



Fostering word fluency of struggling third graders from Germany through motivational peer-tutorial reading racetracks

Anne Barwasser¹ · Karolina Urton^{1,2} · Matthias Grünke¹ · Marko Sperling¹ · David L. Coker³

Accepted: 28 May 2021 / Published online: 6 June 2021
© The Author(s) 2021

Abstract

Automation of frequently used words is a key component in the development of reading fluency. However, acquiring fast word recognition skills is a serious challenge for many children in their early years of formal education. Lagging word recognition leads to general reading problems, as fluency is a vital prerequisite for text comprehension. Recent research shows that the percentage of struggling elementary school readers in Germany is increasing, speaking to the need for widespread implementation of effective word recognition interventions. This pilot study aims to provide preliminary evidence of the effectiveness of peer-tutorial reading race-track training with an integrated motivational system for the sight word fluency of German struggling elementary school students. The intervention comprised twelve 15-min teaching units over a period of three weeks. To encourage reading motivation, the intervention included graphing of performance scores and a group contingency procedure. A control-experimental group design ($N=44$) with pre-, post-, and two follow-up measurements (each after five weeks) was employed to investigate the impact of the treatment on decoding sight words at an appropriate speed. Results demonstrated a significant performance increase in the treatment group, relative to the control group. The effect size can be considered very high (partial $\eta^2 = .76$), indicating that this brief training has the potential to enhance the word recognition of struggling elementary students.

Keywords Effective intervention · Reading fluency · Motivational components · Peer tutoring · Racetracks

✉ Anne Barwasser
anne.barwasser@uni-koeln.de

¹ Department of Special Education and Rehabilitation, University of Cologne, Cologne, Germany

² Department of Educational Sciences, University of Muenster, Muenster, Germany

³ School of Education, University of Delaware, Newark, USA

The importance of remediating reading problems

The ability to decode printed or written symbols to arrive at meaning plays a vital role in academic success and the development of the human mind (Aaron et al., 2008). Tragically, an ever-increasing number of students do not even attain basic reading comprehension skills (Cibulka & Cooper, 2017). Worldwide, about 10 to 15% of the population experience severe difficulties in understanding written text, despite having attained at least a basic level of education (Dyslexia International, 2017; Sprenger-Charolles & Siegel, 2016). Moreover, the percentages of struggling readers seem to increase. In Germany (where this study took place), the Progress in International Reading Literacy Study Survey (PIRLS) indicated that there was an upward trend of weak German fourth-grade readers from 2011 to 2016. The percentage of struggling students rose from 16.9% in 2001 to 18.9% in 2016. Furthermore, the PIRLS survey revealed a decrease of reading motivation, especially for struggling readers (Harju-Luukkainen et al., 2020). The negative consequences of severe reading difficulties are far-reaching. They lead to generally poor performance in school and often a dramatic loss of motivation in all core subjects. In the long run, they increase the risk of school dropout, unemployment, and poverty (Macdonald et al., 2016).

Typically, reading problems are apparent early in the developmental sequence. Students who struggle experience difficulty understanding the relationships between letters and sounds and have trouble decoding phonologically regular words. It takes them much longer than normally achieving peers to acquire a sufficient amount of sight vocabulary, which in turn limits their reading fluency (Kendeou et al., 2009). Sight words are immediately recognized without paying attention to individual sounds that simplify the process of finding the pronunciation and meaning of familiar words automatically (Ehri, 2005). To achieve advanced reading skills, the automatic recognition of words as sight words is of crucial importance (Balass et al., 2010). To support students in their reading development, it is vital to examine how reading skills develop, where exactly problems in reading acquisition may arise, and how to strengthen important skills.

The signification of word recognition skills for text comprehension

Our understanding of the importance of word recognition is informed by contemporary reading models. The dual route theory (DRT) and dual route cascaded model (DRC) of reading (Coltheart, 2005; Coltheart et al., 2001) include two cooperative systems: (a) a lexical route (orthographical decoding) and (b) a non-lexical route (phonological recoding). The lexical route refers to a mental dictionary that helps to recognize words by sight; the non-lexical route enables skilled readers to identify parts of words and connect them to single sounds to decode unfamiliar words. Skilled readers use the non-lexical only for less frequent or unknown words, but struggling readers use it all the time (Coltheart, 2005). Ehri

(2005) also considered sight word reading and a lexical route in her stage-based model of reading, which consists of four phases of reading acquisition: (1) the pre-alphabetic, (b) the partially alphabetic, (c) the fully alphabetic, and (d) the consolidated alphabetic phase. Children in these stages can be distinguished by the extent to which they incorporate phonological awareness and knowledge of grapheme-phoneme correspondences into the building of memory connections between written words and their pronunciation and meaning. Mature readers have succeeded in building up a visual vocabulary that enables them to retrieve words quickly and unconsciously via the lexical route from their mental lexicon (Ehri, 2005). Students who have already reached this state show a greatly facilitated reading process (Morris & Perney, 2018). One similarity across these theories is that sight-word reading (orthographic decoding) is far more efficient than relying on phonological recoding.

Empirical research also has supported the superiority of orthographic decoding. Based on the fundamental tenets of the DRC model, Knoepke et al. (2014) showed that orthographic decoding is better at predicting sentence and text comprehension than phonological recoding in German elementary school children. Further, the results indicated that the ability to recognize words is a significant predictor of general reading skills in all grades. It has been documented repeatedly that high-frequency words are processed more quickly than low-frequency words (Fischer et al., 2014; Kennedy et al., 2013). A study by Masrai (2019) showed that mid-frequency words correlate with reading comprehension in L2 reading. Knowledge of only high-frequency words seems not to be sufficient for skilled reading comprehension (see also Schmitt & Schmitt, 2014; Schmitt et al., 2011). Furthermore, mid-frequency words have not been addressed adequately in pedagogy (Schmitt & Schmitt, 2014). According to other studies, students should receive training in lower frequency words and not only high-frequency words (Calabrèse et al., 2016; Stolowy et al., 2019).

Many struggling students face difficulties using the lexical path. They have restricted access to their mental dictionary and therefore are unable to comprehend the meaning of the text. This occurs because many words are not stored in the mental lexicon, and students cannot access them (Samuels, 2006). As a consequence of inefficient word-reading automaticity, they frequently rely on the non-lexical route (De Jong et al., 2012). This highlights how poor automatization is a key factor for reading fluency, and it may lead to a lack of naming speed and an overall insufficient reading speed (Balass et al., 2010).

As indicated above, reading fluency is a key prerequisite for comprehension. In fact, many studies have shown that the ability to decode correctly and swiftly has a bridging function between transferring the written code into the language and the ability to extract meaning from text (e.g., Kim et al., 2010; National Institute of Child Health and Human Development, 2010; Nese et al., 2013; Roehrig et al., 2008; Schwanenflugel & Kuhn, 2015). A central role of reading fluency is that it alleviates the demands on readers' limited working memory capacity. In terms of reading acquisition, if too many cognitive resources are devoted to decoding, there is little capacity left for higher process capabilities (Guerin & Murphy, 2015; Paige, 2011; Rasinski, 2003).

According to Hayes (2016), the aforementioned considerations illustrated that the ability to retrieve words quickly is a fundamental reading skill and key qualification for further reading development. Volpe et al. (2011) also argued in favor of sight word reading and advocate for implementing sight word reading instruction in the classroom because it is important to establish those skills before students learn to read complete sentences or passages of text. Moreover, becoming better in sight word reading does not solely lead to increased reading fluency and comprehension; it also builds reading confidence and reduces reading frustration across weak readers (Musti-Rao et al., 2015). Other authors have also argued for training sight words in reading because a very important part of teaching reading is teaching sight words and irregular words as they contribute to text comprehension (Sullivan et al., 2013). Studies showed a positive relationship between word knowledge and word pattern knowledge on basic word reading (in poor German readers) (Rothe et al., 2015; Zarić et al., 2020; Zarić & Nagler, 2021) and on sentence-level reading (Zarić & Nagler, 2021). These results underline the necessity of training orthographic knowledge to enhance reading proficiency. Other studies showed a positive effect of sight word training on trained words, untrained words and word reading fluency (McArthur et al., 2015a, 2015b). Due to the significant impact of a limited number of sight words on reading competence, research on the effectiveness of interventions to promote the use of the lexical pathway is central. Furthermore, work in this area can provide teachers with evidence-based practices to develop these skills (Kuhn & Stahl, 2003). Intervening early after the first signs of reading difficulty has been shown to be particularly important to counteract long-term failure (Volpe et al., 2011). This is illustrated by the fact that, without intervention, around 74% of reading impaired children at the age of nine years maintain their deficits in secondary school (Lee & Yoon, 2017).

The necessity of repeated reading for fostering word recognition skills

Once children demonstrate sufficient phonological awareness and adequate decoding skills, word recognition can be promoted in various ways (Hjetland et al., 2017). The most common and most effective method to enhance the reading fluency in struggling learners is repeated reading (National Institute of Child Health and Human Development, 2000). It is defined as an approach “that consists of rereading a short and meaningful passage until a satisfactory level of fluency is reached” (Samuels, 1979, p. 404). The high potency of this method lies in the fact that children need repetition to automate the retrieval of words from the mental lexicon (Mraz et al., 2013).

In their meta-analysis, Chard et al. (2002) demonstrated that repeated reading leads to increased skills to decode accurately and effortlessly in elementary school children with learning difficulties ($d=0.68$). Lee and Yoon’s (2017) meta-analysis, which included empirical studies of the last 25 years to estimate the effects of repeated reading, also confirmed the effectiveness and yielded a total Hedges’ g of 1.41 ($p < .001$) for students with reading disabilities.

Ways to motivate students to engage in repeated reading

However, even very effective interventions like repeated reading miss their intended mark if students are unwilling to get involved in them. For so many children, decoding symbols to arrive at meaning is extremely arduous; they resent it in its entirety (Sabatini et al., 2018). For them, reading interventions must always be complemented with motivational techniques that encourage them to give it a try and to persist (Guthrie & Wigfield, 2000; Harju-Luukkainen et al., 2020; Klem & Connell, 2004).

A number of motivational concepts usually have a marked positive effect on the academic achievement of students, including positive reinforcement, self-monitoring, and praising (Alberto & Troutman, 2008; Copper et al., 2008). In particular, graphing of individual performance scores that allows students to monitor their own progress has demonstrated an especially strong impact on learning positive behaviors (Amato-Zech et al., 2006; Legge et al., 2010) and school performance (Gunter et al., 2003). Often, instruction combines individual motivational techniques. For example, self-monitoring is often used in combination with positive reinforcement (Briesch & Chafouleas, 2009; Joseph & Eveleigh, 2011).

Motivational methods can be applied on an individual level (e.g., positive reinforcement, self-monitoring, praising) and/or on a group level (e.g., group contingencies; Gunter et al., 2003; Stephen & Singh, 2017). Despite the importance of tracking one's own learning progress, when teaching a whole class of approximately 30 students, using motivational techniques on a group level has advantages over using them on an individual level. Group contingencies, which occur when all group members work together to achieve a certain reward (Kerr & Nelson, 2006), emphasize peer influence to reduce problematic and disruptive behavior as well as to trigger positive behavioral attitudes and enhance social behavior (Donaldson et al., 2011; Ginsburg-Block et al., 2006; Hulac & Benson, 2010; Ling & Barnett, 2013; Ling et al., 2011; Wills et al., 2016). In addition, they may foster peer support and community in a classroom (Groves & Austin, 2019). This approach to motivation is especially suitable for use with a whole class of diverse learners and generally has demonstrated a positive impact on students' performance (Bowman-Perrot et al., 2013; Pappas et al., 2010; Rohrbeck et al., 2003).

A cooperative game-based approach to integrate repeated reading into regular inclusive classroom instruction

Unfortunately, no matter how potent a particular approach may be, it seldom finds its way into daily classroom instruction (Johnson & Semmelroth, 2014). A primary reason for the often relatively wide research-practice gap in education lies in the prevalent attitude among teachers that many evidence-based techniques are not compatible with everyday school life and the fact that mostly they are not instructed in using evidence-based practice (Hirschhorn & Geelan, 2008;

Scheeler et al., 2009). It is very challenging to find ways to adequately attend to the individual needs of a particular child without neglecting the rest of the class. However, researchers must meet this requirement and guide teachers if we want schools to successfully practice inclusion.

Therefore, we implemented the incorporation of repeated reading in conjunction with motivational components in the classroom by adopting a cooperative learning game approach. Class-wide peer tutoring, a system in which learners help each other in working pairs, can unburden teachers while providing everyone with intensive practice time (Bond & Castagnera, 2010). This technique is mindful of individual differences in an inclusive setting, so that all students can be involved whether or not they are performing well. It has been demonstrated to be especially useful when trying to foster reading skills in children (Dufrene et al., 2006; Mercer et al., 2011; Spencer, 2006). Class-wide peer tutoring has been found to enhance the reading competency of children with and without behavioral problems (Bowman-Perrott et al., 2013; $d=0.77$), as well as students with differing levels of social competencies (Ginsburg-Block et al., 2006; $d=0.28$) and learning-related behaviors (Ginsburg-Block et al., 2006; $d=0.45$).

To apply repeated reading by letting students teach each other, we used an educational board game, racetracks. Its underlying idea is simple: Flashcards with particular practice words are placed face down on blank cells on a game board, often designed to look like a Formula 1 circuit. The tutee rolls a die and moves the playing piece the respective number of spaces forward. As it lands on a cell, the tutor picks up the corresponding flashcard, presents the word and asks the tutee to read it out loud. In case of a pause of more than three seconds or a mistake, the tutor models reading the word and asks the tutee to repeat it. The game has no winners or losers; its entire purpose is to provide intense sight word practice (Sperling et al., 2019).

Racetracks has shown positive effects on reading fluency in several studies (Barwasser et al., 2021; Erbey et al., 2011; Green et al., 2010; Grünke, 2019; Grünke & Barwasser, 2019; Hopewell et al., 2011; Hyde et al., 2009; Sperling et al., 2019 and it can be used with a wide range of learners (Falk et al., 2003), which makes it useful for students in inclusion settings. However, all respective studies used a one-on-one teaching approach, unsuitable for inclusive classroom instruction.

Research questions

Even though racetracks and the graphing of performance scores, as well as group contingencies techniques, have been evaluated a number of times with samples from different populations, to our knowledge, these interventions have never been implemented in combination and within a framework of class-wide peer-tutoring. Thus, we cannot draw on the findings from previous studies as we specify our research questions, and we must be cautious as we phrase them. Our work is a pilot study that aims to explore whether crucial components of full-scale projects are feasible.

In particular, we wanted to pursue the following questions:

- (1) Does the implementation of a peer-tutorial reading racetracks intervention with motivational components in an inclusive school setting improve the reading fluency of struggling third-grade readers?
- (2) If there are effects of the intervention, do those effects on reading fluency persist five and ten weeks later?

Methods

Design

An experimental control group design with a pre- and post-, as well as a follow-up measurement was applied. The treatment group received a reading racetracks (RR) intervention, and the control group worked with math racetracks (MR). To ensure a high level of internal validity, both groups were taught at the same time, for the same duration, by the same number of special education college students, and using similar materials. The 18 interventionists worked in teams of two, in a control class or treatment class.

The control group received an MR training with multiplication exercises and the treatment groups received an RR intervention with 30 words for reading training. All participants from both groups were compared regarding sight word fluency through a pretest (t1), posttest (t2), and two follow-up measurements (t3 and t4). Both groups met four times a week for 15 min, lasting a total of three weeks. All participants attended 12 intervention sessions which took place with the whole class in a classroom with the class teachers present. The first follow-up happened five weeks after post testing, including two weeks of autumn school holidays, and the second follow-up was conducted ten weeks after post testing.

Participants and setting

The final sample ($N=44$) consisted of third graders from nine classes of five different elementary schools in a high-socioeconomic metropolitan area of Germany. On average, 192 students were enrolled in each school ($SD=8.4$). In every school, except one, two whole classes participated either as experimental group or control group.

Although the RR and MR trainings were implemented in a classroom context with all students present, we focused only on those with the lowest and highest skill levels. To identify our sample, we conducted the Salzburg Reading Screening Tests (SLS; Mayringer & Wimmer, 2014) with 225 students. The SLS measures reading fluency by asking children to decide whether different sentences that are presented to them make sense within an assigned time period. We chose the SLS due to school time constraints, as it is a group screening. Thus, we conducted it with the whole class at once to capture the reading performance of all students of the participating classrooms.

Using the results, we ranked students based on their reading performance from lowest to best performance. Subsequently, we paired the weakest students (tutees) were with the strongest students (tutors). The classification process involved dividing participants into tutors and tutees by taking (a) the ones with the best scores (tutors, reading level at least > 100) and (b) the ones with the weakest scores (tutees, reading level maximum of < 89). Those students with an average reading score were paired together and participated in the intervention, but no data were collected from them and they were therefore not considered part of the sample. Also, no data was collected from the tutors. However, we only collected data from those children selected as tutees ($N=44$), because the weaker readers should not have been able to read the training words, but the tutors, the stronger readers, should have been able to read the words and to give adequate feedback and help. In all, the final tutees had a reading quotient (LQ) of $M=74.6$ ($SD=8.4$) and the tutors an LQ of $M=110.4$ ($SD=8.5$). According to the SLS manual, an LQ of 80–89 is considered below average, 70–79 is considered weak, and less than or equal to 69 is considered very weak. The LQ expresses the extent to which the measured reading ability deviates from the average of the norming sample. The same scaling is used as for intelligence test measurements: where 100 stands for the mean value with an SD of 15 in each case.

Additionally, to gain more information about the tutees, an intelligence test (CFT-20-R, Weiß, 2006; German adapted Version of the Cultural Fair Intelligence Test; Cattell & Cattell, 1963) and the Strength and Difficulties Questionnaire (SDQ; Goodman, 1997) were implemented. The CFT-20-R consists of the following four subtests: series continuation, classifications, matrices, and topological conclusions. According to its manual, the correlations with other German intelligence tests range between $r=0.57$ and $r=0.73$. Additionally, the CFT-20-R is moderately correlated with school grades in mathematics ($r=0.50$). The SDQ consists of five scales (emotional symptoms, behavioral problems, hyperactivity/dislike, peer relationship problems, and prosocial behavior), the first four scales of which can be combined into an overall problem value. Each scale consists of five items with three gradients (0 = *not applicable*, 1 = *reasonably applicable*, and 2 = *definitely applicable*) that are typically completed by teachers. Studies on the psychometric characteristics of the German version of the SDQ indicate good internal consistency. In addition, the German version shows good validity (Becker et al., 2004). Due to time constraints, for the SDQ measurement, the short (16 instead of 25 questions) form was completed by the teachers. For the total problem score, values between 12 and 15 are considered borderline and values between 16 and 40 are considered abnormal. SDQ results for the reading racetrack group were $M=9.8$ ($SD=5.7$) and for the control group $M=9.4$ ($SD=6.8$; Table 2). More specific information about the children (age, gender, special needs, and German L2) was collected through a teacher questionnaire developed by the authors.

To split up the participants into experimental group and control groups, matched pairs of tandems, consisting of a tutor and a tutee, were identified based on the pretest results per school. Each pair was randomly assigned to either the treatment group (working with reading racetracks) or to the control group (working with math racetracks). No significant differences between the experimental and control group,

Table 1 Socio-Demographic Characteristics of Participants Comparing MR and RR

	Maths Racetracks (N=22)		Reading Racetracks (N=22)		<i>p</i>
	<i>N</i>	<i>M</i> (SD)	<i>N</i>	<i>M</i> (SD)	
Age	22	8.6 (0.6)	22	8.6 (0.6)	.87
IQ (CFT)	22	97.1 (15.2)	22	96.2 (11.6)	.82
	<i>N</i>		<i>N</i>		
<i>Gender</i>					.34
Male	9		6		
Female	13		16		
<i>Special Needs</i>					.55
LD	2		1		
Non LD	20		21		
<i>German L2</i>					.76
Yes	10		9		
No	12		13		

Maths Racetracks = control group; Reading Racetracks = treatment group; Learning Disability (LD); Cattell Culture Fair Intelligence Test (CFT); Mean (M), Standard Deviations (SD), participants (N)

Table 2 Results of SDQ and Reading Test Comparing MR and RR

	Maths Racetracks (N=22)		Reading Racetracks (N=22)		<i>p</i>
	<i>N</i>	<i>M</i> (SD)	<i>N</i>	<i>M</i> (SD)	
SDQ	22	9.4 (6.8)	22	9.8 (5.7)	.84
Reading ^(SLS)	22	75.4 (8.8)	22	73.9 (8.1)	.55

SDQ = Strength and Difficulties Questionnaire; Reading = SLS German Reading Screening; Maths Racetracks = control group; Reading Racetracks = treatment group

in terms of gender, age, special needs, German as L2, cognitive abilities, reading proficiency and externalizing problem behavior were found (Tables 1 and 2).

Further, to determine the words for the reading racetracks intervention, word-reading pre-tests were used. This pretest consisted of two PowerPoint presentations. On the slides were individual words, each with three consecutive hashtags, which the children were told to read within 1s of the presentation. The one 1-s cycle was automatically preset, so the slides changed by themselves after 1s. We utilized a list of the 1000 most frequently used German words (<https://wortschatz.uni-leipzig.de/de>) and the database ChildLex (Schroeder et al., 2015) in addition since the prior mentioned list refers more to older students' vocabulary and words mentioned there are of lower frequency for children but still important with regard to the future. The ChildLex database was used to determine the words' overall frequency.

Each PowerPoint presentation contained 70 words to increase the probability of finding 30 words for the intervention that were not familiar sight words for all

subjects. The two presentations were divided into two days to avoid overtaxing the children. The final 30 words of training, which were the same for each participant and were not yet stored as sight words across all children, were of different syllables and a rather mid-frequency of $M=38.4$ ($SD=37.0$) ("Appendix"). This means that the words appeared on average 38 times per million words in a corpus. In comparison, low-frequency words appear five times and high-frequency words more than 100 times per million words (Brysbaert et al., 2018).

Dependent variables and measurement

The dependent variable in this study was automatic recognition of the training words, operationalized by the number of words that the participants read correctly within 1 s after having been presented with it (words read correctly, WRC). As suggested by Ehri (2005), "reading words within one second of seeing them is taken to indicate sight word reading" (p.136). A PowerPoint presentation that contained all 30 training words was used, each on a 1-s timer, to determine whether each word was read by sight. The words were presented in random order for each measurement. All tutees were measured independently in another room to reduce variables of potential influence. The 30 training words were the same for all participants.

Interventions

The intervention was conducted in the regular class setting. Both children of a pair (tandem) sat opposite each other and the racetrack game in the middle of them. At the start signal, the tutee rolled the dice and moved the piece to the appropriate square according to the number (see Fig. 1). Flashcards on this space were turned over and the tutee was asked to read them out loud. The job of the tutor was to attend to the correct pronunciation of the word and provide feedback to the tutee if needed. Tutees had three seconds to correct themselves. If someone did not make it, the tutor read the word aloud correctly, and the tutee was asked to read it again. Flashcards



Fig. 1 Self-made racetrack board

that had been read were left with the front side up to ensure that the other words would be turned around and read so that the tutees could not land on the space of an already read word. After that the tutee rolled the dice again and the procedure was conducted for 10 min. If all flashcards were read before the ten minutes ended, they were shuffled and placed back on the racetrack cells.

After finishing the game, there was a brief word-reading assessment that the tutees graphed. The interventionists used a 1-min timer, and the tutee was asked to read as many words correctly as possible within this minute. Then, using a blank graph, the amount of correctly read words was drawn in the provided cells. The first row was filled in at the first intervention day, and another row was completed on each intervention day. The pairs also received rewards based on their progress over time. A pair was given one point for reading the same number of words correctly as they had the previous intervention. They received two points if there was improvement from the last session. These points were collected in front of the class, where each pair added the amount of points in a bottle to work toward achieving an overall class-wide improvement goal.

The same procedure was used for the control group intervention except the words on the flashcards were replaced with math problems. For the motivational system, the tutee was asked to answer as many problems correctly as possible within two minutes. The interventionists were intensively trained for two days in the implementation of the treatment and control group. In addition, there was a standard treatment protocol with the same instructions on implementation, which the interventionists had to follow.

Material

Materials for the experimental and control groups were exactly equivalent except for the content of the flashcards. Both groups were provided with a racetrack board consisting of 30 empty spaces for the flashcards. The reading racetrack group (experimental) had 30 flashcards with the words on one side, and math racetrack group (control) were provided 30 flashcards with multiplication problems on one side. The tandems which were not included in data collection got different words than the tandems from which data was collected since they were too good readers. Both groups were also given a self-graphing sheet entitled either “Reading Racetrack” or “Math Racetrack.” These sheets consisted of 12 rows with 30 empty cells, each of which represented an intervention day.

Treatment Integrity

To draw valid conclusions regarding the effectiveness of an intervention, treatment integrity is required (Hagermoser et al., 2009). Therefore, a detailed script was provided for the interventionists. The guidelines contained the exact procedure per session regarding the reading intervention and the reward system. All interventionists and the class teachers were instructed in the intervention procedure in addition to

all students of the class (including all tutees and tutors—also those from whom no data were collected). To assess adherence, exposure, quality, and dosage of the intervention, the interventionists were asked to complete a checklist after each session. The questionnaire contained questions like: “Did you, as interventionist, follow the script?”, “Did you adhere to the time frame?”, and “Were all the materials available?”. After each session, the interventionists completed the 18-question questionnaire and submitted a list of students who had participated in the intervention. Additionally, one-third of the sessions (six) were observed by an independent observer who used the treatment integrity sheet to assess the implementation of the session. Interrater agreement equaled 100% between the interventionists and between the external observers. Finally, the intervention was carefully monitored by the first author, who conducted weekly meetings with all interventionists and maintained almost daily contact with all interventionists.

Social validity

Social validity is necessary to determine the acceptance and usefulness of interventions (Briesch et al., 2013; Wolf, 1978). Using the Usage Rating Profile—Intervention by Briesch et al. (2013), we distributed a questionnaire to assess the acceptance, understanding, and feasibility of the intervention among students and teachers. The social validity was assessed by separate seven-item questionnaires for teachers and students. Both questionnaires used a 6-point Likert scale ranging from 0 (*no agreement*) to 5 (*absolute agreement*). Items on the teacher questionnaire were created to assess their understanding (e.g., “The intervention is a good way to improve the reading fluency of students”), acceptance (e.g., “I would use the intervention in my lessons as well”), and perceptions of the feasibility of the intervention (e.g., “The total time required for the intervention procedure was manageable”). Similarly, items on the student questionnaire were designed to assess their acceptance (e.g., “I gladly came to the sessions”) and understanding of the intervention (e.g., “I understood the purpose of the intervention well”). For the students’ survey, the interventionists left the room and the teachers read out loud each scale with all the items in order (a) to avoid bias in the answers if the interventionists had done the questioning, and (b) to ensure that all students understood the questions.

Data analysis

To answer the research question, a 2 (Conditions: Treatment, Control) \times 4 (Time: Pre, Post, Follow-up 1, Follow-up 2) repeated measures analysis of variance (rmANOVA) was conducted to examine the intervention effect on correctly read words per second. Based on the rmANOVA, separate analyses were carried out examining the differences between treatment and control group, as well as between pre-, post- and follow-up data.

If the assumptions of sphericity were not fulfilled (Mauchly test of sphericity), the F estimate was based on the Green House Geisser correction with adjustment

of the degrees of freedom (Field, 2013). The significance of between-subjects and within-subjects effects was tested by using independent and pairwise Welch test comparisons, applying the Bonferroni adjustment (using the mean difference of time_i and time_j). The significance level was set at $p < 0.05$.

Violations of equality of error variances and covariance were tested by means of the Levene test and the Box's test. A robust rmANOVA using the R-package WRS2 (Mair & Wilcox, 2019) was performed to validate the effects.

Results

Preliminary analysis

Prior to the main analysis, the distribution of the dependent variable was tested. The Mauchly test for sphericity was significant ($Mauchly-W=0.33$, $p < .001$). To correct this violation, the Greenhouse–Geisser adjustment was used. When testing for homogeneity of the error variances, as assessed by Levene's test, the results for the post-test ($p < .01$) and the Follow-up 1 condition ($p < .05$) were found to deviate significantly from the null hypothesis, which was evident for the homogeneity of covariances, as assessed by Box's test ($p < .001$). These violations were addressed by the application of a robust rmANOVA using the R package WRS2 (Mair & Wilcox, 2019).

Main analysis

Descriptive results for treatment and comparison groups at pre-, post-, follow-up 1, and follow-up 2 test are summarized in Table 3.

There was no statistically significant pretest difference between students in the reading racetracks and the math racetracks group for WRC $t(41.86)=0.007$, $p=.94$, $d=0.002$.

The rmANOVA to test the interaction effect (Time \times Group) showed a significant result, Greenhouse–Geisser $F(1.71, 71.73)=133.60$, $p < .001$, *partial* $\eta^2=0.76$. The inner-subject main effect for time indicates that this factor has a statistically significant influence on WRC for the reading racetracks group,

Table 3 Descriptive Statistics for Words Read Correctly

Measures	Pretest <i>M</i> (SD)	Posttest <i>M</i> (SD)	Follow-up 1 <i>M</i> (SD)	Follow-up 2 <i>M</i> (SD)
WRC				
Comparison ($N=22$)	1.3 (1.9)	2.2 (2.5)	2.8 (4.2)	4.4 (4.8)
Treatment ($N=22$)	1.3 (1.8)	23.9 (5.3)	23.1 (7.3)	19.8 (7.1)

WRC = Words read correctly, Means (*M*), Standard Deviations (SD)

$F(1.62, 34.02) = 203.96$, $p < .001$, *partial* $\eta^2 = 0.91$. Bonferroni tests for pairwise comparisons revealed a statistically significant increase of WRC after intervention (22.64 , $p < .001$, $d = 5.77$) that remained stable through the follow-up 1 assessment (21.77 , $p < .001$, $d = 4.09$) and follow-up 2 (18.55 , $p < .001$, $d = 3.59$) in comparison to the pretest. Although the performance between the posttest and follow-up 1 remained constant (0.86 , $p = 1.00$, $d = 0.14$), there was a decrease in the WRC from posttest to follow-up 2 (4.09 , $p < .001$, $d = 0.66$), as well as from follow-up 1 to follow-up 2 (3.23 , $p < .001$, $d = 0.45$), albeit with medium effect. Regarding the math racetracks group, there was also a significant influence of the main factor time on WRC, $F(1.64, 34.42) = 4.13$, $p < .05$, *partial* $\eta^2 = 0.16$. The pairwise comparisons assessing the influence of time revealed no significant differences. This refers to the comparison of the pretest with the posttest (0.86 , $p = .31$, $d = 0.39$) and with follow-up 1 (1.50 , $p = 0.26$, $d = 0.46$) and follow-up 2 (2.10 , $p = .15$, $d = 0.57$), as well as between the posttest and follow-up 1 (0.64 , $p = 1.00$, $d = 0.19$) and follow-up 2 (1.23 , $p = .52$, $d = 0.32$), and between follow-up 1 and follow-up 2 (0.59 , $p = 1.00$, $d = 0.13$).

The analysis of the between-subject main effect indicates that there was a significant main effect for group, $F(1, 42) = 137.98$, $p < .001$, *partial* $\eta^2 = 0.77$. Although there was no group difference for the pretest $t(41.86) = 0.007$, $p = .94$, $d = 0.03$, significant group differences were shown for all further measurement times (posttest: $t(29.95) = 308.45$, $p < .001$, $d = 5.29$; follow-up 1: $t(33.63) = 126.17$, $p < .001$, $d = 3.39$; follow up 2: $t(36.81) = 81.06$, $p < .001$, $d = 2.55$).

Due to the significance of the Levene test and the Box's test for homogeneity of the error and for homogeneity of covariances, a robust rmANOVA was conducted, which supports the results of the rmANOVA in terms of the interaction effect Time \times Group ($F(3, 15.48) = 185.10$, $p < .001$, $\eta^2 = 0.97$), as well as the main effect for time ($F(3, 15.48) = 197.50$, $p < .001$, $\eta^2 = 0.97$) and group ($F(1, 19.90) = 160.48$, $p < .001$, $\eta^2 = 0.96$).

Social acceptability of the reading racetrack intervention

To ascertain the social validity of the intervention, results from questionnaires for the teachers (Table 4) and the reading racetracks students (Table 5) are provided. Across all seven items, the six teachers rated the intervention in all three areas (acceptance, understanding, feasibility) as positive, with a general rating between 3 (*agree*) and 5 (*absolutely agree*). The items SocV5 ("I would use the intervention in my teaching") and SocV7 ("The material resources required for this intervention were appropriate") had the highest ratings. In total, questionnaires were returned by six of the nine teachers.

For the reading racetracks students ($n = 22$), the average of all seven items in the questionnaire ranged from 3 (*agree*) to 4 (*strongly agree*). The items SocV1 ("The racetrack helped me to read words correctly"), SocV4 ("I learned a lot during the intervention") and SocV5 ("I came to the sessions with pleasure") were rated highest.

Table 4 Social Validity Questionnaire Teachers

Items	<i>N</i>	<i>M</i> (<i>SD</i>)
SocV1	6	3.8 (1.6)
SocV2	6	3.8 (1.3)
SocV3	6	4.2 (1.3)
SocV4	6	4.2 (1.0)
SocV5	6	4.3 (.8)
SocV6	6	3.2 (2.1)
SocV7	6	4.5 (1.6)

SocV1: Automation is especially important in the context of reading; SocV2: The intervention is a good way to improve the reading fluency of students; SocV3: The intervention is an appropriate way to train reading fluency of sight words; SocV4: The intervention is a good way to overcome problems in reading; SocV5: I would use the intervention in my lessons as well; SocV6: The total time required for the intervention procedure was manageable; SocV7: Resources needed for the intervention were appropriate; 0 (=no agreement) to 5 (=absolute agreement); participants (*N*); Mean (*M*); Standard Deviation (*SD*)

Table 5 Social Validity Questionnaire Students

Items	<i>N</i>	<i>M</i> (<i>SD</i>)
SocV1	22	3.8 (.5)
SocV2	22	3.6 (.7)
SocV3	22	3.6 (.7)
SocV4	22	3.6 (.7)
SocV5	22	3.7 (.8)
SocV6	22	3.5 (1.1)
SocV7	22	3.6 (1.0)

SocV1: The racetrack helped me to read words correctly; SocV2: I think the support also helps other students with reading difficulties; SocV3: I understood the purpose of the intervention well; SocV4: I have learned a lot during the intervention; SocV5: I gladly came to the sessions; SocV6: I enjoyed the intervention; SocV7: I would participate in the intervention again; 0 (=no agreement) to 5 (=absolute agreement); participants (*N*); Mean (*M*); Standard Deviation (*SD*)

Discussion

Main findings

Struggling readers often demonstrate problems in automatic word recognition. As a result, they also have lasting difficulties with higher-order reading processes, such as text comprehension (Ravitch, 2010). The automated decoding of words is a basic prerequisite for the development of advanced reading skills (Hayes, 2016; Knoepke et al., 2014; Tunmer & Chapmann, 2012). It therefore seems important to investigate which interventions can effectively and sustainably help students overcome their challenges

with fluent word reading. The current pilot study examined the extent to which a peer-tutorial reading racetrack training supplemented by motivational components is an effective method to promote reading fluency of students with word-reading difficulties.

Our results indicated that our approach was very effective in increasing the reading fluency of the trained words by students in the experimental group. Even though the participants in the control condition also played a racetracks game (MR), implemented peer-tutoring procedures, and were motivated in exactly the same way as those who practiced reading, they did not show a comparable performance gain. Fortunately, this effect was still clearly evident ten weeks after the end of the intervention. Although data from the second follow-up showed a slight decrease in WRC, the students in the experimental group still read significantly more words than those in the control group. Thus, our results align with the findings of previous studies (e.g. Erbey et al., 2011; Green et al., 2010; Grünke, 2019; Hopewell et al., 2011; Hyde et al., 2009), and give an indication of the long-term effectiveness of the treatment.

In addition, the study provides an indication that words with a mid-frequency can also be trained effectively. The learning of mid-frequency words seems to be of importance for the development of reading competence (Masrai, 2019), as the processing of these words can have a significant slowing effect on lexical processing (Fischer-Baum et al., 2014; Kennedy et al., 2013). Thus, in addition to automated recognition of high-frequency words, acquisition of a comprehensive sight vocabulary of lower frequency words can also strengthen lexical processing (Calabrèse et al., 2016; Stolowy et al., 2019).

Furthermore, the intervention appears to be an equally effective and economical method for inclusive education, as it has been successfully applied using peer-tutoring in an inclusive classroom available to all students. In addition, responses from the teachers' and tutees' social validity questionnaires indicated support for the intervention in the three areas of understanding, acceptance, and feasibility. In future studies, it would also be interesting to look at the evaluation of the social validity of the tutors.

Limitations and further research

The results of this research must be interpreted with caution. A preliminary small-scale study like ours does not allow for far-reaching conclusions about the effectiveness of our approach. This is especially true given that it was conducted in a specific geographical area in Germany and included a constrained number of trained sight words. Thus, the results can only be seen as a first indication of the usefulness of applying reading racetracks with certain motivational techniques in a class-wide peer-tutorial setting. Further studies with a larger sample should be carried out to substantiate the finding. This would allow for the consideration of additional control variables, such as spoken language(s), socioeconomic status, working memory, and home reading environment for matching experimental and control groups, in addition to the control variables used: age, gender, LD, German L2, reading proficiency, and reading behavior. Further, it would be of interest to consider whether differential effects on the effectiveness of the method can be mapped as a function of student characteristics. With regard to the use of diagnostic instruments, it should be noted that, due to time

constraints of the school, the SLS was used in the present study, as it is possible to use this as a group test. However, the test measures reading fluency at sentence level and also assesses sentence comprehension. For future studies, it seems more appropriate to use a diagnostic instrument that measures reading fluency at the word level, such as the Salzburg Reading and Writing Test (SLRT-II; Moll & Landerl, 2010).

Moreover, it remains questionable whether the method is superior to another drill and practice approach. We compared the outcomes of our training with those of a math racetrack intervention. An interesting issue would be to examine the effectiveness of reading racetracks, relative to other methods for strengthening sight-word recognition.

Last, it is still unclear how far-reaching our results are for the reading progress of students since we did not measure transfer effects on unknown words and general reading fluency. However, Knoepke et al. (2014) showed that, according to the dual route cascaded model (Coltheart et al., 2001), the aspect of orthographical decoding skills, addressed in our study is an important prerequisite for reading comprehension and that sight word training can have positive effects on trained and untrained words as well as sentence reading fluency (e.g. McArthur et al., 2015a, 2015b). Thus, we would anticipate a transfer effect on other words, which, of course, we cannot prove on the basis of our data. Future studies should not only examine the effectiveness of reading racetrack interventions on reading fluency at the word level, but also evaluate the impact of the intervention on reading fluency at the sentence and text levels, as well as reading comprehension.

Practical implications

The research presented provides valuable suggestions to support readers who have difficulties in basic reading skills in an inclusive context. This is of particular importance, because a considerable number of students with learning problems are taught in classrooms with very diverse learners. For instance, in Germany, 48.65% of children and youth with special educational needs (i.e., including those with behavioral and language problems) in inclusive schools experience severe academic difficulties (KMK, 2018). The desire to support them in an inclusive context according to their individual needs is often hampered by a lack of resources and the absence of teaching methods that address a heterogeneous population at class level (McLeskey & Waldron, 2011; Schmidt et al., 2002). It is therefore vital to identify methods that are evidence-based, socially valid, easy to apply, and suitable for meeting the needs of different individuals (Mitchell & Sutherland, 2020). Even though many evidenced-based practices are known for struggling students, many general and special education teachers do not use them in their teaching (Maheady et al., 2013). One reason for this may be that these practices have not been adequately taught to teachers and thus they do not know how to use them (Scheeler et al., 2009). Therefore, it is of immense importance to teach the methods to the teachers in a comprehensible way and furthermore to pay special attention to very easy to use and effective interventions, like here the reading racetracks in a peer-tutorial setting. Moreover, the teachers were present during the whole procedure and also received instruction beforehand. Researchers should make practices comprehensible for teachers to

facilitate their implementation in classrooms. As mentioned before, given the scarce resources available in schools, it is important that individual support for students can also take place in class. The present study shows that this is possible, even if the students work together in pairs and speak quietly to each other while reading. As reading comprehension is the key to accessing the curriculum and to academic success, it is irremissible to make sure that students acquire the necessary prerequisite skills, especially sight word recognition. As Hayer (2016) documented, sight-word training increases the ability to decode and comprehend text. Especially for struggling readers, integrating this into the lesson plan seems to be particularly important. In addition to practicing high-frequency words, automating lower frequency words can lead to an increase in reading fluency related to words with irregular sound patterns (Calabrèse et al., 2016; Stolowy et al., 2019). These findings underline the practical importance of training high- and also mid-frequency words.

Due to the low implementation effort, reading racetracks can be integrated easily into everyday school life and thus represents an evidence-based method to support students' ability to recognize sight-words. Through the use of flashcards, it is easy to adapt the words to the students' individual needs beyond sight-word reading. Thus, the intervention can also be adapted to practice basic vocabulary among students with German as a second language (Grünke & Barwasser, 2019; Sperling et al., 2019). In follow-up studies, it would be interesting to systematically assess the optimal intervention dosage. Here, for example, the implementation of controlled single-case studies may be a useful approach, as they allow for a detailed view of the students' learning process (Horner et al., 2005).

The game-based implementation of the training and the use of motivational components at individual and class levels encourages students to engage in the typically monotonous learning of word reading over a longer period of time (Amato-Zech et al., 2006; Lämsä et al., 2018; Legge et al., 2010). In addition to fostering students' skill with word reading or vocabulary, the implementation of peer tutoring is also a way of promoting social integration (Mitchell & Sutherland, 2020). This is of particular importance, as these appear to be significantly lower for students with special educational needs than for those without (Krull et al., 2018).

Conclusion

Despite the limitations, the intervention is highly effective in improving the word-reading fluency of struggling readers. In particular, the low costs and minimal effort required for the intervention make it practical for everyday teaching in inclusive education. Further, the results illustrate the long-term effect of the intervention.

Appendix

See Table 6.

Table 6 Words of Training

Länder (<i>countries</i>)	Anerkennung (<i>acknowledgement</i>)	Katastrophe (<i>catastrophe</i>)	Identität (<i>identity</i>)	rechtzeitig (<i>in time</i>)	zweifeln (<i>to doubt</i>)
zeichnen (<i>to draw</i>)	andererseits (<i>on the other hand</i>)	garantiert (<i>guaranteed</i>)	Langeweile (<i>boredom</i>)	jenseits (<i>beyond</i>)	selbstverständlich (<i>of course</i>)
akzeptiert (<i>accepted</i>)	Ereignisse (<i>events</i>)	anstellen (<i>to queue</i>)	pünktlich (<i>punctual</i>)	motiviert (<i>motivated</i>)	Tätigkeit (<i>activity</i>)
Aufklärung (<i>reconnaissance</i>)	Gerechtigkeit (<i>justice</i>)	gezwungen (<i>forced</i>)	festgelegt (<i>set</i>)	keineswegs (<i>not at all</i>)	Überzeugung (<i>conviction</i>)
erforderlich (<i>required</i>)	Auseinandersetzung (<i>examination</i>)	zuhören (<i>to listen</i>)	keinerlei (<i>none</i>)	Schwierigkeiten (<i>difficulties</i>)	nahezu (<i>almost</i>)

Authors contributions Anne Barwasser developed the idea (conceptualization, material and methodology). Anne Barwasser and Karolina Urton conducted the study. Karolina Urton, Anne Barwasser and Matthias Grünke analyzed the data. Anne Barwasser, Karolina Urton, Matthias Grünke, Marko Sperling and David Coker did the interpretation of data. Anne Barwasser, Karolina Urton, Matthias Grünke, Marko Sperling and David Coker wrote the paper. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. None.

Availability of data and material The data set is available from the first author upon request.

Declarations

Conflict of interest The Authors declare that they have no conflict of interest.

Consent to participate An explicit consent to participate from the parents was not necessary, because by enrolling their children at our cooperation school, they agreed for them to occasionally take part in interventions delivered by graduate special education students.

Consent for publication All authors agreed with the content and all gave explicit consent to submit.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aaron, P. G., Joshi, R. M., & Quatroche, D. (2008). *Becoming a professional reading teacher*. Brookes.
- Alberto, P., & Troutman, A. C. (2008). *Applied behavior analysis for teachers* (8th ed.). Merrill/Prentice-Hall.
- Amato-Zech, N. A., Hoff, K. E., & Doepke, K. J. (2006). Increasing on-task behavior in the classroom: Extension of self-monitoring strategies. *Psychology in the Schools*, 43(2), 211–221. <https://doi.org/10.1002/pits.20137>
- Balass, M., Nelson, J. R., & Perfetti, C. A. (2010). Word learning: An ERP investigation of word experience effects on recognition and word processing. *Contemporary Educational Psychology*, 35(2), 126–140. <https://doi.org/10.1016/j.cedpsych.2010.04.001>
- Barwasser, A., Urton, K., & Grünke, M. (2021). Effects of a peer-tutorial reading racetrack on word fluency of secondary students with learning disabilities and emotional behavioral disorders. *Frontiers in Psychology*, 12, 671385. <https://doi.org/10.3389/fpsyg.2021.671385>.
- Becker, A., Woerner, W., Hasselhorn, M., Banaschewski, T., & Rothenberger, A. (2004). Validation of the parent and teacher SDQ in a clinical sample. *European Child and Adolescent Psychiatry*, 13, 11–16. <https://doi.org/10.1007/s00787-004-2003-5>
- Bond, R., & Castagnera, E. (2010). Peer supports and inclusive education: An underutilized resource. *Theory into Practice*, 45(3), 225–229. https://doi.org/10.1207/s15430421tip4503_4
- Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review*, 42(1), 39–55.

- Briesch, A. M., & Chafouleas, S. M. (2009). Review and analysis of literature on self-management interventions to promote appropriate classroom behaviors (1988–2008). *School Psychology Quarterly*, 24(2), 106–118. <https://doi.org/10.1037/a0016159>
- Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the Usage Rating Profile-Intervention. *Journal of School Psychology*, 51(1), 81–96. <https://doi.org/10.1016/j.jsp.2012.08.006>
- Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45–50.
- Calabrèsse, A., Bernard, J., Faure, G., Hoffart, L., & Castet, E. (2016). Clustering of eye fixations: A new oculomotor determinant of reading speed in maculopathy. *Investigative Ophthalmology and Visual Science*, 57, 3192–3202. <https://doi.org/10.1167/iov.16-19318>
- Cattell, R. B., & Cattell, A. K. S. (1963). *Culture fair intelligence test*. Institute for Personality and Ability Testing.
- Chard, D. J., Vaughn, S., & Tyler, B. (2002). A synthesis of research on effective intervention for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities*, 35(5), 386–406. <https://doi.org/10.1177/00222194020350050101>
- Cibulka, J. G., & Cooper, B. S. (2017). *Technology in school classrooms*. Rowman & Littlefield.
- Coltheart, M. (2005). Modeling reading: The dual-route approach. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 6–23). Blackwell. <https://doi.org/10.1002/9780470757642.ch1>
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascades model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256. <https://doi.org/10.1037/0033-295X.108.1.204>
- Copper, J. O., Heward, W. L., & Heron, T. (2008). *Applied behavior analysis* (2nd ed.). Merrill/Prentice-Hall.
- Dyslexia Action. (2017). *Avoid dyslexia discrimination in the workplace*. <https://dyslexiaaction.org.uk/contact-us/>
- De Jong, C. G., Licht, R., Sergeant, J. A., & Oosterlaan, J. (2012). RD, ADHD, and their comorbidity from a dual route perspective. *Child Neuropsychology*, 18(5), 467–486. <https://doi.org/10.1080/09297049.2011.625354>
- Donaldson, J. M., Vollmer, T. R., Krous, T., Downs, S., & Berard, K. P. (2011). An evaluation of the good behavior game in kindergarten classrooms. *Journal of Applied Behavior Analysis*, 44(3), 605–609. <https://doi.org/10.1901/jaba.2011.44-605>
- Dufrene, B. A., Henington, C., & Townsend, A. E. (2006). Peer tutoring for reading fluency: Student implementation and effects on reading fluency. *Journal of Evidence Based Practice for Schools*, 7, 118–137.
- Ehri, L. C. (2005). Development of Sight Word Reading: Phases and Findings. In M. J. Snowling & C. Hulme (Eds.), *Blackwell handbooks of developmental psychology. The science of reading: A handbook* (pp. 135–154). Blackwell Publishing. <https://doi.org/10.1002/9780470757642.ch8>
- Erbey, R., McLaughlin, T. F., Derby, K. M., & Everson, M. (2011). The effects of using flashcards with reading racetracks to teach letter sounds, sight words, and math facts to students with learning disabilities. *International Electronic Journal of Elementary Education*, 3(3), 213–226.
- Falk, M., Band, M., & McLaughlin, T. F. (2003). The effects of reading racetracks and flash cards on sight word vocabulary of three third grade students with specific learning disability: A further replication and analysis. *International Journal of Special Education*, 18(2), 57–61.
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). SAGE Publications.
- Fischer-Baum, S., Dickson, D. S., & Federmeier, K. D. (2014). Frequency and regularity effects in reading are task dependent: Evidence from ERPs. *Language, Cognition and Neuroscience*, 29(10), 1342–1355. <https://doi.org/10.1080/23273798.2014.927067>
- Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral out-comes of peer-assisted learning. *Journal of Educational Psychology*, 98(4), 732–749. <https://doi.org/10.1037/0022-0663.98.4.732>
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Green, C., McLaughlin, T. F., Derby, K. M., & Lee, K. (2010). Using reading racetracks and flashcards to teach sight words to students with disabilities: Effects for acquisition and response maintenance. *Journal of Educational Research*, 13(2), 84–98.

- Groves, E. A., & Austin, J. L. (2019). Does the good behavior game evoke negative peer pressure? Analyses in primary and secondary school classrooms. *Journal of Applied Behavior Analysis*, 52, 3–16. <https://doi.org/10.1002/jaba.513>
- Grünke, M. (2019). The effects of reading racetracks on the sight word recognition of four elementary school students with learning difficulties. *International Electronic Journal of Elementary Education*, 11(4), 291–297. <https://doi.org/10.26822/iejee.2019450787>
- Grünke, M., & Barwasser, A. (2019). Enhancing sight word fluency of second-language elementary students through reading racetracks. *International Journal of Technology and Inclusive Education*, 8(1), 1373–1378. <https://doi.org/10.20533/ijtje.2047.0533.2019.0167>
- Guerin, A., & Murphy, B. (2015). Repeat reading as a method to improve reading fluency for struggling adolescent and adult literacy. *Journal of Adolescent and Adult Literacy*, 58(7), 551–560. <https://doi.org/10.1002/jaal.395>
- Gunter, P. L., Miller, K. A., & Venn, M. L. (2003). A case study of the effects of self-graphing reading performance data for a girl identified with emotional/behavioral disorders. *Preventing School Failure*, 48(1), 28–31. <https://doi.org/10.1080/1045988X.2003.10871076>
- Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. 403–422). Lawrence Erlbaum Associates Publishers.
- Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review*, 38(4), 445–459.
- Harju-Luukkainen, H., McElvany, N., & Stang, J. (2020). *Monitoring student achievement in the 21st century European policy perspectives and assessment strategies*. Springer Nature Switzerland AG.
- Hayes, C. (2016). The Effects of Sight Word Instruction on Students' Reading Abilities. *Education Masters*, Paper 327. https://fisherpub.sjfc.edu/cgi/viewcontent.cgi?article=1327&context=education_ETD_masters
- Hirschhorn, M., & Geelan, D. (2008). Bridging the research-practice gap: Research translation and/or research transformation. *The Alberta Journal of Educational Research*, 54(1), 1–13.
- Hjetland, H. N., Brinchmann, E. I., Scherer, R., & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: A systematic review. *Campbell Systematic Reviews*, 13(1), 1–155. <https://doi.org/10.4073/csr.2017.14>
- Hopewell, K., McLaughlin, T. F., & Derby, K. M. (2011). The effects of reading racetrack with direct instruction flashcards and a token system on sight word acquisition for two primary students with severe conduct disorders. *Electronic Journal of Research in Educational Psychology*, 9(4), 693–710. <https://doi.org/10.25115/ejrep.v9i24.1462>
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71, 165–179. <https://doi.org/10.1177/001440290507100203>
- Hulac, D., & Benson, N. (2010). The use of group contingencies for preventing and managing disruptive behaviors. *Intervention in School and Clinic*, 45(4), 257–262. <https://doi.org/10.1177/1053451209353442>
- Hyde, C. A., McLaughlin, T. F., & Everson, M. (2009). The effects of reading racetracks on sight word fluency and acquisition for two elementary students with disabilities: A further replication and analysis. *The Open Social Science Journal*, 2, 50–53. <https://doi.org/10.2174/1874945300902010050>
- Johnson, E., & Semmelroth, C. L. (2014). Special education teacher evaluation: Why it matters, what makes it challenging, and how to address these challenges. *Assessment for Effective Intervention*, 39(2), 71–82.
- Joseph, L., & Eveleigh, E. (2011). A review of the effects of self-monitoring on reading performance of students with disabilities. *The Journal of Special Education*, 45(1), 43–53. <https://doi.org/10.1177/0022466909349145>
- Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, 101(4), 765–778. <https://doi.org/10.1037/a0015956>
- Kennedy, A., Pynte, J., Murray, W. S., & Paul, S. A. (2013). Frequency and predictability effects in the Dundee Corpus: An eye movement analysis. *Quarterly Journal of Experimental Psychology*, 66(3), 601–618. <https://doi.org/10.1080/17470218.2012.676054>
- Kerr, M. M., & Nelson, C. M. (2006). *Strategies for addressing behavior problems in the classroom* (5th ed.). Pearson Education.

- Kim, Y.-S., Petscher, Y., Foorman, B., & Zhou, C. (2010). The contributions of phonological awareness and letter-name knowledge to letter sound acquisition: A cross-classified multilevel model approach. *Journal of Educational Psychology*, 102(2), 313–326. <https://doi.org/10.1037/a0018449>
- Klem, A. M., & Connell, J. P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health*, 74(7), 262–273. <https://doi.org/10.1111/j.1746-1561.2004.tb08283.x>
- Knoepke, J., Richter, T., Isberner, M. B., Naumann, J., & Neeb, Y. (2014). Phonological recoding, orthographic decoding, and comprehension skills during reading acquisition. *Zeitschrift Für Erziehungswissenschaft*, 17, 447–471. <https://doi.org/10.1007/s11618-014-0560-z>
- Krull, J., Wilbert, J., & Hennemann, T. (2018). Does social exclusion by classmates lead to behaviour problems and learning difficulties or vice versa? A cross-lagged panel analysis. *European Journal of Special Needs Education*, 33(2), 235–253. <https://doi.org/10.1080/08856257.2018.1424780>
- Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, 95(1), 3–21. <https://doi.org/10.1037/0022-0663.95.1.3>
- Kultusministerkonferenz (2018). Statistische Veröffentlichungen der Kultusministerkonferenz, Dokumentation Nr. 214–Juni 2018. Sonderpädagogische Förderung in Schulen 2007 bis 2016 [Statistical Publications of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany, Documentation No. 214–June 2018. Special education support in schools 2007 to 2016]. www.kmk.org/dokumentation-statistik/statistik/schulstatistik/sonderpaedagogische-foerderung-an-schulen.html
- Lämsä, J., Hämäläinen, R., Aro, M., Koskimaa, R., & Äyrämö, S.-M. (2018). Games for enhancing basic reading and maths skills: A systematic review of educational game design in supporting learning by people with learning disabilities. *British Journal of Educational Technology*, 49(4), 596–607. <https://doi.org/10.1111/bjet.12639>
- Lee, J., & Yoon, S. (2017). The effects of repeated reading on reading fluency for students with reading disabilities. *Journal of Learning Disabilities*, 50(2), 213–224. <https://doi.org/10.1177/0022219415605194>
- Legge, D. B., DeBar, R., & Alber-Morgan, S. R. (2010). The effects of self-monitoring with a MotivAider(R) on the on-task-behaviour of fifth and sixth graders with autism and other disabilities. *Journal of Behaviour Assessment and Intervention in Children*, 1(1), 43–52. <https://doi.org/10.1037/h0100359>
- Ling, S. M., & Barnett, D. W. (2013). Increasing preschool student engagement during group learning activities using a group contingency. *Topics in Early Childhood Special Education*, 33(3), 186–196. <https://doi.org/10.1177/0271121413484595>
- Ling, S., Hawkins, R. O., & Weber, D. (2011). Effects of a classwide interdependent group contingency designed to improve the behavior of an at-risk student. *Journal of Behavioral Education*, 20, 103–116. <https://doi.org/10.1007/s10864-011-9125-x>
- Macdonald, S. J., Deacon, L., & Merchant, J. (2016). “Too far gone”: Dyslexia, homelessness, and pathways to drug use and dependency. *Insights into Learning Disabilities*, 13(2), 117–134.
- Maheady, L., Smith, C., & Jabot, M. (2013). Utilizing evidence-based practice in teacher preparation. In B. G. Cook, M. Tankersley, & T. J. Landrum (Eds.), *Evidence-based practices* (pp. 121–147). Emerald.
- Mair, P., & Wilcox, R. (2019). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, 52, 464–488. <https://doi.org/10.3758/s13428-019-01246-w>
- Masrai, A. (2019). Vocabulary and reading comprehension revisited: Evidence for high-, mid-, and low-frequency vocabulary knowledge. *SAGE Open*, 9(2), 215824401984518. <https://doi.org/10.1177/2158244019845182>
- Mayringer, H., & Wimmer, H. (2014). *SLS 2–9. Salzburger Lese-Screening für die Schulstufen 2–9* [Salzburg Reading Screening for grades 2–9]. Hogrefe.
- McArthur, G., Castles, A., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., & Banales, E. (2015a). Sight word and phonics training in children with dyslexia. *Journal of Learning Disabilities*, 48(4), 391–407. <https://doi.org/10.1177/0022219413504996>
- McArthur, G., Kohnen, S., Jones, K., Eve, P., Banales, E., Larsen, L., & Castles, A. (2015b). Replicability of sight word training and phonics training in poor readers: A randomised controlled trial. *Peer J*, 3, e922. <https://doi.org/10.7717/peerj.922>
- McLeskey, J., & Waldron, N. L. (2011). Educational programs for elementary students with learning disabilities: Can they be both effective and inclusive? *Learning Disabilities Practice*, 26(1), 48–57. <https://doi.org/10.1111/j.1540-5826.2010.00324.x>

- Mercer, C. D., Mercer, A. R., & Pullen, P. C. (2011). *Teaching students with learning problems*. Pearson.
- Mitchell, D., & Sutherland, D. (2020). *What really works in special and inclusive education: Using evidence-based teaching strategies* (3rd ed.). Routledge. <https://doi.org/10.4324/9780429401923>
- Moll, K., & Landerl, K. (2010). *SLRT II 2–9. Lese- und Rechtschreibtest*. (2. korrigiert und erweiterte Aufl.) [Reading and spelling test. (2nd corrected and expanded edition)]. Hogrefe.
- Morris, D., & Perney, J. (2018). Using a sight word measure to predict reading fluency problems in grades 1 to 3. *Reading and Writing Quarterly*, 34(4), 1–11. <https://doi.org/10.1080/10573569.2018.1446857>
- Mraz, M., Nichols, W., Caldwell, S., Beisley, R., Sargent, S., & Rupley, W. (2013). Improving oral reading fluency through readers theatre. *Reading Horizons: A Journal of Literacy and Language Arts*, 52(2). Retrieved from https://scholarworks.wmich.edu/reading_horizons/vol52/iss2/5.
- Musti-Rao, S., Lo, Y., & Plati, E. (2015). Using an iPad® App to Improve Sight Word Reading Fluency for At-Risk First Graders. *Remedial and Special Education*, 36(3), 154–166.
- National Institute of Child Health and Human Development, NIH, DHHS. (2010). *Developing early literacy: Report of the national early literacy panel (NA)*. Washington, DC: U.S. Government Printing Office.
- National Reading Panel. (2000). *Teaching children to read*. National Institute of Child Health and Human Development.
- Nese, J. F. T., Biancarosa, G., Cummings, K., Kennedy, P., Alonzo, J., & Tindal, G. (2013). In search of average growth: Describing within-year oral reading fluency growth across Grades 1–8. *Journal of School Psychology*, 51(5), 625–642. <https://doi.org/10.1016/j.jsp.2013.05.006>
- Paige, D. D. (2011). 16 minutes of “eyes-on-text” can make a difference: Whole class choral reading as an adolescent fluency strategy. *Reading Horizons*, 51(1), 1–20.
- Pappas, D. N., Skinner, C. H., & Skinner, A. L. (2010). Supplementing accelerated reading with class-wide interdependent group-oriented contingencies. *Psychology in the Schools*, 47(9), 887–901. <https://doi.org/10.1002/pits.20512>
- Rasinski, T. V. (2003). *The fluent reader: Oral reading strategies for building word recognition, fluency, and comprehension*. Scholastic.
- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. Basic Books.
- Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Not just speed reading: Accuracy of the DIBELS oral reading fluency measure for predicting high-stakes third grade reading comprehension outcomes. *Journal of School Psychology*, 46(3), 343–366. <https://doi.org/10.1016/j.jsp.2007.06.006>
- Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J., & Miller, T. R. (2003). Peer-assisted-learning interventions with elementary school students-A meta-analytic review. *Journal of Educational Psychology*, 95(2), 240–206. <https://doi.org/10.1037/0022-0663.95.2.240>
- Rothe, J., Cornell, S., Ise, E., & Schulte-Körne, G. (2015). A comparison of orthographic processing in children with and without reading and spelling disorder in a regular orthography. *Reading and Writing*, 28(9), 1307–1332. <https://doi.org/10.1007/s11145-015-9572-1>
- Sabatini, J., Wang, Z., & O'Reilly, T. (2018). Relating reading comprehension to oral reading performance in the NAEP fourth-grade special study of oral reading. *Reading Research Quarterly*, 54(2), 253–271. <https://doi.org/10.1002/rtrq.226>
- Samuels, S. J. (1979). The method of repeated reading. *The Reading Teacher*, 32(4), 403–408.
- Samuels, S. J. (2006). Reading fluency: Its past, present, and future. In T. Rasinski, C. Blachowicz, & K. Lems (Eds.), *Fluency instruction: Research-based best practices* (pp. 7–20). Guilford.
- Scheeler, M. C., Bruno, K., Grubb, E., & Seavey, T. L. (2009). Generalizing teaching techniques to K-12 classrooms: Teaching preservice teachers to use what they learn. *Journal of Behavioral Education*, 18(3), 189–210.
- Schmidt, R. J., Rozendal, M. S., & Greenman, G. G. (2002). Reading instruction in the inclusion classroom. *Remedial and Special Education*, 23(3), 130–140. <https://doi.org/10.1177/07419325020230030101>
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47, 484–503.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26–43. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>

- Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015). childLex—Eine lexikalische Datenbank zur Schriftsprache für Kinder im Deutschen [childLex-A lexical database of written language for children in German]. *Psychologische Rundschau*, 66, 155–165. <https://doi.org/10.1026/0033-3042/a000275>
- Schwanenflugel, P. J., & Kuhn, M. (2015). Reading fluency. In P. Afflerbach (Ed.), *Handbook of individual differences in reading: Reader, text, and context* (pp. 107–119). Routledge.
- Spencer, V. G. (2006). Peer tutoring and students with emotional or behavioral disorders: A review of the literature. *Behavioral Disorders*, 31(2), 204–222. <https://doi.org/10.1177/019874290603100206>
- Sperling, M., Barwasser, A., & Grünke, M. (2019). The effects of a reading racetrack intervention on the sight word fluency of learning-disabled elementary school students with German as second language. *Insights into Learning Disabilities*, 16(1), 79–90.
- Sprenger-charolles, L. L. S., & Siegel, B. (2016). Prevalence and reliability of phonological, surface, and mixed profiles in dyslexia. *Scientific Studies of Reading*, 15(6), 498–521. <https://doi.org/10.1080/10888438.2010.524463>
- Stephen, V. K., & Singh, P. P. (2017). Study of the relationship between student and teachers in terms of reinforcement in primary schools of the Allahabad. *IOSR Journal of Humanities and Social Science*, 22(12), 26–33.
- Stolowy, N., Calabrèse, A., Sauvan, L., Aguilar, C., François, T., Gala, N., Matonti, F., & Castet, E. (2019). The influence of word frequency on word reading speed when individuals with macular diseases read text. *Vision Research*, 155, 1–10. <https://doi.org/10.1016/j.visres.2018.12.002>
- Sullivan, M., Konrad, M., Joseph, L. M., & Luu, K. T. (2013). A comparison of two sight word reading fluency drill formats. *Preventing School Failure*, 57(2), 102–110.
- Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, 45(5), 453–466. <https://doi.org/10.1177/0022219411432685>
- Volpe, R. J., Burns, M. K., Dubois, M., & Zaslofsky, A. F. (2011). Computer-assisted tutoring: Teaching letter sounds to kindergarten students using incremental rehearsal. *Psychology in the Schools*, 48(4), 332–342. <https://doi.org/10.1002/pits.20557>
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2—Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest. [Basic Intelligence Test Scale 2—Revision (CFT 20-R) with vocabulary test and number sequence test.]* Hogrefe.
- Wills, H., Kamps, D., Fleming, K., & Hansen, B. (2016). Student and teacher outcomes of the class-wide function-related intervention team efficacy trial. *Exceptional Children*, 83(1), 58–76. <https://doi.org/10.1177/0014402916658658>
- Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11(2), 203–214. <https://doi.org/10.1901/jaba.1978.11-203>
- Zarić, J., Hasselhorn, M., & Nagler, T. (2020). Orthographic knowledge predicts reading and spelling skills over and above general intelligence and phonological awareness. *European Journal of Psychology of Education*, 36, 21–43. <https://doi.org/10.1007/s10212-020-00464-7>
- Zarić, J., & Nagler, T. (2021). Reading comprehension on word- and sentence-level can be predicted by orthographic knowledge for German children with poor reading proficiency. *Reading and Writing*. <https://doi.org/10.1007/s11445-021-10126-9>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.