Research Papers in Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rred20

Effects of individualised and small-group guided oral reading interventions on reading skills and reading attitude of poor readers in grades 2-4

Ron Oostdam[ab], Henk Blok[c] & Conny Boendermaker[d]

[a] Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands

[b] Knowledge Centre for Teaching and Education, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands

[c] Kohnstamm Institute, University of Amsterdam, Amsterdam, The Netherlands

[d] Teachers College Windesheim, Almere, The Netherlands

Published online: 04 Sep 2014.

To cite this article: Ron Oostdam, Henk Blok & Conny Boendermaker (2014): Effects of individualised and small-group guided oral reading interventions on reading skills and reading attitude of poor readers in grades 2-4, Research Papers in Education, DOI: 10.1080/02671522.2014.953195

To link to this article: http://dx.doi.org/10.1080/02671522.2014.953195

Research Papers in Education, 2014
http://dx.doi.org/10.1080/02671522.2014.953195

Routledge
Taylor & Francis Group

Effects of individualised and small-group guided oral reading interventions on reading skills and reading attitude of poor readers in grades 2–4

Ron Oostdam[a,b]*, Henk Blok[c] and Conny Boendermaker[d]

[a]*Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands;* [b]*Knowledge Centre for Teaching and Education, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands;* [c]*Kohnstamm Institute, University of Amsterdam, Amsterdam, The Netherlands;* [d]*Teachers College Windesheim, Almere, The Netherlands*

(*Received 21 January 2014; final version received 6 August 2014*)

To assess the efficacy of guided oral reading as a remedy for low-achieving readers, two experiments were conducted in the early grades of primary school. In the first, poor-reading students were randomly divided between two treatment groups and a control group. In treatment groups, the intervention was delivered one-to-one, either in a repeated reading (RR) or in a continuous reading format, depending on how often students practised with the same text. In the second experiment, poor-reading students were randomly divided between a group-based guided oral reading condition and a control condition. Groups comprised three students who received an integrated version of continuous and RR. Measures included tests for fluency, reading comprehension (RC), vocabulary (VOC) and reading attitude (RA). The results demonstrate that both the individual and the group variants of guided oral reading are effective for improving fluency and RA. Transfer effects on RC and VOC could not be established.

Keywords: poor reading; guided oral reading; reading fluency; reading attitude; experimental research; primary education

Introduction

Guided oral reading is an established intervention for poor readers in the early grades of primary school. Students who read or reread passages aloud as they receive guidance and feedback are likely to become better readers. As an instructional technique, guided oral reading comes in many formats, depending on, for instance, the amount of rereading, the person offering guidance (teachers, teaching assistants, parents, tutors or peers) and the group size (class, small group, pairs or individuals).

This article focuses on guided oral reading as remedial practice for students who lag behind in reading fluency. These 'dysfluent' readers are likely to encounter difficulties in developing reading comprehension (RC) and vocabulary (VOC). Making sense of a text presupposes virtually effortless and flawless decoding. If reading problems persist, students run the risk of being held back a grade or referred to special education.

*Corresponding author. Email: r.j.oostdam@hva.nl

The National Reading Panel (2000) in the USA concluded that guided oral reading procedures have a consistent and positive impact on word recognition, fluency and RC. The Panel looked at 14 studies published between 1979 and 1999 and reported an average effect size of 0.41. It should be noted that some of the studies included non-impaired readers. However, sample sizes were generally too small to carry out further analyses pertaining to the difference between non-impaired and poor readers.

A much debated question concerns the effectiveness of repeated reading (RR) vs. the continuous reading (CR) format. In RR, students read a text several times – either a fixed number of times or until they meet a predefined mastery criterion (in terms of reading speed or accuracy). In CR, students read a new text without rereading, regardless of their reading speed or accuracy. Some scholars suggest that it is increased reading across a range of materials that generates improved reading, rather than the rereading of text (Allington 2001; Kuhn 2004).

Kuhn and Stahl (2003) reviewed 58 studies on repeated or CR, and found that neither format offered a clear advantage. They concluded that to be effective, an intervention must contain two critical elements: practice and support. In other words, students need guided practice. Wexler et al. (2008) also compared the two formats. They reviewed the results of 19 studies, in this case with poor readers in the age range of 11–21 years old. The subjects presumably included many disappointed students with low levels of self-confidence, who had struggled for many years to become competent readers. For this particular type of student, Wexler and colleagues (op. cit.) found no systematic advantage to be gained from either repeated or CR practice. Therrien (2004) reviewed 18 studies on RR, which corroborated the effectiveness on reading fluency of this particular remedial format. O'Connor, White, and Swanson (2007) asked whether repeated and CR have differential effects for students with and without learning disabilities. They sampled poor readers in grades 2 and 4 with (n = 17) and without (n = 20) learning disabilities, and assigned them randomly to a RR condition, a CR condition or a control condition. No significant differences between practice conditions were found for poor readers or students with learning disabilities.

Although in general most studies comparing repeated and CR reveal no differences in terms of effectiveness on reading fluency, there is still less understanding about the differential effects of both interventions by grade. For practical purposes, it is important to know whether repeated or CR remain equally effective for poor readers in the different grades. Besides, there might be other reasons for preferring one practice condition to the other. One such reason relates to the reading attitude (RA) of children. For poor readers, reading is an activity that requires much effort and motivation. In most literature reading, motivation is closely connected with attitude towards reading. Children with a more positive attitude are generally more motivated to read. Several scholars have reported developmental declines in RA or reading interest for students in primary schools (McKenna, Kear, and Ellsworth 1995; Wigfield et al. 1997). According to McKenna and colleagues (op. cit.), RA – specifically the recreational reading component – declined more sharply for less able readers than for more able readers. A relevant question, therefore, is whether RR and CR affect the RA of poor readers in a similar way. Theoretically, RR may have the advantage that it offers the students success experiences. After a certain amount of practice with the same passage, the student is able to give a near perfect rendering of the text. On the other hand, RR may be boring and it may unintentionally emphasise that reading is a skill that can only be acquired after many unsuccessful attempts. Alternatively, CR may lead to the inverse experience: more laborious

reading, but also less boredom. It remains to be seen which of the two practice conditions affects RA most favourably.

Another argument with respect to the choice between the repeated and the CR format relates to possible transfer effects on RC and VOC. There are theoretical reasons to expect such effects. The more automatic decoding becomes, the more attention readers can pay to the meaning of the text (LaBerge and Samuels1974; Perfetti 1985). Indeed, Klauda and Guthrie (2008) found positive relations between measures of reading fluency and RC, although their cross-sectional data sampling precluded a causal interpretation. Possible transfer effects from reading fluency training on RC are discussed in the review by Kuhn and Stahl (2003): 'Generally, where an increase in fluency was found, there was also an increase in comprehension' (p. 9). Therrien (2004) also paid specific attention to transfer effects on RC in his review of RR interventions. He reports an average effect size of 0.25 for RC, compared with a mean effect size of 0.50 on fluency measures. It should be noted that these estimates are based on within-designs, that is, designs without a control group. They might not be valid for situations in which effect sizes are expressed as a difference between a treated group and an untreated control group.

As mentioned, O'Connor, White, and Swanson (2007) employed a between-design, in which students were randomly divided between two intervention groups (repeated or CR) and an untreated control group. Their hypotheses regarding transfer effects on RC and VOC were dependent on the intervention format. They expected CR to have a stronger effect on RC and VOC than RR. Students who read a wide range of materials are exposed to a larger number of unique words whose meanings are introduced across multiple contexts, while students in the RR condition encounter fewer different words and are less challenged to employ varied strategies for text comprehension. However, O'Connor, White and Swanson (op. cit.) expected no such differences regarding reading fluency. With regard to fluency measures, both practice formats should be equally effective. Indeed, they found repeated and CR to be equally effective with respect to reading fluency. Effect sizes for reading fluency measures ranged from 0.84 to 1.03, but the expected differential effects of repeated vs. CR on comprehension and VOC was not observed. Effect sizes for RC amounted to about 1.00, and for VOC to about 0.50. These latter estimates are remarkably large compared to the outcomes of the above-mentioned research syntheses. For instance, Therrien (2004) reported an average effect size of 0.25 for comprehension measures. Although the length and the intensity of their intervention did not diverge from what is usually found in other studies, the authors do not attempt to explain these large fluency and comprehension effects. Remediation was given three times a week for ten consecutive weeks, and comprised a total of about ten hours of reading help. The one-to-one setting in which the help was offered may have contributed to the impressive effectiveness, but this is only a supposition.

Three recent studies by Vadasy and Sanders (2008a, 2008b, 2009) may shed further light on the potential transfer of fluency training effect on RC and VOC skills. In all three studies, the Quick Reads fluency programme (Hiebert 2003) represented the treatment variable, while the students in the control condition relied on regular classroom instruction. Students in the Quick Reads condition received supplemental fluency tutoring instruction in pairs from trained paraeducator tutors (i.e. teaching assistants) or teachers. The supplemental reading instruction ranged from an average of 25 h in the 2009 study to 36 h in the 2008a study. Thirty-minute reading sessions were held four days per week. Quick Reads prescribes a RR procedure with short,

non-fiction texts on school grade-appropriate topics. Each text was read four times, with students taking turns. The tutors attended to fluency aspects, introduced new VOC and facilitated comprehension, and each text was accompanied by two comprehension questions. At least two texts had to be covered in each session. Participating students were low to moderate achieving readers (from the 10th to 60th percentile) from grades 2 and 3 in the 2008b and 2009 studies, and from grades 4 and 5 in the 2008a study.

In these three studies, the estimated effect sizes were generally much smaller than in the study by O'Connor and colleagues, while the supplemental remedial instruction was two or three times longer (a total of 10 h in the study by O'Connor and colleagues, compared with 25–36 h in the studies by Vadasy and Sanders). In the domain of reading fluency, effect sizes ranged from practically zero in the 2008a study to a statistically significant 0.40 in the two other studies. Compared to other studies the effect sizes were also more modest in the domain of comprehension (between 0.27 and 0.50) and VOC (0.42). In their 2008b study, Vadasy and Sanders found no significant effect sizes on the Grey Oral Reading Test for RC, while the effect size for comprehension in the 2009 study amounted to 0.30 (significant at the 0.05 level). These three studies lead to the conclusion that it cannot be stated with certainty that supplemental guided oral reading instruction produces transfer effects in the domain of RC and VOC. Even more remarkably, the study by Vadasy and Sanders (2008a) failed to demonstrate effects on reading fluency, supposedly, the authors suggest, because the sampled students needed more explicit instruction in alphabetic and decoding efficiency. However, in their later studies with younger, less advanced readers, in which they used the same Quick Reads programme, the authors managed to demonstrate significant growth in reading fluency compared to an untreated control group.

The present study focuses on the possible effects of the two fluency intervention formats (continuous or RR) on reading fluency, RA, comprehension and VOC. Grade was added to the design as a further factor, also in view of Vadasy and Sanders' (2008a) suggestion that grade level might have a moderating effect. The research questions were as follows:

(a) Are the two intervention formats – CR and RR – effective methods for improving the reading fluency of poor readers?
(b) Do these two formats induce favourable transfer effects, specifically on RA, comprehension and VOC?
(c) Is one of the formats better than the other for reading fluency or the transfer variables?
(d) Is there a differential effect of the interventions by grade?

Two experiments are described in this article. In the first, the intervention was delivered one-to-one, while in the second it was delivered to small groups of three students. One-to-one remediation obviously makes large demands on a school's resources. A question of practical importance is therefore whether the same results can be achieved when guided oral reading is offered in a group setting. Both experiments included a non-treated control group. The aim was to compare the outcomes of the experiments in order to establish whether intervention in small groups results in a loss of effectiveness.

Experiment 1

Experiment 1 represents a true experimental design with five measurements: a pre-test, two mid-tests, a post-test and a retention test. Mid-test 1 was given after one-third of the reading sessions had been completed, mid-test 2 after two-thirds had been completed. Retention was measured two months after the post-test. Dependent measures were reading fluency, RC, VOC and RA. Within classrooms, students were assigned to one of three conditions: a CR condition, a RR condition or an untreated control condition. The intervention was delivered one-to-one.

Subjects

Experiment 1 was conducted in 2009–2010 in eight primary schools in the Netherlands. School boards were invited to recommend schools for participation. Only schools with at least six poor readers in each of three grades (grades 2 through 4) were selected for the study. Poor readers were defined as readers who fall within the lowest quintile on a grade-normed test of reading fluency (see below for details).

Blocking was used to assign students to intervention or control conditions. The six poor readers per class were matched in trios according to reading fluency. Students in each trio were randomly assigned to one of the three conditions (CR, RR or control). Matching increased the statistical precision, while randomisation was essential for the internal validity of the experiment.

At the start of the experiment, there were 143 student participants: 18 from each of seven schools and 17 from one school. Before pre-testing began, parental consent was obtained for every student to participate in the study. All 143 students participated through to the end of the experiment. However, data on 17 students originating from six schools are not presented here, because these students did not meet the criterion for poor reading at the pre-test. This criterion was defined by a predefined grade-normed score on the *Technisch Lezen 345678* [Fluency of Word Reading 345678; hereafter FWR 345678] test (Aarnoutse and Kapinga 2005), in which students have to read aloud a list of words as quickly as possible. The score is the number of words read correctly in 90 s. Of the 126 remaining students, 43 were in the CR, 43 in the RR and 40 in the control condition.

Students in the three conditions were more or less uniformly distributed between grades and sex (see Table 1). The percentage of grade-repeating students ranged from 10 to 23%. The percentage of students with one or more dyslexic family members, as reported by the class teacher, ranged from 15 to 26%. The percentage of students for whom Dutch is not the mother tongue ranged from 7 to 19%. Mean age at the start of the experiment was eight years and nine months in all three conditions. The mean differences between students in the three conditions were not statistically significant for any of the variables (alpha set at 0.05), nor were there significant differences on the FWR 345678 test or a test for rapid automatised naming (RAN). The RAN test asks students to name, as quickly as possible, 50 graphemes from a sheet that contains only five graphemes. The score is the number of seconds that a student takes to do this. It has been argued that RAN, particularly of alphanumeric stimuli, is strongly associated with reading success, especially among poor readers (e.g. Bowers and Newby-Clark 2002; Bowers and Wolf 1993; Wolf 1991; Wolf, Bowers, and Biddle 2000; Wolf et al. 2002).

Table 1. Experiment 1: student characteristics; means and standard deviations (between parentheses) in the three conditions.

Measure	Repeated reading (n = 43)	Continuous reading (n = 43)	Control (n = 40)
Proportion in grade 2	0.33 (0.47)	0.28 (0.45)	0.27 (0.44)
Proportion in grade 3	0.33 (0.47)	0.33 (0.47)	0.37 (0.48)
Proportion in grade 4	0.35 (0.48)	0.38 (0.49)	0.37 (0.48)
Proportion of boys	0.49 (0.51)	0.64 (0.49)	0.56 (0.50)
Proportion of students who repeated a grade	0.23 (0.43)	0.10 (0.30)	0.22 (0.42)
Proportion of students with one or more dyslexic family members	0.26 (0.44)	0.19 (0.40)	0.15 (0.36)
Proportion of students for whom Dutch is not the mother tongue	0.12 (0.32)	0.19 (0.40)	0.07 (0.26)
Age at pre-test (months)	105 (11)	105 (10)	105 (13)
FWR 345678 at pre-test	39.37 (17.47)	40.30 (17.32)	40.78 (17.58)
RAN at pre-test (seconds)	38.28 (9.10)	37.86 (10.74)	36.30 (10.78)

Note: FWR = fluency of word reading; RAN = rapid automatised naming.

Measures

Five dependent variables were used to assess treatment effects: FWR, fluency of text reading (FTR), RC, VOC and RA. All the measures that were chosen utilised standardised test taking procedures.

Fluency of word reading

Students took the *Drie-Minuten-Toets* [Three-minutes Test] (DMT; Jongen and Krom 2009). The DMT is a standardised grade-normed test that comprises three sheets with word lists of increasing difficulty (150 words per sheet, arranged in five rows of 30 words each). Students are instructed to read aloud the words on each sheet as accurately and as quickly as possible within one minute. The raw score is the number of words read correctly in one minute. To prevent memory effects, the DMT has three parallel versions. The versions were used alternately, starting with version A at the pre-test. Correlations between the raw scores on the three sheets ranged from 0.92 to 0.96 at the pre-test, indicating that the DMT has high reliability. The three raw scores (one for each of the sheets) were totalled and divided by three; thus, the DMT score represents the number of words correctly read per minute. The DMT was administered five times (pre, mid 1, mid 2, post and retention).

Fluency of text reading

Students took the AVI (Jongen and Krom 2009), a standardised and grade-normed test of FTR. It consists of 11 short passages (100–240 words) of increasing difficulty that students have to read aloud as accurately and as quickly as possible. The number of passages actually mastered (i.e. read within the time and accuracy limits prescribed in the test manual) defines the student's score. The score ranges from 0 (unable to master even the easiest passage) to 11 (indicating that the student masters the most difficult passage). To prevent memory effects, the AVI has two parallel versions. These were alternated, starting with version A at the pre-test. The reliability

(Cronbach's alpha) of the AVI was 0.96 (Krom et al. 2010). The correlation with the DMT score was 0.93 (this sample at pre-test). Both tests combine reading rate and reading accuracy in a single measure. The AVI was administered three times (pre, post and retention).

Reading comprehension

RC was measured with the grade-normed and standardised *Begrijpend Lezen* [Reading Comprehension] test (Krom, Van Berkel, and Jongen 2006–10). RC requires students to read text passages silently and to answer inference questions about the texts. A typical test booklet contains between five and ten passages and 25 multiple choice questions. Test booklets are grade specific. Care was taken to ensure that students took different test booklets at pre- and post-tests. The raw score is the number of correct answers. Raw scores are calibrated among different booklets on a scale ranging from −70 to 120. Reliability estimates (internal consistency) of the test range from 0.76 to 0.89. The RC was administered twice (pre and post).

Vocabulary

VOC was measured with the Dutch version of the Peabody Picture VOC Test (PPVT-III-NL; Schlichting 2005), which is a test of receptive VOC. The raw score was converted into an age-normed quotient score, with an average of 100 and a standard deviation of 15, with age cohorts spanning three months. Reliability coefficients (internal consistency) range from 0.93 to 0.95 for students in the designated grade range (grades 2 through 4). The PPVT was administered twice (pre and post).

Reading attitude

RA was operationalised using a Likert scale with seven items that were specifically developed for this experiment. For example, item 1 reads as follows: 'There are many books I should like to read' [*Er zijn veel boeken die ik graag wil lezen*]. The answer scale has three values, with a positive attitude corresponding to a higher value (1 = not true, 2 = don't know, 3 = true). The raw scores ranged from 7 to 21. Cronbach's alpha ranged from 0.70 (at pre-test) to 0.74 (at post-test). Test–retest reliability, estimated as the correlation between pre- and post-test, was 0.66. As in this case the treatment might have had an effect on RA, this correlation should be interpreted as a lower boundary. The test was administered three times (pre, post and retention).

Interventions

Each of the eight participating schools appointed a part-time (four days a week) teaching assistant to implement the intervention. The assistants were completely crossed with the intervention, implying that each assistant implemented each intervention condition (CR, RR) in his or her school, which allowed possible differences between teaching assistants to be controlled for. The teaching assistants were responsible for administering the tests and holding reading sessions.

Continuous reading

Students in this condition practised reading over 12 consecutive weeks in 20-min sessions, one-to-one with the teaching assistant, four times a week (a total of 48 individual sessions). To accommodate oral reading, the sessions were held in a separate room. Care was taken to ensure that reading sessions did not coincide with the general class instructional reading time. Reading sessions were meant to provide students with additional reading practice, supplementary to classroom practice. Students were allowed to choose their own books, as text difficulty appears not to be related to reading growth (O'Connor, Swanson, and Geraghty 2010). To this end, the teaching assistants were provided with a supply of children's books that represented varying reading levels appropriate for poor readers in the designated grades. Care was also taken to include an adequate number of non-fiction books in the collection, as experience has shown that many poor readers prefer non-fiction to fiction. Students in this condition read a new text every session. The teaching assistants provided attractive and varying instructional formats, including simultaneous reading (student and teaching assistant read aloud together, the latter adjusting his or her reading pace to that of the student), taking turns (the teaching assistant ensuring that the student reads silently when the assistant is reading) and reading to the teaching assistant (the assistant does not read; he or she only listens). The teaching assistant supervised the time-on-task to ensure that sessions were devoted to reading for approximately 20 min. Feedback was provided after every sentence, following the pause, prompt, praise strategy, which earlier research has shown to be a highly significant factor for success in improving reading (Merrett and Thorpe 1996; Wheldall, Merrett, and Colmar 1987). The teaching assistant also ensured that students comprehended what they read by, when appropriate, initiating discussions about the meaning of the text. As a rule, 15 min of each session were devoted to reading and five minutes were spent talking about the meaning of the text. The intervention totalled 16 h per student (48 sessions, 20 min per session).

Repeated reading

Students in this condition received the same treatment as students in the CR condition, with one important difference: they practised reading a text passage from their books until their reading was more or less fluent. RR students chose their books in the same manner and from the same book supply as students in the CR condition. Each session, a text passage was selected with an approximate reading length of approximately four minutes, so that the fragment could be read three or four times in that session. The student continued to read the fragment in the following sessions a maximum of 12 times, or until the teaching assistant judged the reading to be more or less fluent – taking both speed and accuracy into consideration – at which point the student continued with a new fragment of text.

Control

Students in the control condition participated only in the measurement programme. However, for ethical reasons, they were offered CR or RR after post-testing. This arrangement, which unfortunately precludes a valid comparison at the retention stage, was indispensable in order to keep the experiment within the boundary of one

complete school year. Control students followed the school's regular reading programme. This programme could vary from school to school, according to interviews with teachers. Regular classroom practice might include daily independent reading for 20–40 min, whole-class or small-group instruction by the class teacher, peer tutoring or computer training (flashed word reading). Remedial instruction in a one-to-one setting by a reading specialist was discouraged, both during and outside school hours.

Procedures

The teaching assistants were given a written protocol detailing the procedures for administering the tests and holding the reading sessions. The subjects covered included guiding students on book choice, how to employ the pause, prompt, praise technique, the importance of varying instructional formats, how to handle any declining commitment among students, monitoring text comprehension and modelling prosody. The assistants were provided with stopwatches to monitor time-on-task. They were trained in all the relevant procedures during a three-day course at the start of the experiment. They also received on-the-job coaching during two visits by one of the researchers. During each visit, one CR and one RR session were videotaped and then discussed.

The teaching assistants kept a log for each student in CR and RR. Reading sessions were pre-numbered from 1 to 48 to ensure that every student received 48 sessions. Students were expected to catch up with missed sessions (for instance, due to illness). Student logs also helped to determine measurement occasions, that is, mid-test 1 after session 16, mid-test 2 after session 32 and post-test after session 48. For each session, notes were kept about the text that was read, how often a text was practised (only in RR) and whether there were any deviations from the protocol. All logs were collected by the researchers after session 48 and checked. The conclusion is that both interventions were realised with a high degree of fidelity.

At the end of each session, students were asked to rate the effort they had invested in that particular session on a scale of 1–5 (1 = I didn't invest any effort, 5 = I invested a great deal of effort). The teaching assistants were instructed not to influence the student's choice of rating, as student ratings were meant to be the result of an independent self-appraisal. Students' effort ratings proved to be very consistent or homogeneous. Cronbach's alpha, estimated over the 48 effort ratings per student amounted to 0.95. The effort ratings indicated that the effort level was generally high. The mean level, averaged over sessions and students was 4.63 (SD = 0.61). The effort level increased from session 1 (mean = 4.15, SD = 0.78) to session 48 (mean = 4.82, SD = 0.45). There was a slight difference in overall level between CR and RR students in favour of the former (CR students: mean = 4.60, SD = 0.62; RR students: mean 4.66, SD = 0.59). The difference was statistically significant ($t[1,84] = 9.23$, $p < 0.01$), but the effect size expressed as Cohen's d is negligible ($d = 0.10$).

Analysis and results

Table 2 presents the descriptive data for each of the dependent variables for each of the conditions and measurement occasions in terms of means and standard deviation. No statistically significant differences between the three conditions were observed

Table 2. Experiment 1: means and standard deviations on measures and measurement occasions.

	Condition					
	Repeated reading (n = 43)		Continuous reading (n = 43)		Control (n = 40)	
Measure (min–max)/measurement occasion	M	SD	M	SD	M	SD
Fluency of word reading (0–150)						
Pre	31.67	16.04	31.65	16.06	33.53	16.24
Mid 1 (after 16 sessions)	37.40	17.28	37.40	16.06	37.50	16.71
Mid 2 (after 32 sessions)	42.53	17.60	42.35	17.05	41.20	16.83
Post	47.19	19.03	47.47	17.71	44.40	16.82
Retention	51.21	18.39	51.23	18.13	–[a]	–[a]
Fluency of text reading (0–11)						
Pre	2.81	2.19	2.79	2.02	3.05	2.03
Post	4.14	1.93	4.42	2.02	4.00	1.85
Retention	4.98	2.11	5.16	2.32	–[a]	–[a]
Reading comprehension (−70 to 120)						
Pre	2.73	17.75	3.74	19.44	4.09	17.80
Post	11.10	14.69	11.98	18.16	13.27	17.97
Vocabulary (56–144)						
Pre	95.65	13.82	98.69	11.15	95.82	10.89
Post	98.76	11.94	102.60	11.88	98.90	10.40
Reading attitude (7–21)						
Pre	16.09	3.76	15.63	3.93	16.95	3.25
Post	16.95	3.54	16.93	3.71	16.23	3.81
Retention	17.77	2.81	16.91	3.52	–[a]	–[a]

[a]Retention not measured in the control group.

on any of the dependent variables at the pre-test (alpha level set at 0.05). As stated in the subjects section, all students belonged to the lowest quintile on a test for FWR. Retardation on the other dependent measures was less prominent. On the RC measure, 41% of the students scored within the lowest quintile, while on the VOC test only 19% did so. Because the RA measure was not a normed test, it is unknown what percentage of students suffered from low RA.

To assess whether the interventions affected the development of reading skills and attitude, a longitudinal multilevel model was formulated and estimated in MLwiN, version 2.10 (Rasbash et al. 2009), acknowledging the repeated measures and hierarchical character of the database. The basic model had three levels: measurement occasions (four for FWR, two for the other measures), students (n = 126) and schools (n = 8). In the fixed part, the following main effects were entered: measurement occasion (scaled 0, 1, 2 and 3 in the case of FWR, and 0 and 1 in case of the other four dependent measures), two dummies representing the CR and RR conditions contrasted with the control condition, and two dummies representing grade (grades 2 and 4 contrasted with grade 3). Furthermore, dummies were entered for the two- and three-way interactions between measurement occasions, conditions and grades. Two sets of interaction dummies were of special interest in view of the research questions. The condition*measurement occasion dummies represent the hypothesis that students in different conditions show differential growth. The condition*measurement occasion*grade dummies represent the hypothesis that growth is

dependent on both condition and grade. Two other predictors were added to the fixed part of the model, as they appeared to be significantly related to the growth curves of students. Having one or more dyslexic family members had a negative effect on the intercept of the growth curve. This finding is in line with studies that show that children with dyslexic parents run a higher risk of becoming dyslexic (e.g. Van Bergen et al. 2011). Likewise, RAN had a negative effect on the intercept, confirming the view that RAN is negatively associated with reading success (e.g. Bowers and Newby-Clark 2002; Bowers and Wolf 1993; Wolf 1991; Wolf, Bowers, and Biddle 2000; Wolf et al. 2002). In the random part of the basic model, measurement occasion was treated as a random variable at the school and student level, implying that the slope of the growth curves could vary between schools and students.

The first step in the evaluation of the models, one for each of the five measures, was to estimate a simple model containing only main effects (including the predictors). Parameter testing was done by comparing the relevant parameter estimates with their standard errors. If a parameter estimate exceeded the standard error at least twice, the parameter was considered statistically different from zero. This rule of thumb corresponds with an alpha level of about 0.05. Non-significant parameters were removed from the model. In the next step, the two-way interaction dummies were added to the model. Only interaction terms that appeared to be statistically significant were retained. The three-way interaction dummies were evaluated in the same way.

Table 3 summarises the results of the model evaluations, in particular the outcomes that are relevant to the research questions. The condition*measurement occasion parameters were different from zero in the case of FWR, FTR (only the CR contrast) and RA, but not in the case of RC or VOC. This means that students in CR displayed more growth than control students, but only for both fluency measures and RA. Students in RR showed more growth than control students only on the measures for FWR and RA, although the contrast with control students on FTR is nearly significant at the 0.05 level ($p = 0.07$). It further appeared that the parameters for the three-way interaction between condition, grade and measurement occasion were not significantly different from zero for any measure, indicating that the growth per condition is independent of grade.

Retention was only measured for RR and CR students, using FWR, FTR and RA. To establish how students' reading skills developed after the intervention, the effect sizes for two periods were estimated, namely from pre-test to post-test and from post-test to retention. As the control students did not participate in the retention measurement, effect sizes could only be estimated within the conditions. Table 4 shows that student growth continued, but further growth was statistically significant only in the case of RR students on FTR. One should be aware that the two time periods differ in length: the intervention period lasted three months, while the retention measures were given two months after the intervention ended.

Criterion-referenced scores are not very useful for evaluating the treatment effects in terms of the gap between the students and their peers. It was known that the participating students were far behind their peers of the same age at the pre-test, but the extent to which the interventions had reduced this disadvantage was not known. This could be established only for FWR, as this is the only measure used here for which sufficiently detailed grade-normed scores were available. Following the appropriate conversion tables, the raw scores for this test were transformed into

Table 3. Experiment 1: outcomes of multilevel analyses for each of the five dependent variables: fixed effects and standard errors between parentheses; the control condition in grade 3 functions as the reference condition.

	Final model				
Parameter	FTR	FWR	RC	Vocab.	RA
Level 1 (Measurement Occ.)					
Intercept	44.11* (3.49)	4.00* (0.43)	n.s.	102.73* (4.68)	16.19* (1.48)
Grade 2	–14.17* (2.32)	–1.46* (0.22)	–20.01* (2.81)	n.s.	n.s.
Grade 4	15.32* (2.28)	1.92* (0.21)	10.16* (2.68)	n.s.	n.s.
RAN	–0.31* (0.08)	–0.03* (0.01)	n.s.	n.s.	n.s.
Dyslexia in family	–4.65* (1.67)	–0.42* (0.21)	n.s.	n.s.	n.s.
RR condition	n.s.	n.s.	n.s.	n.s.	n.s.
CR condition	n.s.	n.s.	n.s.	n.s.	n.s.
RR condition × Meas. Occ.	1.43* (0.43)	0.34 (0.18)	n.s.	n.s.	1.31* (0.59)
CR condition × Meas. Occ.	1.42* (0.43)	0.57* (0.18)	n.s.	n.s.	1.60* (0.59)
RR × Meas. Occ. × Grade	n.s.	n.s.	n.s.	n.s.	n.s.
CR × Meas. Occ. × Grade	n.s.	n.s.	n.s.	n.s.	n.s.
Level 2 (Student)					
Intercept	3.68* (0.47)	0.94* (0.15)	11.61* (2.11)	n.s.	n.s.

Note: n.s. = not significant; RAN = rapid automatised naming; RR = repeated reading; CR = continuous reading; FWR = fluency of word reading; FTR = fluency of text reading; RC = reading comprehension; Vocab. = vocabulary; RA = reading attitude.
*$p < 0.05$.

Table 4. Experiment 1: intra-individual effect sizes (Cohen's d using the pooled standard deviation) for two time periods, from pre- to post-test, and from post-test to retention.

Measure/condition	From pre-test to post-test	From post-test to retention
Fluency of word reading		
Repeated reading	0.89*	0.21
Continuous reading	0.93*	0.21
Fluency of text reading		
Repeated reading	0.63*	0.45*
Continuous reading	0.80*	0.37
Reading attitude		
Repeated reading	0.24	0.25
Continuous reading	0.37	−0.03

*$p < 0.05$.

norm-referenced scores, expressing the position of the student's reading skill relative to his or her peers in the same grade. FWR norm-referenced scores theoretically range from 1 to 60, expressing the reading age in terms of the number of school months reading instruction has been offered. Assuming 10 effective instructional months per school year and that reading instruction starts in grade 1, a norm score of 10 is equivalent to the reading level of a typical student at the end of grade 1. An index to indicate the success of reading instruction (SRI) was subsequently defined:

$$\mathrm{SRI}_{it} = \mathrm{NRS}_{it}/\mathrm{RIM}_{it}$$

in which SRI_{it} = success of reading instruction index of student i at time t, NRS_{it} = norm-referenced score of student i at time t, RIM_{it} = number of months student i has received reading instruction at time t.

An SRI value of around 1 indicates that the reading development of a student at that point in time coincides with the mean reading development in the population. The lower the SRI, the slower the reading development. An SRI of 0.50 indicates that a student's reading development is proceeding at half speed.

Table 5 shows that at pre-test, students in the intervention conditions displayed a rather low SRI of about 0.40. Compared to their peers, they really were poor readers. After intervention, students in both the RR and the CR conditions succeeded in reducing the gap from 0.38 at the pre-test to about 0.60 at the post-test stage, implicating that the interventions generated accelerated growth, although not enough to completely overcome the reading disadvantage.

Table 5. Experiment 1: means and standard deviations (between parentheses) of the SRI index per condition and measurement occasion, based on the FWR measure.

	Measurement occasion				
Condition	Pre	Mid 1	Mid 2	Post	Retention
Repeated reading (n = 43)	0.38 (0.16)	0.47 (0.17)	0.53 (0.19)	0.60 (0.23)	0.62 (0.26)
Continuous reading (n = 43)	0.38 (0.15)	0.46 (0.17)	0.52 (0.18)	0.59 (0.24)	0.62 (0.27)

Experiment 2

Experiment 2 answers the question whether the same efficacy can be achieved when guided oral reading is offered in a group setting with three students. In many respects, it had the same design as experiment 1, the main difference being that the remediation was offered to groups of three students instead of individually. There were two other differences: the reading sessions were prolonged to 30 min and the distinction between CR and RR was abandoned. The session duration was prolonged on the grounds that 20 min seemed too short a period to accommodate sufficiently long individual reading turns and group discussion of reading strategies and text meaning. The distinction between continuous and RR was abandoned because experiment 1 showed no differences in effectiveness between these two versions. Continuous and RR formats were integrated into a new protocol (see the intervention section).

Like experiment 1, experiment 2 included five measurements occasions: a pre-test, two mid-tests (only for the FWR test; see the measures section), a post-test and a retention test (not for RC and VOC). Again, mid-test 1 was given after one third of the planned reading sessions had been completed, mid-test 2 after two thirds had been completed. The post-test was given after 48 reading sessions. Retention was measured two months after the post-test. Dependent measures were reading fluency, RC, VOC and RA. Students were assigned to one of two conditions: a group-based guided oral reading condition (GGOR) or an untreated control condition.

Subjects

Experiment 2 was conducted in 2010–2011, again in eight primary schools. School boards were invited to recommend schools for participation, preferably schools that had at least six poor readers in each of grades 2, 3 and 4. Poor readers were defined as readers who according to national norms fall within the lowest quintile on the FWR 345678 test, which is a measure of FWR (see experiment 1, subjects section). Although an attempt was made to balance the number of students between grades, it appeared that in the participating schools, the number of poor-reading students declined over the grades, causing an overrepresentation of students in grade 2 and an underrepresentation of students in grade 4 (Table 6). Blocking was used to assign students to the treatment or the control condition. Poor readers were matched in pairs according to their scores on the FWR 345678, as measured at the pre-test. Students in each pair were randomly assigned to either the GGOR or the control condition. In schools with nine poor-reading students in a grade, a second treatment group of three students was formed. In such cases, students were matched in trios, of which two were randomly assigned to the GGOR and one was assigned to the

Table 6. Experiment 2: number of students per grade and condition.

	Grade			
Condition	Grade 2	Grade 3	Grade 4	Total
Group-based guided oral reading	36	27	21	84
Control	28	17	10	55
Total	64	44	31	139

Table 7. Experiment 2: student characteristics; means and standard deviations (between parentheses) in the two conditions.

Measure	Group-based guided oral reading (n = 84)	Control (n = 55)
Proportion of boys	0.61 (0.49)	0.36 (0.49)
Proportion of students who repeated a grade	0.14 (0.39)	0.09 (0.27)
Proportion of students with one or more dyslexic family members	0.10 (0.30)	0.20 (0.40)
Proportion of students for whom Dutch is not the mother tongue	0.17 (0.38)	0.20 (0.40)
Age at pre-test (months)	101 (11)	100 (11)
FWR 345678 at pre-test	37.04 (17.41)	35.53 (18.29)
RAN at pre-test (seconds)	37.25 (10.47)	37.97 (11.41)

Note: FWR = fluency of word reading; RAN = rapid automatised naming.

control condition. Four schools created a second treatment group in grade 2, and one school did so in grade 3. In addition, one school did not have enough poor readers in grade 4 to generate a treatment group consisting of three poor-reading students. Altogether, 28 groups of poor-reading students were included in the experiment: 12 in grade 2, 9 in grade 3 and 7 in grade 4. It should be noted that there was no need to arrange control students into groups of three.

The total number of participating students was 139, of whom 84 were in the GGOR condition and 55 in the control condition. The students' descriptive characteristics are summarised in Table 7 (for the description of the variables, see the subjects section of experiment 1). The proportion of boys was higher in the treatment condition than in the control group (t [1,137] = 8.24, $p < 0.01$). However, on other variables, differences between the conditions were small and statistically non-significant. Considering the unequal representation of boys, it was checked whether sex was related to reading level, as measured with the FWR 345678 at the pre-test. This was found not to be the case (means and standard deviations for boys were 37.51, resp. 18.07; for girls 35.32, resp. 17.39; t [1,137] = 0.53, p = 0.47).

Measures

Five dependent variables were selected to assess treatment effects: FWR, FTR, RC, VOC and RA (see the measures section of experiment 1 for a description of the tests).

Interventions

In the GGOR condition, students practised reading texts in groups of three students during forty-eight 30-minute sessions. The reading sessions were led by trained teaching assistants (see the following section). The sessions were modelled on the sessions in the first experiment. For instance, sessions were held in separate rooms, teaching assistants were provided with a supply of children's books, care was taken to ensure that reading sessions did not coincide with class reading time, and feedback to students was provided following the pause, prompt, praise technique. Approximately, 25 min of the 30-min sessions were devoted to reading, and five were spent talking about the meaning of the text. The remediation programme was

an integrated version of the continuous and RR formats; that is, the students spent most of the time practising the new text in accordance with the CR format. However, as a rule, in each session one text passage was chosen by the teaching assistant to be read several times in accordance with the RR format. Students in the control condition followed the school's regular reading programme and participated only in the measurement programme. For further details, see the interventions section of experiment 1.

Procedures

Each of the eight schools appointed an experienced teaching assistant to conduct the testing and intervention. As in experiment 1, they were given a written protocol detailing procedures for test administration and remediation. They were trained in all the relevant procedures during a three-day course at the start of the experiment. They also received on-the-job coaching during two visits by one of the researchers. During each visit, a reading session was videotaped and then discussed.

Reading sessions were held three or four times a week. The length of the intervention period was between 13 and 16 weeks, depending on daily class schedules (e.g. holidays, celebrations) and the presence of students. The teaching assistants kept a log in which the reading sessions were pre-numbered from 1 to 48. Students had the opportunity to catch up with missed sessions; for instance, if a student missed a session due to illness, he or she had an individual session to make up for it. Logs were also used to determine measurement occasions. All students in the same reading group were generally tested on the same day. Mid-test 1 was given on average after session 16. Due to illness and other disruptions, one student might have attended 15 sessions, while another might have had 17 or 18 sessions. The teaching assistants were instructed to observe the average number of sessions per group. In the same manner, on average, mid-test 2 was administered after reading session 32, and the post-test after session 48. Retention was tested about eight school weeks after post-testing; only GGOR students were tested. Logs were also used to keep notes on the texts that were read and other relevant matters, such as any deviation from the protocol. All logs were collected afterwards and checked for treatment fidelity. The conclusion is that the treatment fidelity was generally high.

Analysis and results

Table 8 shows the descriptive data for each of the dependent variables for each of the conditions and measurement occasions in terms of means and standard deviations. On the dependent variables at the pre-test, there were no statistically significant differences between students in the two conditions (alpha level set at 0.05). All students belonged to the lowest quintile on a test for FWR and retardation on the other dependent measures was less prominent. On the RC measure 36% of the students scored within the lowest quintile, on the VOC test only 18% did so. The percentage of students that suffered from low RA is unknown, because the RA measure was not a normed test.

The data were analysed through specification and testing of longitudinal multilevel models, analogous to experiment 1. The models comprised three levels: measurement occasions (four for FWR, two for the other measures), students (n = 139) and schools (n = 8). In the fixed part of the models, the following variables were

Table 8. Experiment 2: means and standard deviations on measures and measurement occasions.

	Condition			
	Group-based guided oral reading (n = 84)		Control (n = 55)	
Measure (min–max)/measurement occasion	M	SD	M	SD
Fluency of word reading (0–150)				
Pre	30.68	17.16	29.04	15.85
Mid 1 (after 16 sessions)	36.32	18.15	33.00	16.09
Mid 2 (after 32 sessions)	40.46	18.63	36.91	18.00
Post	45.13	19.87	39.93	17.67
Retention	46.66	19.69	–[a]	–[a]
Fluency of text reading (0–11)				
Pre	2.76	1.94	2.64	2.04
Post	4.08	2.26	3.56	2.14
Retention	4.39	2.08	–[a]	–[a]
Reading comprehension (−70 – 120)				
Pre	0.02	19.00	−2.59	18.65
Post	8.15	15.75	6.76	16.10
Vocabulary (56–144)				
Pre	98.67	11.60	98.26	13.11
Post	101.41	9.01	99.51	11.41
Reading attitude (7–21)				
Pre	15.35	3.37	16.33	3.75
Post	17.15	3.25	15.73	3.76
Retention	16.86	3.70	–[a]	–[a]

[a]Retention not measured in the control group.

entered: measurement occasion, a dummy for the students in the GGOR condition, two dummies for grade (grade 2 or 4), a dummy for whether the student had any dyslexic family members, and the continuous distributed variable RAN. Interaction terms were added for the two- and three-way interactions between measurement occasions, conditions and grades. For details, see the analysis and results section of experiment 1.

Table 9 presents the main results of the multilevel modelling, in particular the effects pertaining to the research questions. The condition*measurement occasion parameter was significantly different from zero for three measures: the two fluency measures and RA. In all three cases, GGOR students exhibited the greatest growth. Neither RC nor VOC showed a difference in growth between the conditions. The three-way interaction terms represent the hypothesis that the intervention has differential effects according to the grade. The last column in Table 9 shows that the intervention effect was not dependent on grade. In summary, compared to control students, GGOR students improved in terms of fluency and RA, but not in terms of RC or VOC. This result appeared to be valid for all three grades.

Retention was measured only for GGOR students, using the two fluency measures and RA. The intra-individual growth between post-test and the retention measurement can be deduced from the results in Table 8. Expressed in terms of Cohen's effect size *d,* the growth amounts to 0.08 (for FWR), 0.14 (for FTR) and −0.08 (for RA). None of these figures is statistically different from zero. The

Table 9. Experiment 2: outcomes of multilevel analyses for each of the five dependent variables: fixed effects and standard errors between parentheses; the control condition in grade 3 functions as the reference condition.

	Final model				
	FWR	FTR	RC	Vocab.	RA
Level 1 (Measurement Occ.)					
Intercept	43.80* (3.11)	4.33* (0.42)	6.07* (2.11)	98.00* (1.75)	15.64* (0.28)
Grade 2	–21.10* (1.73)	–2.18* (0.20)	–24.64* (2.32)	n.s.	n.s.
Grade 4	11.18* (1.79)	1.53* (0.20)	11.53* (2.69)	n.s.	n.s.
RAN	n.s.	–0.04* (0.01)	n.s.	n.s.	n.s.
Dyslexia in family	–5.87* (1.96)	–0.44* (0.22)	n.s.	n.s.	n.s.
GGOR	n.s.	n.s.	n.s.	n.s.	n.s.
GGOR × Measurement Occ.	1.16* (0.35)	0.53* (0.15)	n.s.	n.s.	1.67* (0.34)
Meas. Occ. × GGOR × Grade	n.s.	n.s.	n.s.	n.s.	n.s.
Level 2 (Student)					
Intercept	4.26* (0.37)	0.87* (0.21)	n.s.	2.04* (0.90)	n.s.

Note: n.s. = not significant; RAN = rapid automatised naming; GGOR = group-based guided oral reading; FWR = fluency of word reading; FTR = fluency of text reading; RC = reading comprehension; Vocab. = vocabulary; RA = reading attitude.
*$p < 0.05$.

conclusion is therefore that students do not progress after the end of the intervention, but nor do they show a decline in reading attainment.

Finally, as in experiment 1, the SRI index was evaluated to see whether GGOR students had reduced their reading gap as compared to their peers. It should be noted that this index is based solely on the FWR measure. It appeared that GGOR students' SRI increased from 0.39 at pre-test to 0.57 at post-test. Between post-test and retention no further development was visible, as the SRI stayed at 0.57. These figures are similar to what was found in experiment 1 (for a comparison, see Table 5). Considering that the population mean of SRI is 1.00, it must be concluded that GGOR students still display a considerable reading disadvantage.

Summary of results

The conclusions are structured around the research questions and aggregate the outcomes of the two experiments.

(a) Are the two intervention formats – CR and RR – effective methods for improving the reading fluency of poor readers?

Based on experiment 1, the answer is yes. Both RR and CR induced more fluent reading than the control condition, in which students followed only the regular class-reading programme. The advantage of RR and CR was demonstrated on both fluency measures (i.e. for word and for text reading), although on the measure for FTR, the RR students had only a marginally significant advantage considering the adopted alpha level of 0.05. Experiment 2, in which the RR and CR intervention formats were integrated and offered to groups of three students, confirmed the effectiveness of guided oral reading for both fluency measures. The effects of guided oral reading on reading fluency appear to be rather robust, as they occur in different delivery formats (RR or CR, individually or in small groups).

(b) Do these two formats induce favourable transfer effects, specifically on RA, comprehension and VOC?

In both experiments, transfer effects were demonstrated only on the measure for RA. Students in the RR, CR and GGOR conditions realised a more positive attitude development than control students. There were no transfer effects on RC or VOC. The students in intervention and control conditions showed similar growth on RC and VOC.

(c) Is either format better than the other for reading fluency or for the transfer variables?

Based on experiment 1, the effects of RR and CR were practically the same for all measures used, although the effect of RR on the measure for FTR only approached statistical significance, while the effect of CR was significant at the 0.05 alpha level.

(d) Is there a differential effect of the interventions by grade?

No. The effects of guided oral reading, whether in the RR, CR or GGOR format, are independent of the specified grades (grades 2 through 4).

Two further findings deserve attention. Retention measurements two months after intervention withdrawal generally did not show further growth for students in the RR, CR or GGOR conditions, with one exception: RR students demonstrated slightly, but statistically significant higher reading levels on the FTR measure. The other finding is that after intervention, RR, CR and GGOR students still displayed a fluency disadvantage compared to their peers, although the SRI index showed a declining disadvantage. It appears it would take more than a three-month remedial intervention to eliminate the reading fluency gap of the students in the present samples.

Discussion

Three variants of guided oral reading were evaluated in two experiments: namely, RR and CR in individual sessions, and a variant in which the reading sessions were offered to groups of three students. In the last-mentioned variant, students spent most of their time reading new texts, although in each session a fragment was chosen for RR; another difference was that group sessions lasted 30 rather than 20 min.

Remarkably, the efficacy of the three formats, as determined after 48 remediation sessions, appeared to be highly comparable. That is, positive effects were found on the fluency measures, both for word and text reading, and on the measure for RA. Effects on measures for RC and VOC could not be demonstrated. Considering the outcomes expressed in the SRI index, which is based on the FWR measure, all three variants are equally effective. Average index values increased from around 0.40 at the start of the remediation to 0.60 at the post-test.

This equal effectiveness is of practical significance, as it allows teachers to choose the format that best suits them or their students. If, for instance, a particular student is very anxious about making reading errors, he or she might be better suited to the RR format than CR. Furthermore, if there are several poor-reading students in a class, the teacher might choose the group delivery format. Accounting for the above-mentioned difference in session duration (20 vs. 30 min), the scale advantage for groups of three students compared to individual remediation is about 100%: in the same amount of time, twice the number of students can receive reading remediation.

That RR and CR are equally effective remedial techniques for dysfluent readers has been demonstrated in previous research, mostly among English-speaking students. The present study extends this finding to Dutch-speaking students. Interestingly, the Dutch orthography is much more shallow than that of English (Ziegler et al. 2010), meaning that grapheme-phoneme correspondences are more transparent or regular in Dutch than in English. It has been reported that students learning to read in a language that has a consistent orthography are generally accurate readers (De Jong and van der Leij 2003; Landerl, Wimmer, and Frith 1997). As a consequence, poor readers in Dutch are characterised more by slow than by inaccurate reading. The present study indicates that the effectiveness of guided oral reading might be independent of the orthography or a student's specific reading problem, such as inaccuracy or slow decoding. Of course, more research is needed to establish whether the effectiveness of guided oral reading is independent of the type of dysfluency.

The absence of transfer effects on RC and VOC diverges from the results of earlier studies. For instance, O'Connor, White, and Swanson (2007) reported effect sizes of around 1.00 for RC and 0.50 for VOC. Their study shares many similarities with the present one, including one-to-one remediation in either the continuous or the RR format. Their remediation amounted to 10 h over a period of 14 weeks, compared with a total of 16 h over a period of 12 weeks (the present experiment 1) or even a total of 24 h (the present experiment 2). There is no indication that Dutch measures for RC or VOC fall short in terms of reliability or sensitivity. Both measures are trusted and commonly used in the Netherlands, and the test format is not different from that of widely accepted tests, such as the Grey Oral Reading Test (Wiederholt and Bryant 2001) or the Peabody Picture VOC Test (Dunn, Dunn, and Dunn 1997).

The estimates by O'Connor and colleagues (op. cit.) deserve further scrutiny. According to the meta-analysis by Therrien (2004), which was based on 18 studies in which only the RR format was offered, the average effect size for RC measures was 0.25 (se = 0.07). Compared to this estimate, the outcomes reported by O'Connor and colleagues (op. cit.) might possibly be considered outlying values. The reader is also reminded of the studies by Vadasy and Sanders mentioned in the introduction, which produced inconsistent results regarding the transfer effects of fluency training.

The absence of transfer effects on RC and VOC in our study generates new questions. Do transfer effects only appear when students are impaired regarding RC or VOC, which the students in our samples were not? Or do transfer effects require more time to become demonstrable, more than the three months between pre- and post-testing in the present experiments? Or might it be that tests for RC or VOC are less sensitive to student growth than tests for reading fluency? There is still much that is unknown about the effect of reading fluency training on RC and VOC. Further studies should be done to determine the conditions conducive to the transfer effect of fluency training on RC and VOC.

The relationship between reading ability and RA is generally found to be low to moderate (Kirby et al. 2011; McKenna, Kear, and Ellsworth 1995; Petscher 2010; Wigfield et al. 1997). This finding may have led other authors to disregard RA as an important factor in reading remediation. In this light, the fact that all three guided oral reading formats had a positive effect on the RA of poor-reading students is a unique finding. This positive effect on RA is sustained by the retention measure after a period of three months, demonstrating a rather robust impact of the interventions. As far as known, the present study is the first to establish this transfer effect. Of course, replications are necessary in order to determine whether this finding is conclusive. Studies that implement more comprehensive attitude measures to identify the potential underlying processes are recommended. For instance, is it the reading itself that becomes more satisfying for the student, or is it the way students perceive their own reading competences, which Pintrich (2000) refers to as mastery or performance orientation? Transfer to RA has important practical implications. Many teachers know that when it comes to RA, poor readers risk falling into a negative spiral: the longer the reading problem persists, the more negative the RA becomes. Negative attitudes might severely undermine the efforts that students invest in practice.

If formal reading instruction (namely the teaching of the alphabetic principle) starts in grade 1, as it does in many countries, signs of dysfluent reading become

visible at the end of grade 1 or at the start of grade 2. Dysfluency may persist in the following grades, especially if teachers fail to offer adequate instruction and practice. Therefore, dysfluent reading can also be found in grades 3 and 4, and probably even beyond. This raises the question whether students might respond differentially across grades, for instance because classroom instruction in later grades emphasises silent RC more than reading fluency. In neither of the present experiments was differential effectiveness found across grades for any of the measures. This is in agreement with the results reported by O'Connor, White, and Swanson (2007), who included second- and fourth-grade students in their study. However, Vadasy and Sanders did find a grade effect, namely little or no effectiveness for reading fluency in grades 4 and 5 (in their 2008a study), but modest effectiveness in grades 2 and 3 (in their 2008b and 2009 studies). Therefore, further studies into the differential effect of grade are to be welcomed, preferably through direct comparison and including students from different grades in a single design. Meanwhile, reading teachers are advised to act promptly when dysfluent reading persists in students. Such students are in need of supplemental instruction and practice. The sooner remediation starts, the smaller the chance that reading deficiencies will develop.

Notes on contributors

Ron Oostdam is a professor of Educational Studies at the University of Amsterdam, Research Institute of Child Development and Education. He is also a professor of Learning and Instruction and research director of the Knowledge Centre for Teaching and Education at the Amsterdam University of Applied Sciences.

Henk Blok is a senior researcher at the University of Amsterdam, Kohnstamm Institute, the Netherlands.

Conny Boendermaker is an instructor at Teachers College Windesheim, Almere, the Netherlands.

References

Aarnoutse, C. A. J., and T. J. Kapinga. 2005. *Handleiding Technisch Lezen 345678* [Fluency of Word Reading 345678 Manual]. Ridderkerk: Onderwijs Advisering.

Allington, R. 2001. *What Really Matters for Struggling Readers? Designing Research-based Programs*. New York: Longman.

Bowers, P. G., and E. Newby-Clark. 2002. "The Role of Naming Speed within a Model of Reading Acquisition." *Reading and Writing* 15: 109–126.

Bowers, P. G., and M. Wolf. 1993. "Theoretical Links among Naming Speed, Precise Timing Mechanisms and Orthographic Skill in Dyslexia." *Reading and Writing* 5: 69–85.

De Jong, P. F., and A. van der Leij. 2003. "Developmental Changes in the Manifestation of a Phonological Deficit in Dyslexic Children Learning to Read a Regular Orthography." *Journal of Educational Psychology* 95: 22–40.

Dunn, L. M., L. M. Dunn, and D. M. Dunn. 1997. *The Peabody Picture Vocabulary Test*. 3rd ed. Circle Pines, MN: American Guidance Services.

Hiebert, E. H. 2003. *Quick Reads*. Parsippany, NJ: Pearson Learning.

Jongen, I., and R. Krom. 2009. *DMT en AVI* [Three-minutes Test and AVI]. Arnhem: Cito.

Kirby, J. R., A. Ball, B. K. Geier, R. Parrila, and L. Wade-Woolley. 2011. "The Development of Reading Interest and Its Relation to Reading Ability." *Journal of Research in Reading* 34: 263–280.

Klauda, S. L., and J. T. Guthrie. 2008. "Relationships of Three Components of Reading Fluency to Reading Comprehension." *Journal of Educational Psychology* 100: 310–321.

Krom, R., S. van Berkel, and I. Jongen. 2006–10. *Begrijpend Lezen groep 4 – 6* [Reading Comprehension Grades 2–4]. Arnhem: Cito.
Krom, R., I. Jongen, N. Verhelst, F. Kamphuis, and F. Kleintjes. 2010. *DMT en AVI; groep 3 tot en met 8* [DMT and AVI; Grades 1 to 6]. Arnhem: Cito.
Kuhn, M. 2004. "Helping Students Become Accurate, Expressive Readers: Fluency Instruction for Small Groups." *The Reading Teacher* 58: 338–344.
Kuhn, M. R., and S. A. Stahl. 2003. "Fluency: A Review of Developmental and Remedial Practices." *Journal of Educational Psychology* 95: 3–21.
LaBerge, D., and J. Samuels. 1974. "Toward a Theory of Automatic Information Processing in Reading." *Cognitive Psychology* 6: 293–323.
Landerl, K., H. Wimmer, and U. Frith. 1997. "The Impact of Orthographic Consistency on Dyslexia: A German–English Comparison." *Cognition* 63: 315–334.
McKenna, M. C., D. J. Kear, and R. A. Ellsworth. 1995. "Children's Attitudes toward Reading: A National Survey." *Reading Research Quarterly* 30: 934–956.
Merrett, F., and S. Thorpe. 1996. "How Important is the Praise Element in the Pause, Prompt and Praise Tutoring Procedures for Older Low-progress Readers?" *Educational Psychology* 16: 193–206.
National Reading Panel. 2000. *Teaching Children to Read: An Evidence-based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction* (Reports of the Subgroups). Washington, DC: National Institute for Child Health and Human Development.
O'Connor, R. E., H. L. Swanson, and C. Geraghty. 2010. "Improvement in Reading Rate under Independent and Difficult Text levels: Influences on Word and Comprehension Skills." *Journal of Educational Psychology* 102: 1–19.
O'Connor, R. E., A. White, and H. L. Swanson. 2007. "Repeated Reading versus Continuous Reading: Influences on Reading Fluency and Comprehension." *Exceptional Children* 74: 31–46.
Perfetti, C. A. 1985. *Reading Ability*. New York: Oxford University Press.
Petscher, Y. 2010. "A Meta-analysis of the Relationship between Student Attitudes Towards Reading and Achievement in Reading." *Journal of Research in Reading* 33: 335–355.
Pintrich, P. R. 2000. "An Achievement Goal Theory Perspective on Issues in Motivation Terminology, Theory, and Research." *Contemporary Educational Psychology* 25: 92–104.
Rasbash, J., F. Steele, W. J. Browne, and H. Goldstein. 2009. *A User's Guide to MLwiN* (version 2.10). Bristol: University of Bristol.
Schlichting, L. 2005. *Peabody Picture Vocabulary Test III-NL*. Amsterdam: Swets & Zeitlinger.
Therrien, W. J. 2004. "Fluency and Comprehension Gains as a Result of Repeated Reading: A Meta-analysis." *Remedial and Special Education* 25: 252–261.
Vadasy, P. F., and E. A. Sanders. 2008a. "Benefits of Repeated Reading Intervention for Low-achieving Fourth- and Fifth-grade Students." *Remedial and Special Education* 29: 235–249.
Vadasy, P. F., and E. A. Sanders. 2008b. "Repeated Reading Intervention: Outcomes and Interactions with Readers' Skills and Classroom Instruction." *Journal of Educational Psychology* 100: 272–290.
Vadasy, P. F., and E. A. Sanders. 2009. "Supplemental Fluency Intervention and Determinants of Reading Outcomes." *Scientific Studies of Reading* 13: 383–425.
Van Bergen, E., P. F. de Jong, A. Plakas, B. Maassen, and A. van der Leij. 2011. "Child and Parental Literacy Levels within Families with a History of Dyslexia." *The Journal of Child Psychology and Psychiatry* 53: 28–36.
Wexler, J., S. Vaughn, M. Edmonds, and C. K. Reutebuch. 2008. "A Synthesis of Fluency Interventions for Secondary Struggling Readers." *Reading and Writing* 21: 317–347.
Wheldall, K., F. Merrett, and S. Colmar. 1987. "Pause, Prompt and Praise for Parents and Peers: Effective Tutoring of Low Progress Readers." *Support for Learning* 2: 5–12.
Wiederholt, J., and B. Bryant. 2001. *Gray Oral Reading Tests*. 4th ed. Austin, TX: PRO-ED.
Wigfield, A., J. S. Eccles, K. S. Yoon, R. D. Harold, A. Arbreton, C. Freedman-Doan, and P. C. Blumenfeld. 1997. "Change in Children's Competence Beliefs and Subjective Task Values Across the Elementary School Years: A 3-year Study." *Journal of Educational Psychology* 89: 451–469.

Wolf, M. 1991. "Naming Speed and Reading: The Contribution of the Cognitive Neurosciences." *Reading Research Quarterly* 26: 123–141.
Wolf, M., P. G. Bowers, and K. Biddle. 2000. "Naming Speed Processes." *Timing, and Reading: A Conceptual Review, Journal of Learning Disabilities* 33: 387–407.
Wolf, M., A. Goldberg, G. A. O'Rourke, C. Gidney, M. Lovett, P. Cirino, and R. Morris. 2002. "The Second Deficit: An Investigation of the Independence of Phonological and Naming-speed Deficits in Developmental Dyslexia." *Reading and Writing* 15: 43–72.
Ziegler, J. C., D. Bertrand, D. Toth, V. Csepe, A. Reis, L. Faisca, N. Saine, H. Lyytinen, A. Vaessen, and L. Blomert. 2010. "Orthographic Depth and Its Impact on Universal Predictors of Reading: A Cross-language Investigation." *Psychological Science* 21: 551–559.