International Journal of Educational Research 115 (2022) 102037

Contents lists available at ScienceDirect

International Journal of Educational Research

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

Paired Reading with parent or volunteer tutors: Do implementation, process features, or tutor characteristics explain differential effects?

Caroline Villiger [a,*], Loredana Torchetti [a], Erich Hartmann [b], Silke Hauri [c], Annette Tettenborn [d], Catherine Näpflin [d], Alois Niggli [c]

[a] University of Teacher Training Bern, Bern, Switzerland
[b] University of Fribourg, Fribourg, Switzerland
[c] University of Teacher Training Fribourg, Fribourg, Switzerland
[d] University of Teacher Training Lucerne, Lucerne, Switzerland

ARTICLE INFO

Keywords:
Paired Reading
Process features
Implementation quality
Mediation
Extra-curricular learning

ABSTRACT

Within intervention research, there is still little attention paid to *explaining* effects. This paper intends to explain differential effects of a Paired Reading study (RCT) of third grade students struggling with reading ($N = 129$). The study showed differential effects on reading fluency, depending on whether volunteers or parents conducted the training. The mediating role of several implementation and process features, and the role of tutor characteristics, are investigated. Even though volunteers and parents differed in several aspects regarding implementation, process features and tutor characteristics, these differences cannot explain the differential effects. The difficulty uncovering the components responsible for the efficacy of the method is discussed, and implications for future research presented.

1. Introduction

There is a global trend in the growing use of academically-focused learning activities beyond formal schooling (Park et al., 2016). This kind of supplementary education, usually in the form of tutoring, is intended to compensate individual disadvantages, or fill gaps in knowledge, comprehension, or basic skills. The expectations placed on tutoring are high, even though results about the effects have been mixed (according to a literature review of Zhang & Liu, 2022). Furthermore, investigation of the conditions and mechanisms of extra-curricular learning is still needed to better understand the crucial elements of successful tutoring (Munter et al., 2014). To do this, we need more information on the processes and persons involved in tutoring (Zhang & Liu, 2022). For example, Roberts et al. (2018) suggest causal mediation to uncover relevant mechanisms that explain *how* or *why* an educational intervention or program is successful. Insights like these may help to prepare instructors or tutors correctly and, more broadly, contribute to advancing both research and practice.

The present paper follows this idea and applies it to a recent Paired Reading intervention study of third graders, by investigating implementation quality, process features, and tutor characteristics. Paired Reading (PR) is a well-established method aimed at

Abbreviations: PR, Paired Reading.

* Corresponding author.
E-mail address: caroline.villiger@phbern.ch (C. Villiger).

https://doi.org/10.1016/j.ijer.2022.102037
Received 27 April 2022; Received in revised form 23 June 2022; Accepted 25 June 2022
Available online 8 July 2022

fostering reading fluency in a tutoring setting (Topping, 2001). The study, which compared parent and volunteer tutors, showed differential effects: Children who trained with volunteers showed significantly greater increase in reading fluency than children who trained with their parent (Cohen's $d = .21$; Villiger et al., 2019). How can these differential outcomes be explained, even though the two groups attended the same Paired Reading training course and conducted the program over the same time period? The data collected is particularly suitable for investigating the mechanisms of successful tutoring. Published work has focused on whether some students in each condition benefitted more from the PR training than others. It could be demonstrated that within the volunteer condition, the training had beneficial effects for students with higher initial reading fluency; however, no other moderator effect was found (cf. Villiger et al., 2019). The present study focuses on the differential effects of the two conditions and investigates relevant implementation and process features, as well as tutor characteristics suggested by literature on extra-curricular tutoring and learning to better understand the mechanisms underlying the differential effects of parent and volunteer tutors.

1.1. Effects and mechanisms of extra-curricular learning and PR in particular

Because of the growing significance of extra-curricular (academically-focused) learning, this field of education has received increased attention as a research subject in recent decades. A distinction is usually made between private and public forms (e.g. private tutoring vs. after-school programs, etc.), but their commonality is that the learning activities are prepared and/or guided by an instructor or tutor. Whether private or public, each type consists of a myriad of academic learning activities outside of formal education (some more organized than others), which are all intended to improve students' academic skills and knowledge (Park et al., 2016). What empirical evidence is there to support this supposed effectiveness? Generally, the one-to-one tutoring format is known to be effective because it allows intensive and individually adapted support (Elbaum et al., 2000; Slavin et al., 2011). However, it is difficult to accurately assess the effectiveness of extra-curricular learning (e.g., private tutoring) because the students who avail of it may be quite different from those who do not (Hof, 2014). Studies investigating this question have produced mixed results (Bray, 2014), clearly showing that the subject is somewhat complex and multifaceted. The way tutoring is conceptualized from a cultural point of view may also have a bearing on results. However, the majority of research reports a positive relationship between private tutoring and students' academic achievement. A recent meta-analysis about the effectiveness of private tutoring reported medium overall effect sizes (Zhang & Liu, 2022), indicating that it is meaningful in obtaining higher grades (Hattie, 2009). A Swiss study on the subject confirms this result for reading achievement (Hof, 2014). Even though the effects on academic performance are questioned, it is evident that private tutoring might at least reduce the stress caused by underachievement in school (Guill et al., 2020), which is also of considerable importance.

PR is a highly structured tutoring method, aimed at fostering reading fluency (Topping, 2001). It is usually conducted by adults (parents or volunteers) or peer tutors, however, the results reported here only consider PR studies with adult tutors. Earlier findings about the positive effects of the method had a weak methodological base (cf. Topping & Lindsay, 1992), but subsequent studies with sound methodology confirm these reports (e.g., for volunteer tutors: Huemer et al., 2008; Villiger et al., 2019; for parent tutors: Cadieux & Boudreault, 2005; Lam et al., 2013).

When it comes to explaining the effects or mechanisms of tutoring, however, the evidence to date is limited. Zhang & Liu (2022) identified two of 13 moderators examined for tutoring effects, namely *region* and *subject*. The effect sizes of the African studies were much larger than those conducted in Asia and North America, and effect sizes for math tutoring were significantly larger than for other subjects. However, the authors found no moderating effects for variables such as gender, grade, sample size, outcome type, timing (during semester vs. vacation) and duration of tutoring. Furthermore, Guill and colleagues Guill et al. (2020) found that the instructional quality of the tutoring sessions (*structure, challenge*, and *support*) did not moderate their effectiveness.

Diverse theoretical models have been suggested recently to describe tutoring mechanisms on a conceptual level. Inspired by the supply-use model of teaching designed for classroom settings (cf. Helmke, 2009), which currently dominates literature on learning and instruction, Lohaus & Wild (2021) conceived a model that describes relevant features for extra-curricular support and tutoring. This model defines extra-curricular support in the broader sense (not only private tutoring specifically) and takes account of factors such as context, supply structure (intentions of the provider, features of the supply), implementation (quantity, quality), tutor features (e.g., competences), tutees' preconditions, and the extent of the supply's use (e.g., participation). The authors consider a range of immediate and lasting outcomes which go beyond the purely academic outcome. Based on many years of practical and scientific experience, Topping (2020) developed a theoretical model of intergenerational tutoring, which closely refers to the PR method. The model describes processes occurring within and between participants of a tutoring setting (tutor/tutee) containing 16 elements identified in research literature as the most important to training success. These include organization, communication, engagement, social-emotional and technical aspects specific to the PR method, like error management and feedback, assumed to enhance and consolidate reading fluency. Neither theoretical model has yet been empirically verified, but they both outline the relevant features to consider when explaining the effects of extra-curricular learning, and more specifically of PR. These are *personal characteristics and preconditions* of the participants, *implementation* of the supply (quantitative and qualitative aspects), and use or *process factors*, describing the extent and quality of personal involvement (e.g., participation). Even though recent empirical research suggests that aspects such as implementation quality (cf. Guill et al., 2020) or duration of tutoring (cf. Zhang & Liu, 2022) are not significant when explaining tutoring effects, these features are retained in the present study due to their possible relevance.

1.2. Differential effects of volunteers and parents in the LiT[1] study: Due to implementation, process features, or tutor characteristics?

Against the background of the relevant features of extra-curricular tutoring presented above, differential effects can be due to implementation, process features, or personal characteristics/preconditions of the participants. The possible sources of explanation for differential effects are discussed below by drawing on relevant theories and relating them to the current study.

1.2.1. Implementation

It is generally assumed that implementation plays an important role in the effectiveness of a training program (Durlak & DuPre, 2008). Thus, implementation fidelity - defined as a measure indicating the extent to which a method or program is implemented as intended (Powell & Carey, 2012) - appears to be crucial. Investigating implementation fidelity requires the definition of the key components assumed to be responsible for program success (O'Donnell, 2008). Implementation fidelity, and its mediating role, has been successfully investigated in settings other than tutoring (e.g., quality of classroom text discussions in mediating the effects of a literacy program, Matsumura, 2013). The study of Topping et al. (2012) is one of the few to analyze implementation within a tutoring program. The authors investigated whether several process variables - aspects of the Paired Reading technique, and quality of discussion (both based on observational data) and intensity - are correlated with reading progress (measured by standardized tests). The findings were disappointing: No significant relationship was found between intensity and reading gains, nor between process features and achievement. Praise, a core element of PR, was significantly but negatively correlated with reading gains.

De la Rie et al. (2021), who investigated implementation variables in the context of a family literacy program (e.g., the quality of parents' behavior and language during shared book-reading), found a higher quality of implementation for high-SES parents than low-SES and ethnic-minority parents. However, no relationship between implementation and the children's reading achievement was detected, nor were any mediating effects found. Previous studies about tutoring settings show that it is challenging to identify the aspects responsible for learning effects. Nevertheless, the present study attempts to meet this challenge and explain (differential) effects with implementation measures. The variables for which volunteers scored differently should be given particular consideration if the intention is to explain the success of the PR training given by volunteers.

In the present study, volunteers and parents were compared on several aspects of implementation, based on self-reported and observational data: intensity (number of PR training sessions; self-reported), average duration of training sessions (self-reported), timing of a typical training session (observational), and the implementation of various core elements of the PR technique (observational). Of the measures investigated, only a few were significantly different: program intensity, book-related communication, and praise. Even though the children who trained with a volunteer showed more progress in reading, they had met significantly less often with their volunteer tutor than the parent-child dyads (number of training sessions). Furthermore, the volunteers had significantly more book-related communication, and they also praised the child significantly more often than the parents (both aspects based on one video-recorded session; results published in Näpflin et al. 2020). It is assumed that book-related communication (or talk about the content, cf. Topping et al., 2012) and praise (positive feedback and recognition of efforts) can positively influence the progress in reading development and thus explain PR training success (Topping, 2001; Topping et al., 2012).

The duration of a program, also defined as exposure, is generally considered as an important component of implementation (O'Donnell, 2008); therefore it is usually considered when investigating interventions, at least as a control variable, sometimes as a potential moderator (cf. Zhang & Liu, 2022). Programs with a longer duration or greater intensity (higher frequency) have been expected to be more effective in helping students gain reading skills (Wanzek & Vaughn, 2008). However, the relation between duration and training/tutoring outcome is not necessarily linear (the more time spent, the better the outcome). The meta-study of Elbaum et al. (2000) revealed that duration was not associated with outcomes, and that shorter one-to-one tutoring interventions delivered more intensively tended to be more powerful (ES = 0.65 vs. ES = 0.37 for interventions longer than 20 weeks).

1.2.2. Process features

Topping et al. (2012) suggest that the PR method works in ways "other than merely increasing the amount of practice in reading" (Topping et al., 2012, p. 256), including for example social and emotional aspects due to the relationship between tutor and tutee (Bergin, 2001). Such features are known to be powerful in the learning process and must therefore be considered in addition to implementation features, which relate more to the program itself. This paper refers to them as "process features", meaning measures that assess the (inter)actions or individual (emotional) states of the participants during the tutoring process.

The *state of mind* of the tutor and the tutee is a determinant factor for learning and interaction. It is broadly acknowledged that the situational emotions of the tutor and tutee have a distinct effect on the learning outcome: Positive emotions are associated with more successful performance (Pekrun et al., 2011). Furthermore, Lohaus & Wild, 2021 point out that "the utilization of extra-curricular support achieves a substantial impact only if it is used for the self-active processing of the learning impulses presented" (p. 8, translated by author), a statement which applies to learning in any setting. Factors such as participation or *engagement of the tutee* can therefore be considered as measures of "use", meaning the extent to which the tutee is actively involved in the learning activities. According to Helmke (2009), active learning depends on the student's individual preconditions and on context (conducive or inhibitory to learning). It is assumed that whether a tutee is learning with a parent or a volunteer has an impact on process features such as state of mind or engagement. First, the type of relationship with a parent or a volunteer (familiar vs. neutral) may affect the

[1] Declaration of the study name (meaning "Tandem Reading")

tutoring situation, and second, the context (familiar/home vs. formal/at school) may affect the quality of learning (concentration, etc.). Overall, the volunteer condition is more similar to a formal (school-like) teaching situation, and therefore tutees paired with volunteers are more likely to accept this type of situation. As a consequence, they are more engaged and make more effort with learning. In contrast, parental teaching practices and involvement in learning at home are more likely to lead to conflicts (e.g., Neely-Barnes & Dia, 2008), due to the atypical "teaching-learning" situation, which disrupts sensitive parent-child relations (Grolnick, 2003). Therefore, tutees in the volunteer condition might benefit from the more favorable situation, with no pressure, conflict, or specific performance-related expectations (Juel, 1996). This situation leads to an overall *positive state of mind* and greater *engagement*, which facilitates a better learning outcome.

1.2.3. *Personal characteristics/preconditions of the participants*

Tutor characteristics include personal characteristics (age, pedagogic beliefs, self-efficacy beliefs, etc.) and teaching competencies (e.g., familiarity with PR, attendance at training, etc.). Qualifications of tutors have proven to be associated with effect size (prior training received; Elbaum et al., 2000; Ritter et al., 2009). The effects of personal characteristics have been investigated to a lesser extent. As regards the *tutees*, several aspects such as cognitive, motivational, and volitional preconditions, as well as the family background (socio-economic and cultural) have been identified in literature to be important for learning success (cf. Helmke, 2009). Family background, for example, not only influences learning outcome, but highly determines the attitude towards learning (Wilder, 2014).

Previous research conducted by the authors investigated whether some tutees benefitted more from the PR training than others. Only the initial reading fluency showed a moderating effect, meaning that students with a higher reading fluency level at the beginning benefitted more from the training within the volunteer condition than less fluent readers ($d = .47$; comparison group: control). However, this effect was not found within the parent condition. When directly comparing the parent and volunteer conditions, no moderator effects were found, indicating that tutees with specific characteristics did not benefit more from training with the volunteers than with the parents (moderators investigated: initial reading fluency, reading ability, and number of training sessions; cf. Villiger et al., 2019). Thus, the differential effects between parent and volunteer tutors could not be explained by the characteristics of the *tutees* that were investigated. Conversely, differential effects due to *tutor* characteristics still need to be investigated. Descriptive analyses showed that parent and volunteer tutors differed in terms of occupational status ($T = 2.49$*), quantity of books at home ($T = 3.97$***), reading motivation ($T = 3.51$***), and leisure time reading frequency ($T = 5.89$***), the volunteers achieving higher scores in each of these variables. Volunteers also had higher expectations of the benefit of the tutoring program, but they did not differ significantly from those of parents (cf. Villiger et al., 2019).

It is assumed that these types of tutor characteristics might explain the differential effects. For example, one might expect that the more positive the tutors' disposition toward reading (more books at home, higher reading motivation, reading more in spare time), the more they are able to inspire the tutees and activate them during the reading procedure, which ultimately leads to greater training

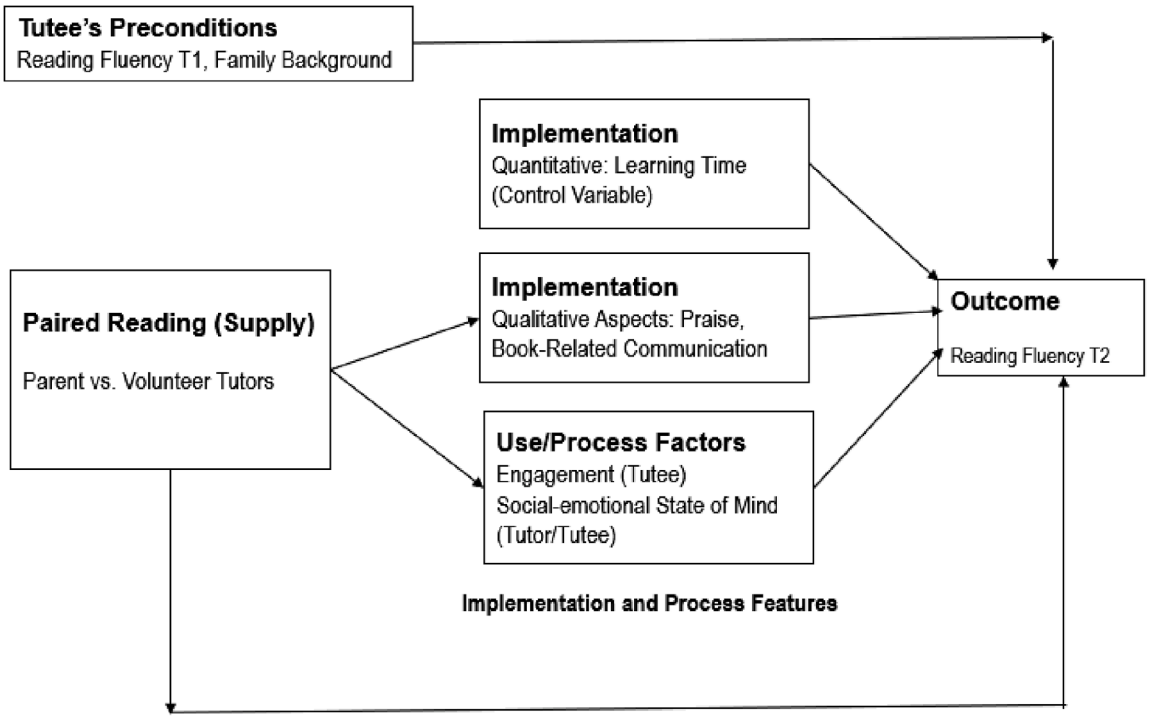

Fig. 1. Mechanisms of extra-curricular learning within tutoring and Paired Reading (PR).

success. Additionally, the expectations of parents *and* teachers are generally known to be powerful factors in predicting academic outcome (Neuenschwander et al., 2007; Rosenthal & Jacobson, 1968; Wilder, 2014). Therefore, higher expectations of the benefit of the tutoring program might produce greater progress in the tutee's reading.

1.3. Research questions and hypotheses

The present study intends to explain the differential effects of a PR study. Since the effects had been investigated in a randomized controlled field trial (comparing volunteers and parents, respectively to the control group), the first aim is to replicate the findings on the differential effects of parents and volunteers (without control group). Second, and most important, the present study addresses the following research questions: How can the differential effects be explained? (1) Do implementation and process features mediate the effect of the PR training on student's reading fluency, and thus explain the differential effects of parent and volunteer tutors? (2) Do tutor characteristics account for differential training effects between parent and volunteer tutors?

The considerations above on the theoretical background and on the study results available so far lead us to a set of assumptions. We hypothesize that: (1a) higher implementation quality among volunteer tutors, and (1b) more positive expression of process features will lead to stronger training effects within the volunteer condition in comparison to the parent condition. To test the hypotheses, we focus on five aspects which are assumed to *mediate* the direct effect of the volunteer condition: book-related communication and praise (aspects of implementation quality), tutors' state of mind, tutees' state of mind, and tutees' engagement (process features). Furthermore, we suppose that specific tutor characteristics are responsible for the training success of the volunteer condition, i.e., (2a) the (positive) general disposition toward reading, and (2b) (positive) expectations toward the training outcome. To test this, these tutor characteristics are introduced as predictors in addition to the tutor group.

Fig. 1 shows the postulated mechanisms of extra-curricular learning within tutoring and PR based on current findings in literature, focusing on the mediation effects.

2. Method

2.1. Sample

Tutees. First, a power calculation for the desired sample size was conducted. Assuming an average expected effect size of Cohen's $d = .30$ for PR programs (c.f., Miller & Kratochwill, 1996; Overett & Donald, 1998), a minimal sample size of $n = 36$ for each condition was defined (repeated measures ANOVA with originally three groups including CG, power of .95, alpha level of .05, correlation among repeated measures $r = .70$). Due to an expected increased drop-out rate, a sample of approximately 180 students was targeted for recruitment.

From the population of 96 third grade classes from two Swiss Cantons ($N = 1307$), students belonging to the 33& lowest-scoring on a standardized screening test (Metze, 2009) were invited to participate in the study ($N = 248$). The participation rate was 81.5% (46 children and their parents in the target group decided not to participate). Four children dropped out during the program, leading to the final sample of $N = 198$. The children and their parents who agreed to participate were randomly assigned either to the parent group (PR training with parent, $N = 67$)[2], to the volunteer group (PR training with a volunteer, $N = 64$), or to the control group (delayed treatment, $N = 67$). Some children withdrew from the experimental group after randomization but agreed to be assigned to the control group. Thus, total randomization applied to 90.6% of the sample. In the present study, only data from the two experimental groups are reported because of the integration of process variables. The data are therefore based on $N = 131$ students. These students were on average 8.87 years old ($SD = 0.54$, Min. = 7.83; Max. = 11.66), and 48 were girls (36.6%).

Tutors. The volunteer tutors were recruited in various ways, including via print media and the Internet. In addition, school principals were asked for addresses of potential volunteers in their municipality and to ensure the trustworthiness of the volunteers who presented themselves in their school district. The occupational background of the volunteers varied (housewives, shop assistants, teachers, business administrators, etc.), but many of them were retired. Nine volunteers had teaching experience at primary school level (among parent tutors: $n = 2$). Parent and volunteer tutors significantly differed in age (parents $M = 39.81$, $SD = 7.89$; volunteers $M = 58.25$, $SD = 12.00$; $T(104) = 10.50$, $p < .001$) and in socio-economic background (parents $M = 44.08$, $SD = 14.75$; volunteers $M = 50.94$, $SD = 16.07$; $T(123) = 2.49$, $p < .05$). Each of the volunteer and parent tutors was assigned only one tutee.

2.2. LiT ("Lesen im Tandem")[2]: a Paired Reading program

This section contains background information about the study on which the data presented here are based. For full details about implementing the PR program, see Table 4 at the end of the article.

Tutor training and requirements for implementation. The present study implemented the PR method (Topping, 2001) which is a tutoring method to enhance reading fluency. The tutors received two evening training sessions of about 1.5 h each (parents and

[2] Teachers informed the project managers 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, this study does not report data about that sample (n=42) because equal distribution between groups would not be ensured (more children with an immigrant background and therefore low vocabulary in the volunteer group).

volunteers were in mixed groups with a maximum of 18 tutors per session). First, the relevance of reading fluency in text comprehension was highlighted. Second, the tutors watched instructional videos illustrating the PR method, and then applied the method step-by-step in pairs following the model shown (Bandura, 1986). Key elements such as reading together (synchronicity), error management, finger tracking, and praise were emphasized. In the second training session, the tutees were present. After viewing a short video about the PR method together, the pairs applied it and reported on their experiences. The tutors were provided with an instruction booklet on the core elements of the method. The training sessions were delivered by five qualified instructors (project managers and fellows) following a detailed script[3].

The pairs were asked to conduct two or three training sessions per week for 15 min each over 20 weeks. Boxes with appropriate books were provided for the participants of the PR program and made available at municipal or school libraries. Pairs with a parent tutor read at home, while pairs with a volunteer tutor read at school, usually after lessons. Coaching was available during the whole implementation period.

Data collection. The data presented here were collected in September 2014 (T1, before the intervention started), and June 2015 (T2, after the intervention). The students completed the questionnaires and tests during regular school time. The assessments were administered by qualified project staff members. The tutors completed a questionnaire at the first training session.

Consideration of ethical standards. Ethical standards (Morgans & Allen, 2005) were taken into account at every stage of the project. Participants were recruited following an informed consent process. Before the project, parent tutors and their children were contacted directly by phone, and volunteer tutors were either invited to an information meeting or contacted by phone, to be notified of the project's aims, implementation of the PR training, and data collection. During implementation, the project's managers were available to provide further information or advice if necessary. After implementation, the main results of the study were communicated to participants via three newsletters written in plain language. No relationship existed between the project's managers and participants which might have led to coerced participation. The parents of children placed in the control group were invited to attend PR tutor training after the study's completion.

2.3. Instruments

2.3.1. Dependent variable

Reading fluency. This outcome variable was measured with a standardized test called LDL – Lernfortschrittsdiagnostik Lesen [assessment of learning progress in reading] (Walter, 2009). The instrument is well-established in German-speaking regions and has a satisfactory parallel test reliability of $r_{tt} = .91$. The person tested reads aloud a given text for 1 min, which is audiotaped. The words read correctly are then counted. This raw score was used for analysis. The assessor followed specific criteria to evaluate the correctness of the words read. The variable was measured at pre- and post-test (T1, T2), using the same text. No memory of the text was expected to remain after five months.

2.3.2. Implementation and process features

Features of implementation fidelity and the process were mainly collected using systematic observation techniques (Hugener et al., 2009) and only in part from self-reporting by tutees and tutors. This observation consisted of videotaping the pair in one reading session around halfway through the training program. As 18 pairs did not consent to being filmed, the sample size is $N = 113$ for some implementation and process features. The observation instruments developed were theory- and data-driven (Seidel et al., 2005), and a detailed description of the dimensions to be observed was compiled into a manual. Two trained raters coded the videos. More detailed descriptions are reported in Näpflin et al., 2020.

Implementation quality. Of several collected measures of implementation quality, two were chosen for analysis that showed significant differences between parent and volunteer tutors: *(a) book-related communication* (dichotomized) representing whether or not time was spent discussing the content of the book during the session (e.g., comprehension-oriented talk, content-related discussion), and (b) *the amount of praise,* which means the number of instances when praise was expressed in different forms (word of praise, positive feedback, nodding assent, etc.).

Duration of tutoring. Tutors reported the duration of each training session in a record booklet. The session durations were added up, and the resulting sum in hours and minutes was the total time of PR training. This measure is used as a control variable, because earlier investigations showed significant differences between parent and volunteer tutors in the number of training sessions conducted (parents investing more time in training overall than volunteers; cf. Näpflin et al., 2020).

State of mind. Three items were employed to capture the self-reported state of mind during the training sessions. Both tutees and tutors answered corresponding versions of the items on a 4-point Likert-type scale (1 = "completely disagree" to 4 = "completely agree"). The tutees' items were: (1) I liked to go to the training session at home/at school, (2) I felt comfortable during the last training session, and (3) My tutor was nice. The tutors' items were: (1) I went into the training with the child with a positive feeling, (2) I felt comfortable during the last session, and (3) I was patient with the child. As participants answered these items three times (at the beginning, in the middle, and at the end of the training program, each time directly after a training session), the items were averaged over the timepoints and then combined to a mean (Cronbach's alpha: tutees = .64; tutors = .81).

Tutees' engagement. Engagement means the extent to which the tutee is actively involved in the learning activities (showing interest

[3] For ethical reasons, the program was also offered to the control group after the experimental study.

Table 1
Descriptive statistics and intercorrelations of the study variables.

		M n	*SD %*	*N*	1	2	3	4	5	6	7	8	9	10	11
1	Reading fluency T1	32.91	9.99	129											
2	Reading fluency T2	48.32	16.44	128	0.77**										
3	Tutees' HISEI	50.66	15.49	127	-.06	.09									
4	Duration of tutoring (hours:minutes)	12:42	3:17	127	-.10	-.11	-.14								
5	Tutor group (1 = parent)	*65*	*50.39*	129	-.02	-.13	.05	.20*							
6	Book-related communication (1 = yes)	*64*	*57.66*	111	-.09	-.09	.09	-.13	-.18^{t}						
7	Praise	3.69	3.22	111	-.04	-.01	-.03	.15	-.28**	.32***					
8	Tutees' engagement	2.94	0.76	107	.19*	.16^{t}	-.04	-.06	.09	.06	-.06				
9	Tutees' state of mind	3.65	0.36	126	.10	.01	-.18^{t}	.13	-.14	.04	.09	.40***			
10	Tutors' state of mind	3.76	0.30	128	-.00	.11	-.10	.00	-.52***	.17^{t}	.11	.13	.20*		
11	Tutors' reading disposition	0.01	0.33	129	-.02	.06	.05	-.13	-.47***	.15	.14	.10	-.06	.21*	
12	Tutors' expectations	3.68	0.47	120	.03	-.06	-.13	.17^{t}	-.15	.22*	.11	.03	.08	.09	.02

Note. For dichotomous variables, instead of *M* and *SD* the number of cases and the percentage of the indicated category are reported (in italic).
$^{t}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

in training and willingness to participate). The video recordings were rated on a 4-point Likert-type scale from 1 to 4 according to tutees' observed engagement. In accordance with a high-inference coding procedure (Seidel et al., 2005), both negative and positive indicators for engagement were defined to build a basis for coding. The more positive indicators observed, the higher the rating given.

2.3.3. Tutee and tutor characteristics

Family background of tutees. Before the intervention started, the tutees' parents were asked to fill out a questionnaire to provide information on the child's family background, e.g., the occupational status of the parents. Each parent was attributed an index according to a standardized classification of occupations (ISEI = International Socio-Economic Index; Ganzeboom & Treiman, 1996). The highest index between the parents was included in the analyses (HISEI). In this study, the variables *cognitive abilities, age*, and *sex* were not controlled for because in the previous study they had no predictive power (Villiger et al., 2019).

Tutor characteristics. Two tutor characteristics were considered in this study. The *tutors' disposition towards reading* consisted of three items: number of books at home (1 = "0-10 books"; 2 = "11-50 books"; 3 = "51-100 books"; 4 = "more than 100 books"), reading enjoyment (1 = "I do not like reading at all" to 4 = "I like reading very much"), and leisure time reading frequency (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 to 2 h each day"; 6 = "more than 2 h each day"). A confirmatory factor analysis combining these three items to a latent construct of tutors' disposition towards reading yielded standardized loadings of .50, .69, and .71, respectively. With three items, the model was just identified, therefore the model fit was perfect. The factor scores were extracted (with a sample mean of $M = 0$) and used for further analyses. The *expectations of the benefit of the training* were also examined. Before the program started, tutors reported on their expectations of the benefit of the PR program ("This program can help to improve reading") on a 4-point Likert-type scale ranging from 1 (completely disagree) to 4 (completely agree).

2.4. Statistical procedure

First, data were screened for multivariate outliers by means of Mahalanobis distances. Two participants from the parent condition were subsequently excluded, resulting in $n = 65$ tutees in the parent condition and $n = 64$ in the volunteer condition. Second, descriptive analyses and group comparisons regarding implementation and process features as well as tutor characteristics were conducted using independent *t*-tests (continuous variables) and χ^2-test (dichotomous variable). Third, a multiple linear regression replicating the differential effect of the tutor conditions on reading fluency at T2 was conducted, controlling for the initial reading fluency at T1, as well as for the tutees' family background and duration of tutoring. Fourth, a path-analytical model was applied to test for the possible mediating effects of several implementation and process features which might explain the differential effects of the tutor conditions. Finally, to test whether tutor characteristics accounted for the differences between the tutor conditions, a different approach than mediation analysis had to be applied, as tutor characteristics cannot be considered as mediating processes. Therefore, multiple linear regressions were conducted: If the group effect were to weaken or even disappear when adding the tutor characteristic, it could be inferred that the characteristic was associated with the differential effect.

Analyses were conducted with the statistical programs SPSS 28 (IBM Corp., 2021) and Mplus 8.7 (Muthén & Muthén, 1998) using the MLR (regressions) or WLSMV estimator (in the mediation analysis, one mediator is dichotomous and therefore, the WLSMV estimator is applied by default). The fit of the path-analytical model was evaluated according to the criteria of Hu & Bentler (1999). Due to the directed hypotheses, one-tailed significance tests (alpha level = .05) were applied. The significance of indirect effects was tested by means of bootstrapped 95% confidence intervals based on 5,000 bootstrap samples. Even though the amount of missing data was minimal for most variables, data were imputed (10 datasets) for the analyses involving regressions. As bootstrap is not implemented with imputed data, missing data in the path-analytical model were handled using the FIML procedure. Standardized results are reported in each case: Continuous coefficients are completely standardized, and dichotomous coefficients are partially standardized (only with regard to the outcome).

3. Results

3.1. Descriptive analyses: Means, standard deviations, and intercorrelations

Table 1 reports the means, standard deviations, and intercorrelations of the variables. As expected, reading fluency increased from T1 to T2 over the sample. The average training duration was about 13 h. Sixty-four (58%) of the 111 observed pairs engaged in book-related communication. On average, the tutees were praised almost four times during the recorded training session. Tutees were observed to be relatively engaged ($M = 2.94$ on a scale of 1 to 4). The state of mind of both tutees and tutors was very positive (tutees: $M = 3.65$, tutors: $M = 3.76$, scale ranging from 1 to 4). Tutors' expectations of the benefit of the PR training were rather high.

There were some associations between the hypothesized mediators: praise and book-related communication were positively associated ($r = .32^{***}$). Moreover, the tutees' state of mind correlated positively with their engagement ($r = .40^{***}$), as well as with the tutors' state of mind ($r = .20^{*}$).

3.2. Descriptive analyses: Group comparison for implementation, process features, and tutor characteristics

Table 2 shows the comparison between the two conditions with respect to key variables used in the analyses below. The parent and volunteer tutors differed significantly on the two implementation features: More volunteers than parents talked about the content of

the book, and volunteers praised the tutees more often than parents. However, significant differences were not found in either the tutees' engagement or their state of mind ($p = .067$). Although the state of mind of all tutors was very positive, volunteers reported a significantly better state of mind than parents. Lastly, volunteer tutors reported a significantly higher disposition toward reading and slightly more positive expectations of the benefits of the PR program than parents ($p = .052$).

3.3. Differential effect of the PR program for the tutor conditions

A multiple linear regression analysis was conducted to replicate the differential effect of the paired reading intervention for parent and volunteer tutors on reading fluency at T2 with the reduced sample (without control group; c.f., Villiger et al., 2019). Reading fluency at T1, family background (HISEI) of the tutees, and duration of tutoring were controlled for (see Table 3; M1). The results confirmed that tutees who trained with their parents gained less in reading fluency than tutees in the volunteer condition ($\beta = -.25$, $p < .05$). While duration of tutoring had no effect ($\beta = .01$, $p = .809$), a higher socio-economic status (HISEI) of the tutees resulted in higher reading fluency at T2 ($\beta = .15$, $p < .01$).

3.3. Mediating effects of the implementation and process features

In the next step, the mediating effects of the implementation and process features were investigated. The fit of the initial path-analytical model was insufficient: $\chi^2(25) = 45.377$, $p < .01$; CFI = 0.85; TLI = 0.77; RMSEA = 0.081; SRMR = 0.190. Inspection of the modification indices revealed two conceptually meaningful amendments: 1) a residual correlation between the implementation features *praise* and *book-related communication*, and 2) a residual correlation between the two process features *tutees' engagement* and *tutees' state of mind*. With these modifications, the resulting model showed an acceptable fit: $\chi^2(23) = 23.55$, $p = .429$; CFI = 0.99; TLI = 0.99; RMSEA = 0.014; SRMR = 0.178. Although the SRMR does not meet the conventional cut-off for good model fit of .08, the model can be considered well-fitting since the χ^2-test of exact fit is not significant (Asparouhov & Muthén, 2018). The results of the path-analytical model are shown in Fig. 2.

Effects of the tutor condition on implementation and process features: Within the postulated model, differences between the parent and volunteer condition emerged on three of the five features analyzed. Parents talked less about the content of the book ($\beta = -.46$, $p < .05$), praised the tutee less often ($\beta = -.67$, $p < .001$), and reported a less positive state of mind ($\beta = -1.04$, $p < .001$). However, no significant differences were found regarding the tutees' observed engagement ($\beta = -.15$, $p = .472$) or the tutees' state of mind ($\beta = -.26$, $p = .174$).

Effects of implementation and process features on reading fluency at T2: Contrary to the hypotheses, none of the five features analyzed showed a significant effect on reading fluency at T2. Thus, a stronger expression of any of these features (book-related communication, praise, engagement, positive state of mind) did not lead to increased reading fluency at T2.

Direct and indirect effects: Within the mediation model, the direct effect of the tutor group on reading fluency was significant by trend ($\beta = -.21$, $p = .066$), indicating that tutees who trained with their parent progressed less in reading fluency than tutees with volunteers.

None of the five indirect effects tested was significant. Considering the bootstrapped 95% confidence intervals (CI), no indirect effect (IE) resulted in CI completely above zero (standardized IE): book-related communication IE = .04 (CI: -.045 to .169); praise IE = .01 (CI: -.113 to .069); tutees' engagement IE = -.01 (CI: -.051 to .036); tutees' state of mind IE = .02 (CI: -.023 to .086); tutors' state of mind IE = -.08 (CI: -.207 to .025).

To sum up, the smaller gains in reading fluency of the tutees in the parent group compared to those who trained with a volunteer could not be explained by the implementation and process features analyzed in the present study.

3.4. Tutor characteristics and differential effects of the tutor groups

Finally, the multiple linear regression (M1) was expanded to test whether the tutors' reading disposition or expectations of the benefits of the PR program would account for the difference in reading gains between tutees who trained with their parent or with a

Table 2
Group comparison for implementation, process features, and tutor characteristics.

	Parents *M (SD)* ($n = 65$)	**Volunteers *M (SD)*** ($n = 64$)	**Statistical comparison**
Implementation			
Book-related comm. (yes)[a]	25	39	$X^2(1) = 3.68$*
Praise	2.75 (2.92)	4.53 (3.28)	$T(109) = 3.00$**
Process features			
Tutees' engagement	2.87 (0.80)	3.00 (0.73)	$T(105) = 0.88$
Tutees' state of mind	3.60 (0.37)	3.70 (0.34)	$T(124) = 1.51^t$
Tutors' state of mind	3.60 (0.33)	3.91 (0.16)	$T(92.76) = 6.85$***
Tutor characteristics			
Tutors' reading disposition	-0.15 (0.36)	0.16 (0.20)	$T(100.14) = 5.97$***
Tutors' expectations	3.61 (0.49)	3.75 (0.44)	$T(117) = 1.64^t$

Note. [a]The number of cases is reported, where book-related communication was observed (dichotomous variable). p-values are one-sided.
^{t}p < .10, *p < .05, **p < .01, ***p < .001.

Table 3
Results of the multiple linear regressions predicting reading fluency T2.

	M1 β *(SE)* (*N* = 124)	M2 β *(SE)* (*N* = 124)	M3 β *(SE)* (*N* = 116)
Reading fluency T1	.78*** (.04)	.78*** (.04)	.78*** (.04)
Tutees' HISEI	.15** (.05)	.14** (.05)	.14** (.05)
Duration of tutoring	.01 (.05)	.02 (.05)	.03 (.05)
Tutor group (1 = parent)	-.25* (.11)	-.24* (.12)	-.28** (.11)
Tutors' reading disposition		.02 (.06)	
Tutors' expectations			-.08 (.06)
R^2	.62	.62	.63

Note. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

volunteer (see Table 3). The results showed that neither the tutors' disposition towards reading (M2) nor their expectations (M3) considerably altered the effect of the tutor group on the reading outcome. Both tutor characteristics themselves were not associated with gains in reading fluency T2 (reading disposition: β = .02, p = .806); expectations: β = -.08, p = .180). Therefore, these tutor characteristics are not accountable for the differences in reading gains observed between tutees of the two conditions.

4. Discussion

4.1. Mechanisms of PR method still unexplained

This study investigates implementation and process features to explain the intervention effects of a PR training with struggling third grade readers. The aim is to throw light onto recent findings about differential effects for two training conditions (volunteer condition: significant effects on students' reading fluency, parent condition: students did not differ from the controls) and to explain the mechanisms of tutoring and the PR method in particular. The findings presented here reconfirm the positive effects of the PR program within the volunteer condition, in line with other findings about the effectiveness of tutoring (Elbaum et al., 2000; Hof, 2014; Zhang & Liu, 2022).

The research to date lacks evidence about the mechanisms responsible for the efficacy of PR. A database comprising diverse sources of implementation and process features (self-reporting, video data) and covering a wide range of measures that describe implementation accuracy served as a solid foundation for this purpose in the present study. However, the analyses presented here cannot elucidate the mechanisms of PR either. The findings reveal that even though in the parent condition less favorable expressions of almost all investigated features were found (except for the tutees' engagement) than in the volunteer condition, those features were not able to explain the differential effects on reading fluency. Therefore, the four hypotheses about mediating effects (1a, 1b, 2a, 2b) found no empirical support. Also, the differential effects could not simply be explained by the tutor characteristics investigated (disposition towards reading, expectations of the benefits of the training). To understand these findings better, several considerations shall be presented.

The intent of this study was to analyze the mechanisms of the effects of a program by investigating many key factors (implementation, process features, and personal characteristics). Even though several features were taken into account, the complexity of the mechanisms (interdependencies of features, relevance of supplementary but ignored features, etc.) could not be depicted according to Topping's statement, "the whole proves greater than its parts" (Topping, 2020, p. 90). In particular, mutual influences or interdependence effects (cf. Laursen, 2005) are likely to be relevant in providing either a positive or negative learning climate, and subsequently impact learning outcomes. It remains a challenge to empirically catch "the whole" or, in other words, what exactly leads to the success of a training program. The higher the ecological validity of a model is in theory (integrating most of the relevant aspects), the more difficult it is to prove isolated effects empirically.

PR is recognized as a highly structured method with a very precisely defined training procedure (Topping, 2001). Previous analyses showed that the majority of parents and volunteers implemented the core elements of PR almost entirely according to the tutor training they received (which speaks for the latter's quality) and in many aspects of the PR program, they did not differ (Näpflin et al., 2020). Thus, the highly structured character of this method might be an additional obstacle to the use of process features to explain effects. It can be assumed therefore that there might be other factors with more explanatory power. The tutors' state of mind revealed significant differences between parents and tutors; this could suggest that the emotional-motivational state of the participants and/or their quality of relationship are variables worthy of exploration in future studies.

The finding that the tutor characteristics investigated did not account for the differential effects leads to the assumption that the success of PR training lies elsewhere, e.g., in the characteristics of the setting/context where the training took place. Since parents conducted the training at home, a place that may hold many distractions and disturbances, and volunteers mostly conducted the training at school in a more studious climate, the setting and group variables are confounded. Furthermore, it can be assumed that the video recording was not able to capture an entirely authentic PR training situation because it took place only once, for about 20 min, and with the person filming present. This timeframe is too short to see completely natural behavior and disregard for the situation, especially within the family context where behavior is expected to be much more spontaneous and emotional than in a formal setting. This might have deformed the assessment of implementation and process features in the parent condition. As a consequence, the

Table 4
Template for Intervention Description and Replication (TIDieR) table (cf. Hoffmann et al., 2014) for describing the Paired Reading (PR) intervention.

Item number	Item	located (article/ section)
1.	**BRIEF NAME** Lesen im Tandem (engl. Tandem Reading)	this article: 2.2
2.	**WHY** PR can increase reading fluency by guided oral reading, error management and feedback. This study investigates whether and why the volunteer tutors' condition is more efficacious than the parent tutors' condition.	this article: Introduction and 1.1
3.	**WHAT** **Materials**: The pairs had access to a box with age-appropriate books situated in the school buildings. 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.	Villiger et al. (2019): 2.2 this article: 2.2
4.	**WHAT** **Procedures**: The PR method proposed by Topping (2001) focuses 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.	Villiger et al. (2019): 2.2
5.	**WHO PROVIDED** **Instructors for tutor training**: Five qualified instructors (project managers and staff members with knowledge of literacy instruction) following a detailed script that had been gone through and discussed together in advance. **Tutors**: parents (of children low-scoring in reading) and volunteers. Analyses are based on $n = 65$ parent tutors and $n = 64$ volunteer tutors. 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). Further information about background of tutors: see Section 2.1 in this article.	Villiger et al. (2019): 2.3 2.1 this article: 2.1
6.	**HOW** **Tutor training** was delivered in groups of 18 tutors maximum (volunteers and parents mixed). Group members met in presence. **PR training** was delivered face to face. Each tutor was assigned only one tutee.	this article: 2.2 2.1
7.	**WHERE** **Tutor training**: The tutor training took place in rooms (classroom or other) made available by schools, bringing together tutors who lived nearby. **PR training**: Volunteer tutors usually delivered the training at school (classroom after lessons, school library, or group room), parent tutors at home (usually living room or kitchen).	Villiger et al. (2019): 2.3
8.	**WHEN AND HOW MUCH** **Tutor training**: from Mid-October to Mid-November 2014, two evening sessions of about 1.5 h each. **PR training**: from Mid-November 2014 to April 2015, two or three times a week, each time for about 20 min, during about 20 weeks (without considering school holiday breaks).	This article: 2.3.2
9.	**TAILORING** No specific tailoring - on the contrary, care was taken to teach the training method in the same way to all tutors.	-
10.	**MODIFICATIONS** From the beginning, the PR training method was extended by including text-focused tutor-child communication, especially at the beginning and at the end of the training session. This element was added to provide fundamental elements of shared reading that enhance understanding and motivation.	Villiger et al. (2019): 2.2
11.	**HOW WELL** (ensuring of implementation fidelity) **planned:** The instructors of the tutor training adhered to a single training protocol (detailed script) to ensure standardized delivery. The tutors were provided with an instruction booklet on the core elements of the method, and they had access to the training videos showing the procedure of the PR method throughout the implementation period. The pairs reported each training session and its duration in a record booklet (training intensity). Each pair was videotaped in one reading session around halfway through the training program to observe how the PR training was implemented (training quality) (except of 18 pairs who did not consent to being filmed).	This article: 2.2
12.	**HOW WELL** (actual implementation fidelity) **actual**: 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 sessions (volunteers: 76.2%, parents: 83.6%). 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. The data confirmed that the majority of the pairs implemented the method as intended (more details: Näpflin et al., 2020).	Villiger et al. (2019): 2.4 Näpflin et al. (2020): *Results (german: Ergebnisse), pp. 9-12* this article: 2.3.2

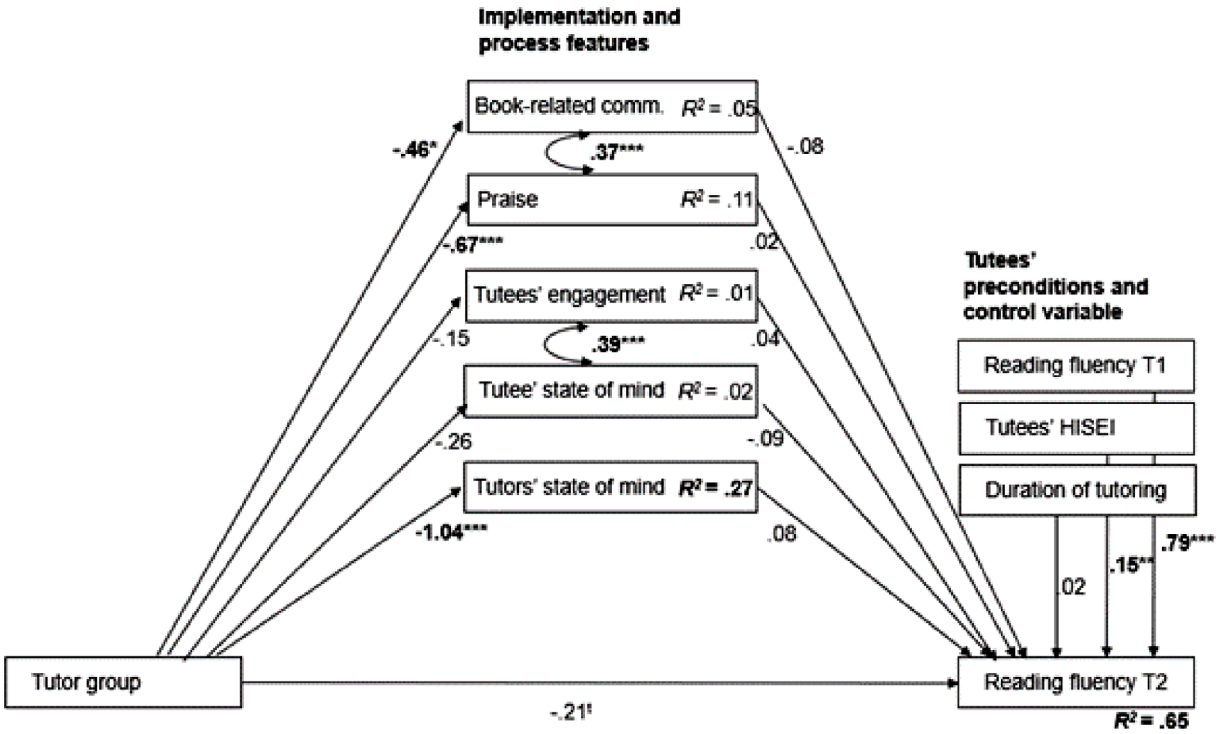

Fig. 2. Results of the path-analytical model testing the mediating effects of tutor group via implementation and process features on reading fluency T2

Note. $N = 125$. Coefficients are standardized; significant effects (one-tailed) are reported in bold; error terms omitted for ease of presentation. tutor group: 0 = volunteer, 1 = parent; book-related communication: 0 = no, 1 = yes.

specific character and influence of the setting variable on training efforts could not be determined correctly. It is possible that the children in the parent condition especially showed more engagement and effort when being recorded than at other times during the sessions. Despite the fact that video recording is still considered as state-of-the-art for analyzing learning situations (Seidel & Shavelson, 2007), the distorting influence of the presence of film staff, especially for short durations as in this study, must be considered.

Overall, the present study highlights the challenges of assessing the process and implementation features accurately and associating them to the outcome with the aim of explaining differential effects. However, for learning within tutoring settings, it is crucial to determine which components are responsible for the successful implementation of methods (Roberts et al., 2018).

4.2. The mystery about ineffective tutoring within the family

It is generally acknowledged that family background has a major impact on academic achievement (OECD, 2001). The present study confirms this by the fact that children whose parents have a higher occupational status made higher gains in reading fluency, regardless of whether their PR training was with a parent or volunteer. However, the gains in reading fluency of children in the parent condition did not differ from the control group (cf. Villiger et al., 2019), which means that they corresponded to the expected, development-related growth, and the findings presented here do not provide reasons for this. Thus, the ineffectiveness of the parent condition remains a mystery.

In the present study, tutees were children who scored low in a reading screening test, and who had therefore been recommended by their teacher to participate. This differs from previous studies on PR with parent tutors, which reported the positive effects of the method (cf. Cadieux & Boudreault, 2005; Lam et al., 2013). Even when the reasons for the ineffectiveness of the parent condition are unknown, it can be assumed that children with reading disabilities do not benefit from following a training program with their parent. The fact that parents praised their child less often and had less communication about the book suggests that the training situation was less "relaxed" and pleasant. Moreover, parents took advantage of the possibility of conducting the training during the children's spare time and at weekends which was not possible for volunteer tutors. That being said, the effort of executing additional academic work might be even more arduous if it takes place during leisure time. Based on this data, however, no conclusive recommendation against parental tutors can be made; the general willingness of the child to train with a parent is certainly decisive in any case.

4.3. Limitations and implications for future research

The sample size of the present study represents a severe limitation. It is possible that a larger sample may have shown mediation

effects to emerge more easily. However, this is a general difficulty of intervention studies because they are extremely time-consuming and costly, meaning that intervention samples are rarely large. Moreover, this study only selected students with confirmed difficulties in reading fluency, which reduces sample size possibilities even more. Other considerations emphasize that sample size alone is not crucial, but a carefully selected sample that allows for experimental manipulation of relevant moderators is equally important (Bryan et al., 2021).

Another limitation might be the fact that most of the process variables are based on only one measurement point (one video recording). This may not take sufficient account of possible variations in implementation. On the other hand, studies on the generalizability of observational data show that cross-content factors, in particular of instructional quality in teaching/learning processes, exhibit high stability (Praetorius et al., 2014), which supports the procedure chosen here. Nevertheless, as mentioned above, in the context of short video recording and tutoring, a larger number of video recordings might have increased the ecological validity of the data. Future research therefore needs to focus on how core elements of the implementation and learning processes may be accurately captured. Greater attention should be paid to qualitative instead of quantitative aspects, because a higher incidence is not always better; what is crucial is applying the correct measure (e.g., for praise).

Finally, the present study did not attain its aim of explaining the mechanisms of effective tutoring. However, it does reveal the difficulty of highlighting the crucial components of effective extra-curricular learning. More empirical studies are definitely needed to test theoretical models which try to map the complexity of learning situations such as these. However, different types of tutoring should be distinguished; whereas in highly structured settings the accuracy of implementation is of primary importance, in open settings more general measures of teaching quality - expertise, activation, responsiveness - are likely to be purposeful when explaining learning outcomes. However, the interpersonal relationship might be relevant in any type of setting, especially given the one-to-one format of tutoring.

5. Conclusions

The present PR study is one of the few (besides Topping et al., 2012) that includes implementation, process features and personal characteristics of tutors in the analyses and tries to throw light onto the mechanisms of the method. Although the features considered did not allow the fulfillment of this goal, the study goes some way to identifying the relevant aspects of effective tutoring and extra-curricular learning. Therefore, it may encourage further research into meeting and overcoming the challenges of this study, for example, by using larger samples. Moreover, the present study highlights once more the difficulty of assessing implementation and process features accurately. Further research needs to focus on this challenge and search for features that promise a high explanatory power. The aim of these efforts is to better understand the circumstances in which the PR method or any extra-curricular learning approach proves to be efficacious, and which aspects must be given particular attention when training tutors.

Funding

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

Acknowledgments

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 Amanda Mannix who was involved in editorial assistance, as well as the testing and data entry staff for their valuable work.

References

Asparouhov, T., & Muthén, B. (2018). *SRMR in Mplus*. Version May 2, 2018. https://www.statmodel.com/download/SRMR2.pdf.

Bergin, C. (2001). The parent-child relationship during beginning reading. *Journal of Literacy Research, 33*(4), 681–706, 10.1080%2F10862960109548129.

Bray, M. (2014). The impact of shadow education on student academic achievement: Why the research is inconclusive and what can be done about it. *Asia Pacific Education Review, 15*, 381–389. https://doi.org/10.1007/s12564-014-9326-9

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behaviour science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour, 5*, 980–989. https://doi.org/10.1038/s41562-021-01143-3

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.

De la Rie, S., van Steensel, R., van Gelderen, A., & Severiens, S. (2021). Effects of a Dutch family literacy program: The role of implementation. *Education Sciences, 11*(50). https://doi.org/10.3390/educsci11020050

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3-4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Elbaum, B., Vaughn, S., Hughes, M. T., & Watson Moody, S. (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*(4), 605–619. https://doi.org/10.1037/0022-0663.92.4.605

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

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

Guill, K., Lüdtke, O., & Köller, O. (2020). Assessing the instructional quality of private tutoring and its effects on student outcomes: Analyses from the German National Educational Panel Study. *British Journal of Educational Psychology, 90*(2), 282–300. https://doi.org/10.1111/bjep.12281

Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität – Diagnose, Evaluation und Verbesserung des Unterrichts*. [*Quality of instruction and teacher professionalism – Diagnosis, evaluation and improvement of teaching*]. *Klett-Kallmeyer*.
Hof, S. (2014). Does private tutoringwork? Theeffectiveness of private tutoring: A nonparametric bounds analysis. *Education Economics, 22*(4), 347–366. https://doi.org/10.1080/09645292.2014.908165
Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., et al. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *British Medical Journal (BMJ), 348*. g1687 https://www.bmj.com/content/348/bmj.g1687.
Huemer, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency among poor readers of German: Many ways to the goal. *Annals of Dyslexia, 58*, 115–137. https://doi.org/10.1007/s11881-008-0017-2
IBM Corp. (2021). *IBM SPSS Statistics for Windows*, Version 27.0. IBM Corp.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structures analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118
Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly, 31*(3), 268–289. https://doi.org/10.1598/RRQ.31.3.3
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
Laursen, B. (2005). Dyadic and group perspectives on close relationships. *International Journal of Behavioral Development, 29*(2). https://doi.org/10.1080/01650250444000450, 97–11.
Lohaus, A., & Wild, E. (2021). Extracurriculare Förderangebote für benachteiligte Kinder und deren Eltern: Ein Angebot-Aneignungs-Modell zur Inanspruchnahme und Wirkung. [Extracurricular support services for disadvantaged children and their parents: A supply-use model of take-up and impact]. *Zeitschrift für Pädagogische Psychologie, 35*(1), 1–10. https://doi.org/10.1024/1010-0652/a000268
Matsumura, L. C., Garnier, H. E., & Spybrook, J. (2013). Literacy coaching to improve student reading achievement: A multi-level mediation model. *Learning & Instruction, 25*, 35–48. https://doi.org/10.1016/j.learninstruc.2012.11.001
Metze, W. (2009). *Stolperwörter-Lesetest* (Unpublished document).
Morgans, A., & Allen, F. (2005). Education. Getting ethics committee approval for research: A beginners guide. *Journal of Emergency Primary Health Care, 3*(3). https://ajp.paramedics.org/index.php/ajp/article/view/322.
Munter, C., Wilhelm, A. G., Cobb, P., & Cordray, D. S. (2014). Assessing fidelity of implementation of an unprescribed, diagnostic mathematics intervention. *Journal of Research on Educational Effectiveness, 7*, 83–113. https://doi.org/10.1080/19345747.2013.809177
Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide 1998-2017* (8th edition). Muthén & Muthén.
Näpflin, C., Frommelt, M., Hugener, I., Tettenborn, A., Villiger, C., Hauri, S., & Hartmann, E. (2020). Implementationsqualität unter der Lupe: Unterscheiden sich Eltern und Lesecoachs in der Umsetzung eines Trainings zur Förderung der Leseflüssigkeit? [Scrutinizing implementation quality: Do parent tutors and volunteer tutors differ in the implementation of a reading fluency training?]. *Psychologie in Erziehung und Unterricht, 67*(2), 95–111. https://doi.org/10.2378/peu2019.art17d
Neely-Barnes, S. L., & Dia, D. A. (2008). Families of children with disabilities: A review of literature and recommendations for interventions. *Journal of Early and Intensive Behavior Intervention, 5*(3), 93–107. https://doi.org/10.1037/h0100425
Neuenschwander, M. P., Vida, M., Garrett, J. L., & Eccles, J. S. (2007). Parents' expectations and students' achievement in two western nations. *International Journal of Behavioral Development, 31*(6), 594–602. https://doi.org/10.1177/0165025407080589
O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*, 33–84. https://doi.org/10.3102/0034654307313793
OECD. (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. OECD. https://doi.org/10.1787/9789264195905-en
Park, H., Buchmann, C., Choi, J., & Merry, J. J. (2016). Learning beyond the school walls: Trends and implications. *Annual Review of Sociology, 42*, 231–251. https://doi.org/10.1146/annurev-soc-081715-074341
Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology, 36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002
Powell, D. R., & Carey, A. (2012). Approaches to program fidelity in family literacy research. In B. H. Wasik (Ed.), *Handbook of family literacy* (2nd ed., pp. 387–400). Routledge.
Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction, 31*(1), 2–12. https://doi.org/10.1016/j.learninstruc.2013.12.002
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
Roberts, G., Scammacca, N., & Roberts, G. J. (2018). Causal mediation in educational intervention studies. *Behavioral Disorders, 43*(4), 457–465. https://doi.org/10.1177/0198742917749560
Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the Classroom: Teacher expectation and pupils' intellectual development*. Holt: Rinehart & Winston.
Seidel, T., Prenzel, M., & Kobarg, M. (2005). How to run a video study. Technical report of the IPN video study. *Waxmann*.
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
Topping, K. J. (2001). Thinking Reading Writing. A practical guide to paired learning with peers, parents and volunteers. *Continuum*.
Topping, K. J. (2020). A theoretical model of intergenerational tutoring. *Journal of Intergenerational Relationships, 18*(1), 88–105. https://doi.org/10.1080/15350770.2019.1646182
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
Villiger, C., Hauri, S., Tettenborn, A., Hartmann, E., Näpflin, C., Hugener, I., & Niggli, A. (2019). Effectiveness of an extracurricular program for struggling readers: A comparative study with parent tutors and volunteer tutors. *Learning & Instruction, 60*, 54–65. https://doi.org/10.1016/j.learninstruc.2018.11.004
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*, 126–142. https://doi.org/10.1177/0022219407313426
Wilder, S. (2014). Effects of parentalinvolvement on academic achievement: Ameta-synthesis. *Educational Review, 66*(3), 377–397. https://doi.org/10.1080/00131911.2013.780009
Zhang, E., & Liu, Y. (2022). Effects of private tutoring intervention on students' academic achievement: A systematic review based on a three-level meta-analysis model and robust variance estimation method. *International Journal of Educational Research, 112*, 1–21. https://doi.org/10.1016/j.ijer.2022.101949