Learning and Instruction 60 (2019) 54–65

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Effectiveness of an extracurricular program for struggling readers: A comparative study with parent tutors and volunteer tutors

Caroline Villiger[a,*], Silke Hauri[c], Annette Tettenborn[b], Erich Hartmann[d], Catherine Näpflin[b], Isabelle Hugener[e,†], Alois Niggli[c,1]

[a] University of Teacher Training Bern, Fabrikstrasse 8, CH-3012, Bern, Switzerland
[b] University of Teacher Training Lucerne, Töpferstrasse 10, CH-6004, Lucerne, Switzerland
[c] University of Teacher Training Freiburg, Murtengasse 36, CH-1700, Freiburg, Switzerland
[d] University of Freiburg, Rue St. Pierre Canisius 21, 1700, Freiburg, Switzerland
[e] University of Teacher Training Lucerne, Museggstrasse 37, CH-6004, Lucerne, Switzerland

ARTICLE INFO

Keywords:
Paired reading
Parent tutors
Volunteer tutors
Struggling readers
Randomized controlled trial

ABSTRACT

This study reviews the effectiveness of an extracurricular *paired reading* program to enhance the reading of struggling readers. For the first time, two program conditions are compared within one study: parent tutors and volunteer tutors. The program was implemented within a randomized controlled field trial; its effects on reading fluency and reading ability were investigated on a sample of 198 Swiss third graders with reading difficulties. The findings revealed that volunteers outperformed parents: Children who trained with volunteers developed significantly better reading fluency after 20 weeks (d = .21). However, the main effects on reading fluency did not last at follow-up and no effects on general reading ability were found. Children with higher reading fluency at the pretest benefitted significantly more than very poor readers (post-test: d = .47; 5 month FU: d = .39). The study highlights the benefit of volunteer tutoring and the necessity of ongoing, adaptive support for very poor readers.

1. Introduction

Current theory about reading acquisition processes stresses that word decoding and fluency are important for releasing the cognitive capacities that enable readers to understand reading content (Perfetti, 1985; Samuels, 2006) and to develop reading competence. However, the acquisition of reading competence, beginning with the consolidation of basic, lower-order processes such as word decoding and fluency, is not self-evident for some students. They need intensive and evidence-based training as soon as difficulties are recognized (Tier 2 intervention or secondary prevention; Foorman, 2003). However, teachers do not always have the necessary time to offer intensive and individualized training during class instruction. Cooperation with other partners, such as parents or volunteers, who provide additional training can be a possible solution to this problem (Epstein et al., 2002). In this study, an extracurricular program was implemented using an established fluency training method, *paired reading* (PR; Topping, 1988; 2001), to evaluate the effectiveness of non-professional tutors. By means of a randomized controlled field trial, the study investigated the effects of PR training given by parent tutors and volunteer tutors (who were strangers to the students) in comparison with a control group (who received no additional training out of school). The effects on reading fluency and general reading ability were evaluated at the end of the intervention period (20 weeks) as well as five months later (at the follow-up).

1.1. The efficacy of paired reading – a brief literature review

PR, developed by Keith Topping (1988), is a method that focuses on training reading fluency. This method has remained convincing to date because it meets several criteria that are known to be beneficial for promoting the reading ability of beginners and advanced readers. First, reading fluency is considered a prerequisite for reading comprehension, which is a core competence in reading and learning (Perfetti, 1985). Several studies have shown that the training of reading fluency has

* Corresponding author. University of Teacher Training Bern, Fabrikstrasse 8, CH-3012, Bern, Switzerland.
E-mail addresses: caroline.villiger@phbern.ch (C. Villiger), hauris@edufr.ch (S. Hauri), annette.tettenborn@phlu.ch (A. Tettenborn), erich.hartmann@unifr.ch (E. Hartmann), catherine.naepflin@phlu.ch (C. Näpflin), alois.niggli@bluewin.ch (A. Niggli).
[1] retired.
[†] Diseased author.

https://doi.org/10.1016/j.learninstruc.2018.11.004
Received 12 February 2018; Received in revised form 16 November 2018; Accepted 20 November 2018
Available online 02 December 2018

positive effects on reading comprehension (Chard, Vaughn, & Tyler, 2002; Kuhn & Stahl, 2003; National Reading Panel, 2000). Moreover, guided oral reading is critical for building a bridge between listening comprehension and reading comprehension, which releases decoding processes (National Reading Panel, 2000; Topping, 2001). Second, in a meta-analysis of effective programs for struggling readers, the one-to-one tutoring format of PR and of other methods was generally found to be beneficial (Slavin, Lake, Davis, & Madden, 2011); however, the effect sizes were not specified for different outcome variables. One-to-one tutoring allows intensive support adapted to a child's individual difficulties in reading, which makes it successful – especially if tutors receive training in advance (Elbaum, Vaughn, Hughes, & Moody, 2000; Ritter, Barnett, Denny, & Albin, 2009). Third, the PR method integrates modeling and scaffolding practices, which support the development of reading fluency (Chard et al., 2002). Finally, the method also takes into account the role of motivation in learning by offering the child the possibility to initiate reading sequences him- or herself (self-determination theory of Deci & Ryan, 2002). This is particularly important given that the procedure is highly structured. Furthermore, tutors provide positive feedback (praise) whenever a child reads a difficult word successfully, which enhances learning (see the Pause Prompt Praise approach; McNaughton, Glynn, & Robinson, 1987).

The literature on PR is substantial (Topping, 2001). This literature review only considers PR studies with adult tutors (parents, and other non-professional tutors), but not peer tutors. Topping and Lindsay (1992) performed a systematic review of studies with parent tutors and volunteer tutors who conducted a PR program with elementary school students (neither specifying for each tutor type, nor for reading ability of the children). The authors reported positive effects on reading accuracy and comprehension (in terms of mean ratio gains); however, they also highlighted that many of the reviewed studies were "… riddled with methodological weaknesses" (p. 220), for example, due to self-selected samples and a lack of random assignments to conditions. Although the authors omitted studies on children with severe learning problems, the methodological discrepancies among the studies remain; thus, it is difficult to draw a conclusion on the effectiveness of the PR method. More precise information is needed on the implementation (i.e., intensity) to compare and interpret the effects of PR.

Scholars have investigated these points more thoroughly in recent studies about PR with adult tutors. These studies have a sounder methodological basis; most of them have an experimental design and consider more measures than only reading gains (sometimes including family background, characteristics of the setting, and/or aspects of implementation). Most of these studies also found diverse positive effects of the PR method (although without providing effect sizes): Huemer, Landerl, Aro, and Lyytinen (2008) reported better global word reading fluency than the control group in their study with non-professional volunteer tutors (fourth grade children struggling with reading); Overett and Donald (1998) found increased reading accuracy and comprehension for a program with parent tutors and fourth graders struggling with reading; Lam, Chow-Yeung, Wong, Lau, and Tse (2013) reported better word recognition and reading fluency when parent tutors trained with preschool children; and Cadieux and Boudreault (2005) detected effects on general academic abilities and phonological awareness but no effects on reading ability for a program with parent tutors and kindergarten children (before formal reading instruction). A study by Miller and Kratochwill (1996) with parents as tutors and their second, third, or fourth graders struggling with reading only found effects on reading accuracy, rate and comprehension for children who completed the training as intended (n = 7); however, the evidence for this finding is rather scanty. Still, this latter study also collected data about implementation using tape recordings, which is valuable to our understanding of the effectiveness of PR. However, the small sample size of this study and most of the studies cited above did not allow to conduct more complex analyses that integrate diverse variables. The study of Lam et al. (2013) is the only one with a sample of a critical size (N = 195). However, because the sample of this Chinese study is composed of preschoolers, it is problematic to draw conclusions for Western countries, where reading fluency first becomes an educational objective at the primary school level. Moreover, all these studies neglected to analyze long-term effects.

Nonetheless, it is worth investigating whether a program shows effects beyond the intervention time (e.g., Mueller et al., 2015). The immediate effects might be due to reasons other than the method, such as the tutor's attention to the child (Hager, Huebner, & Hasselhorn, 2000). Furthermore, a lack of previous research about PR requires a direct comparison of the effectiveness of diverse tutors (e.g., parents vs. volunteers). This approach promises to generate more differentiated knowledge about the program's effectiveness in diverse circumstances (Topping & Lindsay, 1992).

1.2. Parents or volunteers as tutors – does it make a difference?

To meet the individual needs of students who are at risk and offer them intensive training before they fail, education usually falls back on paraprofessionals (O'Keeffe, Slocum, & Magnusson, 2011). In most PR studies, the parents were the tutors (e.g., Lam et al., 2013; Overett & Donald, 1998); in some studies, volunteer tutors were recruited (Huemer et al., 2008). To date, no comparative study has analyzed whether it matters which adult person gives the PR training. However, one might expect it to make a difference whether a parent or a volunteer provides the training because of the different types of relationships they have with the child and the different settings in which the training is conducted. In the following, some differences are described in detail.

The main difference between a parent and a volunteer tutor lies in the *familiarity* with the child. Although familiarity between a volunteer and a child may grow over time, the intimacy that family members share from the beginning means that many patterns of interaction are already established and facilitate communication. Existing relational schemas allow better communication because even when things are not explicitly said, mechanisms of inference help supply missing information (Koerner & Fitzpatrick, 2002). However, implicit communication may also give rise to misunderstanding. Moreover, familiarity sometimes leads to increased impulsiveness, which can result in conflicts. Indeed, research on homework indicates that conflicts arise more frequently in families with struggling students, sometimes due to bad grades or excessive expectations (Moroni, Dumont, & Trautwein, 2016). Generally, it is assumed that conflicts and pressure arise especially due to atypical "teaching-learning" situations at home, which disrupt sensitive parent-child relations (Grolnick, 2003). In contrast, when children receive support from volunteer tutors, they rarely know their tutors beforehand; consequently, no relationship is established yet. Tutors do not have precise expectations about the child's academic achievement, and because the "teaching-learning" situation is more easily accepted in this condition, conflict is unlikely to emerge. The training therefore offers the opportunity to build a relationship on a "neutral" basis without pressure (Juel, 1996). Finally, research on homework has shown that conflicts with parents can have a negative impact on the development of a child's achievement (Cooper, Lindsay, & Nye, 2000).

In general, meta-analyses investigating the effects of interventions on enhancing parental involvement have revealed moderate effects on academic outcomes (See & Gorard, 2015). Nevertheless, interventions focusing on *structuring homework support* are more likely to have positive effects (e.g., Villiger, Niggli, & Wandeler, 2010), and this also applies to children with learning difficulties (Bryan, Burstein, & Bryan, 2001). Research about volunteer tutoring and its effects is much sparser. A recent meta-analysis including 21 randomized field trials reported positive effects of volunteer tutoring on several measures of student achievement (in comparison to students without tutoring; Ritter et al., 2009). The authors found no differences between the types of tutor (parent tutors vs. college-age tutors vs. community volunteers). However, studies with parent tutors are clearly outnumbered by studies

with other types of tutor. Furthermore, given the unique character of the tutoring program of each study, the evidence of the findings is limited.

Another difference between parents and volunteers concerns the *setting*. Parents usually provide the training at home, whereas volunteers do it in a public but peaceful place (e.g., at school) in order to avoid noise and other distractions. Thus, the volunteers' setting has a much more formal and non-personal character. The home setting strengthens the climate of intimacy and familiarity; however, parents may have to simultaneously address requests from other family members, phone calls and so on, which generate considerable distraction. Additionally, families are usually burdened with busy day-to-day schedules and time pressure, which makes it difficult to implement training in a beneficial way (McElvany & van Steensel, 2009).

In sum, the reported findings show that it might make a difference whether the training is conducted by a parent or a volunteer tutor. Presumably, in terms of the effectiveness of the PR method, the type of relationship with the child and the home setting could be to the disadvantage of children training with parent tutors.

1.3. Variables influencing the effectiveness of paired reading

In addition to its focus on comparing parents and volunteers as tutors, the present study investigates whether some children benefited more from the program than others (aptitude-treatment-interaction approach, Cronbach & Snow, 1977). In educational studies, and more specifically in reading research, cognitive correlates of reading ability as much as environmental and social factors that correlate with reading skills are usually considered to explain individual differences (Schatschneider & Petscher, 2011). The question whether children with low versus high initial knowledge benefit differently from an intervention has been investigated in numerous studies (Pfost, Hattie, Doerfler, & Artelt, 2014; Stanovich, 1986), but not yet in PR studies. Lam et al. (2013) are probably the only ones that examined moderation effects within a PR study with parent tutors by testing the interaction of the intervention and family income. The authors found that families with low incomes and those with high incomes both benefited from the program. In the present study, the following aspects are tested as potential moderators: the initial reading level (before the intervention), vocabulary knowledge, cognitive abilities, and family background (occupational status of the parents).

Furthermore, treatment fidelity is given special consideration in this study. Treatment fidelity usually encompasses data about the intensity of training, i.e. duration, and the quality of training delivery. Especially the quality of training delivery has often been neglected in earlier literacy research (De la Rie, van Steensel, & van Gelderen, 2017). When attributing the training effects to the program, we need to be sure that the participants included in the analyses have carried out the training in the way intended by the program. Furthermore, it is of high interest to examine whether treatment fidelity and training success are associated (Lyon & Moats, 1997; McElvany & van Steensel, 2009). Previous studies (not necessarily PR studies) showed that such an association is not easy to establish (Topping, Thurston, McGavock, & Conlin, 2012, in relation to a peer tutoring study; Wanzek & Vaughn, 2008, a study with adult tutors).

1.4. The present study

The present study focuses on the immediate and follow-up effects of the extracurricular PR program on (1) reading fluency and (2) reading ability. The following research questions are addressed:

a) Do the effects of the PR program differ based on whether the tutors are parents or volunteers?
b) Do child characteristics (i.e., initial reading level, vocabulary knowledge, cognitive abilities, and family background) moderate the training effects?
c) Does the number of training sessions (as an indicator of treatment fidelity) predict or moderate reading gains?

We hypothesized that because of the highly structured procedure, the program would have significant effects on *reading fluency* at the end of the intervention in both conditions (parents/tutors) relative to a control group (Hypothesis 1). Furthermore, we hypothesized that due to the intensity of the intervention (about 20 weeks, 2 to 3 training sessions per week), these gains would be maintained at the follow-up (Hypothesis 2). Similarly, we expected that the intervention would affect general *reading ability*. However, since the PR program did not primarily aim to foster reading ability (e.g., comprehension), we expected those effects to be weaker (Hypothesis 3) and eventually delayed at the follow-up (Hypothesis 4). Generally, based on the findings of homework research, we assumed that the effects of volunteer tutors would be stronger than those of parent tutors (Hypothesis 5). Furthermore, we expected all students within the volunteer condition to benefit similarly from the program (Hypothesis 6). Within the parent condition, we expected readers with very low initial reading level to benefit less because according to previous research, conflicts are likely to arise in the teaching-learning situations of struggling students at home (Hypothesis 7). We examined the interaction effects between other child characteristics (vocabulary knowledge, cognitive abilities, and occupation of parents) and the experimental condition without any expectations. Finally, we hypothesized that the number of training sessions would predict gains, but not necessary moderate gains in reading outcomes (Hypothesis 8[2]).

2. Method

2.1. Recruitment and final sample

Students. The target group was children struggling with reading fluency. First, a total of 1307 students (from 96 s grade classes) were assessed by means of a standardized screening test (Stolperwörter-Lesetest; Metze, 2009) in two cantons of Switzerland (Lucerne and Fribourg). Both cantons are largely rural areas; their capital towns have under 100,000 residents. The preliminary sample ($N = 278$) consisted of students who scored lower than 33 percent on the screening test and for whom teachers expected the program to be helpful ("The child should participate in the program", yes/no). We excluded 29 students who had dyslexia and therefore were already receiving special education. The remaining students were then randomly assigned to the parent group, the volunteer group, or the control group (delayed treatment)[3]. In the next step, parents were contacted by phone to obtain their consent to participate. Forty-six children and their parents withdrew from participation in the project for different reasons (no time available, no need, no interest, etc.). Among the 46 children, 25 were assigned to the parent condition, 12 to the volunteer condition, and nine to the control group[4]. Four children dropped out during the program (three from the volunteer group, one from the parent group), three of them because the tutors were in a severe accident that hindered them from continuing the PR program. One child was excluded from the sample because of failure

[2] Moderation would mean that children with a higher number of training sessions would benefit differently in the two treatment conditions.

[3] Teachers informed the people responsible for the project whether parents were unable to do the training due to a lack of language knowledge (reading in German). In this case, their children were randomly assigned either to the volunteer group or to the control group. However, in this study, we do not report data about this particular sample ($n = 42$) because it would interfere with the equal distribution between groups (more children with an immigrant background and therefore low vocabulary in the volunteer group).

[4] We cannot completely exclude selection effects, especially within the parent group, because of the effort required to deliver the training.

to comply with the implementation instructions. As the drop-out reasons were very diverse and not systematic for any of the conditions, we can reject the assumption of differential, treatment-correlated attrition (Shadish, Cook, & Campbell, 2002).

Finally, 198 third grade students (Lucerne $n = 95$, Fribourg $n = 103$) from 50 schools were considered for the analyses (parent $n = 67$, volunteer $n = 64$, control $n = 67$; total randomization for the three conditions: 90.6% [5]). The students were on average 8.90 years old ($SD = 0.51$; Min. = 7.83; Max. = 11.66).

Volunteer tutors. The volunteer tutors were recruited in various ways, including local radio, newspaper, flyers, and the Internet. In addition, school principals were contacted to obtain the addresses of potential volunteers and to ensure the integrity and trustworthiness of the self-announced volunteers in their school district. Finally, 64 volunteers agreed to participate after receiving a briefing about the program. Time availability was a central criterion for the agreement. The occupations of the volunteers varied, ranging from housewives, shop assistants, commercial employees, and teachers to business administrators. The number of volunteer tutors with teaching experience at the primary school level was slightly higher ($n = 9$) than the number of parent tutors with such experience ($n = 2$) due to the recruitment procedure (school principals sometimes suggested retired teachers of their school.)

2.2. The LiT[6] program

The present study implemented Topping's (1987) PR method, which focuses on fluency, and extended it by including text-focused tutor-child communication, especially at the beginning and at the end of the training session. In line with the arguments of Overett and Donald (1998) and Baker (2003), this element was added to provide fundamental elements of shared reading that enhance understanding and motivation. Tutors were instructed to spend no more than 15 min per session reading with the child and to devote a maximum of five additional minutes to text-focused talk. The pairs were asked to conduct two or three training sessions per week, and the intervention lasted about 20 weeks (without considering school holiday breaks). Pairs with a parent tutor read at home, whereas pairs with a volunteer tutor read at school, usually after lessons (school headmasters or teachers were asked to provide an appropriate room). The language of instruction was German.

The method proposed by Topping (1987) focuses especially on reading together aloud in close synchrony. The tutor monitors the reading process by pointing at the text with his or her finger. If the tutee makes a mistake or struggles, the tutor stops his or her finger and leaves the tutee to read the word correctly for three to 5 s. If the tutee succeeds, the tutor gives praise, and both continue reading. If the tutee fails, the tutor provides the correct word, the tutee repeats it, and they continue reading. The tutee can give a signal when he or she wants to read alone for some time. Then, the tutor stops reading while still sliding his or her finger along the text. If the tutee makes a mistake, the tutor provides the correct word, and the pair continues reading.

The pairs were provided with a series of age-appropriate books. Each pair had access to a book box situated in the school buildings. Before the intervention, the books were chosen by a book expert (criteria: suitable topic for children, appropriate text difficulty, basic language level, appropriate topic for communication) and then reviewed by reading experts. The texts were largely narrative; some were expository. Usually, the pairs read from the same book over several training sessions until it was finished. The children were not allowed to continue reading the book individually between two sessions.

2.3. Procedure

The following figure gives an overview of the procedure of the present study (see Fig. 1).

Tutor training. Parent tutors and volunteer tutors received two evening training sessions of about 1.5 h each. Parents and volunteers were mixed in groups with a maximum of 18 tutors per session. First, they were explained the importance of reading competence and, in particular, reading fluency for understanding text (Pikulski & Chard, 2005). Subsequently, the tutors watched videos illustrating the PR method. Step by step, they tried to apply the method by working with another tutor and following the models shown in the videos. Key elements such as reading together (synchrony), correction of mistakes, finger monitoring, and positive feedback were highlighted. The underlying concept was based on social cognitive theory (Bandura, 1986); the tutors' role in the PR method was modelled in videos of a competent tutor, and they were guided away from focusing on mistakes and toward concentrating on the child's development. Moreover, the training fostered the tutors' ability to adapt to educational contexts, as not all children require the same amount of support or praise (Corno & Snow, 1986). In the second training session, the children also participated, which allowed them to get to know their (volunteer) tutor. Furthermore, after viewing a short video about the PR method, the pairs were invited to apply it and report on their experiences. The tutors were provided with an instruction booklet that helped them remember the content of the training sessions and apply the method consistently. The training sessions were delivered by five qualified instructors (project managers and staff members with knowledge of literacy instruction) following a detailed script. The sessions occurred in several school buildings, bringing together tutors who lived nearby.

Data collection. The data were collected in September 2014 (T1, before the intervention started), June 2015 (T2, after the intervention), and November 2015 (T3). The students completed the questionnaires and tests during regular lesson time at the three points of measurement (group or individual testing, duration: max. two lessons). The assessments were administered by qualified project staff members following detailed instructions. The parents of all children involved in the study and the volunteer tutors completed a questionnaire before the intervention started (T1).

Measures against drop-out. Several measures were implemented to prevent drop-out during the intervention. Before the intervention, the participants were asked to sign a confirmation of participation, which included acknowledgement of the program requirements. The aim of this was that participants would take the study seriously, thus enhancing their commitment. Furthermore, at the beginning of the intervention, the children were informed that they would receive a diploma at the end of the PR program and that there would be a draw for nine mini-tablets. Halfway through the program, the participants (students, tutors, and parents) were invited to attend a cultural program in which an author of children's books read aloud some of his own short stories.

2.4. Treatment fidelity

To check whether the pairs fulfilled the program requirements, tutors were asked to note the following items in a record book: (1) the date and time when the training sessions were conducted, (2) the duration of each training session, (3) the books they read, and (4) any remarks about special conditions or disturbances. From the 130 pairs, we received 117 record books. The number of total training sessions ranged from 23 to 75 ($M = 46.56$, $SD = 9.31$). Eighty percent of the pairs met the basic requirement of having conducted at least 40 training

[5] The allocation of some of these children needed to be changed after randomization because of the (non-)availability of volunteers in some municipalities. The new allocation still complied with randomized procedure. The remaining 9.4% of the children withdrew from the experimental group but agreed to be assigned to the control group. Sensitivity analyses without those $n = 19$ children revealed no different results to our research questions.

[6] LiT = German abbreviation for "Tandem Reading" (Lesen im Tandem).

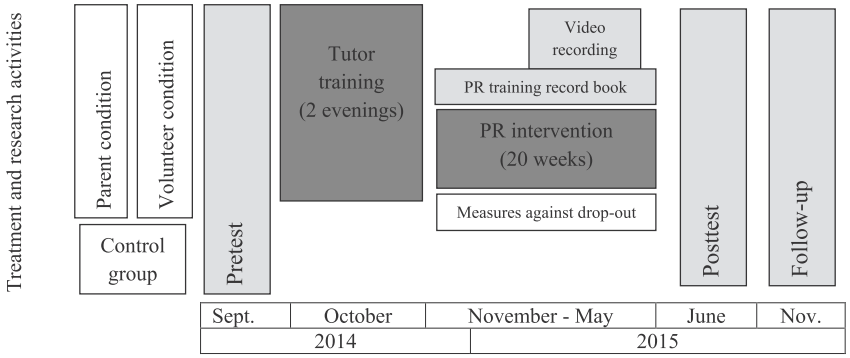

Fig. 1. Procedure of the present study.

sessions (volunteers: 76.2%, parents: 83.6%). However, as the minimum number of training sessions required to generate effects is not known, all pairs were used for analyses. Furthermore, a video of one training session of almost each pair was available (n = 113). Several aspects of treatment fidelity and interaction quality were coded by means of low inference and high inference category systems: fidelity to PR method, conversation activities, praise (yes/no), treatment of reading errors, and warmth. Two independent and reliable coders (inter-coder agreement: > 85.0%) or raters (generalizability coefficient: > 0.92) were involved (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). The data confirmed that the majority of the pairs implemented the method as intended. No differences were found between the parent and volunteer tutors, except for praise, which means that significantly more volunteers than parents offered praise to the children during reading. However, this variable did not prove to be a predictor for reading development. In the presented analyses, the number of training sessions is used as an indicator of treatment fidelity.

2.5. Reading instruction at school

Reading fluency is a curricular topic in grade 3 (and beyond). To collect information on the amount of fluency instruction provided at school, teachers were given a questionnaire after the intervention period in which they reported how often they worked with the children on reading fluency (1. “I worked about x weeks on reading fluency during the school year”, 2. “During those weeks, I worked about x minutes on reading fluency”). No significant differences were found between the groups (parent tutors, volunteer tutors, and control group).

2.6. Instruments

Reading fluency (dependent variable). This proximal measure was assessed with a standardized test called LDL – Lernfortschrittsdiagnostik Lesen [English: assessment of learning progress in reading] (Walter, 2009) – a well-established instrument in German-speaking regions (parallel test reliability r_{tt} = .91 for second to fourth grade). The tested person reads a given text for 1 min, which is audiotaped. Subsequently, the correctly read words within 1 min are counted. This number constituted the raw score, which was used for analyses in the present study. The assessor followed specific criteria for evaluating the correctness of the read words. Reading fluency was assessed at the three measurement points (T1, T2, and T3). The children read the same text at each measurement point. Since there was a time lag of at least five months between the measurement points, we did not expect memory effects.

Reading ability (dependent variable). This variable was measured with a standardized test using maze selection (Walter, 2013). In the maze selection test, students must select the correct word from a set of three word choices in each sentence. In comparison to reading fluency, reading ability is a more distal measure that takes into account the processes of comprehension and phrasing/word order. This method is commonly used for measuring reading ability at school (Deno et al., 2009). The general reading level was assessed at the three measurement points (T1, T2, and T3). The children read the same maze-text at each measurement point. The instrument has satisfactory parallel test reliability (r_{tt} = .78 for grade 3; r_{tt} = .80 for grade 4) and internal consistency (maximum likelihood (ML) reliability α = .93; Walter, 2013).

Cognitive abilities (control variable). Cognitive abilities were measured with an instrument called CFT 1-R (Weiss & Osterland, 2013), a non-verbal test that measures perception-based performance under time pressure and figural reasoning. The global factor indicates general intelligence. The test reliability is satisfactory, with Kuder-Richardson 20 = .90. This variable was measured at the first measurement point only (T1). For the analyses, the raw score was used.

Vocabulary (control variable). Vocabulary was assessed to control for language ability, which has been identified as an important prerequisite for reading ability (Kintsch, 1998). For this purpose, a subtest of the standardized SET 5–10 instrument developed by Petermann (2012) was used. The children were given 40 pictures of ordinary objects (e.g., “strawberry”) and actions (e.g., “painting a wall”) and asked to name them. The raw score of the number of correctly named pictures was used for the analyses. According to Petermann (2012), the internal consistency of the vocabulary test is satisfactory, with Kuder-Richardson 20 = .83. For our analyses, the value of the first measurement point (T1) was used.

Age (control variable). Students' birth date was provided in the parents’ questionnaire. Subsequently, we transformed it into the variable “age at the beginning of intervention” (November 2014).

HISEI (occupation of parents; control variable). Before the intervention started, parents completed a questionnaire at home to provide information on the child's family background. This included specifying the profession of each parent. Each parent was attributed an index according to a standardized classification of occupations (International Socio-Economic Index of Ganzeboom & Treiman, 1996). In a final step, the highest index between the parents was included in the analyses (HISEI).

Number of training sessions. The tutors provided this information by means of a record book where they noted each training session. The total number of training sessions was subsequently counted.

Personal characteristics of tutors. Before the intervention started, the tutors provided personal information, such as their occupational status and age. The quantity of books in the home was assessed using a similar procedure to that developed by Moser and Tresch (2003). Tutors responded to the question “About how many books do you have at home?” with four response categories: 1 = 0–10 books, 2 = 11–50 books, 3 = 51–100 books, 4 = more than 100 books. Furthermore, they reported on how much they liked reading on a 4-point Likert-type scale (reading enjoyment (cf., Schiefele, Schaffner, Möllner, & Wigfield, 2012): from 1 = “I do not like reading at all” to 4 = “I very much like reading”) and how often they read in their leisure time on a 6-point

Likert-type scale (1 = never or almost never, 2 = 1 to 3 times a week, 3 = up to 30 min each day, 4 = between 30 and 60 min each day, 5 = 1–2 h each day, 6 = more than 2 h each day). Finally, the tutors reported on their expectations regarding the LiT program ("The LiT program can help to improve reading" on a 4-point Likert-type scale ranging from 1 (*completely disagree*) to 4 (*completely agree*).

2.7. Statistical procedure

To evaluate the effects of the intervention, we ran regression analyses with SPSS 23 (Statistical Package for the Social Sciences) to assess changes in reading fluency and general reading ability, controlling for the initial level. Dichotomous variables were retained in the original metrics (parent condition and volunteer condition each as a dummy variable, with the control group as a reference group). The continuous variables were z-standardized ($M = 0$, $SD = 1$) to enhance the interpretability of the resulting regression coefficients. Thus, the coefficients indicate the proportion of the standard deviation by which the dependent variables (reading fluency and reading ability) increase or decrease if the predictor changes by one standard deviation. To test for moderator effects, the interaction terms were created using z-standardized values. For significant intervention and moderator effects, Cohen's *d* effect sizes are reported following the recommendations of Feingold (2013). Significant interactions were followed up with simple slope tests at 1 SD above and below the mean of the moderator (using unstandardized values).

For the regression analyses, the average amount of missing data per variable was 0.5% (maximum: 1.5%). Because this number was very low, missing values were not estimated. Consequently, the concerned subjects were not considered for analysis (procedure by SPSS: listwise deletion of records). The hypotheses concerning the main effects of the program were tested by a one-tailed significance test.

3. Results

3.1. Descriptive analyses: group comparisons

In a first step, the groups were compared with respect to the variables sex, cognitive abilities, vocabulary, parents' occupational status (HISEI), age, and reading level at T1. No significant differences were found for these variables between the three groups at baseline (see Table 1). In all groups, boys were overrepresented. The mean value of almost 62 in cognitive abilities represents an average of 68.9% correct answers (total raw score: 90). The comparison with a German norm sample (raw score: 64.8) shows that the mean value of our sample is slightly lower (Weiss & Osterland, 2013). In vocabulary, the students named about 31 pictures correctly, on average, of a total of 40. This fairly high score can be explained by the fact that the sample largely consisted of German-speaking children (see paragraph "recruitment and final sample"). The average HISEI value of 50.15 corresponds to that of a qualified technical expert or a secretary with customer contact.

The tutors were tested for possible differences in key variables such as occupational status (HISEI), quantity of books at home, age, reading enjoyment, reading amount, expectations regarding the LiT program, and number of training sessions (see Table 2). Parent tutors and volunteer tutors differed in many aspects. The volunteer tutors generally had a higher occupational status; they usually had more books at home, were older, and reported higher reading enjoyment and reading amount in leisure time. Parents completed almost six training sessions more than volunteers on average. The parents' and volunteers' initial expectations regarding the LiT program did not differ.

Table 3 displays the means and standard deviations for reading fluency at each measurement time (T1, T2, T3) by group (parents, volunteers, and control group).

The fluency scores of each group clearly increased over time. The increase between T2 and T3 was somewhat lower, likely because of the shorter time interval. Moreover, the standard deviation increased over time for all groups, but it increased to a greater extent in the parent tutor group and an even greater extent in the volunteer tutor group. The reading ability scores increased over time too, and with this, the standard deviation of each group; this finding indicates higher dispersion over time. Additional analyses showed that the children of the two cantons (Lucerne and Fribourg) do not differ significantly in reading fluency or reading ability at any measurement point ($p > .13$).

3.2. Descriptive analyses: means, standard deviations, and intercorrelations

Table 4 displays the means and standard deviations for all measures used in the regression analyses as well as the correlations between the variables. Vocabulary and cognitive abilities were significantly associated ($r = .17$, $p < .05$), and vocabulary and cognitive abilities were correlated with reading ability. Additionally, reading ability at T2 was significantly correlated with sex and HISEI. None of the control variables correlated in a significant way with reading fluency. As expected, the intercorrelations of reading fluency and reading ability over time (T1–T2–T3) were rather high, indicating a high stability of fluency ($r = .76/.73/.82$, $p < .01$) and a rather high stability of reading ability ($r = .57/.61/.75$, $p < .01$). Reading fluency and reading ability became increasingly more associated over the three measurement points ($r = .44/.65/.67$, $p < .01$).

3.3. Predicting reading fluency at T2 and T3, moderator effects

To address our first research question, we conducted separate regression analyses with reading fluency at T2 and T3 as dependent variables (see Tables 5 and 6). Model 1 (M1) investigates the predictive power of the initial fluency score; Model 2 (M2) integrates all predictors; Model 3 (M3) includes the two treatment group variables (dummy variables: treatment vs. control group), showing the respective effects of the two treatments relative to the control group; finally, Model 4 (M4) investigates the moderator effects of initial reading fluency.

Predicting reading fluency at T2

Model 1 shows a significant effect of reading fluency at T1, which remained relatively stable across the subsequent models (Table 5). When including control variables (M2), the parents' HISEI emerged as the only significant predictor. When the two treatment variables were

Table 1
Group comparison 1 (child characteristics).

Control variables	Parent tutors ($n = 67$)	Volunteer tutors ($n = 64$)	Control group ($n = 67$)	Total sample ($N = 198$)	Statistical comparison
Sex (1 = male)	70.1%	56.2%	59.7%	62.1%	$\chi^2_{(2,195)} = 2.93$, ns
Cognitive abilities *M (SD)*	60.84 (11.10)	63.47 (8.90)	61.75 (8.59)	61.99 (9.62)	$F_{(2,195)} = 1.26$, ns
Vocabulary *M (SD)*	31.79 (5.32)	31.61 (4.56)	31.55 (5.16)	31.65 (5.01)	$F_{(2,195)} = 0.04$, ns
HISEI parents *M (SD)*	51.75 (14.14)	49.89 (16.94)	48.85 (15.13)	50.15 (15.40)	$F_{(2,193)} = 0.60$, ns
Age *M (SD)*	8.84 (.49)	8.90 (.59)	8.95 (.44)	8.90 (.51)	$F_{(2,194)} = 0.75$, ns
Reading fluency T1 *M (SD)*	33.09 (9.20)	33.06 (11.20)	34.81 (11.42)	33.66 (10.62)	$F_{(2,195)} = 0.59$, ns
Reading ability T1 *M (SD)*	6.12 (3.73)	7.08 (3.86)	6.81 (3.43)	6.66 (3.68)	$F_{(2,195)} = 1.19$, ns

Table 2
Group comparison 2 (tutor characteristics).

Key characteristics of tutors	Parent tutors (n = 67)	Volunteer tutors (n = 64)	Total sample (N = 131)	Statistical comparison
ISEI *M (SD)*	44.08 (14.75)	50.95 (16.06)	47.56 (15.75)	$T_{(1,124)} = 2.49$*
Quantity of books at home *M (SD)*	3.07 (.89)	3.61 (.63)	3.34 (.82)	$T_{(1,119)} = 3.97$***
Age *M (SD)*	39.81 (7.29)	58.25 (12.00)	49.03 (13.54)	$T_{(1,104)} = 10.50$***
Reading motivation *M (SD)*	3.55 (.68)	3.88 (.33)	3.71 (.56)	$T_{(1,95)} = 3.51$***
Reading amount *M (SD)*	3.11 (1.19)	4.22 (.94)	3.66 (1.21)	$T_{(1,121)} = 5.89$***
Expectations regarding LiT Program *M (SD)*	3.62 (.49)	3.75 (.44)	3.68 (.47)	$T_{(1,120)} = 1.51$, ns
Number of training sessions *M (SD)*	49.26 (11.09)	43.73 (5.84)	46.56 (9.31)	$T_{(1,127)} = 13.19$**

Note. ***$p < .001$, **$p < .01$,*$p < .05$.

Table 3
Means and standard deviations of groups for reading fluency and reading ability (T1, T2, T3).

		Parent tutors (n = 67)	Volunteer tutors (n = 64)	Control group (n = 67)	Total sample (N = 198)
Reading fluency *M (SD)*	T1	33.09 (9.20)	33.06 (11.20)	34.81 (11.42)	33.66 (10.62)
	T2	46.87 (14.78)	50.40 (18.31)	48.88 (14.41)	48.68 (15.86)
	T3	55.88 (16.74)	58.11 (20.84)	59.59 (15.65)	57.86 (17.81)
Reading ability *M (SD)*	T1	6.12 (3.73)	7.08 (3.86)	6.81 (3.43)	6.66 (3.68)
	T2	11.18 (4.58)	12.03 (5.74)	11.45 (4.40)	11.55 (4.92)
	T3	14.83 (6.49)	15.92 (6.53)	16.02 (4.94)	15.59 (6.02)

included (M3), the volunteer condition significantly predicted fluency, whereas the parent condition did not. Expressed in the metric of Cohen's d (1988), the effect of the volunteer condition ($B = 0.21$) was $d = 0.21$ (Feingold, 2013). In the next step, moderator effects were investigated for several variables (initial reading fluency, vocabulary, cognitive abilities, and HISEI) to see whether some students benefitted from the treatment more than others. An effect was found for the interaction term "volunteers x reading fluency T1". To decompose the moderation, simple slope tests were performed at high (+1 *SD* above the mean) and low (−1 *SD* below the mean) levels of initial reading fluency (Feingold, 2013). The results indicated that the PR program had beneficial effects for students with higher initial reading fluency ($B = 7.44$, $SE = 2.44$, $p < .01$, $d = .47$). Students with low starting reading fluency did not benefit from the program ($B = -0.85$, $SE = 2.48$, $p = .73$) (see Fig. 2).

Predicting reading fluency at T3

With regard to predicting reading fluency five months after the training, the effect of the treatment in the volunteer condition was no longer present. Nevertheless, the moderator effect of reading fluency at T1 on the volunteer condition was still significant (Table 6). The slope tests indicated that students with a higher initial reading fluency level improved more ($B = 4.64$, $SE = 2.90$, $p = .11$, $d = .39$), and students with a lower initial reading fluency did not benefit from the program over the longer term ($B = -3.11$, $SE = 2.98$, $p = .30$) (see Fig. 3).

3.4. Predicting reading ability at T2 and T3

When addressing our second research question, we investigated the effects of the two treatments on reading ability following the same procedure as used in the analyses with reading fluency.

3.4.1. Predicting reading ability at T2

Reading ability at T1 significantly predicted the dependent variable (Table 7, M1). Except for cognitive abilities, none of the control variables appeared to be a significant predictor (M2). The coefficients of the control variables were stable across the subsequent models, even when treatment variables were included: Neither the parent group nor the volunteer group significantly differed from the control group in terms of reading ability gains (M3). Furthermore, no moderator effect was found for any of the tested variables (M4).

Table 4
Means, standard deviations, and intercorrelations for all measures.

		M	*SD*	1	2	3	4	5	6	7	8	9	10	11
1	Sex (1 = male)	–[a]	–[a]											
2	Vocabulary	31.65	5.01	.00										
3	Cognitive abilities	61.99	9.62	-.08	.17*									
4	HISEI parents	50.15	15.40	.01	.14	.12								
5	Age	8.90	0.51	.01	.04	.13	.04							
6	Reading fluency T1	33.66	10.62	.07	-.03	.05	-.10	.06						
7	Reading fluency T2	48.68	15.86	.09	.08	.11	.05	.09	.76**					
8	Reading fluency T3	57.86	17.81	.05	.09	.09	.05	.00	.73**	.82**				
9	Reading ability T1	6.66	3.68	.09	.19*	.28**	.08	.08	.44**	.51**	.53**			
10	Reading ability T2	11.55	4.92	.14*	.17*	.30**	.16*	.04	.53**	.65**	.64**	.57**		
11	Reading ability T3	15.59	6.02	.09	.20**	.34**	.10	.07	.49**	.59**	.67**	.61**	.75**	
12	Parent condition	–[a]	–[a]	.12	.02	-.09	. .07	-.08	-.04	-.08	-.08	-.11	-.05	-.09
13	Volunteer condition	–[a]	–[a]	-.08	-.01	.11	-.01	.01	-.04	.07	.01	.08	.07	.07
14	Control group	–[a]	–[a]	-.04	-.01	-.02	-.06	.07	.08	.01	.07	.03	-.01	.11

Note. $N = 198$
**$p < .01$,*$p < .05$.
[a] Dichotomous variables.

Table 5
Predicting reading fluency (RF) at time 2: Results from regression analysis with moderation.

	M1		M2		M3		M4	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
RF T1	.79***	.05	.79***	.07	.80***	.05	.65***	.08
Sex (male)			.08	.09	.09	.09	.11	.09
Vocabulary			.08	.05	.08	.05	.07	.05
HISEI parents			.11*	.05	.11*	.05	.11*	.05
Age			.03	.05	.03	.04	.05	.04
Cognitive abilities			.03	.05	.02	.05	.03	.05
Parent tutors (vs. CG)					-.01 [a]	.11	-.01 [a]	.11
Volunteer tutors (vs. CG)					.21* [a]	.11	.21* [a]	.11
Parent tutors x RF T1							.18	.12
Volunteer t. x RF T1							.27*	.10
R^2	.57		.60		.61		.62	

Note. $N = 198$.
***$p < .001$, **$p < .01$,*$p < .05$.
[a] One-tailed significance test, according to the directed hypothesis.

Table 6
Predicting reading fluency (RF) at time 3: Results from regression analysis with moderation.

	M1		M2		M3		M4	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
RF T1	.77***	.05	.79***	.08	.79***	.05	.66***	.08
Sex (male)			.03	.10	.04	.10	.05	.10
Vocabulary			.09	.05	.09	.05	.08	.05
HISEI parents			.12*	.05	.12*	.05	.12*	.05
Age			-.05	.05	-.06	.05	-.04	.05
Cognitive abilities			.02	.05	.02	.05	.02	.05
Parent tutors (vs. CG)					-.08	.12	-.08 [a]	.12
Volunteer tutors (vs. CG)					.05	.12	.05 [a]	.12
Parent tutors x RF T1							.17	.13
Volunteer t. x RF T1							.23*	.12
R^2	.54		.57		.57		.58	

Note. $N = 198$.
***$p < .001$, **$p < .01$,*$p < .05$.
[a] One-tailed significance test, according to the directed hypothesis.

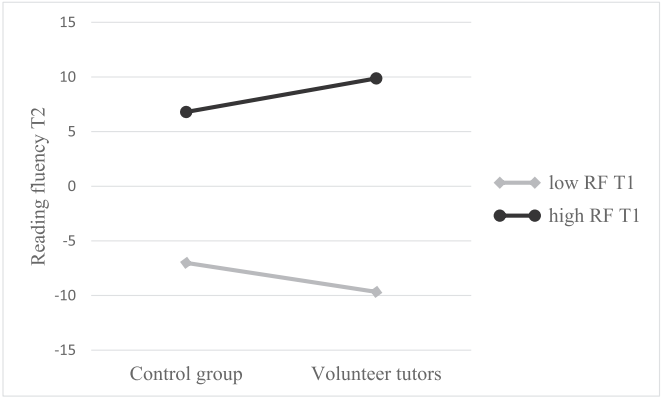

Fig. 2. Moderation effect between volunteer condition and reading fluency T1.

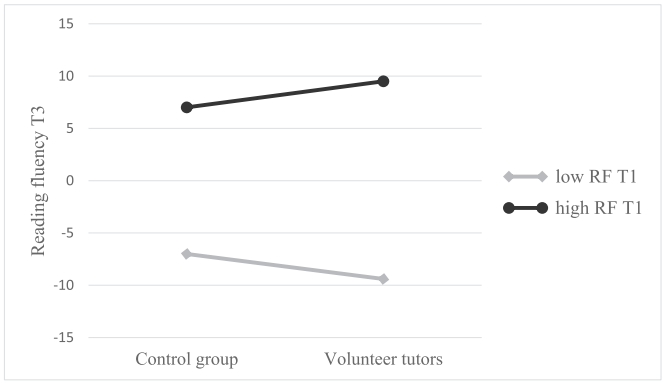

Fig. 3. Moderation effect between volunteer condition and reading fluency T1.

Table 7
Predicting reading ability (RA) at time 2: Results from regression analysis with moderation.

	M1		M2		M3		M4	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
RA T1	.57***	.06	.50***	.06	.50***	.06	.50***	.11
Sex (male)			.23	.12	.23	.12	.23	.12
Vocabulary			.04	.06	.04	.06	.04	.06
HISEI parents			.10	.06	.10	.06	.10	.06
Age			-.03	.06	-.03	.06	-.03	.06
Cognitive abilities			.15*	.06	.15*	.06	.15*	.06
Parent tutors (vs. CG)					.01	.14	.00[a]	.15
Volunteer tutors (vs. CG)					.05	.14	.05[a]	.14
Parent tutors x RA T1							-.04	.15
Volunteer t. x RA T1							.04	.15
R^2	.33		.37		.37		.37	

Note. $N = 198$.
***$p < .001$,*$p < .05$.
[a] One-tailed significance test, according to the directed hypothesis.

Table 8
Predicting reading ability (RA) at time 3: Results from regression analysis with moderation.

	M1		M2		M3		M4	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
RA T1	.61***	.06	.54***	.06	.54***	.06	.42***	.11
Sex (male)			.14	.12	.14	.12	.14	.12
Vocabulary			.06	.06	.06	.06	.06	.06
HISEI parents			.03	.06	.03	.06	.02	.06
Age			-.01	.06	-.01	.06	-.01	.06
Cognitive abilities			.18**	.06	.18**	.06	.18**	.06
Parent tutors (vs. CG)					-.08	.14	-.08[a]	.14
Volunteer tutors (vs. CG)					-.07	.14	-.08[a]	.14
Parent tutors x RA T1							.20	.14
Volunteer t. x RA T1							.15	.14
R^2	.34		.41		.42		.42	

Note. $N = 198$.
***$p < .001$, **$p < .01$,*$p < .05$.
[a] One-tailed significance test, according to the directed hypothesis.

Table 9
Predicting reading fluency T2/T3 and reading ability T2/T3: Results from regression analysis with number of training sessions (NTS).

	Reading Fluency T2		Reading Fluency T3		Reading Ability T2		Reading Ability T3	
	M1		M2		M3		M4	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Reading fluency T1	.83***	.10	.82***	.11				
Reading ability T1					.47***	.11	.61***	.10
Sex (male)	.11	.12	.11	.13	.15	.17	.16	.15
Vocabulary	.04	.06	.04	.07	.02	.09	.09	.08
HISEI parents	.13*	.06	.16*	.06	.11	.08	.01	.07
Age	.07	.06	-.04	.06	-.10	.08	-.09	.07
Cognitive abilities	.00	.06	-.02	.06	.11	.08	.23**	.07
Number of training sessions (NTS)	.02	.07	.03	.08	.04	.10	-.07	.09
Volunteer tutors (vs. parent tutors)	.25*[a]	.13	.15	.14	.03 [a]	.17	-.11 [a]	.16
Volunteer t. x NTS	.02	.16	-.04	.17	-.16	.21	-.15	.20
Volunteer t. x RF T1	.11	.13	.08	.14				
Volunteer t. x RA T1					.11	.15	-.08	.14
R^2	.65		.61		.37		.48	

Note. $N = 131$.
***$p < .001$, **$p < .01$,*$p < .05$.
[a] One-tailed significance test, according to the directed hypothesis.

3.4.2. Predicting reading ability at T3

Table 8 shows the results from the regression analyses when predicting reading ability at the follow-up. Again, no effects were found for the treatment groups in comparison to the control group (M3), and no moderator effect occurred (M4).

To investigate the number of training sessions as a possible predictor or moderator of reading gains, we ran additional regression analyses with the variable "number of training sessions" and the two program conditions (volunteers = 1, parents as reference group). Table 9 shows the results for reading fluency T2/T3 and reading ability T2/T3 as dependent variables (same modeling as before). The findings reveal that the number of training sessions is not a significant predictor of reading outcomes in any of the models. Furthermore, no moderator effect could be found. The variable does not contribute to explaining the variance between the children in terms of reading gains in a considerable way (M1: $R^2 = 0.4\%$; M2: $R^2 = 1.2\%$; M3: $R^2 = 0.1\%$; M4: $R^2 = 0.4\%$). Additionally, we repeated the analyses for those dyads that completed the required number of training sessions (≥40). We found similar results – again, the variable "number of training sessions" did not predict nor moderate reading gains. Moreover, treatment effects did not considerably change when excluding those who had not completed the required training sessions.

4. Discussion

The present study examined the effectiveness of an extracurricular PR program with a sample of third graders with reading difficulties. PR has been widely investigated, though not always very rigorously in terms of methodology (e.g., sampling procedure, randomization, implementation check), and in most of the cases, with relatively small samples that did not allow for more complex analyses. The present study addresses those issues more thoroughly, is based on a larger sample ($N = 198$), and investigates a randomized controlled trial at post-test and follow-up. Furthermore, the present study brings into direct comparison a parent and a volunteer condition, which is a new aspect in PR research and of high interest for educational research in the broader sense. Namely, it raises the question of how the tutoring setting may influence learning outcomes.

4.1. Short-term effects on students' reading fluency – differential effects of volunteers and parents

Our first research question focused on the effects of the two PR program conditions on reading fluency. Although parents and volunteers had received the same tutor training, students in the volunteer group developed their reading fluency significantly more during the intervention. Furthermore, against all odds, implementation (number of training sessions) did not explain the success of the volunteer group, and the first analyses of the video data showed almost no differences in the way parents and volunteers conducted the training (see section 2.4). Rather, it seems plausible that the two conditions represent different types of settings (formal vs. informal), which supports the idea that parent tutoring (at home) is not as conducive to teaching and learning than after-school volunteer tutoring is (Cooper et al., 2000; Grolnick, 2003). Hence, the effectiveness of the volunteer setting could be explained by its similarity to a "real" instructional setting, given the time (immediately after school) and place (in most cases, at school). The finding that the parent condition did not show any effects in comparison to the control group is surprising though, and it is inconsistent with earlier studies. We can only assume that this is due to the sampling procedure (no self-selection, except at the moment of compliance to participate) and the related fact that the reading difficulties in our sample were more pronounced than in other studies. In line with findings of homework research (e.g., Moroni et al., 2016), we can assume that the "parent as tutors" setting was not beneficial for our particular sample (children with reading difficulties). Although the PR method is highly structured, which should favor positive interactions while learning (Bryan et al., 2001), the hindering effects of a possibly burdened parent-child learning situation could not be canceled out. Previous research that investigated in more depth the parent-child interactions during the training supports this assumption. A case study revealed that a child made negative comments to her mother and occasionally became frustrated while reading (Miller & Kratochwill, 1996).

According to conventional standards, the intervention effect for volunteer tutors (ES $d = 0.21$) can be interpreted as moderate (Cohen, 1988). However, it must be considered that the total time of training on average (about 15 min x 46.56 training sessions) accounted for just over 11 h of training over a period of about 20 weeks. Given this low intensity, the effect can be estimated as satisfactory.

4.2. Effects at the follow-up

Contrary to our expectations, we observed no lasting main effects of the PR program on reading fluency at the follow-up (five months later). It can be assumed that it was necessary to continue reading aloud exercises to maintain the acquired reading fluency at the end of the 20-week training period. Basic skills such as reading fluency require consolidation, which is not necessarily achieved by continuing with autonomous (silent) reading after the PR program. It takes time to transfer the techniques learned within the training to daily routines (automatization of decoding processes) – this is especially the case for children who worked hard to acquire them (Mueller et al., 2015). In comparison, a program such as Reading Recovery (Clay, 1991), which aims to develop self-monitoring in reading, is more likely to have long-term effects because it enables students to make use of meta-cognitive strategies in their future reading (Hurry & Sylva, 2007; Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994). Second, the children of the volunteer group who had achieved significant gains in reading fluency at the posttest no longer saw their tutor, which might have had a detrimental effect on their development because they no longer benefitted from his or her positive attention (Hager et al., 2000). Finally, more personalized interventions that continuously tailor resources and activities to the individual student have shown longer-term effects; however, they are much more demanding for tutors (Burns, Senesac, & Silberglitt, 2008). Another way to increase program effectiveness – not less demanding for tutors – could be to intensify the program. Torgesen et al. (2001) have provided evidence that a shorter but daily training of reading ability might be more effective in the short term and even the long term. However, higher program intensity might be very challenging for volunteer tutors such as those in the present study.

4.3. Moderating effects on reading fluency

According to the aptitude-treatment-interaction approach (Cronbach & Snow, 1977), one research question focused on the differential effects of the program. We found interaction effects between the volunteer condition and the initial fluency level: the better their initial reading fluency, the more the children benefitted from the training with volunteers. The effect size was even higher at the follow-up. Thus, it was not the weak readers who benefitted most but the readers with less pronounced difficulties (Stanovich, 1986). We can assume that the very weak readers were still occupied with word-level processes (Gough & Tunmer, 1986), and therefore, their reading fluency did not improve as much. Phonics or sight word training might have been more successful in light of their developmental state (Chall, 1983; Suggate, 2016). Finally, we must take into account the very weak readers' lack of motivation to read. They might have perceived such training as a threatening activity (Nielen, Mol, Sikkema-de Jong, & Bus, 2016), and consequently resisted to the treatment (Torgesen, 2000).

4.4. No effects on general reading ability

In the present study, we found no effects of the PR program on reading ability (Research question 2). This could be because the program does not *directly* focus on semantics and syntax at the sentence level, which is the focus of the maze test used for assessing reading ability. Another explanation could be that effects on general reading ability are more pronounced with higher age, meaning that they could more easily be found in children doing the training from grade 4 on (see Overett & Donald, 1998). If the training takes place in the early elementary years (up to third grade), lower-order processes such as word decoding and fluency first need to be consolidated before the impact of the program at the sentence level (reading ability) can be manifested (Cadieux & Boudreault, 2005; Chall, 1983).

Finally, the lack of effects on reading ability could also be due to the fact that, with increasing age, reading ability (and more specifically comprehension) is affected by other abilities than fluency alone, such as vocabulary, cognitive ability (see intercorrelations in this study), general world knowledge, or higher order processing skills (Perfetti, Landi, & Oakhill, 2005). Presumably, struggling readers might need additional support on those specific dimensions in order to improve general reading ability.

4.5. Limitations and future research

Although this study followed the high standards of a randomized controlled trial and included verifications of the treatment's integrity, it has some limitations. First, because of the effort required to conduct the PR program, there are possible selection effects in the parent group. Obviously, only parents who had the necessary time available to conduct the training were willing to participate. Second, it is evident that the parents and volunteers constituted two distinctly different groups (age, ISEI, attitude toward reading, etc.). However, this problem could not be avoided by recruiting a volunteer group that was comparable to the parent group. In reality, parents usually do not volunteer to give training to a child other than their own. The reasons for this are time availability and motivation. Additionally, the recruitment of volunteers was not random. Thus, our volunteers consisted of people with above-average motivation. However, in field research, it is an ethical obligation to recruit tutors who are positively disposed to the task. Third, our underlying data supply only a few elements that help us explain the differential effects of parents and volunteers. Against our expectation, the intensity of the program did not contribute to explaining training success. Obviously, we still need to search for other factors that may influence the efficacy of PR and similar tutoring settings in a considerable way (see Overett & Donald, 1998).

5. Conclusions

To summarize, the results of our study confirm earlier findings about the positive effects of PR programs on reading fluency and point out the benefit for struggling readers to follow such a program with a volunteer tutor rather than with a parent tutor. Presumably, the volunteer setting, which is more formal and thus comparable with an instructional setting, is critical for this. The lack of long-term main effects highlights the necessity to continue supporting struggling readers by providing exercises and attention in order to consolidate the acquired abilities. Apparently, a single training program of 20 weeks is not sufficient to improve reading fluency, nor reading ability, especially for students with a very low initial reading level. However, struggling students with a higher initial reading level took advantage of the program, even in the long term. Those findings imply that the PR program, in the way it has been conducted in our study, is not necessarily beneficial for every struggling reader. The choice of a specific program, program intensity (more than 20 weeks for some students), and initial reading level presumably need to be harmonized at the student level.

Funding

This research was financed by the Swiss National Science Foundation (project no. 149560); the Canton of Lucerne, and the Canton of Freiburg, Switzerland. The opinions expressed are those of the authors and do not represent views of the funding bodies.

Acknowledgements

The authors wish to thank the tutors and students for their participation and thus for making this research possible. We would also like to thank people who were involved in editorial assistance, Loredana Torchetti for statistical support as well as the testing and data entry staff for their valuable work.

References

Baker, L. (2003). The role of parents in motivating struggling readers. *Reading & Writing Quarterly, 19*(1), 87–106. https://doi.org/10.1080/10573560308207.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bryan, T., Burstein, K., & Bryan, J. (2001). Students with learning disabilities: Homework problems and promising practices. *Educational Psychologist, 36*(3), 167–180. https://doi.org/10.1207/S15326985EP3603_3.

Burns, M. K., Senesac, B. J., & Silberglitt, B. (2008). Longitudinal effects of a volunteer tutoring program on reading skills of students identified as at-risk for reading failure: A two-year follow-up study. *Literacy Research and Instruction, 47*, 27–37. https://doi.org/10.1080/19388070701750171.

Cadieux, A., & Boudreault, P. (2005). The effects of a parent-child paired reading program on reading abilities, phonological awareness and self-concept of at-risk pupils. *Reading Improvement, 42*(4), 224–237.

Chall, J. S. (1983). *Stages of reading development.* New York: McGraw-Hill Book Compagny.

Chard, D. J., Vaughn, S., & Tyler, B. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35*, 386–406. https://doi.org/10.1177/002221940203500101.

Clay, M. M. (1991). *Becoming literate: The construction of inner control.* Pourtsmouth, NH: Heinemann.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd. Ed.). Hillsdale, NJ: Erlbaum.

Cooper, H., Lindsay, J. J., & Nye, B. (2000). Homework in the home: How student, family, and parenting-style differences relate to the homework process. *Contemporary Educational Psychology, 25*, 464–487. https://doi.org/10.1006/ceps.1999.1036.

Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In C. Wittrock (Ed.). *Handbook of research on teaching* (pp. 605–629). New York: Macmillan Publishing Company.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements.* New York: John Wiley.

Cronbach, L. J., & Snow, R. W. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington.

Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research.* Rochester, NY: University of Rochester Press.

Deno, S. L., Reschly, A. L., Lembke, E. S., Magnusson, D., Callender, S. A., Windram, H., et al. (2009). Developing a school-wide progress-monitoring system. *Psychology in the Schools, 46*, 44–55.

De la Rie, S., van Steensel, R. C. M., & van Gelderen, A. J. S. (2017). Implementation quality of family literacy programmes: A review of literature. *The Review of Education, 5*(1), 91–118. https://doi.org/10.1002/rev3.3081.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92*, 605–619. https://doi.org/10.1037/0022-0663.92.4.605.

Epstein, J. L., Sanders, M. G., Salinas, K., Simon, B., van Voorhis, F., & Jansorn, N. (2002). *School, family and community partnerships: Your handbook for action* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Feingold, A. (2013). A regression framework for effect size assessments in longitudinal modeling of group differences. *Review of General Psychology, 17*(1), 111–121. https://doi.org/10.1037/a0030048.

Foorman, B. R. (Ed.). (2003). *Preventing and remediating reading difficulties: Bringing science to scale.* York Press.

Ganzeboom, H. B. G., & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 international standard classification of occupations. *Social Science Research, 25*, 201–239. https://doi.org/10.1006/ssre.1996.0010.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6–10. https://doi.org/10.1177/074193258600700104.

Grolnick, W. S. (2003). *The psychology of parental control. How well-meant parenting backfires.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hager, W., Huebner, S., & Hasselhorn, M. (2000). Zur Bedeutung der sozialen Interaktion bei der Evaluation kognitiver Förderprogramme [The meaning of social interaction in evaluations of cognitive training programs]. *Zeitschrift für Padagogische Psychologie, 14*(2/3), 106–115. https://doi.org/10.1024/1010-0652.14.23.106.

Huemer, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency among poor readers of German: Many ways to the goal. *Annuals of Dyslexia, 58*, 115–137. https://doi.org/10.1007/s11881-008-0017-2.

Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Research in Reading, 30*(3), 227–248. https://doi.org/10.1111/j.1467-9817.2007.00338.x.

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly, 31*(3), 268–289. https://doi.org/10.1598/RRQ.31.3.3.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge: Cambridge University Press.

Koerner, A. F., & Fitzpatrick, M. A. (2002). Toward a theory of family communication. *Communication Theory, 12*(1), 70–91. https://doi.org/10.1111/j.1468-2885.2002.tb00260.x.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95*, 3–21. https://doi.org/10.1037/0022-0663.95.1.3.

Lam, S., Chow-Yeung, K., Wong, B. P. H., Lau, K. K., & Tse, S. I. (2013). Involving parents in paired reading with preschoolers: Results from a randomized controlled trial. *Contemporary Educational Psychology, 38*, 126–135. https://doi.org/10.1016/j.cedpsych.2012.12.003.

Lyon, G. R., & Moats, L. C. (1997). Critical conceptual and methodological considerations in reading intervention research. *Journal of Learning Disabilities, 30*, 578–589. https://doi.org/10.1177/002221949703000601.

McElvany, N., & van Steensel, R. (2009). Potentials and challenges of family literacy interventions: The question of implementation quality. *European Educational Research Journal, 8*(3), 418–433. https://doi.org/10.2304/eerj.2009.8.3.418.

McNaughton, S., Glynn, T., & Robinson, V. (1987). *Pause prompt and praise: Effective remedial reading tutoring.* Birmingham, UK: Positive Products.

Metze, W. (2009). *Stolperwörter-Lesetest.* (Unpublished document).

Miller, B. V., & Kratochwill, T. R. (1996). An evaluation of the paired reading program using competency-based training. *School Psychology International, 17*, 269–291. https://doi.org/10.1177/0143034396173003.

Moroni, S., Dumont, H., & Trautwein, U. (2016). Keine Hausaufgaben ohne Streit ? Eine empirische Untersuchung zu Prädiktoren von Streit wegen Hausaufgaben [No homework without conflict? An empirical study on predictors of conflict because of homework]. *Psychologie in Erziehung und Unterricht, 2*, 91–106. https://doi.org/10.2378/peu2016.art12d.

Moser, U., & Tresch, S. (2003). *Best Practice in der Schule. Von erfolgreichen Lehrerinnen und Lehrern lernen [Best practice in school. Learning from successful teachers].* Buchs, Switzerland: Aargauer Lehrmittelverlag des Kantons AG.

Mueller, B., Mayer, A., Richter, T., Krizan, A., Hecht, T., & Ennemoser, M. (2015). Differential effects of reading trainings on reading processes: A comparison in grade 2. *Zeitschrift für Erziehungswissenschaft, 18*, 489–512. https://doi.org/10.1007/s11618-015-0648-0.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development Clearinghouse.

Nielen, T. M. J., Mol, S. E., Sikkema-de Jong, M. T., & Bus, A. G. (2016). Attentional bias toward reading in reluctant readers. *Contemporary Educational Psychology, 46*, 263–271. https://doi.org/10.1016/j.cedpsych.2015.11.004.

Overett, J., & Donald, D. (1998). Paired reading: Effects of a parent involvement programme in a disadvantaged community in South Africa. *British Journal of Educational Psychology, 68*, 347–356. https://doi.org/10.1111/j.2044-8279.1998.tb01296.x.

O'Keeffe, B. V., Slocum, T. A., & Magnusson, R. (2011). The effects of a fluency training package on paraprofessionals' presentation of a reading intervention. *The Journal of Special Education, 47*(1), 14–27. https://doi.org/10.1177/0022466911404072.

Perfetti, C. A. (1985). *Reading ability.* New York: Oxford University Press.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skills. In M. J. Snowling, & C. Hulme (Eds.). *The science of reading: A handbook* (pp. 227–247). London, UK: Blackwell.

Petermann, F. (2012). *SET 5-10. Sprachstandserhebungstest für Kinder im Alter zwischen 5 und 10 Jahren [Test to assess the literacy skills of children aged 5 to 10 years].* Göttingen: Hogrefe.

Pfost, M., Hattie, J., Doerfler, T., & Artelt, C. (2014). Individual differences in reading development. A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research, 84*(2), 203–244. https://doi.org/10.3102/0034654313509492.

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher, 58*(6), 510–519. https://doi.org/10.1598/RT.58.6.2.

Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly, 29*(1), 8–39. https://doi.org/10.2307/747736.

Ritter, G. W., Barnett, J. H., Denny, G. S., & Albin, G. R. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research, 79*(1), 3–38. https://doi.org/10.3102/0034654308325690.

Samuels, S. J. (2006). Toward a model of reading fluency. In S. J. Samuels, & E. Farstrup (Eds.). *What research has to say about fluency instruction* (pp. 24–46). Newark, DE: International Reading Association.

Schatschneider, C., & Petscher, Y. (2011). Statistical modeling in literacy research. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Vol. Eds.), *Handbook of reading research: Vol. IV*, (pp. 54–65). New York, NY: Taylor & Francis.

Schiefele, U., Schaffner, E., Möller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly, 47*(4), 427–463. https://doi.org/10.1002/RRQ.030.

See, B. H., & Gorard, S. (2015). Does intervening to enhance parental involvement in education lead to better academic results for children? An extended review. *Journal of Children's Services, 10*(3), 252–264. https://doi.org/10.1108/JCS-02-2015-0008.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Belmont, CA: Wadsworth.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review, 6*(1), 1–26. https://doi.org/10.1016/j.edurev.2010.07.002.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 22*, 360–407. https://doi.org/10.1598/RRQ.21.4.1.

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities, 49*(1), 77–96. https://doi.org/10.1177/0022219414528540.

Topping, K. J. (1987). Paired reading: A powerful technique for parent use. *The Reading Teacher, 40*, 608–614.

Topping, K. J. (1988). *Paired reading training pack.* Huddersfield: Kirklees Paired Learning

Project.
Topping, K. J. (2001). *Thinking Reading Writing. A practical guide to paired learning with peers, parents and volunteers.* London, New York: Continuum.
Topping, K. J., & Lindsay, G. A. (1992). Paired reading: A review of the literature. *Research Papers in Education, 7*(3), 199–246. https://doi.org/10.1080/0267152920070302.
Topping, K. J., Thurston, A., McGavock, K., & Conlin, N. (2012). Outcomes and process in reading tutoring. *Educational Research, 54*(3), 239–258. https://doi.org/10.1080/00131881.2012.710086.
Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice, 15*(1), 55–64. https://doi.org/10.1207/SLDRP1501_6.
Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*(1), 33–58. https://doi.org/10.1177/002221940103400104.
Villiger, C., Niggli, A., & Wandeler, C. (2010). Fördern statt einmischen: Evaluation eines Kurzzeit-Elterntrainings zur Betreuung von Lesehausaufgaben [Support instead of interference: Evaluation of a short-time parental homework training for assisting with reading]. *Psychologie in Erziehung und Unterricht, 57*, 257–272. https://doi.org/10.2378/peu2010.art18d.
Walter, J. (2009). *LDL Lernfortschrittsdiagnostik lesen. Ein curriculumbasiertes verfahren. [Diagnosis of learning progress in reading. A curriculum-based method].* Göttingen: Hogrefe.
Walter, J. (2013). *VSL Verlaufsdiagnostik sinnerfassenden lesens. [Diagnosis of progress in reading comprehension].* Göttingen: Hogrefe.
Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *Journal of Learning Disabilities, 41*(2), 126–142. https://doi.org/10.1177/0022219407313426.
Weiss, R. H., & Osterland, J. (2013). *CFT-1R. Grundintelligenztest Skala 1 (Revision).* Göttingen: Hogrefe.