Effectiveness of reciprocal teaching for reading comprehension: A two-year study in an authentic setting with low-achieving adolescents.

WORK IN PROGRESS, CONFIDENTIAL< DO NOT CITE

M. Okkinga\textsuperscript{a}, A.J.S. van Gelderen\textsuperscript{c}, E. van Schooten\textsuperscript{d}, R. van Steensel\textsuperscript{b} & P.J.C. Sleegers\textsuperscript{e}

\textsuperscript{a}University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
+3164>1554929 (voice), +31 53 489 4579 (fax)
\texttt{m.okkinga@utwente.nl} (corresponding author)

\textsuperscript{b}Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
+31 10 408 8603 (voice)
\texttt{roelvansteensel@gmail.com}

\textsuperscript{c} Hogeschool Rotterdam, Museumpark 40, 3015CX The Netherlands
+31 10 794 5475 (voice),
\texttt{A.J.S.van.gelderen@hr.nl}

\textsuperscript{d}Kohnstamm Institute, P.O. Box 94208, 1090 GE Amsterdam, The Netherlands
+31 20 525 1353 (voice), +31 20 525 1201 (fax)
\texttt{evanschooten@kohnstamm.uva.nl}

\textsuperscript{e}University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
+31 53 489 4821 (voice), +31 53 489 4579 (fax)
\texttt{p.j.c.sleegers@utwente.nl}
RECI PROCA L TEACHING IN CLASSROOMS

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 authentic classroom settings and to what extent intervention 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 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 (grade 7). Our study shows that reciprocal teaching had a negative and modeling a positive effect on adolescent low-achievers’ growth in reading comprehension. In addition, results suggest that the quality of implementation of reciprocal teaching in authentic classroom settings should receive more attention.

Keywords: reciprocal teaching, reading comprehension, reading strategies, low achieving students, teacher implementation
Effectiveness of reciprocal teaching for reading comprehension: A two-year study in an authentic setting with low-achieving adolescents.

Reading comprehension is an essential skill students attain during their school career. 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 other school subjects besides language classes. 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. This study is an extension of earlier research in which we analyzed the effects of an intervention aimed at improving reading comprehension of low-achieving adolescents (Okkinga, Van Steensel, Van Gelderen, & Sleegers, submitted), based on principles of reciprocal teaching as introduced by Palincsar and Brown (1984), for a period of one school year. In this study, we analyze the effects of a two-year longitudinal intervention, thus extending the duration of the intervention. We examined its implementation in the everyday practice of language teachers, teaching low-achieving adolescents in Dutch secondary schools. Moreover, we analyzed the association between the quality of instruction and student 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,
RECIproCAL TEACHING IN CLASSROOMS

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 (Rosenshine & Meister, 1994; Kelly, Moore, & Tuck, 2001; 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. However, 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. As such, 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 given by researchers or research assistants. Only a few studies in their 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.

Thus, it is unclear from the review (Rosenshine & Meister, 1994) whether reciprocal teaching is as effective in authentic settings as it is small-group tutoring. In our view, authentic settings in education have two characteristics: students are present in their
classroom and instruction takes place within the classroom (whether this is whole-class instruction or small-group instruction) and instruction is given by their own teacher (as opposed to researcher-led interventions).

Qualitative studies in which teachers were followed during the implementation of reciprocal teaching or similar interventions suggest that the quality of implementation is indeed a problem in authentic settings (Duffy, 1993; Seymoor & Osana, 2003; Hacker & Tenent, 2002). Results of those studies show 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); that teachers revealed that their knowledge about reading strategies increased substantially during training, but their understanding of didactic principles was not developed optimally (Seymour & Osana, 2003); and that when teachers implemented reciprocal teaching in their classrooms, they found it hard to stick to the original format as students exhibited poor group discourse skills and poor application of reading strategies (Hacker & Tenent, 2002). To deal with those problems, the teachers extended whole-class instruction of reading strategies and they provided more scaffolding of strategy use in different kinds of contexts while at the same time providing scaffolding of the collaborative process (Hacker & Tenent, 2002). 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. The results of our experimental study, which lasted for one school year, (Okkinga, Van Steensel, Van Gelderen, & Sleegers, submitted) confirm these problems. 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. However, we did find an interaction between teacher implementation of strategy-instruction and the intervention. This effect implied that in the
RECIPROCAL TEACHING IN CLASSROOMS

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. A few conclusions can be drawn from these results. First, our results underscore the importance of including teacher implementation in the analyses and neglecting such variation can result in overlooking meaningful effects. Second, no interaction effects were found for other components of teacher implementation (modeling and group work). For these components it may be necessary for teachers to have more time to become accustomed and familiar with how to implement modeling and group work in an effective way. Both the aforementioned qualitative studies and the results of our experimental study point into the direction that implementing reciprocal teaching in authentic settings is very complex and effective implementation takes longer that one school year to master. It may be assumed that modeling and group work in an authentic setting pose more challenges to the teachers compared to small-group tutoring. To deal with those challenges, teachers need more support and more time.

The present study

Our study aims to contribute to existing knowledge in three ways. First, we examined whether the principles of reciprocal teaching—originally developed for small-group tutoring (Palincsar & Brown, 1984)—can be successfully used in authentic settings in prevocational education, in which regular 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. Third, we offered teachers a coaching programme over the
RECIproCAL TEACHING IN CLASSROOMS

course of two school years, resulting in a long-term intervention in which teachers have time
to master the application of reciprocal teaching.

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 in a long-term
experimental study. This allows insight into the conditions under which the treatment will be
effective in authentic reading education of low achieving adolescents.

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

1. Is reciprocal teaching provided by students’ regular teachers authentic 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 both control variables on the student level
(gender, IQ, vocabulary knowledge, metacognitive knowledge, and age) and the class level
(teacher replacement and cancelled classes). Finally, three classroom variables (strategy
instruction, modeling, group work) were hypothesized as moderator variables.
Gender is included as a control variable because girls generally show greater reading skills than boys (Logan & Johnston, 2009; Schaffner, Philipp, & Schiefele, 2014). Theoretical models suggest that reading comprehension draws heavily on IQ, vocabulary knowledge and metacognitive knowledge (Just & Carpenter, 1976, 2004; LaBerge & Samuels, 1974; Rumelhart, 2004; Samuels, 2004) and empirical evidence confirms this assumption (e.g. Ouelette & Beers, 2010; Van Gelderen et al., 2004, 2007; Verhoeven & Van Leeuwe, 2008; Trapman, Van Gelderen, Van Steensel, Van Schooten, & Hulstijn, 2014). Therefore, IQ, vocabulary knowledge and metacognitive knowledge are included as control variables. Age is also included as a control variable.

Teacher replacement and cancelled classes were included as class-level control variables because of major changes in the classroom. Six teachers (three treatment and three control teacher) were replaced during the study (See Attrition and teacher replacements). 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: direct 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, pre-university education—and 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 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 ($59 \text{ vs } 41\%$) in the treatment condition. The distribution in the control condition, however, was more equal ($49 \text{ vs } 51\%$). 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 ($n = 15 \text{ vs. } n = 5$), 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, $\text{min} = 1, \text{max} = 38$) 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 the treatment lasted, six of a total of twenty teachers were replaced during the study. One teacher (1) in the control condition found a new job halfway the first school year. She was immediately replaced. Another teacher in the control condition (2) had suffered a stroke toward the end of the second school year and was immediately replaced. A third control teacher (3) was replaced after the start of the second school year because of scheduling issues. One treatment teacher (4) became terminally ill halfway the first school year. It took the school some time to find a replacement for him. Another treatment teacher (5) was diagnosed with cancer at the start of the second school year. It took the school a few weeks to find a replacement for her. Finally, one treatment teacher (6) became pregnant toward the end of the first school year and after her leave, she changed jobs. It was not
possible to replace her 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. Seven students switched classes within their school and therefore dropped out ($n_{\text{treatment}} = 4$; $n_{\text{control}} = 3$), and three students were ill for a long period of time, 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 was statistically equal, $\chi^2(3) = 4.78, p = .19$. Together with the dropped out class, a total of 292 students (nootje: 1 leerling valt in twee categorieen) remain, of which 140 students in the treatment condition and 152 students in the control condition.

**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, 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 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):

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 intervention 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). 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 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.

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. 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. Figure 1 provides for each reading strategy an example of an assignment.
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 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 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.

Teachers were given a template for the lessons that would help them keeping focused on the reading strategies (See Figure 2).
RECIPROCAL TEACHING IN CLASSROOMS

conducted by the trainer during an intervention lesson, followed by a feedback meeting 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 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 central 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,
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 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. Vocabulary knowledge was administered two times (at pretest and at the end of the first school year) and the average of both was used as a measure for vocabulary knowledge. The Cronbach’s alpha coefficients were .86 and .85, respectively.

**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.

**Metacognitive knowledge.**

Metacognitive knowledge was measured by a questionnaire consisting of 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 test consisted of 45 items and was administered at the end of the first school year. The Cronbach’s alpha coefficient was .51.

**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 (Okkinga, Van Steensel, Van...
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).
RECIPROCAL TEACHING IN CLASSROOMS

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 four classroom observations.

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 by two coders. Across these 30 observations, 93.89% observed agreement was obtained.

Analyses

After the collection of all data, the dataset contained data of 292 students. For one class (n=24 students) it was not possible to do class observations in the first school year. These students were not included in the analyses. To prevent additional loss of students in the analyses,
missing data was imputed at the item level (never exceeding 7% of the cases per item), using SPSS missing value analysis. If a student was not present during a test session, no missing values were imputed of variables measured in that test session. After imputation, the dataset contained a total of 238 students without missing values. Of these, 110 students belonged to the treatment condition and 128 students to 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 following time points: pretest in September 2011, posttest in June 2012, retention reading comprehension in September 2012, and posttest in June 2013. As there are four measurements for reading comprehension, the dataset is multiplied by four, resulting in 952 ‘cases’. Thus, growth is measured as a repeated measures.

We tested whether a) 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 (modeling, strategy-instruction, 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 correct for these variables. Fourth, we tested whether the student-level predictors gender, IQ, age, vocabulary knowledge, and metacognitive knowledge significantly improved model fit. Dichotomous independent class (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). Fifth, the treatment variable and the interaction between treatment and occasion were added to the model to answer the first research question. A treatment effect implies a greater learning gain in the treatment group and thus a significant interaction effect between occasion and treatment. To answer the second research question, three new sets of analyses were performed, starting with a model containing the significant predictors of the first two sets. Separately for each of the three moderator variables (strategy-instruction, group work and modeling), we checked whether adding the moderator variable and its interactions with the occasion and treatment variables (two two-way and one three-way interaction) 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) would indicate a differential effect on growth of the moderator variable on students in the experimental and the control group.

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

---

1 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, which can be seen as the effect of the intervention.
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\(^2\).

Regression coefficients for class level variables were tested with number of classes as sample size (df = number of classes – number of predictors – 1) (Hox, 2010).

**Results**

**Descriptive statistics**

Table 1 shows the mean student scores for all continuous variables, as well as correlations (all four reading comprehension tests, IQ, vocabulary, and metacognitive knowledge).

No significant differences are found between the treatment and the control condition. The highest correlation is found between posttest reading comprehension at time two and time three \((r=.76, p<.01)\). No significant differences were found between the control and experimental classes at pretest for vocabulary knowledge, IQ, metacognitive knowledge, and reading comprehension, supporting the assumption that both conditions were comparable at the start of the treatment. In addition, no significant (uni-level) differences were found between the control and experimental classes at all posttests for reading comprehension.

In Table 2, 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

\(^2\)The difference in \(-2\)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.
classrooms. The difference between both groups is statistically significant on the .05 level for all variables, except for strategy-instruction. Bar charts (See Figures 3-5) for each classroom variable show that only in a few instances treatment teachers scored maximally.

Insert Table 2 and Figures 3-5 about here

**Multilevel analyses**

As no significant random variance was found at the school level, models with three levels were tested (occasion-, student-, and class level). Random slopes were applied for both the class and student level (See Appendix A). Next, the teacher-level control variables were entered in the model. Both ‘teacher replacement’ and ‘cancelled classes’ did not significantly contribute to the model and were omitted from all further analyses (See Appendix B). 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, whereas vocabulary knowledge, metacognitive knowledge, and IQ did (See Table 3). Model 4 (See Table 3) is therefore the model referred to as model 0 in subsequent analyses.

Insert Table 3 about here

In the next step, no effect of the treatment on the two years growth in reading comprehension was found, in other words, the interaction between occasion and treatment is not significant (See Table 4, Model 2; ΔIGLS=2.131, df=1, p>.05).

Moderator effects of the teacher variables were tested subsequently. In Table 4, the results are depicted for the effect of strategy-instruction. We did not find a significant
influence of strategy-instruction on growth in reading comprehension (a non-significant interaction between occasion and strategy-instruction; see Table 4, 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 (a non-significant interaction between strategy-instruction, occasion and treatment; see Table 4, Model 6; \( \Delta IGLS = .025, df=1, p > .05 \)). Likewise, no influence of group work on growth in reading comprehension, nor a moderator effect of group work was found (See Table 5, Model 3; \( \Delta IGLS = .315, df=1, p > .05 \) and Model 4; \( \Delta IGLS = .007, df=1, p > .05 \)). Thus both the level of strategy-instruction and the level of support of group work did not influence the growth in reading comprehension, nor did it influence the difference in growth between treatment and control group.

Insert Table 4 & 5 about here

Modeling, however, did moderate the effect of the treatment over time (See Table 5). It appeared that in the treatment condition more elaborate modeling positively contributed to students’ growth in reading comprehension (See Table 5, model 8; \( \Delta IGLS = 6.821, df=1, p < .01; b = .403, SE = .141, p < .05 \)). The moderator effect of modeling (modeling*occasion*treatment) explains 13.69% of the variance at the class level and 43.75% of the variance in slopes at class level. The slope variance\(^3\) at class level does remain significant however (\( \Delta IGLS = 3.882, df=1, p < .05 \)). Interestingly, when accounting for this three-way interaction, the main effect of the treatment over time becomes significantly negative (See Table 5, model 8; \( b = -.345, SE = .094, p < .05 \)).

The interpretation of the combination of these two effects becomes clear when looking at regression lines for different combinations of scores on the independent variables (Hox,

\(^3\)For testing the significance of the slope variance, the covariance between slope and intercept was fixed at zero.
We chose to calculate regression lines separately for the control and the treatment condition for teachers who score 1 SD below average on modeling, for teachers who score average on modeling, and for teachers who score 1 SD above average on modeling (See Figures 6-8). Accordingly, based on the regression equation we calculated six regression lines. This means that growth in reading comprehension for control students is higher than for treatment students when we take differences in modeling into account. In treatment classes where modeling is more elaborate, the negative effect of the treatment is diminished.

Insert Figures 6-8 about here

Thus, a main effect of the intervention could not be established without including aforementioned moderator effects. Without the moderator effect of modeling, students in the experimental condition did not outperform students in the control condition after two years of treatment. However, a higher than average score on modeling appeared to significantly add to growth in reading comprehension in the treatment condition, thus revealing a negative main effect of the treatment over time.

Discussion

Our study set out to analyze how reciprocal teaching can improve low achieving adolescents’ reading comprehension in authentic classroom settings and to what extent intervention effects are dependent on teacher behavior. Apart from analyzing the overall effects of the treatment in an authentic 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 growth of reading comprehension. Answering our second research question, we did not find significant results for strategy-instruction and group work. However, modeling moderated the effect of the treatment over time. It appears that in the treatment condition more elaborate modeling of strategy use positively contributes to students’ growth in reading comprehension.

Interestingly, when accounting for this three-way interaction (modeling*treatment*occasion), the main effect of the treatment is significantly negative. This means that growth in reading comprehension for control students is higher than for treatment students. However, in treatment classes where modeling is more elaborate, the negative effect for the treatment is diminished compared to control classes. At the same time, when implementation fidelity is low (i.e. modeling is not implemented as intended), the lower the level of reading comprehension compared to the control condition.

*Discussie moet nog verder uitgewerkt worden met onderstaande punten*

- An explanation why the treatment did not improve reading comprehension compared to the control is that the students’ own teachers experienced problems incorporating the new method, as opposed to researchers (Hacker & Tenent, 2002), resulting in a less than optimal implementation of the new approach.

- Furthermore, reciprocal teaching was originally designed for small-group tutoring (Palincsar & Brown, 1984). Reciprocal teaching may not be quite as applicable in whole-classrooms, because in such situations it is hard for teachers to control students’ group activities.

*Literatuur die gebruikt kan worden verklaring resultaten:*

- “Our expectation was that basal-comprehension approach would produce discussions that ranged further from text content than did the experimental approaches. We based
that position on our familiarity with conventional basal lessons, which tend to target
questions to a variety of issues, including direct inquiries about text content,
establishing purpose for reading, character feelings and motivations, connections to
personal experience, applying strategies to text ideas, and questions beyond guiding
comprehension, such as word recognition, grammar, and progress-monitoring
assessments.” “Our findings suggest that getting students to actively build meaning
while reading does not necessitate knowledge of and focus on specific strategies, but,
rather it may require attention to text content in ways that promote attending to
important ideas and establishing connections between them. We acknowledge that the
consensus in the field is that strategies instruction is useful. But an overall positive
picture does not provide insight into what it is about strategy instruction that may
enhance student reading.” “Focusing on strategies during reading may leave students
less aware of the overall process of interacting with text, especially in terms of the
need to connect ideas they encounter and integrate those ideas into a coherent whole.”
(McKeown, Beck, & Blake, 2009).

- “In fact, it may not be the strategies themselves that engender changes in
comprehension, but possibly some other factors that strategy instruction fosters, such
as deeper engagement with the text and awareness of the need to monitor
comprehension.” (Compton, Miller, Elleman, & Steacy, 2014).

- “Grouping format is particularly salient when implementing CSR because
fundamental to the success of CSR (at least as it is hypothesized to work), is that
students of mixed ability are placed in small, cooperative groups with specific roles
and responsibilities, that, while they rotate, are maintained long enough to
demonstrate effective group functioning. These roles require considerable knowledge
and expertise (e.g., “gist or main idea expert) and require students to question, model,
and support the learning of others in the group. “” However, without the range of reading ability in the group, we suspect that the overall group functioning was less successful with students providing fewer of the “teacher” behaviors that make improvements in students’ reading comprehension within cooperative groups possible.” (Vaughn et al., 2013, p. 159).
References


RECI PRO CA L TEACHING IN CLASSROOMS


RECIPROCAL TEACHING IN CLASSROOMS


RECI PROCA L TEACHING IN CLA SSR O OMMS


RECIPROCAL TEACHING IN CLASSROOMS


RECIPROCAL TEACHING IN CLASSROOMS


### Figures

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicting</td>
<td>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.</td>
</tr>
<tr>
<td>Summarizing</td>
<td>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.</td>
</tr>
<tr>
<td>Self-questioning</td>
<td>Read the text. Note at least five questions that spring to mind while reading.</td>
</tr>
<tr>
<td>Clarifying</td>
<td>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.</td>
</tr>
<tr>
<td>Interpreting</td>
<td>Read the text. Underline the signal words. Answer the questions, while noting the signal words:</td>
</tr>
<tr>
<td>cohesive ties</td>
<td>• Which contrast is explained in lines 16-17? [signal word = however]</td>
</tr>
<tr>
<td></td>
<td>• Why are energy boosters unfit as sports drinks? [signal word = hence]</td>
</tr>
</tbody>
</table>

*Figure 1. Examples of strategy assignments.*
Figure 2. Template for the lessons

| Introduction | 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 | Instruct the students to work together in groups of two or three. Let them work on the remainder of the work sheet (example in Appendix).  
|           | 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 | 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?  

Figure 3. Bar chart for modeling, for both the control and treatment teachers.

Figure 4. Bar chart for strategy-instruction, for both the control and treatment teachers.
Figure 5. Bar chart for modeling, for both the control and treatment teachers.

Figure 6. Regression lines for the experimental and control condition for 1 SD below average modeling.
Figure 7. Regression lines for the experimental and control condition for average modeling.

Figure 8. Regression lines for the experimental and control condition for 1 SD above average modeling.
### Tables

#### Table 1
**Descriptives Student-Level Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment (n=110) Mean(SD)</th>
<th>Control (n=128) Mean(SD)</th>
<th>t-value</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Posttest reading comprehension (4)</td>
<td>37.77(8.46)</td>
<td>39.28(8.53)</td>
<td>1.36</td>
<td>.71*</td>
<td>.70*</td>
<td>.60*</td>
<td>.57*</td>
<td>.22*</td>
<td>.34*</td>
</tr>
<tr>
<td>2. Posttest reading comprehension (3)</td>
<td>36.85(7.10)</td>
<td>36.93(8.60)</td>
<td>.08</td>
<td>-</td>
<td></td>
<td>.76*</td>
<td></td>
<td>.61*</td>
<td>.56*</td>
</tr>
<tr>
<td>3. Posttest reading comprehension (2)</td>
<td>37.72(6.81)</td>
<td>36.72(8.69)</td>
<td>.97</td>
<td>-</td>
<td></td>
<td>.68*</td>
<td></td>
<td>.57*</td>
<td>.29*</td>
</tr>
<tr>
<td>4. Pretest reading comprehension (1)</td>
<td>35.47(7.21)</td>
<td>34.67(8.38)</td>
<td>.79</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>.59*</td>
<td>.29*</td>
</tr>
<tr>
<td>5. Vocabulary</td>
<td>49.66(6.80)</td>
<td>49.56(7.85)</td>
<td>.97</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>.22*</td>
<td>.39*</td>
</tr>
<tr>
<td>6. IQ</td>
<td>36.01(5.08)</td>
<td>35.36(5.24)</td>
<td>1.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.18*</td>
</tr>
<tr>
<td>7. Metacognitive knowledge</td>
<td>26.26(4.19)</td>
<td>25.59(4.53)</td>
<td>1.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<.01

#### Table 2
**Descriptives Teacher-Level Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment (n=8) Mean(SD)</th>
<th>Control (n=10) Mean(SD)</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy-instruction</td>
<td>1.81(.80)</td>
<td>1.13(.65)</td>
<td>2.02</td>
<td>.061</td>
</tr>
<tr>
<td>Modeling</td>
<td>.84(.65)</td>
<td>.33(.35)</td>
<td>2.15</td>
<td>.047</td>
</tr>
<tr>
<td>Group work</td>
<td>2.00(1.14)</td>
<td>.30(.33)</td>
<td>4.51</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*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.
Table 3
Multilevel analyses with reading comprehension (repeatedly measured) as dependent variable to establish influence of student-level variables (IQ, gender, vocabulary knowledge, metacognitive knowledge, and age (N=952 cases/238 students).

<table>
<thead>
<tr>
<th>Model:</th>
<th>model 0</th>
<th>model 1</th>
<th>model 2</th>
<th>model 3</th>
<th>model 4</th>
<th>model 5</th>
<th>model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed part</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>34.593</td>
<td>34.651</td>
<td>33.628</td>
<td>34.550</td>
<td>35.019</td>
<td>35.058</td>
<td>35.057</td>
</tr>
<tr>
<td>Occasion (in months)</td>
<td>.156 (.040)</td>
<td>.155</td>
<td>.155 (.040)</td>
<td>.155 (.040)</td>
<td>.155</td>
<td>.155</td>
<td></td>
</tr>
<tr>
<td>IQ</td>
<td>.391 (.074)</td>
<td>.374</td>
<td>.235 (.063)</td>
<td>.240 (.063)</td>
<td>.219</td>
<td>.219</td>
<td></td>
</tr>
<tr>
<td>Gender (0=male, 1=female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.521 (.047)</td>
<td>.531 (.047)</td>
<td>.479 (.048)</td>
</tr>
<tr>
<td>Metacognitive knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.274 (.079)</td>
<td>.274 (.079)</td>
</tr>
<tr>
<td>Age (in days/365)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.084 (.570)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random part (variances)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td>18.218(7.174)</td>
<td>15.791</td>
<td>16.015</td>
<td>8.311 (3.485)</td>
<td>8.185 (3.466)</td>
<td>7.628</td>
<td>7.694</td>
</tr>
<tr>
<td>Class slope var occasion</td>
<td>.022 (.010)</td>
<td>.022</td>
<td>.022 (.010)</td>
<td>.022 (.010)</td>
<td>.022</td>
<td>.022</td>
<td></td>
</tr>
<tr>
<td>Class covar slope*intercept</td>
<td>-.355 (.209)</td>
<td>-.291</td>
<td>-.286</td>
<td>-.318 (.157)</td>
<td>-.321 (.157)</td>
<td>-.325</td>
<td>-.325</td>
</tr>
<tr>
<td>Student slope var occasion</td>
<td>.018 (.010)</td>
<td>.018</td>
<td>.018 (.010)</td>
<td>.018 (.010)</td>
<td>.018</td>
<td>.018</td>
<td></td>
</tr>
<tr>
<td>Student covar slope*intercept</td>
<td>.142 (.144)</td>
<td>.126</td>
<td>.093</td>
<td>.092 (.126)</td>
<td>.107 (.125)</td>
<td>.117</td>
<td>.117</td>
</tr>
<tr>
<td>Occasion (rep. measures)</td>
<td>17.861 (1.158)</td>
<td>17.861</td>
<td>17.861</td>
<td>17.861</td>
<td>17.861</td>
<td>17.861</td>
<td>17.861</td>
</tr>
</tbody>
</table>

Deviance

-2\*loglikelihood (deviance) 6054.507 6028.44 6022.51 5926.841 5928.564 5916.94 5916.92

Difference between 26.064 5.930 95.692 1.723 11.619 .022

-2\*loglikelihood

Difference df 1 1 1 (-1) 1 1

p <.001 <.05 <.001 >.05 <.001 >.05

Compared to model:

0 1 2 3^ 4^ 5

^ Model 0 is model 6 from Appendix A
^ Model 4 is more parsimonious than model 3, therefore gender is omitted
^ Model 5 is more parsimonious than model 6, therefore Model 5 is used as model 0 in subsequent analyses
## Table 4

*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=952/238)*

<table>
<thead>
<tr>
<th>Model:</th>
<th>model 0</th>
<th>model 1</th>
<th>model 2</th>
<th>model 3</th>
<th>model 4</th>
<th>model 5</th>
<th>model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>35.058 (.739)</td>
<td>35.377 (.889)</td>
<td>34.671 (.986)</td>
<td>34.463 (.971)</td>
<td>34.494 (1.005)</td>
<td>34.174 (1.003)</td>
<td>34.137 (1.030)</td>
</tr>
<tr>
<td>Occasion (in months)</td>
<td>.155 (.040)</td>
<td>.155 (.040)</td>
<td>.207 (.051)</td>
<td>.207 (.051)</td>
<td>.207 (.051)</td>
<td>.207 (.051)</td>
<td>.207 (.051)</td>
</tr>
<tr>
<td>IQ</td>
<td>.219 (.062)</td>
<td>.219 (.062)</td>
<td>.221 (.062)</td>
<td>.221 (.062)</td>
<td>.221 (.062)</td>
<td>.221 (.062)</td>
<td>.221 (.062)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.479 (.048)</td>
<td>.480 (.048)</td>
<td>.479 (.048)</td>
<td>.477 (.048)</td>
<td>.477 (.048)</td>
<td>.477 (.048)</td>
<td>.477 (.048)</td>
</tr>
<tr>
<td>Metacognitive knowledge</td>
<td>.274 (.079)</td>
<td>.276 (.079)</td>
<td>.277 (.079)</td>
<td>.278 (.079)</td>
<td>.278 (.079)</td>
<td>.278 (.079)</td>
<td>.278 (.079)</td>
</tr>
<tr>
<td>Treatment (1=experiment, 0=control)</td>
<td>-.710 (1.033)</td>
<td>-.710 (1.033)</td>
<td>1.322 (1.491)</td>
<td>1.322 (1.491)</td>
<td>1.322 (1.491)</td>
<td>1.322 (1.491)</td>
<td>1.322 (1.491)</td>
</tr>
<tr>
<td>Treatment*occasion</td>
<td>-1.144 (0.076)</td>
<td>-1.144 (0.076)</td>
<td>1.155 (0.076)</td>
<td>1.155 (0.076)</td>
<td>1.155 (0.076)</td>
<td>1.155 (0.076)</td>
<td>1.155 (0.076)</td>
</tr>
<tr>
<td>Strategy-instruction gmc</td>
<td>-.659 (0.742)</td>
<td>-.659 (0.742)</td>
<td>-.562 (1.102)</td>
<td>-.562 (1.102)</td>
<td>-.562 (1.102)</td>
<td>-.562 (1.102)</td>
<td>-.562 (1.102)</td>
</tr>
<tr>
<td>Treatment*strategy-instruction</td>
<td>-1.179 (1.489)</td>
<td>-1.179 (1.489)</td>
<td>-.173 (1.486)</td>
<td>-.173 (1.486)</td>
<td>-.173 (1.486)</td>
<td>-.173 (1.486)</td>
<td>-.173 (1.486)</td>
</tr>
<tr>
<td>Occasion*strategy-instruction</td>
<td>.079 (0.052)</td>
<td>.079 (0.052)</td>
<td>.088 (0.077)</td>
<td>.088 (0.077)</td>
<td>.088 (0.077)</td>
<td>.088 (0.077)</td>
<td>.088 (0.077)</td>
</tr>
<tr>
<td>Occasion<em>strategy-instruction</em>treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Random part (variances)**

| Class | 7.628 (3.251) | 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 var occasion | .022 (.010) | .022 (.010) | .019 (.009) | .019 (.009) | .019 (.009) | .015 (.007) | .015 (.007) |
| Class covar slope*intercept | -.325 (.154) | -.346 (.160) | -.284 (1.143) | -.270 (.135) | -.270 (.135) | -.231 (.121) | -.231 (.121) |
| Student slope var occasion | .018 (.010) | .018 (.010) | .018 (.010) | .018 (.010) | .018 (.010) | .018 (.010) | .018 (.010) |
| Student covar slope*intercept | .117 (.123) | .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) | 17.861 (1.158) |

**Deviance**

-2*loglikelihood (deviance) | 5916.945 | 5916.529 | 5914.398 | 5913.722 | 5913.708 | 5911.543 | 5911.518 |

Difference between | | | | | | | |
-2*loglikelihood | | | | | | | |
Difference df | | | | | | | |
*p* | >.05 | >.05 | >.05 | >.05 | >.05 | >.05 | >.05 |

**Note.** Bold=p<.05; italics=p<.01; bold and italics=p<.001

1 Model 0 is Model 5 from Table 3
### Table 5

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=952/238)

<table>
<thead>
<tr>
<th>Model:</th>
<th>model 0 $^*$</th>
<th>model 1</th>
<th>model 2</th>
<th>model 3</th>
<th>model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>34.671(0.986)</td>
<td>34.242(1.090)</td>
<td>31.946(1.859)</td>
<td>31.667(1.922)</td>
<td>31.799(2.480)</td>
<td>34.798(1.033)</td>
<td>35.578(.962)</td>
<td>35.381(.968)</td>
<td>36.099(1.965)</td>
</tr>
<tr>
<td>Occasion (in months)</td>
<td>.207(.051)</td>
<td>.207(.051)</td>
<td>.207(.051)</td>
<td>.228(.063)</td>
<td>.218(.134)</td>
<td>.206(.051)</td>
<td>.206(.051)</td>
<td>.178(.055)</td>
<td>.276(.057)</td>
</tr>
<tr>
<td>IQ</td>
<td>.219(.062)</td>
<td>.215(.062)</td>
<td>.219(.062)</td>
<td>.219(.062)</td>
<td>.219(.062)</td>
<td>.218(.062)</td>
<td>.218(.062)</td>
<td>.218(.062)</td>
<td>.218(.062)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.479(.048)</td>
<td>.478(.048)</td>
<td>.482(.048)</td>
<td>.482(.048)</td>
<td>.482(.048)</td>
<td>.480(.048)</td>
<td>.470(.048)</td>
<td>.470(.048)</td>
<td>.471(.048)</td>
</tr>
<tr>
<td>Metacognitive knowledge</td>
<td>.277(.079)</td>
<td>.279(.079)</td>
<td>.289(.079)</td>
<td>.289(.079)</td>
<td>.289(.079)</td>
<td>.273(.079)</td>
<td>.260(.079)</td>
<td>.260(.079)</td>
<td>.261(.079)</td>
</tr>
<tr>
<td>Treatment (1=experiment, 0=control)</td>
<td>.871(1.471)</td>
<td>1.917(1.819)</td>
<td>3.839(2.211)</td>
<td>4.462(2.826)</td>
<td>4.01(1.587)</td>
<td>2.69(1.369)</td>
<td>6.91(1.404)</td>
<td>2.57(1.347)</td>
<td>2.57(1.347)</td>
</tr>
<tr>
<td>Treatment*occasion</td>
<td>-.114(.076)</td>
<td>-.115(.076)</td>
<td>-.115(.076)</td>
<td>-.161(.122)</td>
<td>-.152(.153)</td>
<td>-.114(.076)</td>
<td>-.115(.076)</td>
<td>-.156(.082)</td>
<td>-.345(.094)</td>
</tr>
<tr>
<td>Group work gmc</td>
<td>-.565(1.65)</td>
<td>-.565(1.65)</td>
<td>-.565(1.65)</td>
<td>-.565(1.65)</td>
<td>-.565(1.65)</td>
<td>-.565(1.65)</td>
<td>-.565(1.65)</td>
<td>-.565(1.65)</td>
<td>-.565(1.65)</td>
</tr>
<tr>
<td>Treatment*group work</td>
<td>.334(.219)</td>
<td>.334(.219)</td>
<td>.334(.219)</td>
<td>.334(.219)</td>
<td>.334(.219)</td>
<td>.334(.219)</td>
<td>.334(.219)</td>
<td>.334(.219)</td>
<td>.334(.219)</td>
</tr>
<tr>
<td>Occasion*group work</td>
<td>.028(.049)</td>
<td>.028(.049)</td>
<td>.028(.049)</td>
<td>.028(.049)</td>
<td>.028(.049)</td>
<td>.028(.049)</td>
<td>.028(.049)</td>
<td>.028(.049)</td>
<td>.028(.049)</td>
</tr>
<tr>
<td>Occasion<em>group work</em>treatment</td>
<td>.014(.171)</td>
<td>.014(.171)</td>
<td>.014(.171)</td>
<td>.014(.171)</td>
<td>.014(.171)</td>
<td>.014(.171)</td>
<td>.014(.171)</td>
<td>.014(.171)</td>
<td>.014(.171)</td>
</tr>
<tr>
<td>Modeling gmc</td>
<td>.543(1.077)</td>
<td>3.808(2.03)</td>
<td>2.983(2.168)</td>
<td>5.950(2.392)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment*modeling</td>
<td>-.490(2.36)</td>
<td>-.490(2.36)</td>
<td>-.490(2.36)</td>
<td>-.490(2.36)</td>
<td>-.490(2.36)</td>
<td>-.490(2.36)</td>
<td>-.490(2.36)</td>
<td>-.490(2.36)</td>
<td>-.490(2.36)</td>
</tr>
<tr>
<td>Occasion*modeling</td>
<td>.083(.075)</td>
<td>.083(.075)</td>
<td>.083(.075)</td>
<td>.083(.075)</td>
<td>.083(.075)</td>
<td>.083(.075)</td>
<td>.083(.075)</td>
<td>.083(.075)</td>
<td>.083(.075)</td>
</tr>
<tr>
<td>Occasion<em>modeling</em>treatment</td>
<td>.403(1.41)</td>
<td>.403(1.41)</td>
<td>.403(1.41)</td>
<td>.403(1.41)</td>
<td>.403(1.41)</td>
<td>.403(1.41)</td>
<td>.403(1.41)</td>
<td>.403(1.41)</td>
<td>.403(1.41)</td>
</tr>
<tr>
<td><strong>Random part (variances)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td>7.477(3.182)</td>
<td>7.064(3.046)</td>
<td>6.602(2.896)</td>
<td>6.533(2.874)</td>
<td>6.527(2.872)</td>
<td>7.772(3.298)</td>
<td>4.80(2.297)</td>
<td>4.572(2.225)</td>
<td>3.946(2.015)</td>
</tr>
<tr>
<td>Class slope var occasion</td>
<td>.019(.009)</td>
<td>.019(.009)</td>
<td>.019(.009)</td>
<td>.018(.008)</td>
<td>.018(.008)</td>
<td>.018(.009)</td>
<td>.018(.009)</td>
<td>.018(.009)</td>
<td>.018(.009)</td>
</tr>
<tr>
<td>Class covar slope*intercept</td>
<td>-.298(1.43)</td>
<td>-.288(1.40)</td>
<td>-.288(1.37)</td>
<td>-.282(1.46)</td>
<td>-.309(1.46)</td>
<td>-.204(1.16)</td>
<td>-.182(1.08)</td>
<td>-.115(0.83)</td>
<td>-.115(0.83)</td>
</tr>
<tr>
<td>Student slope var occasion</td>
<td>.018(.010)</td>
<td>.018(.010)</td>
<td>.018(.010)</td>
<td>.018(.010)</td>
<td>.018(.010)</td>
<td>.018(.010)</td>
<td>.018(.010)</td>
<td>.018(.010)</td>
<td>.018(.010)</td>
</tr>
<tr>
<td>Student covar slope*intercept</td>
<td>.116(.124)</td>
<td>.116(.124)</td>
<td>.116(.124)</td>
<td>.115(.123)</td>
<td>.115(.123)</td>
<td>.116(.123)</td>
<td>.114(.123)</td>
<td>.113(.124)</td>
<td>.115(.123)</td>
</tr>
<tr>
<td>Occasion (rep. measures)</td>
<td>17.861(1.158)</td>
<td>17.861(1.158)</td>
<td>17.861(1.158)</td>
<td>17.861(1.158)</td>
<td>17.861(1.158)</td>
<td>17.861(1.158)</td>
<td>17.861(1.158)</td>
<td>17.861(1.158)</td>
<td>17.861(1.158)</td>
</tr>
<tr>
<td>Deviance</td>
<td>5914.398</td>
<td>5913.699</td>
<td>5911.51</td>
<td>5911.195</td>
<td>5911.188</td>
<td>5914.165</td>
<td>5911.381</td>
<td>5910.219</td>
<td>5903.398</td>
</tr>
<tr>
<td>Difference between -2*loglikelihood (deviance)</td>
<td>.699</td>
<td>2.189</td>
<td>.315</td>
<td>.007</td>
<td>.233</td>
<td>2.784</td>
<td>1.162</td>
<td>6.821</td>
<td></td>
</tr>
<tr>
<td>Difference df</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>&gt;.05</td>
<td>&gt;.05</td>
<td>&gt;.05</td>
<td>&gt;.05</td>
<td>&gt;.05</td>
<td>&gt;.05</td>
<td>&gt;.05</td>
<td>&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Compared to model</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>
RECIPROCAL TEACHING IN CLASSROOMS

Note. Bold = p < .05; italics = p < .01; bold and italics = p < .001; * Model 0 is Model 2 from Table 4
Appendices

Appendix A

Multilevel analyses with reading comprehension (repeatedly measured) as dependent variable to establish multilevel structure (N=952 cases/238 students).

<table>
<thead>
<tr>
<th>Model:</th>
<th>model 0</th>
<th>model 1</th>
<th>model 2</th>
<th>model 3</th>
<th>model 4</th>
<th>model 5</th>
<th>model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed part</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>35.232</td>
<td>34.661</td>
<td>34.385</td>
<td>34.633</td>
<td>34.635</td>
<td>34.608</td>
<td>34.593</td>
</tr>
<tr>
<td>Occasion (in months)</td>
<td>.150</td>
<td>.150</td>
<td>.150</td>
<td>.150</td>
<td>.150</td>
<td>.152</td>
<td>.156</td>
</tr>
<tr>
<td>Random part (variances)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>8.497</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class slope var occasion</td>
<td>.019</td>
<td>.019</td>
<td>.019</td>
<td>.019</td>
<td>.019</td>
<td>.019</td>
<td>.019</td>
</tr>
<tr>
<td><strong>Deviance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2*loglikelihood (deviance)</td>
<td>6132.566</td>
<td>6100.015</td>
<td>6097.977</td>
<td>6077.942</td>
<td>6077.883</td>
<td>6058.501</td>
<td>6054.507</td>
</tr>
<tr>
<td>Difference between</td>
<td>32.551</td>
<td>2.038</td>
<td>22.037</td>
<td>.059</td>
<td>19.382</td>
<td>3.994</td>
<td></td>
</tr>
<tr>
<td>Difference df</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>p value</td>
<td>&lt;.001</td>
<td>&gt;.05</td>
<td>&lt;.001</td>
<td>&gt;.05</td>
<td>&lt;.001</td>
<td>&lt;.05</td>
<td></td>
</tr>
<tr>
<td>Compared to model:</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

*= fixed at zero
Appendix 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=952/238)

<table>
<thead>
<tr>
<th>Model:</th>
<th>model 0*</th>
<th>model 1</th>
<th>model 2</th>
<th>model 3</th>
<th>model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed part</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>35.93 (1.094)</td>
<td>35.292 (1.914)</td>
<td>34.940 (1.227)</td>
<td>34.961 (1.149)</td>
<td>34.666 (1.121)</td>
</tr>
<tr>
<td>Occasion (in months)</td>
<td>.156 (.040)</td>
<td>.157 (.040)</td>
<td>.178 (.044)</td>
<td>.157 (.040)</td>
<td>.173 (.038)</td>
</tr>
<tr>
<td>Teacher replacement (yes=1)</td>
<td>-3.160 (2.153)</td>
<td>-1.558 (2.619)</td>
<td>-1.01 (.094)</td>
<td>-6.995 (3.997)</td>
<td>-1.392 (4.889)</td>
</tr>
<tr>
<td>Teacher replacement*occasion</td>
<td>.156 (.040)</td>
<td>.157 (.040)</td>
<td>.178 (.044)</td>
<td>.157 (.040)</td>
<td>.173 (.038)</td>
</tr>
<tr>
<td>Cancelled classes</td>
<td>-6.995 (3.997)</td>
<td>-1.392 (4.889)</td>
<td>-3.333 (.169)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancelled classes*occasion</td>
<td>.156 (.040)</td>
<td>.157 (.040)</td>
<td>.178 (.044)</td>
<td>.157 (.040)</td>
<td>.173 (.038)</td>
</tr>
</tbody>
</table>

Random part (variances)

| Class | 18.218 (7.174) | 18.306 (7.196) | 17.829 (7.044) | 19.654(7.650) | 18.106 (7.136) |
| Class slope var occasion | .022 (.010) | .022 (.010) | .020 (.009) | .022 (.010) | .017 (.008) |
| Class covar 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 var occasion | .018 (.010) | .018 (.010) | .018 (.010) | .018 (.010) | .018 (.010) |
| Student covar 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) |

Deviance

| -2*loglikelihood (deviance) | 6054.507 | 6052.474 | 6051.360 | 6051.972 | 6048.400 |
| Difference between | 2.033 | 1.114 | 2.535 | 3.572 |
| Difference df | 1 | 1 | 1 | 1 |
| p | >.05 | >.05 | >.05 | >.05 |
| Compared to model: | 0 | 1 | 0 | 3 |

*=Model 6 from Appendix A