van allerlei databestanden en de ontelbare analyses die we gedraaid hebben, hadden we soms ook nog tijd voor een kijkje in het museum om de hoek. Jouw kennis van de wereld en jouw gevoel voor humor zijn vele malen bronnen geweest van nieuwe inzichten *of* van daverende lachsalvo's. Jouw geestigheid en snelle denken zijn daarvoor een gouden combinatie. Dank je wel voor je leerzame lessen.

Lidia Arends, als meta-analyse expert ben jij in de laatste fase gevraagd om mee te helpen aan hoofdstuk 5. Dank je wel voor alle analyses die je hebt gedaan. Het was altijd prettig praten bij jou op de kamer.

### **BELEX-partners**

Dit proefschrift was er niet geweest zonder de inzet van de scholen die hebben meegedaan aan BELEX. Ik wil alle scholen dan ook hartelijk bedanken om samen met mij het diepe in te springen en de docenten om mij toe te laten in hun lessen. In het bijzonder wil ik Simonique bedanken. Jouw manier van lesgeven, samen met de gesprekken die we hebben gevoerd, hebben mij gevormd als onderwijsonderzoeker.

Ook de leerlingen wil ik op deze plek bedanken. Jullie onbevangen en eerlijke reacties waren verhelderend. Juist opmerkingen als 'Ben jij er weer? Nee he, ik *haat* lezen!', hebben mij de noodzaak doen inzien van goed onderwijs in begrijpend lezen. Jullie zijn de kern van dit proefschrift en zonder jullie medewerking was dit proefschrift er niet geweest. Dank jullie wel daarvoor!

De trainers van Nieuwsbegrip wil ik hier ook graag bedanken voor hun inzet, het meedenken, en het afreizen naar alle scholen om 'hun' docenten te trainen en te coachen: Tamar Israel, Hanneke Maasland, Ilona de Milliano, Marlon Ruwette en Femke Zweekhorst.

Mirjam de Bruijne, jij bent niet alleen belangrijk geweest voor het BELEX-project als CED-medewerker om ons in te wijden in Nieuwsbegrip, maar daarnaast ben jij erg belangrijk voor mij om te sparren over onze onderzoeken. Jij hebt het hele traject van begin tot einde meegemaakt en daarom vind ik het heel fijn dat je mijn paranimf wilt zijn. Dank je wel!

De dataverzameling op 10 scholen had ik absoluut niet alleen kunnen doen. Ik wil dan ook alle student-assistenten bedanken die hieraan hun steentje hebben bijgedragen. Ik dank jullie voor het vroege opstaan in de ochtend (meermaals vóór 6 uur) en het oplossen van alle praktische problemen die gaandeweg altijd op de proppen kwamen.

Gedurende de dataverzamelingsfase hebben drie stagiaires en één scriptiestudent bijgedragen aan BELEX. Dank jullie wel Anne-Loes Boon, Miou Meer en Nihayra Leona. Jullie bijdragen hebben verdieping gebracht aan het project. In het bijzonder wil ik Rosa Teepe bedanken. Van een fantastische eerste stagiaire ben je uitgegroeid tot een dierbare vriendin. Dank je wel voor alle gesprekken, zowel werk als niet-werk gerelateerd, voor onze reis naar Cyprus, voor ons laatste congresbezoek in Finland. En voor alle taartjes, entetjes en eindeloze telefoongesprekken.

### **Collega's**

Doordat ik drie werkplekken heb gehad (op de UT, op het Kohnstamm Instituut, en op de Hogeschool Rotterdam) en elke week op elke plek te vinden was, was ik omringd door

veel collega's. Eén van de eerste keren dat ik me volledig voelde opgenomen in een groep collega's, was toen ik met de SALSA-aio's en Amos en Roel de straten van Montréal onveilig heb gemaakt. Dank je wel Claudia van Kruistum, Ilona de Milliano, en Mirjam Trapman! Door jullie weet ik nu het succesvolle recept van congresbezoek: ga naar de sessies die je leuk lijkt, maar geniet ook van de stad waar je bent en de mensen om je heen. Mirjam, ik vond het heel fijn om later met jou in Amsterdam af te kunnen spreken om te praten over al onze kleine en grote onderzoeksproblemen.

Naar de UT ging ik meestal op woensdagen. Vroeg eruit, laat thuis. Van de UT wil ik in de eerste plaats 'de 5 van OWK' bedanken. Ondanks dat onze projecten inhoudelijk ver uit elkaar lagen, was het heerlijk om collega's te hebben die in hetzelfde schuitje zitten. Onze schrijfweek zal ik dan ook niet snel vergeten. Dank je wel Ilona Veer, Marijn Wijga, Rike Bron, en Arnoud Oude Groote Beverborg. Ook de overige leden van de vakgroep wil ik bedanken. De onderzoeksgerelateerde bijeenkomsten waren altijd nuttig en leerzaam. In het bijzonder wil ik Carola Bouwens bedanken, die secretaresse was bij Onderwijskunde. Dank je wel voor alle hulp en de gezellige praatjes.

Op de Hogeschool Rotterdam was ik het vaakst waardoor deze plek het meest vertrouwd voelde. De prettige werksfeer die gecreëerd is in de loop der jaren na de start van het kenniscentrum heeft daaraan zeker bijgedragen. Ik wil graag alle promovendi van het kenniscentrum Talentontwikkeling bedanken. In het bijzonder Sanneke de la Rie en Marlon Ruwette. Sanneke, jij hebt me wegwijs gemaakt op de HR. Dank je wel! Marlon, in de laatste fase heb jij me weten te motiveren door onze afspraken tijdens mijn bevallingsverlof. Dank je wel daarvoor.

Halverwege het promotietraject was ik onderdeel van de *JURE 2014 committee*. Wij organiseerden een congres voor junior onderzoekers van *Learning & Instruction*. Met veel plezier heb ik samen gewerkt met Marios, Rosa, Thomas, Irene, Elena en Jelena. Thank you for a wonderful journey together, I learned a lot from you!

## Vrienden & familie

Als laatste wil ik, zoals traditie betaamt, vrienden en familie bedanken. Niet dat zij nu een directe rol hebben gespeeld bij de totstandkoming van dit proefschrift, maar zonder wordt het leven toch wel saai en zinloos. Als eerste wil ik het gezin bedanken waar ik uit voortkom. Dank je wel papa en mama en zusje Nanda. Terugkijkend op mijn jeugd kan ik vaststellen dat jullie mij hebben gestimuleerd om altijd het beste uit mezelf te halen, niet op te geven als het even tegenzit en vertrouwen te hebben in wat ik (nog niet) kan.

Mijn schoonouders Ad en Trudy wil ik bedanken voor hun niet aflatende interesse in mijn onderzoek. Jullie bemoedigende woorden en belangstellende vragen waren erg belangrijk voor mij. Trudy, dank je wel dat je in de laatste schrijffase afgelopen zomer op

mijn 'mamadagen' op Kilian wilde passen. Dit gaf mij net weer wat meer ruimte om het proefschrift af te ronden.

Mijn oude huisgenoten uit Utrecht met aanhang wil ik bedanken voor de heerlijke 'oud-huisgenoten diners' die wij roulerend bij elkaar houden. Eveline, Thomas, Busz (Michiel), Marijn, Marije en Meike: dank je wel voor alle leuke gesprekken over jullie eigen ervaringen in het werken in het onderwijs, het doen van onderzoek en het ouderschap.

Janet, jou heb ik leren kennen op een vakgroep etentje in Enschede waar we erachter kwamen dat we bij elkaar in de straat woonden in Voorburg. Dat kan geen toeval zijn geweest. Ik koester onze fijne theemomenten, lunches, en lekkere taartjes, waarbij we altijd kunnen praten over de luchtige en minder luchtige dingen die ons bezig houden.

Lily-Ann, jouw kijk op het leven en jouw vertrouwen in mensen vind ik bewonderenswaardig. Dank je wel voor het zijn van een goede vriendin.

Karin, Lieke en Maron en wil ik bedanken voor de buitenritten die we in vroeger tijden maakten, de fijne etentjes en de gezellige weekenden weg. Jullie zijn me dierbaar.

Paardrijden doe ik nog steeds op de maandagavond en ik wil dan ook graag de 'ladies van Wittebrug 21u' bedanken voor het samen kunnen delen van onze hobby: Danique, Ellen, Esther, Julie, Mijntje en Simone.

Mijn oude bestuursgenoten van Labyrintbestuur 6!, Ruud, Serge, Melanie, Alex, en Kirsten: dank je wel dat jullie jezelf zijn. We zijn zo'n heerlijk clubje anders samen. In het bijzonder wil ik Melanie bedanken. Door het bestuursjaar hebben wij elkaar gevonden. Onze levens lopen vrij parallel aan elkaar, aangezien we allebei begonnen zijn aan een promotietraject en we allebei in dit traject moeder zijn geworden. Ik ben daarom erg blij dat jij mijn paranimf wilt zijn. Dank je wel!

Lieve Bas, woorden schieten te kort om te beschrijven wat jij voor me betekent. Van ons tweeën ben jij dan ook de prater en schrijver, ondanks het proefschrift dat er nu ligt. Jouw liefde, geduld, logica en humor hebben mij door de zware tijden gesleept die ik gekend heb tijdens het promotietraject. Maar ook heeft dit traject ons tweemaal naar Canada doen afreizen, waar we allebei genoten hebben van de natuur, de steden, het kamperen met een tentje, vuurtjes maken en *sockeye* zalm roosteren. Dank je wel, lieve Bas. Ik hou van jou.

En als laatste is tijdens het promotietraject een derde gezinslid erbij gekomen. Lieve Kilian, jij hebt mijn wereld onomkeerbaar veranderd. Het is een genot om jou te zien opgroeien tot de persoon die je bent. Daarnaast ben ik extra trots als je weer een sprong hebt gemaakt in je taalontwikkeling. Ik hoop dat jij later net zoveel van lezen zult houden als ik!





REFERENCES

SAMENVATTING IN HET NEDERLANDS

DANKWOORD (ACKNOWLEDGEMENTS)

**ABOUT THE AUTHOR | PUBLICATIONS  
AND PRESENTATIONS**

ICO DISSERTATION SERIES



## About the author

Mariska Okkinga was born on December 8, 1983, in Franekeradeel, the Netherlands. She grew up in different parts of the Netherlands (Friesland, Dodenwaard, Arnhem), Luxembourg (Bridel, Pontpierre) and Belgium (Balen). After obtaining her Baccalaureate at the European School in Mol, Belgium, in 2002, she moved back to the Netherlands where she started studying Humanistics at the University for Humanistic Studies in Utrecht. She obtained her Bachelor of Arts degree in 2006. She switched to studying Psychology at the University of Leiden and obtained her Bachelor of Science degree in 2008. Next, she finished her Research Master, focusing on Developmental Psychology, in 2011. She specialized in studying metacognition in



young children during her research internship at the Faculty of Education, University of Cambridge, and adolescents for her master thesis "Metacognitive skillfulness in students with different educational levels and its relation to fear of failure and motivation".

In 2011 Mariska started as a PhD candidate at University of Twente. Her research project focused on fostering prevocational adolescents' reading comprehension with a treatment that aimed to teach students how to use reading strategies. During her time as a PhD candidate, she was a member of the JURE (**JU**nior **R**esearchers of **E**ARLI - European Association for Research on Learning and Instruction) committee in 2013-2014 and organized the JURE conference in Cyprus, 2014. In terms of teaching, she taught the course Research Studio for several years at the University of Twente. She presented her research at national (ORD) as well as international conferences (EARLI, ECER).

Currently, Mariska is employed as postdoc researcher at the Rotterdam University of Applied Sciences. She conducts research into the design and effects of an ICT-learning environment to support reading to learn in both teacher training students and prevocational adolescents.

## Publications and Presentations

### Scientific papers

- Okkinga, M.**, Van Steensel, R, Van Gelderen, A.J.S., & Slegers, P.J.C. (2016). Effects of reciprocal teaching on reading comprehension in low achieving adolescents. The importance of specific teacher skills. *Journal of Research in Reading*.
- Okkinga, M.**, Van Schooten, E., Van Gelderen, A.J.S., Van Steensel, R., & Slegers, P.J.C. (submitted). Effects of reciprocal teaching on reading comprehension: A two-year study in an authentic setting with low-achieving adolescents.
- Okkinga, M.**, Van Schooten, E., Van Gelderen, A.J.S., Van Steensel, R., & Slegers, P.J.C. (submitted). Do prior vocabulary and metacognitive knowledge matter in the effectiveness of instructing reading strategies? An aptitude\*treatment study on reading comprehension of low-achieving adolescents.
- Okkinga, M.**, Van Steensel, R, Van Gelderen, A.J.S., Arends, L.R., Van Schooten, E., & Slegers, P.J.C.(submitted). Effectiveness of teaching reading strategies in classrooms: A meta-analysis.

### Book chapters

- Teepe, R., & **Okkinga, M.** (2013). Leesstrategieën van eentalige en meertalige zwakke adolescente lezers: een oriënterende studie. In Schram, D. (Eds). *De aarzelende lezer over de streep* (p.123-150). Delft: Eburon.

### Paper presentations and symposia

- Okkinga, M.**, Van Steensel, R., Van Gelderen, A.J.S, Van Schooten, E., Slegers, P.J.C., Arends, L. (2017, August). *Effectiveness of reading strategy interventions in classrooms: A meta-analysis*. Pater presented at the meeting of European Association for Research on Learning and Instruction Conference, Tampere, Finland, 29 August – 2 September.
- Okkinga, M.**, Van Schooten, E., Van Gelderen, A.J.S., Van Steensel, R., & Slegers, P.J.C. (2016, August). *Effects of reciprocal teaching on reading comprehension of low achieving adolescents. Differential effects with vocabulary and metacognitive knowledge*. Paper presented at the 7th International Biennial Conference of EARLI SIG 16 Metacognition, Nijmegen, 23-26 August.

- Van Steensel, R., **Okkinga, M.**, Van Schooten, E., Van Gelderen, A.J.S., & Slegers, P.J.C. (2015, August). Effects of reciprocal teaching: A two-year intervention study among low achieving adolescents. In A.J.S. van Gelderen (Chair), *Implementing Reciprocal Teaching in Classrooms: Promoting Reading Strategies and Comprehension*. Symposium conducted at the meeting of European Association for Research on Learning and Instruction Conference, Limassol, Cyprus, 25-29 August.
- Okkinga, M.**, Van Schooten, E., Van Gelderen, A.J.S., Van Steensel, R., & Slegers, P.J.C. (2015, April). *Does reciprocal teaching contribute to low achieving adolescents' reading comprehension?* Paper presented at the meeting of The Reading Network, Leiden, April 10.
- Okkinga, M.**, Van Gelderen, A.J.S., Van Schooten, E., Van Steensel, R., Slegers, P. (2014, September). *Effects of strategy-instruction and reciprocal teaching: a two-year longitudinal study among low achieving adolescents*. Paper presented at the meeting of European Conference on Educational Research, Porto, 1-5 September.
- Okkinga, M.**, Steensel, R, van, Gelderen, A.J.S. van, Van Schooten, E., & Slegers, P.J.C. (2014, June). Effecten van Nieuwsbegrip in het vmbo: een tweejarige interventiestudie. In: L. van den Bulk (Chair), *Lezen en schrijven met Nieuwsbegrip: onderzoek naar interventies, leesprocessen, opvattingen en de lespraktijk van docenten*. Symposium conducted at the meeting of Onderwijs Research Dagen, Groningen, 11-13 June.
- Okkinga, M.**, Van Steensel, R, Van Gelderen, A.J.S., & Slegers, P.J.C. (2013, August). Does promoting reading engagement foster low achieving adolescents' reading ability? In: J. Razakowski (Chair), *Promoting reading ability and vocabulary. Evidence from recent intervention studies*. Symposium conducted at the meeting of European Association for Research on Learning and Instruction Conference, Munich, 27-31 August.
- Okkinga, M.**, & Veenman, M.V.J. (2013, August). *Production vs availability deficiency of metacognition: comparing low and high achieving students*. Paper presented at the meeting of European Association for Research on Learning and Instruction Conference, Munich, 27-31 August.
- Okkinga, M.**, Van Steensel, R, Van Gelderen, A, & Slegers, P.J.C. (2013, May). Bevorderen van begrijpend lezen in het vmbo: een klassenexperiment. In: A. Wilschut (Chair), *De lees- en schrijfontwikkeling van vmbo-leerlingen: onderzoek naar verschillen, determinanten en interventies*. Symposium conducted at the meeting of Onderwijs Research Dagen, Brussels, 29-31 May.

Steensel, R. van, **Okkinga, M.**, Gelderen, A. van, & Slegers, P. (2013, August). *Does reciprocal teaching contribute to adolescent low achievers' reading comprehension?* Paper presented at the European Conference on Reading, Jönköping, Sweden, 6-9 August.

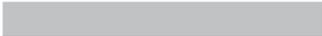
### **Presentations for the educational field**

**Okkinga, M.**, Van Steensel, R., Van Gelderen, A.J.S., Van Schooten, E., & Slegers, P.J.C. (2016, November). *Nieuwsbegrip in het vmbo. Het belang van instructie op maat.* Presentation at Stichting Lezen Congres, Amersfoort, 10 November.

**Okkinga, M.** (2014, October). *Voed de wetenschap. Nieuwsbegrip in het vmbo: het belang van instructie op maat.* Presentation at the CED conferentie Begrijpend Lezen, Rotterdam, 1 October.

Teepe, R. **Okkinga, M.**, Van Steensel, R. (2012, November). *Leesstrategieën van één- en meertalige zwakke lezers.* Presentation at Stichting Lezen Congres, Amsterdam, 8 November.





REFERENCES

SAMENVATTING IN HET NEDERLANDS

DANKWOORD (ACKNOWLEDGEMENTS)

ABOUT THE AUTHOR | PUBLICATIONS

AND PRESENTATIONS

**ICO DISSERTATION SERIES**



## ICO Dissertation Series

*ICO (Interuniversity Centre for Educational Sciences) dissertations 2016*

352. Osagie, E.R. (14-12-2016). Learning and Corporate Social Responsibility. A study on the role of the learning organization, individual competencies, goal orientation and the learning climate in the CSR adaptation process. Wageningen: Wageningen University.
351. Beckers, J. (09-12-2016). With a little help from my e-portfolio; supporting students' self-directed learning in senior vocational education. Maastricht: Maastricht University.
350. Oonk, C. (07-12-2016). Learning and Teaching in the Regional Learning Environment: Enabling Students and Teachers to Cross Boundaries in MultiStakeholder Practices. Wageningen: Wageningen University.
349. Kock, W.D. de. (24-11-2016). The effectiveness of hints during computer supported word problem solving. Groningen: University of Groningen.
348. Jansen in de Wal, J. (18-11-2016). Secondary school teachers' motivation for professional learning. Heerlen: Open University of the Netherlands.
347. Claessens, L.C.A. (30-09-2016). Be on my side, I'll be on your side: Teachers' perceptions of teacher– student relationships. Utrecht: Utrecht university.
346. Bohle Carbonell, K. (28-9-2016). May I ask you....? The influence of Individual, Dyadic, and Network Factors on the Emergence of Information in Exchange Teams. Maastricht: Maastricht University.
345. Scheer, E.A. (23-09-2016). Data-based decision making put to the test. Enschede: University of Twente. 344. Hsiao, Y.P. (23-09-2016). Peer Support to Facilitate Knowledge Sharing on Complex Tasks. Heerlen: Open University of the Netherlands.
343. Hubers, M.D. (08-09-2016). Capacity building by data team members to sustain schools' data use. Enschede: University of Twente.
342. Reijnders, P.B.G. (02-09-2016). Retrieval as a Cognitive and Metacognitive Study Technique to Learn from Expository Text. Heerlen: Open University of the Netherlands.
341. Bouwer, I.R., & Koster, M.P. (02-09-2016). Bringing writing research into the classroom: The effectiveness of Tekster, a newly developed writing program for elementary students. Utrecht: Utrecht University.
340. Geel, M.J.M. van, & Keuning T. (08-07-2016). Implementation and Effects of a Schoolwide DataBased Decision Making Intervention: A Large-Scale Study. Enschede: University of Twente.

339. Louws, M.L. (06-07-2016). Professional learning: what teachers want to learn. Leiden: Leiden University.
338. Gerken, M. (29-06-2016). How do employees learn at work? Understanding informal learning from others in different workplaces. Maastricht: Maastricht University.
337. Trinh Ba, T. (28-6-2016). Development of a course on integrating ICT into inquiry-based science education. Amsterdam: Vrije Universiteit Amsterdam.
336. Kock, Z.D.Q.P. (23-06-2016). Toward physics education in agreement with the nature of science: Grade 9 electricity as a case. Eindhoven: Eindhoven University of Technology.
335. Vrugte, J. ter (16-06-2016). Serious support for serious gaming. Enschede: University of Twente.
334. Bos, L.T. (15-06-2016). Moving Beyond Words. Supporting Text Processing Using a Situation Model approach. Amsterdam: Vrije Universiteit Amsterdam.
333. Draaijer, S. (10-06-2016). Supporting Teachers in Higher Education in Designing Test Items. Amsterdam: Vrije Universiteit Amsterdam.
332. Ebbeler, J. (12-05-2016). Implementing data use in schools: effects on the professional development of educators and the role of school leaders in data teams. Enschede: University of Twente.
331. De Beer, H.T. (11-05-2016). Exploring Instantaneous Speed in Grade Five: A Design Research. Eindhoven: Eindhoven University of Technology.
330. Kok, E.M. (01-04-2016). Developing visual expertise; from shades of grey to diagnostic reasoning in radiology. Maastricht: Maastricht University.
329. Wolff, C. (16-02-2016). Revisiting 'withitness': Differences in teachers' representations, perceptions, and interpretations of classroom management. Heerlen: Open University of the Netherlands.
328. Bogert, N.J. van der (21-01-2016). On teachers' visual perception and interpretation of classroom events using eye tracking and collaborative tagging methodologies. Eindhoven: Eindhoven University of Technology.
327. Ouweland, K.H.R. (15-01-2016). Effects of Observing and Producing Deictic Gestures on Memory and Learning in Different Age Groups. Rotterdam: Erasmus University Rotterdam.
326. Lucieer, S.M. (06-01-2016). Selecting Students for Medical Education: Exploring Novel Approaches. Rotterdam: Erasmus University Rotterdam.

# Teaching reading strategies in classrooms

## *Does it work?*

Reading comprehension is a necessary skill in today's knowledge-based economy.

However, many children and adolescents have trouble understanding the meaning of texts, which may hinder their school careers and future professions. Since the 1980's, reading programs have focused on teaching reading strategies in order to foster reading comprehension. Most of this research was done with small tutoring groups with researchers as instructors.

However, there are indications that the approach to instructing reading strategies is not always successful in improving reading comprehension in whole-classroom settings with teachers as instructors of reading strategies. The general aim of this dissertation is to gain further insight into how teaching reading strategies to students in whole-classroom settings promotes reading comprehension.

The dissertation consists of the results of an experimental study directed at testing the effects of a popular Dutch program for teaching reading strategies called 'Nieuwsbegrip' in the context of low-achieving adolescent students.

The main goal of this study was to find evidence whether the principles used in that program work in the context of whole-classroom instruction for these low-achieving students. In addition, the results of the experimental study triggered a separate meta-analysis of studies for the effects of interventions in reading strategy instruction specifically taking place in whole-classroom contexts.

Overall, the results suggest that reading-strategy programs can be beneficial in whole-classroom settings for low-achieving adolescents to improve reading comprehension. But, attention for implementation quality is crucial to create the best circumstances in which such programs can thrive. For schools to implement reading-strategy programs such as 'Nieuwsbegrip' it is beneficial to invest in teacher training, specifically focused on knowledge of the nature and characteristics of reading comprehension processes, how to diagnose the reading problems of their students, and how to instruct reading strategies and guide group work in whole-classroom settings.