

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

Mariska Okkinga

University of Twente, Enschede, The Netherlands

Roel van Steensel

Erasmus University, Rotterdam, The Netherlands

Amos J. S. van Gelderen

Kohnstamm Institute of University of Amsterdam, Amsterdam, The Netherlands
and Rotterdam University of Applied Sciences, Rotterdam, The Netherlands

Peter J. C. Slegers

University of Twente, Enschede, The Netherlands

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 behaviour. Over the course of 1 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 behaviour were focused on instruction of reading strategies, modelling 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.

Highlights

What is already known about this topic

- Reciprocal teaching is a 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; (b) expert modelling, scaffolding and fading; and (c) students practising and discussing reading strategies with other students, guided and coached by the teacher.
- High quality of implementation of reciprocal teaching by teachers in classrooms is difficult.

What this paper adds

- After 1 year of implementing reciprocal teaching, no main effects of the treatment were established.
- Intervention effects were moderated by quality of instruction: strategy instruction led to higher scores on reading comprehension in the treatment condition but not in the control condition.
- Implementation of the instructional principles was by no means optimal: teachers were unable to provide detailed guidance to students working in small groups and modelling of strategies requires more experience and theoretical insight in the use and nature of reading strategies.

Implications for practice and/or policy

- Extensive training and coaching are needed for teachers to become experts in reciprocal teaching.
- Teachers need hands-on tools to be able to guide students in their collaborative group work and to fade the teachers' role in order to allow more individual self-regulation by students in their use of strategies.
- Implementation quality has to be taken into account when doing effectiveness research and when adopting new, theory-based didactic approaches.

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. Because 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 programmes that target low-achieving adolescents are thus of vital importance. In this study, we analyse 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 analysed 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, summarising and clarifying; (b) expert modelling, scaffolding and fading; and (c) students practising 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 modelling the use of reading strategies (Rosenshine & 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 modelling 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), 16 studies were analysed. The authors found an overall positive effect on reading comprehension, with a median Cohen's effect size value ($d=.32$) for standardised 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 nonsignificant 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 nonsignificant 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 with 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 programme, focusing on instructing reading strategies. During the study, the teachers were interviewed several times. A major conclusion from this study is that teachers realised 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 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 (Hacker & Tenent, 2002). They examined the way 17 teachers implemented reciprocal teaching and adapted the method to their own teaching practice over the course of 3 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 2 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-centred to a student-centred approach, which hampered the implementation of collaborative group work in discussing and practising reading strategies. Second, Hacker and Tenent (2002) found that the students had difficulties with using all four reading strategies (predicting, questioning, summarising and clarifying). Not all strategies were used, and the strategies that were used (summarising 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 aforementioned 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 pre-vocational education to improve reading comprehension, in which the students’ teachers are delivering the lessons (Woolley, 2011). Second, we analysed 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 analysing 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, modelling and group work) moderate effects on reading comprehension?

Method

Sample selection and description

Our study focused on adolescent low achievers. Our operationalisation of low achievement was based on educational track. The Netherlands has a tracked system of secondary education. After primary education, students are placed in one of three tracks – pre-vocational secondary education, senior general secondary education and pre-university education – on the basis of their scores on a general attainment test (Ministry of Education, Culture, & Science, 2006). Because students in pre-vocational education are generally characterised 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 randomisation procedure, implying that (a) if their class was assigned to the treatment condition, they were prepared to take part in our training and coaching programme 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 programme nor discuss its contents with the colleague in the treatment condition.

- Control teachers were requested to use their regular language programme 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. Randomisation was carried out 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 male teachers 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 pre-test–post-test randomised controlled trial (Shadish, Cook, & Campbell, 2002). The design included one independent variable (treatment vs control) and one dependent variable (reading comprehension at post-test). We included four control variables: gender, reading comprehension at pre-test, vocabulary knowledge at pre-test and IQ at pre-test.

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 & 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), 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 modelling 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 emphasised 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 modelling. 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 7 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. *Summarising*. Students are instructed to summarise 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, analyse 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 example, by using ‘signal words’ (different types of connectives).

The treatment was offered in the context of an existing program called ‘Nieuwsbegrip’®, 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) every week, starting Monday evening.

Lessons were provided in sequences of 6 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 practised in a central strategy assignment that was provided on a worksheet. In addition, students could work on other assignments (i.e. answering questions about the text) on the worksheet. In the final lesson of each sequence, all strategies were practised 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 worksheets that were used during the treatment.

Training and coaching of treatment teachers

Treatment teachers took part in an extensive training and coaching programme 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 1-hour training sessions. In *Session 1*, they received general, practical information about the programme (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 modelling 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 modelling 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,

Table 1. Examples of strategy assignments, translated from several assignment sheets from the program ‘Nieuwsbegrip’.

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: Which contrast is explained in lines 16–17? [signal word = however] Why are energy boosters unfit as sports drinks? [signal word = hence]

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 keep focused on the reading strategies (Table 2). 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 with the one used by the researchers (see section on 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 modelling 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 analysed according to the three principles of instructional strategies in the treatment condition: instruction of reading strategies, modelling 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

Table 2. Template for the lessons that the treatment teachers used.

Introduction	<ul style="list-style-type: none"> ☒ Write the subject of the text and the central strategy of the lesson on the blackboard. ☒ Introduce the subject and the central strategy with a whole-class approach and activate prior knowledge. ☒ Write down questions students have about the text during orientation. ☒ Read the first paragraph together and model the central strategy. ☒ 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.
Processing	<ul style="list-style-type: none"> ☒ Instruct the students to work together in groups of two or three. Let them work on the remainder of the work sheet. ☒ 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.
Reflection	<ul style="list-style-type: none"> ☒ Reflect with the students on the reading process as well as the content. ☒ Together with the students, answer the questions they had before reading the text. Did reading the text answer those questions?

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'.

to cohesive ties. Self-questioning did not occur, and little attention was given to summarising.

No attention was given to modelling 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, which 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 pre-test and post-test 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, 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 pre-test was .86.

IQ. Intellectual ability was measured by administering the Raven Progressive Matrices, a nonverbal IQ test. The total test consists of 60 items, divided into five 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 pre-vocational 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 programme 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 modelling and group work (Palincsar & Brown, 1984). We constructed these variables in the following manner, resulting in three 4-point scales (0–3) to be used for further analysis:

1. Direct instruction of reading strategies. We distinguished four categories of behaviour:
 - 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 modelling. Here also, we distinguished four categories of behaviour:
 - a. Teachers did not use any modelling of strategy use (0 points).
 - b. Teachers modelled strategy use (1 point).
 - c. Teachers modelled 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 modelled strategy use, asked students to think aloud and provided them with feedback (3 points).
3. Group work, with four categories of behaviour:
 - 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 pre-test and the vocabulary and IQ tests were administered in the fall of 2011, just before the start of the treatment, and the reading comprehension post-test 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 standardised 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 2 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 programme. 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 towards 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.¹

There was some attrition among students, mainly because of transfers to different schools (seven 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 and nested in schools). Because there was significant random variability at the class level, we performed multi-level 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 carried out in the following order (Hox, 2010). First, all control variables were added (gender, reading comprehension at pre-test, vocabulary at pre-test and IQ at pre-test), with the final three variables centred around the grand mean (Hox, 2010; Snijders & Bosker, 1999). Second, the moderator variables (strategy instruction, modelling and group work) were entered. It was not necessary to centre the moderator variables around the grand mean, as zero was meaningful in the scoring of the classroom variables (see section on 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 owing to missing values, either with missing values within a test or questionnaire or because students were not present at one of the test sessions because of illness (despite the fact that at each school, at least one extra test session was organised). 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 $N=338$).

Results

Descriptive statistics

Table 3 shows the mean student scores for all continuous variables, as well as correlations (post-test reading comprehension, pre-test reading comprehension, IQ and vocabulary).

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

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 classroom modelling, 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 modelling. Given the small sample ($N=19$), the nonsignificant difference in the case of modelling 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 section on Classroom variables and treatment fidelity for an explanation of each variable).

Table 3. Comparison of treatment and control students in terms of reading comprehension (post-test and pre-test), 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. Post-test reading comprehension	37.37 (7.20)	36.22 (8.91)	1.30	.69*	.54*	.25*
2. Pre-test reading comprehension	35.20 (7.24)	34.51 (8.53)	0.80		.57*	.29*
3. Vocabulary	48.02 (8.28)	47.35 (9.17)	0.71			.54*
4. IQ	35.95 (5.62)	35.19 (5.20)	1.28			

* $p < .01$.

Table 4. Comparison of treatment and control teachers in terms of classroom observations: strategy instruction, modelling and group work.

Variable	Treatment (<i>n</i> = 9)		Control (<i>n</i> = 10)	
	Mean (<i>SD</i>)	Mean (<i>SD</i>)	<i>t</i> -value	<i>p</i> -value
Strategy instruction	1.89 (0.82)	0.80 (0.75)	−3.02	.008
Modelling	1.11 (0.86)	0.50 (0.47)	−1.95	.068
Group work	1.94 (1.21)	0.50 (0.67)	−3.27	.005

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

Multi-level analyses

As a first step, we examined whether the data had a multi-level structure. A model with only a student level had an *IGLS* of 2373.309. A model with both a student and a class level had a significantly better fit (*IGLS* = 2321.862; Δ *IGLS* = 51.447, *df* = 1, *p* < .001). Adding a school level did not improve model fit (*IGLS* = 2320.221; Δ *IGLS* = 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 pre-test positively contributed to post-test reading comprehension. There was also an effect of gender: boys scored significantly lower on post-test reading comprehension than girls. The effect of IQ was nonsignificant, however: it appeared that pre-test reading comprehension and vocabulary already accounted for the variance in IQ. Therefore, IQ was dropped from the model. The resulting model (Model 1; Table 5) represented a significant increase in fit compared with Model 0 (Δ *IGLS* = 210.156, *df* = 3, *p* < .001).

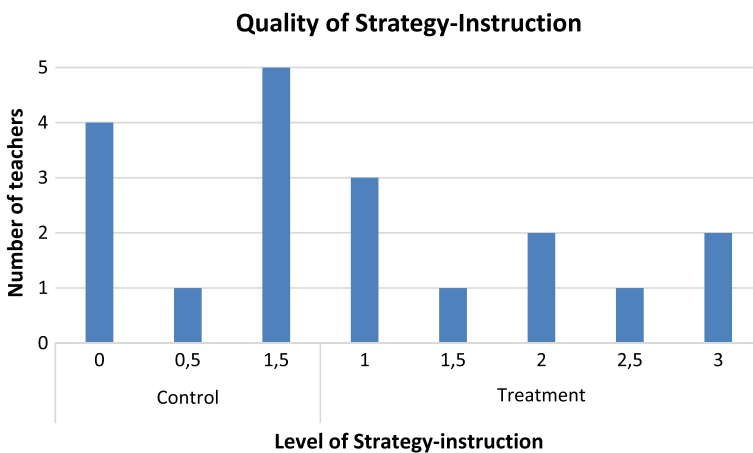


Figure 1. Bar chart for strategy instruction, for both the control and treatment teachers. Scores are calculated as the mean over two classroom observations per teacher. [Colour figure can be viewed at wileyonlinelibrary.com]

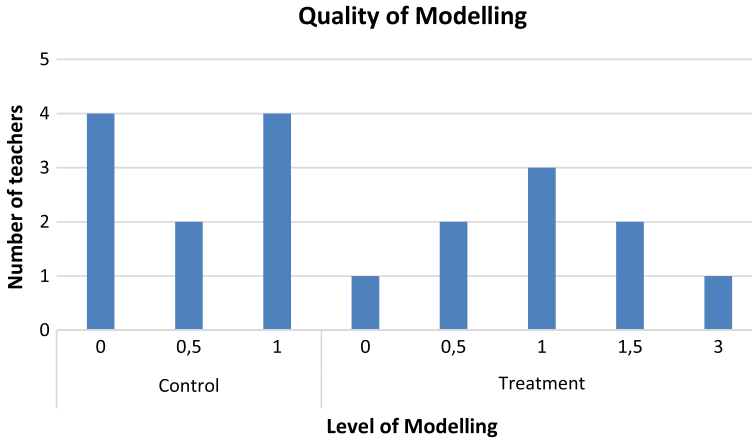


Figure 2. Bar chart for modelling, for both the control and treatment teachers. Scores are calculated as the mean over two classroom observations per teacher. [Colour figure can be viewed at wileyonlinelibrary.com]

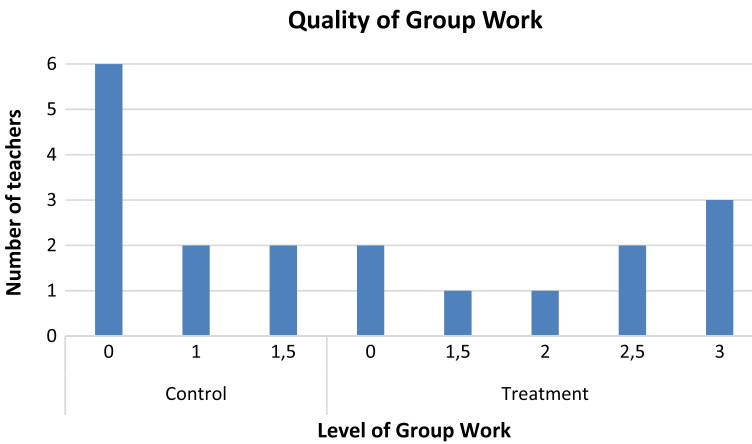


Figure 3. Bar chart for group work, for both the control and treatment teachers. Scores are calculated as the mean over two classroom observations per teacher. [Colour figure can be viewed at wileyonlinelibrary.com]

In Model 2 (Table 5), the moderator variables (strategy instruction, modelling and group work) were entered. This did not result in a significant increase in model fit ($\Delta IGLS=3.095$, $df=3$, $p > .05$). 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 ($\Delta IGLS=0.458$, $df=1$, $p > .05$). 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, modelling and group work). Of these three models, only Model 4a resulted in a significant improvement of fit compared with Model 3 ($\Delta IGLS=5.033$, $df=1$, $p < .05$), implying that the strategy instruction variable was a significant moderator of the treatment effect on reading comprehension ($B=3.183$,

Table 5. Results of the multi-level analyses, predicting post-test reading comprehension by gender, pre-test reading comprehension, pre-test vocabulary (as control variables), strategy instruction, modelling, group work (as moderator variables) and treatment (yes/no).

	Model 0	Model 1	Model 2	Model 3	Model 4a	Model 4b	Model 4c
<i>N</i> = 338							
Fixed effects							
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-test		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-test		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)
Modelling			-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 * modelling						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 (<i>IGLS</i>)	2321.862	2111.706	2108.611	2108.153	2103.120	2107.180	2107.394
Difference (Δ <i>IGLS</i>)		210.156***	3.095	0.458	5.033*	0.973	0.759
<i>df</i>		3	3	1	1	1	1

Note: *IGLS*, Iterative Generalized Least Squares.

**p* < .10.

***p* < .05.

****p* < .01.

*****p* < .001.

$SE = 1.311$, $df = 14$, $p < .05$). 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% of class-level variance.

Discussion

Our study set out to analyse how reciprocal teaching can improve low-achieving adolescents' reading comprehension in whole-classroom settings and to what extent intervention effects are dependent on teacher behaviour. Apart from analysing 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 modelling and promoted more student modelling, 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 post-test. 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% 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 behaviours 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 making students to model reading strategies during group work was a challenge. Similar observations were made by Hacker and Tenet (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 modelling 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 behaviour, teachers found it hard to gradually transfer control to

students. These classroom management issues were nonexistent 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 modelling or group work. There are at least two explanations for this observation. First, the difference between instruction of reading strategies and modelling 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 modelling and group work are relatively unknown areas and therefore harder to master.

Second, modelling and group work are dependent on initiatives afforded to students: higher scores on the former imply that more modelling is being carried out 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 modelling (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 modelling 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 seem to support this. The maximum score for modelling 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 modelling and direct guidance. This was particularly true in classes with many struggling readers.

It appears that the training and coaching offered in our study were 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, modelling 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 programme, we believe that even more training and coaching are 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 Tenen (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 randomisation 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 randomisation 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 practise 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.

Notes

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

References

- CED-Groep. (2011). *Algemene Handleiding Nieuwsbegrip* [Manual "News comprehension"].
- Curtis, M. E. (2002, May). *Adolescent reading: A synthesis of research*. Washington, DC: National Institute of Child Health and Human Development. Department of Education. <http://216.26.160.105/conf/nichd/synthesis.asp>.
- 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.
- 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.
- Duffy, G.G. (1993). Teachers' progress toward becoming expert strategy teachers. *The Elementary School Journal*, 94(2), 109–120 Retrieved from <http://www.jstor.org/stable/1001963>.

- 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.
- 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.
- Gille, E., Loijens, C., Noijens, 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.
- 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*, (pp. 403–422). New York: Erlbaum.
- Hacker, D.J. & Tenent, A. (2002). Implementing reciprocal teaching in the classroom: Overcoming obstacles and making modifications. *Journal of Educational Psychology*, 94(4), 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(2), 145–163. DOI:10.1093/applin/17.2.145.
- Hox, J.J. (2010). *Multilevel analysis: Techniques and applications*. (Second edn). New York: Routledge.
- 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 edn), (pp. 1182–1218). Newark, DE: International Reading Association.
- 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.
- 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.
- Logan, S. & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where the differences lie. *Journal of Research in Reading*, 32(2), 199–214. DOI:10.1111/j.1467-9817.2008.01389.x.
- 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.
- Organisation for Economic Co-operation and Development [OECD] (2004). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.
- OECD (2014). *PISA 2012 results: What students know and can do – student performance in mathematics, reading and science (volume I, revised edition, February 2014)*. PISA: OECD Publishing. doi: 10.1787/9789264201118-en.
- 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(2), 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(2), 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(3–4), 231–253. DOI:10.1080/00461520.1987.9653051.
- Pressley, M. & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.
- 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.
- Rosenshine, B. & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479–530. DOI:10.3102/00346543064004479.
- 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 edn). 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 edn), (pp. 1127–1148). Newark, DE: International Reading Association.
- Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading*, 3(3), 257–279. DOI:10.1207/s1532799xssr0303_4.
- 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.
- Simmons, D., Fogarty, M., Oslund, E.L., Simmons, L., Hairell, 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.
- Snijders, T.A.B. & Bosker, R.J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Spörer, N., Brunstein, J.C. & Kieschke, U. (2009). Improving students' reading comprehension skills: Effects of strategy instruction and reciprocal teaching. *Learning and Instruction*, 19(3), 272–286. DOI:10.1016/j.learninstruc.2008.05.003.
- Trapman, M., Gelderen, A. van, Schooten, E. van, & Hulstijn, J. (in press). 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*. doi 10.1080/10573569.2016.1183541
- 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., 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), 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. & Schooten, E. (2014). 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*. DOI:10.1111/1467-9817.12042.
- 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. & Leeuwe, V. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, 22(3), 407–423. DOI:10.1002/acp.1414.
- Woolley, G. (2011). *Reading comprehension: Assisting children with learning difficulties*. Dordrecht, The Netherlands: Springer International.

Mariska Okkinga studied psychology at Leiden University. She did a research internship for her Research Master in Developmental Psychology at Cambridge University, after which she earned her MSc degree in 2011. She is currently working as a PhD researcher at the University of Twente.

Dr. Roel van Steensel is an assistant professor at the Department of Pedagogical Sciences at Erasmus University Rotterdam. His research interest are emergent literacy development, adolescent reading and writing development, literacy education and family literacy.

Dr Amos van Gelderen is professor at Rotterdam University of Applied Sciences and senior researcher at the Kohnstamm Institute of the University of Amsterdam. Areas of interest are language learning and education in L1/L2, language awareness and interrelations between the abilities of speaking, writing, reading and listening.

Dr Peter Sleegers is Professor of Educational Organization and Management at the University of Twente. Dr. Sleegers has published extensively on leadership, innovation and educational policy in more than 60 referred journal articles and several edited books. Current research projects are studies into the effects of educational leadership on student motivation for school, longitudinal research into sustainability of reforms and design studies into professional learning communities.

Received 18 February 2015; revised version received 29 June 2016. Accepted 01 July 2016.

Address for correspondence: Mariska Okkinga, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands. E-mail: *m.okkinga@utwente.nl*