

Intervention combining cooperative learning and instruction in reading comprehension strategies in heterogeneous classrooms

Nina Klang,^{1,2*} Johanna Åsman,¹ Maria Mattsson,¹
Claes Nilholm¹ & Jenny Wiksten Folkryd¹

¹Uppsala University, Sweden; ²Mälardalen University, Sweden

Abstract

The aim of this study was to investigate the effects of intervention, combining instruction in cooperative learning and reading comprehension strategies on students' reading comprehension in grade 5. The teachers in the experiment group implemented the intervention while the teachers in the control group received training in reading comprehension strategies and taught as usual. Students in the experiment group and control group participated in tests of reading comprehension before and after the intervention. The results showed that being a part of the experiment group did not lead to greater gains in reading comprehension above the control group. Students of teachers who did not fully implement the intervention attained higher scores on reading comprehension than students in the control group. Students of teachers who fully implemented the intervention, on the other hand, received lower scores. The results are discussed with regard to research on teachers' integration of intervention into their instructional routines.

Keywords: *cluster-randomised experiment; cooperative learning; heterogeneous classrooms; reading comprehension*

Responsible editor: Gustaf B. Skar

Received: December, 2020; Accepted: December, 2021; Published: February, 2022

Introduction

Despite the increased number of proficient readers globally and literacy levels that are above the global average in Sweden, Norway, Finland and Denmark, the outcomes are not the same for all students. This is evident from international evaluations that reveal gender gaps, as female students achieve better results than male students. The

*Correspondence: Nina Klang, e-mail: nina.klang@edu.uu.se

evaluations also reveal socioeconomic disparities, as students from disadvantaged backgrounds are less likely to attain minimum levels of reading proficiency (Mullis et al., 2017; Schleicher, 2019). Thus, the challenge is to provide reading instruction in which students with different prerequisites learn and participate. For students who experience difficulties in reading comprehension, the transition to middle school is particularly challenging as reading requirements increase as texts become longer and contain more subject-specific words (Snyder, 2010). Thus, it is important that research on reading comprehension interventions in heterogeneous classrooms in middle school be enhanced. The current study combined the instructional approach of Cooperative Learning (CL) with explicit instruction in reading comprehension strategies through the approach of Reciprocal Teaching (RC) to promote students' reading comprehension in heterogeneous classrooms in grade 5.

Interventions that promote reading comprehension

A number of interventions have been proposed for promoting reading comprehension. Many of these interventions share a common feature – explicit instruction in reading comprehension strategies. The interventions are based on the assumption that skilled readers take a strategic approach to the task of reading, predicting and reflecting on their expectations prior to reading, clarifying unfamiliar concepts and posing questions in the process of reading (Pressley & Gaskins, 2006). One form of intervention is Reciprocal Teaching. Originally developed by Anne Marie Palinscar and Ann Brown (1984), it focuses on explicit instruction in four comprehension-fostering activities: predicting the content of a text, formulating questions, clarifying unfamiliar words and summarising the main idea of a text. These activities are embedded in dialogic teaching and small-group discussions, in which the responsibility for using reading strategies is gradually handed over from the teacher to the student. This approach has led to positive results in students' reading comprehension (Lee & Tsai, 2017; Rosenshine & Meister, 1994).

In recent years, this approach to instruction has been complemented by interventions that promote small group interactions. The researchers responsible for these interventions argue that it is important to support group dynamics as it may affect the students' success in reading comprehension (Law, 2014). The CL approach involves structuring group work to promote group cohesion at the same time as each group member's contribution is highlighted and valued (Gillies, 2016; Johnson et al., 2009). Two approaches to cooperative learning that focus on Reciprocal Teaching have been researched: Collaborative Strategic Reading (CSR; Boardman et al., 2016a) and Cooperative Integrated Reading Composition (Stevens et al., 1991). In these approaches, reading comprehension strategies are embedded in group activities that are organised in accordance with the Cooperative Learning approach (CL; Gillies, 2016).

Despite the extensive research base used in small group instruction in reading comprehension, some questions still remain. One such question concerns distinguishing between contributions of CL and contributions of instruction in reading strategies. A review by Puzio and Colby (2013) revealed higher comprehension achievement in classes that practised cooperative and collaborative learning. However, it was not possible to calculate the unique effects of these approaches to instruction on reading comprehension as they often formed a part of multicomponent interventions. Thus, it may be difficult to disentangle the effects of CL from the effects of instruction in reading comprehension strategies.

An additional question concerns the differential effect of interventions for students who have varying reading comprehension abilities from the outset. For example, in a study by Law (2014), structured group work favoured students who achieved below average rather than students who achieved above average in reading comprehension. The authors hypothesised that high-achieving students may not have the same need for structured discussions as low-achieving students. Thus, there is a need for studies that use more fine-grained analysis that allow these effects to be separated.

Finally, the interventions reported above were conducted in a North American context, except for one study, conducted in an Asian context (Law, 2014). A review of randomised controlled trials (RCTs) in educational research in a Scandinavian context (Pontoppidan et al., 2018) revealed that although there has been an increase in RCTs in the educational field in Denmark and Norway, this increase was not observed in the same degree in the Swedish context. Thus, the current study may be seen as a contribution to the experimental study of interventions with a focus on cooperative learning and reading comprehension in a Scandinavian context.

Implementing cooperative learning and reading comprehension interventions

It is generally acknowledged that the success of an intervention is strongly related to the quality of its implementation (Durlak & DuPre, 2008). Students of teachers who implement interventions with higher fidelity tend to make greater gains than students of teachers who do not adhere to intervention guidelines to a similar degree (Boardman et al., 2016b). Previous studies of CL approach revealed challenges in teacher implementation efforts (Buchs et al., 2017; Gillies & Boyle, 2010). The teachers in these studies experienced difficulties integrating the CL approach into the existing curriculum and responding to the challenges that arise in group work (Buchs et al., 2017; Gillies & Boyle, 2010). This instructional approach may require a considerable change not only in teacher practices but also in teacher views of student learning, valuing child-centred pedagogy and collaboration among students (Ghaith, 2018). Thus, effects of such a complex intervention may be more evident in studies with a longer duration (Slavin, 2015).

The studies, reported above, suggest the need to include indicators of implementation quality in the evaluation of the CL-RC intervention. However, the relationship between implementation fidelity and student outcomes is not straightforward. For example, in a study of a reading comprehension intervention (Klingner et al., 2004), students of a teacher who implemented the intervention with low fidelity, received the highest gains in reading comprehension. The researchers hypothesised that these unexpected results could be attributed to this teacher's experience and instructional skills. Moore et al. (2019) studied teacher implementation of a reading comprehension intervention in relation to integration of the intervention with the content of instruction. The researchers found that teachers who implemented the intervention with a high degree of fidelity did not actually integrate it with the content of instruction. Thus, the intervention was an "add-on" activity, not a practice integrated with the instructional content and existing routines.

In addition, there may be a tension between implementation fidelity and the need to adapt interventions to teacher instructional routines. Kim et al. (2017) compared a standard literacy intervention with an intervention, which allowed the teachers to make adaptations, based on their knowledge and experience. The results showed that adapted intervention led to higher gains in students' reading comprehension than the standard intervention. With regard to the complex relationship between implementation fidelity, teacher adaptations and student outcomes, there is a need to consider teacher implementation when studying the effect of CL-RC intervention on student outcomes.

The current study

The aim of this study was to explore the effect of the CL approach, combined with instruction in reading comprehension strategies (CL-RC), on students' reading comprehension. The aim of the study is specified in the following two research questions:

- (a) What is the effect of CL-RC on students' reading comprehension and does the effect differ for students who perform below and above median at pre-test?
- (b) Does the effect of CL-RC differ with regard to the level of implementation of the intervention?

Method

The study utilised a cluster-randomised experimental design using an experiment group and a control group. Students in the experiment group received instruction according to the CL-RC approach and data on the students' reading comprehension in both groups were collected before and after the intervention. Ethical approval was obtained from the Swedish Ethical Review Authority (Dnr 2017/372) prior to the start of the study.

Participants

The participants were 56 teachers of 958 students in grade 5. Power analyses revealed the need to recruit 1,020 students in 51 classes, assuming an effect size of 0.30 and power of 80% and an intra-class correlation of 0.10 in classes of approximately 20 students each. Information about the project was sent to all the compulsory schools receiving students in grade 5 in five municipalities. The information was sent via e-mail and was followed up by telephone conversations with the heads of the schools. In addition, the information was posted on the Uppsala University website and via Facebook in cooperative learning interest groups. After recruitment and submittal of informed consent from teachers, students and students’ legal guardians, the teachers were randomly assigned to either the experiment or the control group. Prior to randomisation, the teachers were assigned to three groups based on their reports of experience and knowledge of the CL approach: (a) having knowledge and experience of CL; (b) having some knowledge and experience of CL; (c) having no knowledge or experience of CL. Randomisation was conducted within each group and approximately an equal proportion of teachers in both the experiment and control groups had knowledge and experience of CL.

Table 1 reports data on the characteristics of the participating classes in the CL-RC intervention study. As shown in Table 1, the classes comprised an average of 27 students and there were 33–36% of students in need of extra adaptations or special support in 75% of the classes. According to the Swedish National Educational Agency [SNEA] (2014), the need for special support is not tied to a medical diagnosis, but rather to students’ achievement in accordance with the national curriculum. Two types of special support are provided at school – extra adaptations and special support, documented in an individual educational plan (IEP). A special letter of consent was sent to students’ guardians asking for permission to gain access to information of whether the students needed extra adaptations and special support. Guardians of 12 students gave their permission. Therefore, information on the need for extra adaptations and special support was retrieved through teacher reports on group level.

Table 1. Background characteristics of classes and teachers in the intervention and control groups

	CL group	Control group
Number of classes	27	25
Mean number of students per class	27	27
Proportion of children with special educational needs (SEN) per class		
1st quartile	.17	.27
2nd quartile	.27	.33
3rd quartile	.33	.36
Teachers who reported using the CL approach in at least three lessons per week	18 (20 responses)	
Classes in which teachers reported using the CL approach in reading comprehension in at least one lesson per week	14 (16 responses)	
Teachers in the control group who reported working with reading comprehension in at least one lesson per week		11 (15 responses)

Figure 1 shows the number of students recruited from the outset as well as the number of students participating in the pre- and post-test. As seen in the figure, 958 students were recruited in the study, but tests at pre- and post-measurement could be obtained only for 614 (attrition rate of 34%) students and 605 students (attrition rate of 37%) for the two reading comprehension tests. The reasons behind attrition

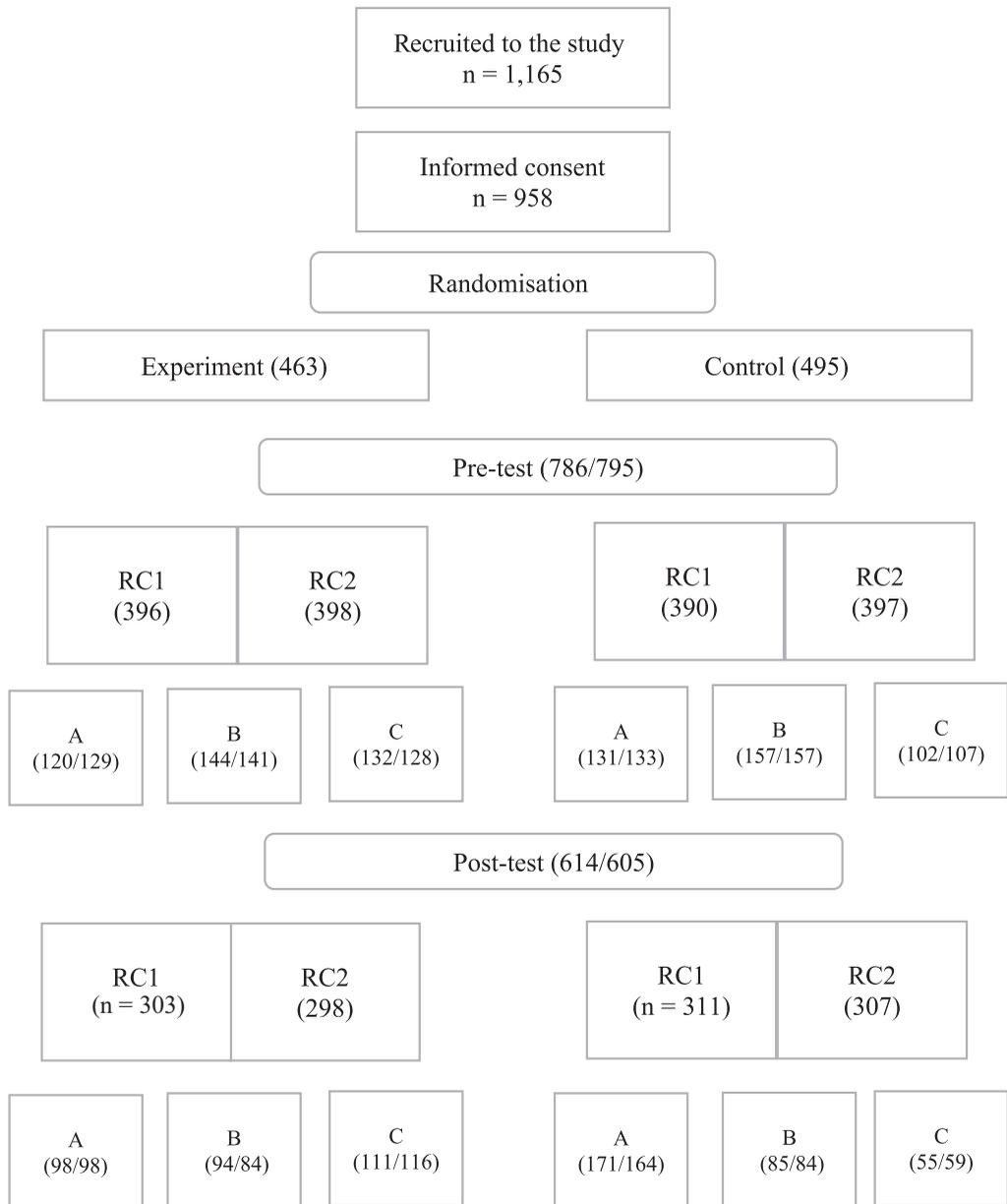


Figure 1. Number of participants at recruitment and number of participants per test at pre- and post-measurement.

were teacher dropout from the study due to sick leave or personal circumstances (five teachers in the experiment group and two teachers in the control group). Furthermore, some teachers did not send questionnaires to the researchers and some students were absent on the day of data collection. Little’s MCAR test (Little, 1988) was conducted to investigate missing data patterns in the two reading comprehension tests. The results showed that the null hypothesis that data are missing completely at random could not be supported ($\chi^2 = 13.84, p = .001; \chi^2 = 8.99, p = .011$), indicating that there are patterns of missingness in the data that may only be explained by unobserved variables (Enders, 2010).

In Table 2, the results of independent sample T-tests are presented, comparing the group of students with data at pre- and post-measurement with students for whom the data were missing. The variables in the table include pre-test scores on reading comprehension and mathematical problem-solving tests, as well as sociometric nominations, indicating students’ peer acceptance and friendships, retrieved from the overall project (Klang et al., 2020). As seen in the table, students in the experiment group for whom the data were missing had significantly lower pre-test scores in reading comprehension, mathematical problem solving, lower peer acceptance and fewer reciprocated friendships. Thus, participating in the project may have been more challenging for classes characterised by lower achievement and less social acceptance and friendships from the start.

Table 2. Mean values and standard deviations of student scores on tests of reading comprehension and mathematical problem-solving as well as sociometric nominations, reported separately for participants with missing values and participants without missing values

	Control group		CL group	
	Missing	Not-missing	Missing	Not-missing
Reading comprehension (RC1) at pre-test	16.27 (5.89)	16.96 (6.19)	15.57 (6.25)**	17.68 (5.59)**
Reading comprehension (RC2) at pre-test	11.79 (4.49)	11.67 (4.32)	9.67 (4.14)**	11.81 (3.97)**
Mathematical problem-solving at pre-test	9.38 (3.97)**	10.95 (3.91)**	7.15 (4.44)**	8.65 (3.69)**
Nominated by classmates (peer acceptance)	.47 (.16)**	.52 (.16)**	.39 (.17)	.41 (.17)
Reciprocated nominations (friendships)	.29 (.17)**	.37 (.19)**	.22 (.16)**	.26 (.17)**

Note: **p ≤ .001

CL-RC intervention

The CL-RC intervention lasted for 15 weeks and comprised two parts. The first part focused on the CL approach to support group processes and lasted for seven weeks. The CL approach was introduced to teachers during two days of training. The

activities were structured in accordance with the five principles of the CL approach (Gillies, 2016; Johnson et al., 2009): positive interdependence, promotive interaction, individual accountability, explicit instruction in social skills and group processing. To promote positive interdependence, the activities focused on group-building activities by creating group names and logos. In addition, the teachers emphasised the group's performance rather than individual performances. For promotive interaction, the group members were seated close to each other and shared materials. For individual accountability, the teachers asked individual students to report on the results of the group work so that each student would share responsibility for the group's work. Explicit instruction in social skills was provided by focusing on one or two social skills for a longer period of time, modelling the skill and providing activities for practicing the skill. Finally, to ensure group processing, the teachers allocated time at the end of each lesson for the groups to evaluate their work and plan future actions.

In the second part of the intervention, CL was embedded in curricular activities in reading comprehension and mathematical problem solving. During one day of training, the teachers were presented with research on reading comprehension strategies, inspired by Reciprocal Teaching (Palinscar & Brown, 1984). To support the teachers in embedding CL in their instruction, lesson modules based on two books, *The Invisible Child* by Tove Jansson (1962) and *The Breadwinner* by Deborah Ellis (2012), were created by the researchers in collaboration with the teachers who participated in a pilot study that preceded the start of the experimental study. These teachers received training in CL and developed and piloted the materials in their classes prior to the start of the study.

Each module comprised three lessons: The first lesson was devoted to activities that involved predicting the contents of the books and answering questions of varying complexity – from gathering and searching for information to making inferences from larger portions of text. Two subsequent lessons in each module focused on explaining unfamiliar words, students answering questions and also formulating their own questions and summarising the text contents. In this way, the responsibility for formulating questions was successively handed over from the teacher to the student. The teachers in the intervention group were supposed to continue their instruction using the CL-RC approach after having used the lesson modules. Thus, the instruction focused on four comprehension-fostering activities (Palinscar & Brown, 1984) conducted in small groups and structured in accordance with the five principles of the CL approach (Gillies, 2016; Johnson et al., 2009). An example of instructional activities is presented in appendix. Further, an in-depth analysis of the reading comprehension activities in the intervention is given in Klang et al. (forthcoming).

The fidelity of the implementation was investigated by teacher self-report data and observations of two lessons. The teachers in the intervention group provided information on time devoted to the CL approach on five occasions during the intervention period. Teachers for 18 out of 27 classes reported that they used the CL approach in 3–4 lessons per week (Table 1). Observations were conducted twice

during school visits, followed by a coaching conversation and written feedback to each teacher. An observation protocol based on the principles of the CL approach was used (Johnson et al., 2009). Teachers from 21 out of 27 classes used the CL principles in their teaching, thereby meeting the fidelity of implementation requirements (Klang et al., 2020). In addition, at the end of the intervention period, the teachers were asked how much time they had devoted to embedding the CL approach in reading comprehension activities (CL-RC approach). As seen in Table 1, teachers from 14 classes reported implementing the CL approach in these activities in at least one lesson per week.

Control group

During the first part of the intervention, the teachers in the control group were instructed to teach as they would otherwise teach with no intervention. In the second phase, the teachers received two days of training in promoting the reading comprehension and mathematical problem-solving of their students. Similar to the intervention group, the training focused on reading comprehension strategies. The teachers were also supported with lesson modules. In contrast to the intervention group, the training and modules did not contain CL activities. No observations were made in the control group. At the end of the study, teachers from 11 out of 25 control classes reported that they had taught reading comprehension strategies in at least one lesson per week (see Table 1).

Outcome measures

Reading comprehension is a complex cognitive process in which the reader interacts with a given text in order to construct meaning. In this process, important factors are the reader's prior knowledge, motivation and interest, purpose of reading and the characteristics of a given text (e.g. Pearson & Hamm, 2005). In order to measure these complex processes, we used two reading comprehension tests involving texts of fiction from the Swedish national tests of literacy for the 6th year. The tests will hereinafter be referred to as *reading comprehension test one* (RC1) and *reading comprehension test two* (RC2). The tests measure reading comprehension, grouped into four reading comprehension processes: (a) finding the requested information, (b) drawing simple conclusions, (c) consolidating, interpreting the information and reflecting, (d) evaluating and critiquing the contents, language and textual elements (Department of Nordic Languages [DNL], 2012, 2013, 2014).

To minimise the risk of a learning effect (Shadish et al., 2002), during both pre- and post-measurement, we used three versions of tests used in Sweden in 2012 (test A), 2013 (test B) and 2014 (test C). Furthermore, to reduce the risk of the tests being non-equivalent in difficulty, the tests were used in blocks so that an equal number of classes received each test version. As seen in Figure 1, the number of students who completed versions of tests A, B and C was approximately 30–40% of the students for

each test version. However, at post-test, the number of students who completed test A was disproportionately large in the control group as well as the experiment group of teachers who fully implemented the CL-RC intervention. This was partially due to attrition.

In previous research the stability of Swedish national tests for grade 9 has been problematised, as the tests differed in length and type of tests as well number of items per test (Tengberg, 2017). Although stability over time is described as desirable in the system framework for Swedish national tests (SNEA, 2017), statistical reports for the three versions of tests for grade 6 (SNEA, 2021) point to reservations and caution in comparing results between the tests. In Table 3, the number of items referring to each of the reading processes is reported, including Cronbach alpha values and the data on the proportion of students who achieved acceptable levels of reading comprehension. As seen in the table, the number of items per reading process and the maximum composite scores varied. Generally, Cronbach alpha values were below .90, indicating a risk that the items aggregated to the composite score may not be measuring the same construct (Crocker & Algina, 2008). The proportion of students with acceptable levels of reading comprehension at the national level when the tests were administered was lower for test A.

Table 3. Data on reading comprehension tests

	A 2012		B 2013		C 2014	
	RC1	RC2	RC1	RC2	RC1	RC2
Reading processes						
1 Finding requested information	2	3	2	2	2	2
2 Drawing simple conclusions	1	3	2	3	3	1
3 Consolidating, interpreting the information, and reflecting	3	2	4	1	3	2
4 Evaluating and critiquing the contents, language, and textual elements	4	0	1	1	1	2
Maximum composite score	35	21	28	22	36	22
Cronbach alpha without deletion of items						
pre-test	.498	.387	.675	.610	.670	.468
post-test	.576	.412	.658	.657	.666	.539
Cronbach alpha after deletion of items (two versions)						
pre-test	.445/ .542	.314/ .325	-	.596/ .573	.670/ .602	.374/ .408
post-test	.392/ .488	.352/ .361	-	.590/ .641	.613/ .598	.479/ .482
Proportion of students who achieved acceptable levels of reading comprehension at national level	92.7%		94%		94%	

The students' test responses were graded by six university students in their last year of teacher training. Prior to the grading, the students received two days' training from a teacher experienced in grading national and international reading comprehension tests. To ascertain interrater reliability, 20% of the students' responses to each type of test were graded by two students. The item-by-item correlation between the graders varied from .787 to .995, indicating a high level of consensus. In addition, the Kendall coefficient of concordance (Gibbons & Chakraborti, 2011) between graders was estimated for the total scores of each test. There was statistically significant agreement between the graders for all tests, with the exception of test A, for which significant levels of consensus were not achieved between all the graders. Before each grading session (pre- and post-measurement), time was allocated to discuss any disagreements among the graders. Also, before entering the test scores into the database, disagreements were resolved as each pair of graders discussed their scores and agreed on a common score for each test item.

After grading, the item scores were aggregated into composite scores for each test, by summing up the item scores in accordance with scoring guidelines of the tests (DNL, 2012, 2013, 2014). The maximum scores, which differed across tests A, B and C, were aligned by deleting the students' responses to separate items so that the maximum scores were 28 points for RC1 and 20 points for RC2. In the deletion process, careful consideration was given to the equal representation of items referring to different reading comprehension processes. The Cronbach alpha values after deleting separate items are reported in Table 3.

Data analysis

The data analysis was conducted with the R program for statistical computing, using the lme4 package (Bates et al., 2015). Multilevel mixed linear regression analysis was used, as it enabled studying the nested sources of variability in students' reading comprehension scores, which could be attributed to students' classes and teachers (Snijders & Bosker, 2012). The analysis distinguishes between fixed coefficients (that are held constant across the entire sample) and random coefficients (that are allowed to vary by specific variables, hypothesised to add variability to a given effect). In this way, the analysis renders a more nuanced understanding as to the effect of an intervention on student outcomes. In the first model, the students' reading comprehension scores were studied as a function of time (pre- and post-measurement) and group belonging (experiment and control group), including an interaction between time and group. To take into account any variation in the students' reading proficiency at pre-test on both the classroom and the individual level, two random intercepts were included, varying by class and the individuals' identification codes.

In two additional models, the results were studied separately for students who performed above and below the median at pre-test. This was conducted in order to test the differential effects of the intervention on the performance of these students at

post-measurement. Finally, to test the effect of the degree of implementation on students' outcomes, a similar model was utilised with random intercepts, which differed in the variable of group belonging (control, CL partially implemented, and CL fully implemented). All the models were run using two versions of reading comprehension tests (see outcome measures). As long as the results did not differ by type of test, only results for one version of the tests are reported. The missing data were handled using listwise deletion (Enders, 2010). To allow for comparability of estimates across studies, intra-class correlation (ICC), rendering the proportion of variance in reading comprehension scores explained by students' class membership, was reported as an indicator of random effect (Lorah, 2018; Snijders & Bosker, 2012). To report fixed effects, standardised coefficients, expressed in standard deviations as units of measurement (Snijders & Bosker, 2012), have been suggested (Lorah, 2018). In this study, as the independent variable was binary, partially standardised coefficients are reported, obtained by standardising the dependent variable, students' reading comprehension scores, prior to analyses (Lorah, 2018).

Drawing conclusions from the multilevel mixed linear regression analyses presumes that the assumptions of the model have been met (Demos & Salas, 2021; Snijders & Bosker, 2012). First, the assumption of linearity was tested by visual analysis of graphically plotted model residuals (Demos & Salas, 2021). Judging by retrieved plots, the assumption of linearity was not violated for either of the models, indicating that the linear model fits the data. The assumption of homogeneity of variances was tested by Levene's test of homogeneity of variance using extracted model residuals (Demos & Salas, 2021). The test results were not significant, indicating that the variance of the residuals is equal across groups and may not bias the results.

The assumption of the normal distribution of residuals was controlled by visually inspecting quantile-quantile plots (QQ-plots), in which quantiles from the data in the model are plotted against theoretical normal distribution (Demos & Salas, 2021). The visual analyses of the plots showed that the curve deviated somewhat from the straight line, indicating deviation from normality. For models, including subgroups of students who scored below and above median, the curve deviated from a straight line to a larger degree, indicating a risk of less precision in the random coefficient estimates (Schielzeth et al., 2020). Testing the assumptions of the model required ggplot2 package (Wickham, 2016), car package (Fox & Weisberg, 2019) and lattice package (Sarkar, 2008).

Results

Below, the results of analyses as to the effect of the CL-RC intervention on students' reading comprehension are reported separately for each of the research questions.

What is the effect of CL-RC on students' reading comprehension and does the effect differ for students who perform below and above median at pre-test?

As seen in Table 4, judging by the mean values, students in both experiment and control groups made minor improvements during the intervention period. However, judging by partially standardised regression coefficients, being part of the experiment group that received instruction using the CL-RC approach did not result in significantly greater improvements in reading comprehension compared to the control group.

Table 4. Mean scores (standard deviation in parentheses), partially standardised multilevel regression estimates (β_1) and 95% confidence intervals (CI), intraclass correlation coefficients (ICC) for tests of reading comprehension (comparison between experiment and control group)

	Control group		CL group		Effect of CL	Below median at pre-test	Above median at pre-test	ICC
	n(RC1) = 304 n(RC2) = 307		n(RC1) = 312 n(RC2) = 298					
	Pre	Post	Pre	Post	partially standardised β_1 (95% CI)	partially standardised β_1 (95% CI)	partially standardised β_1 (95% CI)	
RC1	16.95 (6.19)	17.41 (4.95)	17.68 (5.58)	17.77 (5.96)	-.07 (-.24, .12)	-.29* (-.57, .004)	.14 (-.15, .43)	.27
RC2	11.77 (4.32)	11.66 (4.16)	11.70 (3.98)	12.00 (4.08)	.10 (-.12, .32)	-.17 (-.46, .12)	.23 (-.05, .51)	.34

Note: * $p \leq .05$

In Table 4 the effect of the CL-RC approach on students’ reading comprehension is also reported separately for the students who achieved below and above median at pre-test. As shown in the table, for students who achieved below median at pre-test, being a part of the experiment group had a significantly negative effect in comparison with students in the control group. However, judging by the width of confidence intervals encompassing both negative and positive values, this effect needs to be interpreted with caution. Regarding students who achieved above median at pre-test, no significant differences between experimental and control group were found.

Does the effect of CL-RC differ with regard to the level of implementation of the intervention?

In Table 5, the effect of the intervention on students’ reading comprehension is reported for three groups – the control group, students of CL-RC teachers whose instruction did not meet all the required implementation criteria (CL-RC partial implementation) and CL teachers whose instruction met all the criteria for implementation (CL-RC full implementation). The results show that students of teachers who partially implemented the CL-RC approach made gains in their reading comprehension scores compared to students in the control group. The students in the CL-RC partial implementation group could be expected to score 0.26 and 0.70 standard deviations higher than students in the control group.

On the other hand, for students of teachers who fully implemented the CL-RC approach, a negative effect was noted in comparison to students in the control group. Thus, students in the CL-RC full implementation group could be expected to score 0.37 and 0.46 standard deviations lower than the students in the control group.

Table 5. Mean scores (standard deviation in parentheses), partially standardised multilevel regression estimates (β_1) and 95% confidence intervals (CI), intraclass correlation coefficients (ICC) for tests of reading comprehension (comparison between experiment and control group)

	Control group		CL group partial implementation		CL group full implementation		CL partial implementation	CL full implementation	ICC
	Pre	Post	Pre	Post	Pre	Post	partially standardised β_1 (95% CI)	partially standardised β_1 (95% CI)	
RC1	16.92 (6.19)	17.41 (4.95)	16.77 (5.16)	18.68 (5.78)	18.54 (5.85)	18.68 (6.01)	.26* (.04,.47)	-.37** (-.58 -.16)	.27
RC2	11.77 (4.32)	11.66 (4.16)	10.14 (3.84)	12.94 (3.75)	13.14 (3.56)	11.14 (4.18)	.70** (.44,.96)	-.46** (-.71, -.21)	.34

Note: * $p \leq .05$, ** $p \leq .001$

Discussion

Providing reading comprehension instruction in heterogeneous classrooms has been the focus of research for several decades (Lee & Tsai, 2017; Rosenshine & Meister, 1994). Explicit instruction in reading comprehension strategies has shown to be a promising way of promoting reading comprehension (Palinscar & Brown, 1984). In recent years, this instruction has been complemented by approaches, such as cooperative learning, that promote positive group interactions (Boardman et al., 2016a; Stevens et al., 1991). This study aimed to contribute to previous research by studying a cooperative learning intervention (Gillies, 2016; Johnson et al., 2009) with a focus on reading comprehension strategies (CL-RC). The intervention was conducted in heterogeneous classrooms, 75% of which included 33–36% of students with special educational needs. The results of the study showed that the CL-RC intervention did not lead to gains in students’ reading comprehension when the reading comprehension scores of students in experiment and control group were compared over time.

In contrast to previous studies on the effect of the CL approach on reading comprehension (Boardman et al., 2016a; Stevens et al., 1991), this study found no significant effects. It attempted to address the challenges in distinguishing between the effects of CL instruction and the effects of instruction in reading comprehension (Puzio & Colby, 2013) by providing training and lesson modules in reading comprehension to both the intervention group and the control group. However, challenges in implementation of the intervention prevented the authors from drawing conclusions about the effects of the CL approach, *per se*. Previous studies have revealed that

teachers may face challenges when implementing the CL approach in their classrooms (Buchs et al., 2017; Gillies & Boyle, 2010). Future studies that seek to distinguish between the effects of CL and the effects of reading comprehension instruction may need to focus on supporting teachers in implementing the intervention and to provide observation data from the instruction in both the intervention group and the control group.

Furthermore, to study the effect of the degree of implementation on students' outcomes, we compared the achievement gains of students in three groups – control group, partial implementation, and full implementation. The results showed that the intervention may have had a positive effect on students of teachers who partially implemented the CL-RC approach and a negative effect on students of teachers who fully implemented the approach. This counterintuitive finding could be attributed to a disproportional use of test A at post-test in classes of teachers who implemented the CL approach fully in comparison with classes of teachers who implemented the approach partially. The finding can also be discussed with regard to the relationship between implementation fidelity and student outcomes. While implementation fidelity has traditionally been viewed as a prerequisite of intervention success (Durlak & DuPre, 2008), recent studies have pointed to the need of additional indicators of how interventions are integrated into existing teacher practices (Kim et al., 2017; Moore et al., 2019).

An in-depth study of how teachers integrated a reading comprehension intervention with the content of instruction (Moore et al., 2019) showed that teachers who were deemed to have a high degree of implementation fidelity only moderately integrated the instructional approach with the content of instruction. For these teachers the intervention was rather an “add-on” than an integrated part of their teaching (Moore et al., 2019). Thus, another possible explanation of the counterintuitive finding in the study is that teachers who devoted time to implementing all the components of the CL-RC approach, supportive of group work, had less time to integrate it with the content of instruction and their existing routines. A review of studies of CL approach (Slavin, 2015) called for longer duration of interventions with regard to the complexity of this instructional method.

The current study was also an attempt to investigate the effects of the CL-RC approach on the reading comprehension of students with varying abilities at the start of the intervention. According to Law (2014), the CL approach may benefit students who achieved below average but not students who achieved above average. The results of this study indicate that the CL approach did not result in improvements in reading comprehension for the subgroups. A negative effect was found for a subgroup of students who achieved below median at pre-test. However, the width of confidence intervals encompassing both negative and positive values precludes drawing further conclusions. In addition, as is evident from the methods section, the assumption of normality was violated when only subgroups of students were included in the analysis. This violation entails a risk of lower precision of estimates in regression

models (Schielzeth et al., 2020). If a larger number of students had been included in the study, this may have enabled a more fine-grained analysis of the effects of the CL approach on the subgroups of students.

Finally, the choice of outcome measures for the study may have influenced the results obtained. In this study, reading comprehension tests from Swedish national tests of literacy were used. These tests were chosen due to their suitability for large scale administration, as individual testing of students was not deemed to be feasible. Besides, the tests were well-known for many teachers of literacy and, according to our contacts with teachers, were a preferable measure to use.

The choice of the tests entailed several challenges. In order to avoid learning effects (Shadish et al., 2002), different versions of the tests were used at pre- and post-measurement. Previous research has identified a lack of comparability in the Swedish national tests for grade 9 over time (Tengberg, 2017). Swedish National Educational Agency reports for each of the years of administration also point to the lack of comparability over time, especially concerning test A (SNEA, 2021). Choosing different tests may have entailed different levels of difficulty and a lack of equivalence among the tests.

As seen in Table 3, the tests contained different numbers of items measuring separate reading comprehension processes. In seeking to achieve equivalence in maximum scores, we deleted separate items measuring specific reading processes and performed regression analyses on different variants of the test composite scores. As seen in Table 3, the Cronbach alpha values for these composite scores were low, indicating that the separate items may not measure the same underlying construct (Crocker & Algina, 2008). The test scores were summarised in accordance with the scoring manuals of the tests, aggregating the item scores into a total score. A more nuanced model for aggregation of results involving modern test theory (Crocker & Algina, 2008) may have rendered different results.

Furthermore, national tests constitute global measures of reading comprehension and may therefore result in lower effect sizes compared to researcher-designed measures (Shadish et al., 2002; Slavin & Madden, 2011). Future studies might need to include both treatment-inherent and treatment independent measures for students' reading comprehension (Slavin & Madden, 2011).

The results of the study must be interpreted with caution due to the methodological limitations involved. Firstly, the results are based on data with a considerable attrition rate, which may have had a negative impact on the precision of estimates of standard errors in the regression analyses (Snijders & Bosker, 2012). Further analyses of patterns of missing values revealed an attrition bias as students from classes characterised by lower achievement and lower peer acceptance and friendships did not participate in the study. The teachers of these classes may have needed more support when implementing the intervention.

Secondly, the recruitment of teachers focused on teachers who were interested in the CL approach and the subsequent randomisation led to the inclusion of teachers

who were knowledgeable and experienced in the CL approach in the control group. A more sophisticated procedure with random sampling of teachers before recruitment efforts may have been preferable. In addition, a quasi-experimental design including teachers not interested in CL approach might be more viable with regard to the aim of the study. The combination of the previously reported lack of test equivalence and challenges in implementation limits the possibility of drawing conclusions as to the effect of CL-RC intervention on students' reading comprehension.

Acknowledgements

We would like to express our gratitude to Caroline Liberg, Åsa af Geijerstam and Cecilia Bergwall for their support in the design of the study. We further appreciate teacher students' efforts and engagement in grading of tests. We would also like to express our gratitude to Rights and Brands Moomin and Bonnier Carlsen Publisher for permission to use the books *Invisible Child* and *The Breadwinner*. This work was supported by the Swedish Research Council under Grant 2016-04679.

Author biographies

Nina Klang, Associate Professor of Education, Special Needs Education, Uppsala University and Mälardalen University. Nina teaches courses in teacher education and special needs education. Her research focuses on instruction for students with disabilities from the perspective of inclusion.

Johanna Åsman is a mother tongue teacher and currently works as a special needs educator. She graduated from the programme of special needs education at Uppsala University. Johanna's research interest is reading comprehension with a focus on how students talk about texts.

Maria Mattsson is a mother tongue teacher and currently works a special needs educator. She graduated from the programme of special needs education at Uppsala University. Maria's research interests include reading learning, inclusion and differentiated instruction.

Claes Nilholm, Professor of Education, special needs education, Uppsala University. Claes Nilholm is a professor of education at Uppsala University. His main research interests are inclusive education and theoretical issues in special needs education.

Jenny Wiksten Folkeryd, Professor of Education, Uppsala University. Her research focuses educational linguistics including reading and writing development, language across the curriculum, text analysis and literacy assessment.

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Boardman, A. G., Vaughn, S., Buckley, P., Reutebuch, C., Roberts, G., & Klingner, J. (2016a). Collaborative strategic reading for students with learning disabilities in upper elementary classrooms. *Exceptional Children*, 82(4), 409–427. <https://doi.org/10.1177/0014402915625067>
- Boardman, A. G., Buckley, P., Vaughn, S., Roberts, G., Scornavacco, K., & Klingner, J. K. (2016b). Relationship between implementation of collaborative strategic reading and student outcomes for adolescents with disabilities. *Journal of Learning Disabilities*, 49(6), 644–657. <https://doi.org/10.1177/0022219416640784>
- Buchs, C., Filippou, D., Pulfrey, C., & Volpé, Y. (2017). Challenges for cooperative learning implementation: Reports from elementary school teachers. *Journal of Education for Teaching: Cooperative Learning: Exploring Challenges, Crafting Innovations*, 43(3), 296–306. <https://doi.org/10.1080/02607476.2017.1321673>
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage learning.
- Department of Nordic Languages. (2012). *Ämnesprov, läsår 2011/2012. Svenska och svenska som andraspråk* [Subject tests, 2011/2012. Swedish and Swedish as a second language]. Swedish National Educational Agency.
- Department of Nordic Languages. (2013). *Ämnesprov, läsår 2012/2013. Svenska och svenska som andraspråk* [Subject tests, 2012/2013. Swedish and Swedish as a second language]. Swedish National Educational Agency.
- Department of Nordic Languages. (2014). *Ämnesprov, läsår 2013/2014. Svenska och svenska som andraspråk* [Subject tests, 2013/2014. Swedish and Swedish as a second language]. Swedish National Educational Agency.
- Demos, A., & Salas, S. (2021). *A language, not a letter: Learning statistics with R*. <https://ademos.people.uic.edu/index.html>
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3), 327–350. <https://doi.org/10.1007/s10464-008-9165-0>
- Ellis, D. (2012). *Den osynliga flickan* [The invisible girl]. Bonnier Carlsen Publisher.
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.
- Ghaith, G. M. (2018). Teacher perceptions of the challenges of implementing concrete and conceptual cooperative learning. *Issues in Educational Research*, 28(2), 385–404.
- Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric statistical inference* (5th ed.). Routledge.
- Gillies, R. M. (2016). Cooperative learning: Review of research and practice. *Australian Journal of Teacher Education*, 41(3), 39–54. <https://doi.org/10.14221/ajte.2016v41n3.3>
- Gillies, R. M., & Boyle, M. (2010). Teachers' reflections on cooperative learning: Issues of implementation. *Teaching and Teacher Education*, 26(4), 933–940. <https://doi.org/10.1016/j.tate.2009.10.034>
- Jansson, T. (1962). *Berättelsen om det osynliga barnet* [The story of the invisible child]. © Tove Jansson 1962 Moomin Characters™.
- Johnson, D. W., Johnson, R. T., & Holubec, E. J. (2009). *Circle of learning: Cooperation in the classroom*. Interaction Book Company.
- Kim, J. S., Burkhauser, M. A., Quinn, D. M., Guryan, J., Kingston, H. C., & Aleman, K. (2017). Effectiveness of structured teacher adaptations to an evidence-based summer literacy program. *Reading Research Quarterly*, 52(4), 443–467. <https://doi.org/10.1002/rrq.178>
- Klang, N., Olsson, I., Wilder, J., Lindqvist, G., Fohlin, N., & Nilholm, C. (2020). A cooperative learning intervention to promote social inclusion in heterogeneous classrooms. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.586489>
- Klang, N., Asman, J., Mattsson, M., & Wiksten Folkeryd, J. (forthcoming). Envisioning texts through cooperative learning – peer-led and teacher-led discussions. Submitted to *Nordic Journal of Literacy Research*.
- Klingner, J. K., Vaughn, S., Arguelles, M. E., Tejero Hughes, M., & Ahwee Leftwich, S. (2004). Collaborative strategic reading: “Real-world” lessons from classroom teachers. *Remedial and Special Education*, 25(5), 291–302. <https://doi.org/10.1177/07419325040250050301>
- Law, Y. (2014). The role of structured cooperative learning groups for enhancing Chinese primary students' reading comprehension. *Educational Psychology*, 34(4), 470–494. <https://doi.org/10.1080/01443410.2013.860216>

- Lee, S. H., & Tsai, S.-F. (2017). Experimental intervention research on students with specific poor comprehension: A systematic review of treatment outcomes. *Reading and Writing, 30*(4), 917–943. <https://doi.org/10.1007/s11145-016-9697-x>
- Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>
- Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education, 6*(1), 1–11. <https://doi.org/10.1186/s40536-018-0061-2>
- Moore, B. A., Boardman, A. G., Lasser, C. J., Schmidt, K. M., Smith, C. E., & Schwarz, V. S. (2019). Integrating evidence-based reading practices into middle-school content instruction: Exploring a facet of sustainability. *Teacher Development, 23*(3), 287–306. <https://doi.org/10.1080/13664530.2019.1610038>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016. International results in reading*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA). <http://pirls2016.org/download-center/>
- Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117–175. https://doi.org/10.1207/s1532690xci0102_1
- Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices – past, present, and future. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13–69). Lawrence Erlbaum.
- Pontoppidan, M., Keilow, M., Dietrichson, J., Solheim, O. J., Opheim, V., Gustafson, S., & Andersen, S. C. (2018). Randomised controlled trials in Scandinavian educational research. *Educational Research, 60*(3), 311–335. <https://doi.org/10.1080/00131881.2018.1500194>
- Pressley, M., & Gaskins, I. W. (2006). Metacognitively competent reading comprehension is constructively responsive reading: How can such reading be developed in students? *Metacognition and Learning, 1*(1), 99–113. <https://doi.org/10.1007/s11409-006-7263-7>
- Puzio, K., & Colby, G. T. (2013). Cooperative learning and literacy: A meta-analytic review. *Journal of Research on Educational Effectiveness, 6*(4), 339–360. <https://doi.org/10.1080/19345747.2013.775683>
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*(4), 479–530. <https://doi.org/10.3102/00346543064004479>
- Sarkar, D. (2008). *Lattice: Multivariate data visualization with R*. Springer.
- Schielzeth, H., Dingemans, N. J., Nakagawa, S., Westneat, D. F., Allogue, H., Teplitsky, C., Réale, D., Doehrmann, N. A., Garamszegi, L. Z., Araya-Ajoy, Y. G., & Sutherland, C. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution, 11*(9), 1141–1152. <https://doi.org/10.1111/2041-210X.13434>
- Schleicher, A. (2019). *PISA 2018. Insights and interpretations*. Organization for Economic Cooperation and Development. <https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.
- Slavin, R. E. (2015). Cooperative learning in elementary schools. *Education, 43*(1), 5–14. <https://doi.org/10.1080/03004279.2015.963370>
- Slavin, R., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness, 4*(4), 370–380. <https://doi.org/10.1080/19345747.2011.558986>
- Snyder, L. (2010). Reading expository material: Are we asking the right questions? *Topics in Language Disorders, 30*(1), 39–47. <https://doi.org/10.1097/TLD.0b013e3181d098b3>
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. Sage.
- Stevens, R., Slavin, R., & Farnish, A. (1991). The effects of cooperative learning and direct instruction in reading-comprehension strategies on main idea identification. *Journal of Educational Psychology, 83*(1), 8–16. <https://doi.org/10.1037//0022-0663.83.1.8>
- Swedish National Educational Agency. (2014). *Support measures in education – on leadership and incentives, extra adaptations and special support* [Stödinsatser i utbildningen – om ledning och stimulans, extra anpassningar och särskilt stöd]. Swedish National Educational Agency.
- Swedish National Educational Agency. (2017). *The system framework for national tests by Swedish National Educational Agency* [Skolverkets systemramverk för nationella prov]. Typoform.

Intervention combining cooperative learning and instruction

- Swedish National Educational Agency. (2021). *Statistics on preschool, school and adult education*. Retrieved from <https://www.skolverket.se/skolutveckling/statistik/sok-statistik-om-forskola-skola-och-vuxenutbildning?sok=SokC&verkform=Grundskolan&omrade=Nationella%20prov&lasar=2013%2F14&run=1> on the 10 of September 2021.
- Tengberg, M. (2017). National reading tests in Denmark, Norway, and Sweden: A comparison of construct definitions, cognitive targets, and response formats. *Language Testing*, 34(1), 83–100. <https://doi.org/10.1177/0265532215609392>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.

Appendix

An example of instructional activities in the CL-RC intervention

Reading comprehension	Cooperative learning
Predicting Teacher poses a question: “What do you think the story is going to be about?” (two pictures from the book are on the whiteboard)	First, the students write down their thoughts. Second, students work in dyads and take turns in sharing their thoughts. Both students should be prepared to give an account of the dyad’s thoughts in class.
Asking and answering questions The teacher poses a question to the class: “What explanation have we got for Ninni becoming invisible?”	First, the students share their thoughts in a dyad. Second, two dyads build a group and share their thoughts.
Clarifying The students read a passage from the novel, underline and discuss difficult words and phrases.	Students work in dyads. They have two roles: one student reads the text aloud and the other underlines words. They write down the underlined words on pieces of paper. All the pieces of paper are shared by the group of two dyads in common discussion.
Summarizing The teacher presents a task: “Describe the little My. Who is she? What is her role in the story? Why?”	The students are given a big sheet of paper, divided into four parts on each side and a part in the middle. Each student writes down his or her thoughts in one part on the side of the sheet. After discussion, the group’s secretary formulates the group’s common thoughts in the middle of the sheet.