The Effects of Peer Racetracks on the Reading Performance of German L2 Students With and Without Behavioral Difficulties

Anne Barwasser, Susanne Hisgen, and Matthias Grünke
University of Cologne, Germany

Abstract

Reading difficulties in German students are steadily increasing, while reading motivation is decreasing, yet reading is one of the most important aspects of literacy. Complicating matters, reading instruction is challenged by an enormous heterogeneity among the student body, which includes students learning German as a second language and students with behavioral problems. Thus, many teachers are faced with trying to narrow the large gap between strong and weak readers by providing reading support, which often has to start with basic skills, and at the same time meeting the individual students' needs. The present single-case study assessed the effects of a simple peer-tutored motivational reading racetrack intervention on word fluency within a short period of time for students with and without behavioral problems (N = 9) for whom German is an L2. The results are promising with high overlap indices and the followup data showing stable improvements. Limitations and implications are discussed.

Keywords: German second language, literacy intervention, peer tutoring, reading racetracks, word recognition, behavioral difficulties

The German education system is increasingly faced with the challenge of dealing with linguistically heterogeneous students due to increased migration and a growing number of bilingual students acquiring German as an L2 (GL2) (Busse, 2013). In particular, many students have difficulty accessing the German language due to the language systems of their countries of origin, including languages that are transparent or semitransparent and languages that use alphabetic or non-alphabetic written language. In transparent languages, each letter corresponds to one sound, whereas in semitransparent languages (e.g., English) each letter represents different sounds depending on the combination of letters (Buetler et al., 2014; Rodriguez et al., 2016). Alphabetic script is a phonographic script, whose characters primarily represent phonetic segments of spoken language (Gnanadesikan, 2017). The German language is a transparent alphabetic writing system (Perfetti & Dunlap, 2008).

The 2018 Program for International Student Assessment (PISA) showed that students in Germany without an immigrant background scored 63 points higher in reading literacy than students with an immigrant background (Kaya & Elster, 2018), placing Germany among the countries with the largest gap in literacy between L1 and L2 speakers among all Organisation for Economic Cooperation and Development (OECD) countries (OECD, 2010). Furthermore, the Progress in International Reading Literacy Study Survey (PIRLS) reported a significant upward trend in the number of struggling fourth-grade readers in Germany, including GL2. In addition, overall reading motivation decreased across both native and non-native younger and older students (Lafontaine et al., 2018).

Generally speaking, students whose native language is not the language of instruction face a higher risk of school problems than their L1 peers (Golinkoff et al., 2019; Kieffer et al., 2008; Kieffer & Vukovic, 2013; Lesaux & Kieffer, 2010). It is important to learn the German language appropriately in order for these students to have equal educational opportunities in German society (Han, 2012) and, therefore, it is urgent to promote the literacy skills of children learning German.

The first important aspect in second-language reading is word recognition. Without this basal competence, L2 readers have greater difficulty in rapid

reading than L1 readers, and these difficulties often lead to problems in L2 comprehension (Barwasser et al., 2021c; Barwasser et al., 2021d; Dijkstra & Van Heuven, 2002; Schroeter & Schroeder, 2018; Qu et al., 2018;). For example, Cirino et al. (2013) demonstrated the challenges of students with reading difficulties related to basic reading skills (e.g., word recognition). Thus, word recognition also plays a fundamental role in reading comprehension (Perfetti & Stafura, 2014). According to the Dual Route Cascaded Model (DRC; Coltheart et al., 2001), word recognition can occur via two different routes: a phonological route (indirect), in which each word must be recoded letter by letter based on grapheme-phoneme correspondence rules, and an orthographic route (direct), in which written words can be mapped directly onto mental representations of word forms. For the German language, Zari and Nagler (2021) found that orthographic knowledge is a better predictor of comprehension than phonological knowledge, underscoring the importance of strengthening whole-word reading.

Being able to read sight words plays a fundamental role (Balass et al., 2010; Kendeou & O'Brien, 2018) and is crucial in terms of reading fluency and reading comprehension (Barwasser et al., 2021a; Musti-Rao et al., 2015). Sight words can be defined as words that are read within one second of their appearance without phonological recoding (Ehri, 2005). McArthur et al. (2015) showed positive effects of sight word training on practiced and unpracticed words and on fluency in word and sentence reading. Automation in word reading can be particularly challenging for L2 learners, who often require more working memory capacity to read a text in the nonnative language (Fraser, 2007). When working memory becomes overloaded, learning is often impossible.

Students with behavioral difficulties, in particular, need special support when faced with such a challenge in a new written language task because they often have serious deficits in achieving reading proficiency (Wanzek et al., 2014). Furthermore, these students are less likely to engage in literacy activities (Becherer et al., 2020; Roberts et al., 2020). Further, students with lower reading skills are more likely to exhibit behavioral challenges than students without (Lin et al., 2013). In terms of L2, language difficulties have also been linked to behavior issues (Jansen et al., 2020), as poorer language skills are a risk factor for developing behavior difficulties (Chow & Wehby, 2018; Hollo et al., 2014). Petersen and LeBeau (2021) reported that social skills, language skills, and behavior difficulties are interrelated and that language ability plays a fundamental role in the development of externalizing behavior difficulties. In sum, the combination of academic and behavioral difficulties/disorders can make it difficult for professionals to provide effective support (Kauffman, 2005), including, in this context, helping struggling students in order to achieve a higher language proficiency level (Grigorenko et al., 2020; Harn et al., 2014).

How to Foster Sight Word Reading

Interventions that can be implemented economically and beneficially are particularly useful (Solis et al., 2017). To address the aforementioned challenges, there are a number of evidence-based interventions that support less proficient readers in inclusive settings (Grigorenko et al., 2020; Harn et al., 2014). Sperling et al. (2019) suggested that a mix of two or more components (e.g., peer tutoring and repeated reading) appears to be most effective in supporting heterogeneous readers equally.

Repeated Reading Through Reading Racetracks

Repeated reading (RR) is necessary to achieve greater reading fluency (Chard et al., 2009), as borne out in a number of studies (e.g., Chard et al., 2002; Ring et al., 2013; Zimmermann et al., 2021). Further, several studies have shown positive effects of repeated word reading on both word reading outcome (e.g., Martens & De Jong, 2008) and general word decoding skills – also in L2 (Shimono, 2018; Van Gorp et al., 2017).

Gamified educational components have been found effective in engaging students in literacy interventions and bringing about positive outcome in reading (Lämsä et al., 2018). One example of integrating RR into an intervention and making it game-like is Reading Racetracks (RT) – a circular board with squares on which cards can be placed with inputs whose content can be trained automatically, such as sight words. RT have been shown to be effective for lower-performing readers in second-language (e.g., Barwasser et al., 2021c; Grünke & Barwasser, 2019; Sperling et al., 2019) and first-language reading (e.g., Barwasser et al., 2021a; Barwasser et al., 2021b; Davenport et al., 2019), as well as for students with learning disorders (LD) and emotional and behavioral disabilities (EBD) (e.g., Barwasser et al., 2021b; Barwasser et al., 2021d).

Peer Tutoring as an Inclusive Tool

To turn an intervention into an inclusion tool where children with different characteristics can participate successfully, a peer-tutorial (PT) element may be added. PT is defined as a method where students work together on a specific exercise (Dufrene et al., 2010). Adding PT to interventions has beneficial effects (Mercer et al., 2011), and several studies have demonstrated its effectiveness across multiple age groups in terms of both academic gains and

social-emotional well-being (e.g., Bowman-Perrott et al., 2014; Moeyaert et al., 2021). With regard to academic achievement, PT has been found successful for reading (Dufrene et al., 2006; Hattie, 2008; Moeyaert et al., 2021). Positive effects have also been demonstrated for students learning an L2 with and without LD (Cole, 2014; Klingner et al., 2014). Moreover, a literature synthesis by Okilwa and Shelby (2010) revealed that PT had a positive impact on academic achievement in a wide range of subject areas for 6- to 12-year-olds, regardless of the type of impairment.

Self-Graphing as a Motivational Boost

Motivation plays a leading role in learning (Marinak & Gambrell, 2008). Thus, it is crucial to make an intervention as engaging as possible, especially for struggling students. Despite its game-like character, RT benefits from self-management components such as self-graphing, whereby students can track their progress and get individual feedback (Sutherland & Snyder, 2007). Beyond motivation, being able to monitor their own progress has positive impacts on students' behavior (Amato-Zech et al., 2006; Legge et al., 2010) as well as academic performance (Sutherland & Snyder, 2007) because they can compare themselves to themselves rather than their peers (Menzies et al., 2009). For example, Guzman et al.'s (2018) meta-analysis of K-12 students' self-monitoring of reading achievement showed a large positive effect on reading achievement as a result of the use of self-graphing.

Research Aim

In light of the fact that the number of less proficient L2 readers in Germany is increasing along with additional difficulties such as behavioral problems, there is an urgent need for interventions that meet the individual needs of the heterogeneous student body. The current study used a peer-tutorial racetrack intervention that is adapted to individual needs through its use of multiple components while promoting one of the most important areas of reading: reading of sight words. Specifically, we looked at the effects of the intervention on third graders with GL2 with and without behavioral problems who showed severe difficulties in literacy, including the students' and teachers' views of the intervention in the context of social validity. Our research questions were as follows:

1. Does a combined RT intervention have positive effects on the sight word reading of struggling GL2 readers with and without behavior difficulties?
2. Are the positive effects maintained after six weeks without intervention?
3. How is the intervention rated by the participants and their teachers?

Methods

Participants and Setting

The study was conducted at an inclusive elementary school in North Rhine-Westphalia, Germany, with students from two third-grade classes. Before the study started, consent forms were distributed to the legal guardians of potential participants. Screenings were conducted to assess students' level of proficiency to choose the final participants and as a basis for dividing the students into tutors and tutees.

The first instrument was the Salzburg Reading Screening Test (SLS 2-9; Mayringer & Wimmer, 2014), which was administered to all students of both classes. In addition, a German vocabulary test, the Integrated Teacher Report Form (Weiß, 2006), was administered – a German version for externalizing behavior problems (Integrated System Teacher Report Form [ITRF-G]; Volpe et al., 2018). Finally, a word pretest was used to crystallize the final words for the intervention. Student characteristics (e.g., date of birth, gender, first and second language) were collected through a teacher questionnaire.

Assessment

Salzburg Reading Screening (SLS 2-9)

The SLS 2-9 was administered as a group test to assess reading fluency. Students are presented with a list of meaningful and nonsensical sentences, with a prompt to check off the meaningful sentences (e.g., *Lemons taste sour*. a) true, b) not true). The evaluation is based on the number of correctly selected sentences, which is converted into a reading quotient (RQ). The RQ is the extent to which the measured reading ability deviates from the average of a norming sample where 100 stands for the mean value with a standard deviation of 15. Norms (reading quotients) for 11,900 students are available for the SLS 2-9. An RQ of 80-89 is considered below average, 70-79 weak, and less than or equal to 69 very weak. With regard to test validity, the test scores are closely related to speed in reading word lists aloud ($r = $ .80 to .90) (Wimmer & Mayringer, 2014). The reliability of the SLS ranges between .87-.95. Participants scoring below RQ 79 were chosen for the study.

ITRF-G

The teachers were asked to complete the German short version of the Integrated System Teacher Report Form (ITRF-G; Volpe et al., 2018) about the potential tutees. The ITRF-G is a universal screening procedure using a 4-point- scale to identify behavior problems in (a) learning-related behavior (APD) with 8 items and

Table 1
Characteristics of Student Participants

Student	Gender	Age/Grade	SN	OPP	APD	Reading RQ	Words (PR)	German L2	L1
Lava	f	8/3	EBD	10	1	69	42	Yes	Serbian
Emyl	m	8/3	/	0	2	67.5	58	Yes	Arabic
Gülcan	f	9/3	/	1	6	66	10	Yes	Kurdic
Ferhat	m	8/3	/	15	13.5	65.5	12	Yes	Turkish
Brav	m	10/3	/	8.5	10.5	65	10	Yes	Romanian
Tugce	f	8/3	/	0	1	64	18	Yes	Turkish
Gül	f	9/3	EBD	14	19.5	63.5	21	Yes	Turkish
Vaneza	f	9/3	/	0	1.5	71	15	Yes	Kurdic
Metül	m	8/3	EBD	12	5	69	42	Yes	Arabic

Note. SN = special needs; f = female; m = male; LD = learning disabilities; EBD = emotional behavioral disorder; L1 - first language; L2 = second language; OPP = disruptive behavior according to the ITRF(OPP); APD = learning-related behavior according to the ITRF(APD); RQ = reading quotient; Words PR = percentile on German Vocabulary Test.

(b) oppositional/disruptive behavior (OPP) with another 8 items. The cutoffs are as follows: 10 for APD, 5 for OPP 5, and 13 for the overall problem value. The ITRF has a high internal consistency (APD: α = .91; OD: α = .87; total: α = .91) (Volpe et al., 2018). We implemented the ITRF because problem behavior can negatively impact reading and writing as well as overall school performance (Campbell et al., 2018).

Word Pretesting

The researcher-developed word pretest was administered to identify words for the intervention that the majority of the participants were not already storing as sight words. It was composed of two PowerPoint presentations with 70 words each and was conducted on two consecutive days. Each word was displayed on a separate slide for one second followed by three slides with hashtags to slow down the speed. The word list was generated on the basis of the German childLex (Schroeder et al., 2015), which displays word frequency. The students' task was to read the words correctly aloud within 1 sec of their appearance (Ehri, 2005). A total of 30 words were selected that could not be read across the majority of participants.

German Vocabulary Test

The German version of the Culture Fair Intelligence Test and the subtest for German vocabulary (CFT-20-R; Weiß, 2006) (cutoff percentile rank <15) was administered as a group screening containing 30 items with words from semantic domains and abstract concepts based on a representative sample of $N = 2724$;

it has good reliability ($r = .87$) and a high internal consistency, .86 to .96 (Weiß, 2006). The students are given a series of words, each accompanied by a choice of other words and have to decide which of the words most closely reflect the main words. This task was chosen because vocabulary knowledge is a well-known predictor of success in L2 literacy (Schmitt et al., 2011).

Final Participants

The final participants assumed the role of tutees, each of whom was assigned a tutor, who was identified as having a high reading level on the SLS 2-9 (>120). Students were assigned to pairs by first sorting students according to their RQ. No data was collected on the tutors, except that they had to be able to read all the training words correctly. The final participants were nine tutees ($N = 9$), all were GL2 and had started learning German with the entry of kindergarten at age 3-4. The exception was Brav, who had entered school at age 7.5. According to the ITRF screening, four children showed problem behavior, and three were diagnosed with EBD. Behavioral difficulties in this context means problem behavior independent of an official EBD diagnosis. However, all tutees showed severe difficulties in reading.

We also surveyed the social validity of the intervention as viewed by the two teachers, because it is important to find out whether the method would also be used in schools by teachers. The two teachers were the class teachers of the children, with teacher 1 (female, age 35) teaching five of the children and the other teacher (female, age 39) teaching the other four.
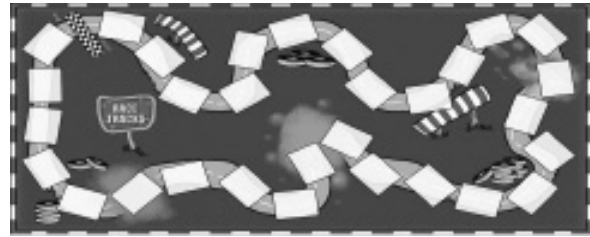
Design

A multiple-baseline design was used across individuals, with the start of the intervention occurring at different times for individuals (Morgan & Morgan, 2009). The participants were randomly assigned to three groups with different phase lengths to fulfil the requirement for a multiple-baseline design of at least three tiers (Tate et al., 2016). The intervention took place three days per week for 10 minutes over a period of seven weeks. Before the start of the intervention, baseline data was collected from the tutees. Then, in the following sessions, the RT was applied followed by data collection. Since the study was conducted during regular class time, the students were pulled out of their classes by the interventionists for the intervention period. Six master's-level students in special education served as test leaders and interventionists. In two configurations, four students were always together in pairs in one group, rotating through the groups in order to prevent an assessor effect. The remaining two students assessed the data. Group 1 had five baseline measurements, Group 2 had six, and Group 3 had seven, which were randomly assigned. Three followup measurements (Phase E) using the same measuring instrument were administered six weeks after the intervention, ended including two weeks of Christmas holidays.

Dependent Variables and Measurement

The dependent variable was the number of correctly read training words. The total was used to determine the extent of the students' word recognition regarding the words of training. The word test was a PowerPoint presentation composed of the 30 training words selected during the pretest. It was constructed the same way as the word pretest and was identical to it in terms of structure and mode of operation. The order of the individual words in the presentation was varied for each test in order to avoid a practice effect. The 30 words each appeared on a separate slide in an automatic 1-sec rhythm (Ehri, 2005), and two interventionists guided the tests. Each word that was correctly read within one second was counted as correct.

Materials

At the beginning of the intervention phase, the student pairs (tutor + tutee) were presented with the RT field and the flashcards. The self-made game field was 30x60cm and contained 30 empty rectangular cells, which were distributed at even intervals on the racetrack (see Figure 1). Each pair of students received a playing field and a game piece. For each of the 30 cells, there was a flashcard showing one of the 30 words to be trained.

Figure 1
Racetrack Gameboard

For the self-graphing procedure, a sheet was designed that included 12-14 rows (depending on the number of intervention sessions) with 30 small squares for the maximum number of words to be read correctly (see Figure 2). The rows were one below the other so that students could easily track their learning visually. Words for the intervention were of high frequency (mean: 2878) (childLex; Schroeder et al., 2015), meaning that they appear at a frequency of 100 per million words in a corpus (Brysbaert et al., 2018).

Figure 2
Self-Graphing Sheet

Note. Leserennstrecke = Reading Racetracks; Trainingsbogen = Training Sheet.

Procedures

Baseline

Before the actual intervention (Phase B) began, a baseline phase (Phase A) was conducted to determine the current status of the dependent variable and thus be able to estimate the effectiveness of the intervention. However, instead of determining the dependent variable only after each baseline session, children were engaged for 10 minutes beforehand to counter the argument that any improvement during the intervention was due only to a Hawthorne effect – change in behavior by the subjects of a study due to their awareness of being observed. Keeping students engaged and paying attention during the baseline period made the Hawthorne effect less likely. During each baseline session, pairs of students were engaged for 10 minutes in a dif-

ferent program that consisted of cognitive tasks, including logical continuation of a sequence and determining which object did not match the others in the sequence. After each baseline session, students were assessed on the dependent variable.

Intervention

Before the intervention phase began, the tutors were trained for 1.5 hours on how to give feedback in order to fulfill their role as tutors. For this purpose, they were given sample sentences and sample situations while working in pairs.

In the intervention phase, the RT game was played, and prior to its start, it was thoroughly explained to all participants. The pairs remained constant throughout the intervention phase. The RT game was played for 10 min always with the 30 training words, which were printed on individual flashcards and spread across the game board. The tutees moved along the flashcards with their game piece figure one after the other, and the respective card was revealed by the tutor. The tutee read the corresponding word aloud. If it was incorrect, the tutor corrected it after 3 sec to allow for self-correction. If a correction was made, the tutee repeated the correct form of the word aloud. The revealed cards remained face up.

If the tutee read all the cards within the 10 min, they were reshuffled and distributed face down on the field and a new game started for the remainder of the time. After each measurement from the intervention phase on, the tutees recorded the number of correctly read words on their self-graphing sheet. The number of correctly read words was reported to them by the interventionists. The tutees were then allowed to color in the number of boxes in the appropriate rows for the current session on their self-graphing sheet.

Treatment Fidelity

Treatment fidelity aims to improve the accuracy and consistency of an intervention to ensure correct delivery (Smith et al., 2007). The implementation was monitored by means of a treatment fidelity questionnaire, which was completed by the interventionists after each unit and by an external person after one third of the intervention time. To ensure faithful implementation of the training, the treatment sheet listed various components, including the environment where the support was given, the provision of materials, the process of support and feedback, and the recording of the handling of student behavior. Example questions were as follows: "Did the support take place without external disturbances?," "Were the materials ready for support?," and "Did the tutor have 3 seconds for self-correction?." Overall interrater reliability was 98% across the interventionists as well as between the interventionists and the external raters.

Social Validity

In order to increase the social validity of the study, two questionnaires were distributed to class teachers and the participants after the end of the intervention, asking them to record their acceptance of the intervention. The statement items both questionnaires are listed in Table 2. The items were rated on a 5 point-Likert scale from 0 = "totally not agree" to 4 = "totally agree." For the student survey, the interventionists were not present to avoid biased results. Further, to prevent students' low reading level from influencing the results, the teachers went over the statements individually with each student.

Table 2
Social Validity Items

Students	Racetrack helped me to be able to read words correctly.
	I think the intervention also helps other students with difficulties in reading.
	I understood well the meaning of the intervention.
	I learned a lot during the intervention.
	I was happy to come to the intervention.
	I enjoyed the intervention.
	I would participate again.
Teachers	Automation is especially important in the context of reading.
	The intervention is a good way to improve students' reading fluency.
	The intervention is an appropriate way to train reading fluency of sight words.
	I would use the intervention in my classroom.
	The total time required to complete the intervention was manageable.
	The material resources required for this intervention are adequate.

Results

For the data analysis, the statistical program R and the SCAN package for single-case analysis (Wilbert & Lueke, 2021) were used. First, visual inspection was applied followed by overlap measures and regression analysis at Level 2, once for each group and once across all groups, with a particular focus on possible Phase A trends, the increase from Phase A to Phase B (slope), and the direct increase at the onset of the intervention (level). The Non-Overlap of All Pairs (NAP; Parker et al., 2011a), the Percentage Exceeding the Median Trend (PEM; Ma, 2006), the Percentage of All Non-Overlapping Data (PAND; Parker et al., 2007), and Tau-U (Parker et al., 2011b; A vs. B + TrendB − TrendA) were selected as overlap measures.

The NAP is the percentage of all pairwise comparisons between baseline and the intervention phase and displays improvement across phases (Parker et al., 2011a). Thus, it can be used when data shows some variation in the phases. The PEM is the percentage of data points exceeding the baseline phase median, which is a good overlap index when there are floor or ceiling effects in the baseline phase. PAND refers to the total number of data points that do not overlap between phases while using all data from both phases, which makes it more concrete and reliable (Parker et al., 2007). Tau-U is a combination of Kendall's rank correlation and Mann-Whitney U while being able to control for possible baseline trends (Parker et al., 2011b). To directly test treatment effects, a regression analysis on level 2 (across all students) was performed to estimate level and slope effect from Phase A to Phase B (see Tate et al., 2016). First an analysis was performed for each group and afterwards across all groups. Finally, the SCAN package (Wilbert & Lueke, 2021) can control for autocorrelation single-case analysis.

Visual Analysis and Descriptive Statistics

The baselines could be characterized as not having much variation across participants, which can also be seen when looking at the standard deviation (*SD*) in Phase A, which ranged from 0.5 to 2.19 with mean values ranging from 0.40 to 12.00. Nevertheless, the data for Lava (*M* = 2.60, *SD* = 2.19), Tugce (*M* = 3.50, *SD* = 2.17), and Gül (*M* = 1.71, *SD* = 1.11) went slightly upward towards the end of the baseline while Brav's seemed to go a bit down. Emyl started with higher values (*M* = 12.00, *SD* = 1.00) than the other participants, and quickly reached the maximum number of words to be read in the intervention phase. Lava also reached the maximum number very rapidly, although her baseline data was significantly lower. Brav (*M* = 0.40, *SD* = 0.55), Ferhat (*M* = 1.33, *SD* = 0.82), and Gül had the lowest mean values at baseline (*M* = 1.71, *SD* =1.11). Metül (*M* = 5.57, *SD* = 0.79) and Vaneza (*M* = 5.42, *SD* = 0.79) also had flat baselines, which seemed to go down for a short time, but then stabilized again towards the end of the baseline.

Regarding Phase B, all children improved. Brav showed a rather low mean value, due to the fact that he needed a certain amount of time for the data to increase. Although he was the one with the least increase, he still reached a maximum value of 20 in Phase B. Lava was very strong in Phase B and, like Emyl, achieved the maximum possible value of 30. Tugce also displayed a vast increase and reached a value of 29.00. Gülcan (*M* Phase A = 5.00; *M* Phase B = 18.71), Metül (*M* Phase A = 5.57; *M* Phase B = 20.50), and Vaneza (*M* Phase A

= 5.43; *M* Phase B = 19.25) had a similar increase from Phase A to Phase B in mean values. Gül (*M* Phase A = 1.71; *M* Phase B = 20.09) and Ferhat (*M* Phase A = 1.33; *M* Phase B = 14.92) also had similar Phase A mean values but Gül showed a stronger increase than Ferhat.

Overall, all participants showed a clear increase without a followup drop with visible trends in Phase B and level effects from Phase A to Phase B. Thus, the followup measurements proved to be stable, showing that the students could still read the words after six weeks without intervention. Only Vaneza (23, 26, 24), Lava (29, 30, 28), and Gül (24, 26, 24) displayd a slight drop at the third followup measurement (see Figure 3 and Table 3).
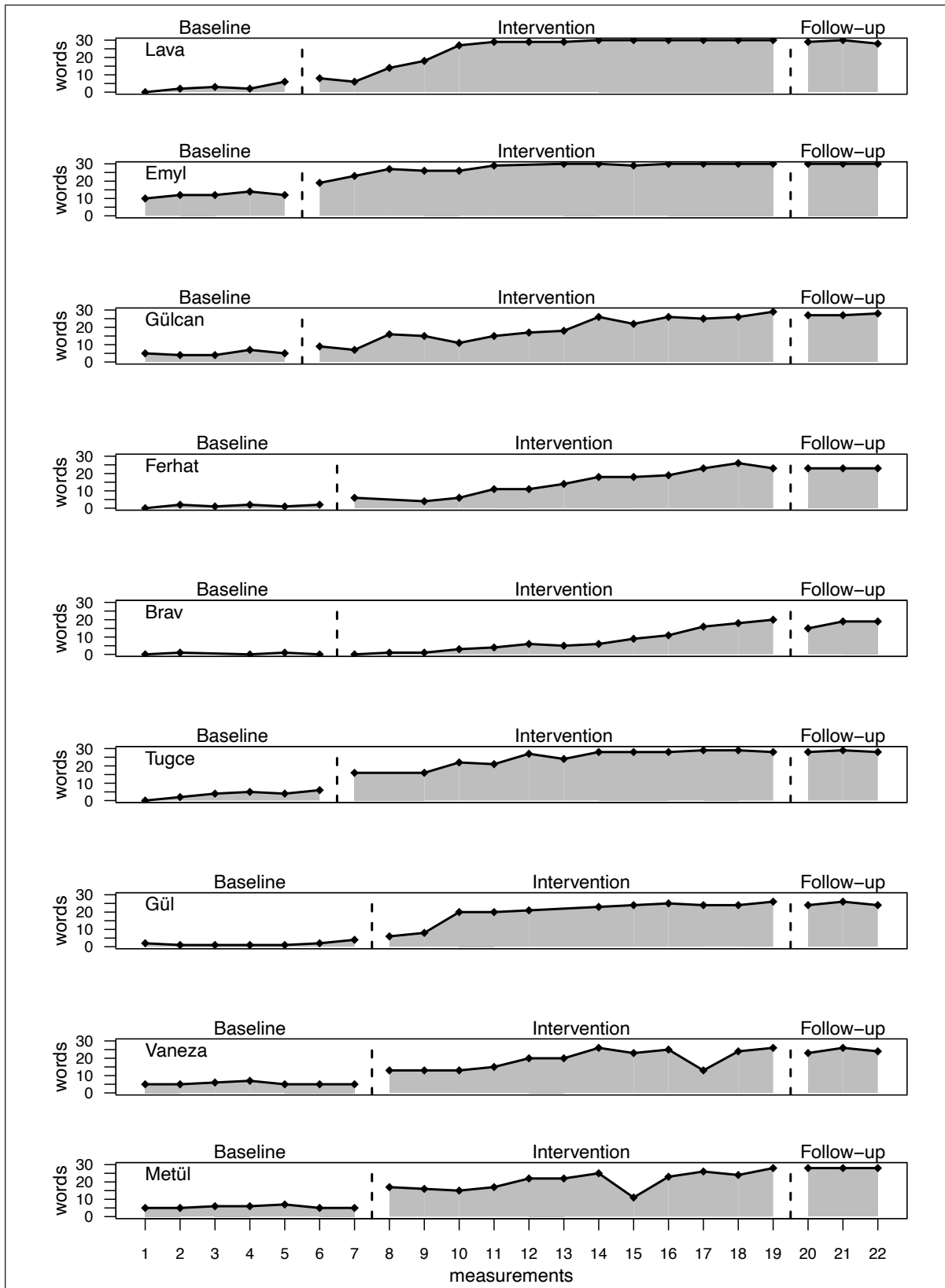
Overlap Indices

According to the NAP, all subjects achieved a strong effect (99.00-100.00; *p*<.001) except for Brav, who displayed a moderate effect (92.00; *p*<.01). The PEM showed a highly effective treatment for all. The PAND displayed strong effects (94.74-100.00) for all subjects except Brav, who had a moderate effect (77.78). In terms of the Tau-U values (considering Phase A trends), all participants presented large changes (>.60, *p*<.001) except Gülcan, who exhibited a very large change (0.81, *p*<.001).

With regard to the regression analysis, first per group, a significant level effect from Phase A to Phase B (*p*<.01) could be seen for Group 1 but no significant slope effect. Overall, there was no Phase A trend. In Group 2, there was a significant slope effect (*p*<.01) from Phase A to Phase B with an average improvement of 1.139 words per session. There was also no Phase A trend. Group 3 displayed a significant level effect from Phase A to Phase B (*p*<.001) and a significant slope effect (*p*<.01), with a beta coefficient of 1.078. There was also a significant level effect from Phase B to Phase E (*p*<.01). The regression analysis across all groups showed a significant level effect (*p*<.001) as well as slope effect (*p*<.001) from A to B and a level effect from B to E (*p*<.001). The students were able to read an average of 1.056 words per session.

Social Validity

Students

The results of the social validity survey for the students were as follows. All items were rated between "partly agree" and "strongly agree." First, all children found that the RT helped them read words correctly and that the intervention can help other children as well. Lava and Gülcan were the only ones to "partly agree" with this statement. Second, the students understood the intervention well and felt that they have learned a lot. Overall, the intervention was perceived as very pleasant. Again, Lava and Gülcan were the only

Figure 3
Words Read Correctly

Table 3
Descriptive Data and Overlap Indices for Each Participant for Words Read Correctly

	Lava	Emyl	Gülcan	Ferhat	Brav	Tugce	Gül	Vaneza	Metül
N(A)	5	5	5	6	6	6	7	7	7
N(B)	14	14	14	13	13	13	12	12	12
N(E)	3	3	3	3	3	3	3	3	3
M(A) (SD)	2.60 (2.19)	12.00 (1.00)	5.00 (1.22)	1.33 (0.82)	0.40 (0.55)	3.50 (2.17)	1.71 (1.11)	5.43 (0.79)	5.57 (0.79)
M(B) (SD)	24.29 (8.84)	27.62 (3.40)	18.71 (7.04)	14.92 (7.38)	7.69 (6.69)	24.67 (4.85)	20.09 (6.77)	19.25 (5.53)	20.50 (5.18)
Max(B)	30.00	30.00	29.00	26.00	20.00	29.00	26.00	26.00	28.00
M(E) (SD)	29.00 (1.00)	30.00 (0.00)	27.33 (0.58)	23.00 (0.00)	17.67 (2.31)	28.33 (0.58)	24.67 (1.15)	24.33 (1.53)	28.00 (0.00)
Max(E)	30.00	30.00	28.00	23.00	19.00	29.00	26.00	26.00	28.00
NAP (p)	99.00 (<.001)	100.00 (<.001)	99.00 (<.001)	100.00 (<.001)	92.00 (<.001)	100.00 (<.001)	100.00 (<.001)	100.00 (<.001)	100.00 (<.001)
PEM	100.00	100.00	100.00	100.00	92.31	100.00	100.00	100.00	100.00
PAND	94.74	100.00	94.74	100.00	77.78	100.00	100.00	100.00	100.00
Tau-U (p)	0.78 (<.001)	0.68 (<.001)	0.81 (<.001)	0.73 (<.001)	0.75 (<.001)	0.64 (<.001)	0.69 (<.001)	0.72 (<.001)	0.70 (<.001)

Note. N = measurements; *M* = mean; *SD* = standard deviation; Max = maximum value; A = Phase A; B = Phase B; E = follow-up; *p* = *p* value; PEM = Percentage Exceeding the Median Trend; PAND = Percentage of All Non-Overlapping Data; NAP = Non-Overlap of All Pairs.

ones to state "partly agree." All would like to participate again; Lava was unsure, however.

Teachers

The teachers also rated the intervention as positive. All were of the opinion that the intervention is a good way to train reading fluency and sight words. In addition, teachers responded "agree" and "totally agree" to the statement that their students had become better at reading overall. All teachers responded that they would use the intervention in their own classes. The materials seemed appropriate as did the timing across the board. Only one teacher stated "partly agree" when asked about time management. However, this did not refer to the intervention itself, but to the individual measurements after each session.

Discussion

Main Findings

The aim of the present study was to evaluate the effect of a peer-tutorial racetrack game on automating sight words in primary school students with and without behavioral difficulties who have reading difficulties

in their GL2. The intervention was specifically designed to meet individual student needs within a heterogeneous study body such as is found in an inclusion classroom (meaning that different students can learn together with one method). The results show a clear overall increase for each child with flat baselines in a very short time and stable data at followup.

Looking at the results in more detail, the fact that there is no significant slope effect for Group 1 is partly because Emyl already had higher values in the baseline and that he, as well as Lava, rapidly reached the maximum possible number of words. Otherwise, all showed an increase and also stable followup data in Group 1. Gülcan improved significantly more slowly than those of other group members. In the German vocabulary screening, she only had a percentile of 10, which is low, and might have had an impact on her reading performance (e.g., Schroeter & Schroeder, 2017). Lava indicated "partly agree" on the social validity survey in response to whether she enjoyed the intervention and would participate. Since she reached the maximum number of words quickly, boredom might explain her response. The ITRF screening showed that she has severe problems with disruptive behavior and had been diagnosed with EBD. However, problem behavior did not seem to have an impact on her results.

Group 2 displayed a steady increase. Tugce even reached the upper range quite quickly and reacted positively to the onset of the intervention even though she was one of the weakest readers. Compared to the others, Brav needed more time until he could correctly read several words. He had a percentile in German vocabulary of 10 and was the only students with Romanian as L1. His low vocabulary knowledge and language background might have played a role for the slower increase (Lesaux & Kieffer; 2010; Schroeter & Schroeder, 2017). Brav also exhibited problems in OPP and APD, which relates to concentration and, thus, could have been a factor. All three students rated the intervention positively. Ferhat also had problems in the areas of OPP and APD. However, this did not seem to affect the success of the intervention much for him. The followup data was stable overall; Brav was the only one who dropped, but only minimally. However, since he was the student who seemed to have the most difficulty in memorizing the words, it was possible that he needed more automation.

Across the board, Group 3 also showed significant improvement. Vaneza and Metül responded positively to the intervention while Gül made a huge jump at the third intervention session. Followup data was stable for each child. Gül was the weakest reader according to the prior reading screening, had an EBD diagnosis, and problems in OPP as well as APD. Perhaps she had to settle in before she could focus on the intervention and/or perhaps memorizing sight words was still difficult for her at first because she was trying to read via phoneme-grapheme correspondence, which is typically for German readers (Landerl & Wimmer, 2008).

Overall, the data does not show that the students with behavioral difficulties had more problems than their peers benefiting from the intervention. Looking specifically at the three students with an official EBD diagnosis, it is not clear that EBD had a negative impact on the intervention, as all three benefited from the intervention in a very short period of time, even though they started with very low baseline scores. Interesting, these findings differ from other researchers arguing that students with behavioral difficulties have greater problems benefiting from literacy interventions and that teachers have difficulty finding appropriate reading interventions for them (Becherer et al., 2020; Forlin & Chambers, 2011; Nelson et al., 2003; Roberts et al., 2020). Motivation plays a major role in behavior, so perhaps the playful and motivating nature of the racetracks helped ensure that students with behavioral difficulties benefit to the same extent as those without behavioral difficulties. Likewise, we cannot fully determine whether students' different L1s had an impact, because the progress of the students cannot be differentiated by L1. In terms of gender, Brav and Ferhat were the weakest overall, but this does not apply to Emyl and Metül. Therefore, it cannot be said with certainty that there was a difference between the subjects in terms of gender.

Based on the fairly clear results, we can conclude that PT RT, which was only conducted for 10-minute sessions, after a very short time, had a positive effect on the learning of sight words of less proficient GL2 primary school readers with Kurdish, Turkish, Arabic, Serbian, and Romanian language backgrounds, with and without behavioral difficulties. Followup data also showed that the children were able to recall the words after six weeks. In addition, there was no difference between children with and without behavioral problems, and no difference by age.

The finding that problem behavior did not seem to play a major role is consistent with Kempe et al. (2011), who argued that reading problems and externalizing behavior should be seen as independent of each other. The findings are also consistent with findings from PT studies (e.g., Moeyaert et al., 2021; Okilwa & Shelby; 2010) and previous studies using reading racetracks (Barwasser et al., 2021a; Barwasser et al., 2021b; Grünke & Barwasser, 2019; Sperling et al., 2019). Further, the results also support studies with a focus on self-graphing (e.g., Legge et al., 2010; Sutherland & Snyder, 2007).

Limitations

Despite the promising results, several limitations of the present study warrant mentioning. As with all individual case studies, the sample size was small, making it difficult to draw generalized conclusions. However, the advantage of such studies lies in the opportunity they offer to directly see individual responses to an intervention and, therefore, be better able to assess and adapt it to specific characteristics. In addition, they allow researchers to see through the learning process how long it takes for an intervention to elicit responses in the students. Furthermore, it is unclear what influence the children's languages of origin played. According to Zigler and Goswami (2005), learning to read is a process of understanding orthography-phonology mapping, which may explain differences in reading in different languages. Overall, While there was no clear difference in this respect between the children overall, the influence of the various L1s on GL2 is unknown. Further, we did not measure reading in children's L1 – thus, they might have problems in L2 reading but not in L1 – maybe as a result of undergoing a language learning process.

Additionally, at this point we can only draw conclusions about children with the language backgrounds

mentioned above and, of course, only about German as L2. Thus, our findings cannot automatically be transferred to other languages and sight word reading, even if a transfer could be expected due to prior studies focusing on sight word training in, for example, English (McArthur et al., 2015).

For Emyl, the already quite high values in the baseline were of interest, with a ceiling effect expected to follow. Nevertheless, his direct response to the intervention and the initially strong increase show that the intervention seemed to be effective for him as well. Plus, we cannot be sure which effects the RT with PT had vs. the self-graphing, since self-monitoring is also considered to promote reading (Menzies et al., 2009). However, since most instruction in Germany classrooms uses reinforcers and research has shown the importance of motivation, we decided to also reinforce while combining the RT with self-graphing.

As a last limitation, it should be mentioned that the very positive effects may also be due to the small group size and, therefore, the more intensive teaching. And yet, in a previous study, Barwasser et al. (2021b) also found strong effects in a classroom setting of a racetrack intervention. However, the latter study was conducted with children without special needs support.

Implications

Based on the results and limitations, our study has several implications for practice and research. As a next step, one could look at other language backgrounds as well as other languages in which sight vocabulary is promoted in PT. In addition, since an L1 influences the L2 (Shum et al., 2016), it would be interesting to measure L1 reading beforehand to see connections. Reading motivation could also be surveyed before and after using questionnaires to see correlations since reading and motivation are strongly related (Marinak & Gambrell, 2008). Further, many children fail in reading because they have not yet fully internalized the L2 alphabetic principle. On the basis of a previous screening, more information could be obtained in order to assess whether a given intervention might be too difficult.

Regarding the intervention itself, one could choose a larger pool of training words simply to minimize a training effect. To avoid a training effect in the current study, the words in the measurements were shown in randomized order each time, but this does not completely rule out a training effect. Moreover, use of a larger pool of training words would minimize the like-lihood of ceiling effects, besides offering the children an opportunity to store a larger sight vocabulary. However, since not all of the children in the study were able to memorize all of the words at the end, the time of the intervention and possibly the number of training weeks would have to be increased in order to create automation. Especially children with weaker skills need a high degree of automation and to have words presented to them with high frequency (Chard et al., 2009).

In future studies, we would also like to focus on the tutors, both in terms of how they liked the intervention and to determine if their skills improved. For this purpose, we could use another questionnaire and general reading screenings before and after the intervention – and also with the tutees – in order to detect influences of the intervention on their general reading fluency. Plus, one could embed the training words in texts and, by measuring before and after, determine if the text can be read faster as a result. In addition, one could separate the RT from the self-graphing within a single-case design with A B BC phases or an experimental group design with the control group receiving only the Racetracks. Moreover, one could contrast one intervention with another and generate a larger sample to make more generalized statements about the effectiveness of PT RT in GL2. Referring back to the limitation of the small group size, a further study could be implemented in a classroom with students with special needs to see if these students need to be in a small group to improve or if a classroom setting would lead to the same positive effects.

Conclusion

Overall, this study showed that it is possible, in a short time and in a simple way, to improve the sight vocabulary of poorly reading primary students in the second language, some of whom also show behavioral difficulties. Training sight vocabulary is essential for reading (Musti-Rao et al., 2015) and sets the trajectory for reading fluency and comprehension. Our intervention adds another piece to the puzzle of reading research and through peer tutoring makes Reading Racetracks a tool that can be used in an inclusive way, allowing all children to participate. Use of the intervention can contribute to narrowing the large gap between students with and without German L2 in reading and give everyone the same educational opportunities while enhancing their language skills.

References

Barwasser, A., Urton, K., & Grünke, M. (2021a). Effects of peer-tutorial reading racetracks on sight word reading of secondary school students with learning disabilities and emotional behavioral disorder. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2021.671385

Barwasser, A., Urton, K., Grünke, M., Sperling, M., & Coker, D. L. (2021b). Fostering word fluency of struggling third graders from Germany through motivational peer-tutorial reading racetracks. *Reading & Writing, 35*, 29–53.

Barwasser, A., Schmitz, S., & Grünke, M. (2021c). A sub-lexical patterns training with Racetracks on trained and untrained words in low-literacy German students with behavior problems with and without learning disabilities. *Learning Disabilities: A Contemporary Journal.*

Barwasser, A., Nobel, K., & Grünke, M. (2021d). Peer-reading racetracks for word reading of low-achieving graduating students with learning disabilities and behavioural problems. *British Journal of Special Education.*

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

Becherer, J., Koeller, O., & Zimmermann, F. (2020). Externalizing behaviour, task-focused behaviour, and academic achievement: An indirect relation? *British Journal of Educational Psychology, 91*(1), 27–45.

Bowman-Perrott, L., Burke, M. D., Zhang, N., & Zaini, S. (2014). Direct and collateral effects of peer tutoring on social and behavioral outcomes: A meta-analysis of single-case research. *School Psychology Review, 43*(3), 260–285. https://doi.org/10.1080/02796015.2014.12087427

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. https://doi.org/10.1177/0963721417727521

Buetler, K. A., de León Rodríguez, D., Laganaro, M., Müri, R., Spierer, L., & Annoni, J. M. (2014). Language context modulates reading route: an electrical neuroimaging study. *Frontiers in Human Neuroscience, 8*, 83.

Busse, V. (2013). An exploration of motivation and self-beliefs of first year students of German. *System, 41*(2), 379–398. https://doi.org/10.1016/j.system.2013.03.007

Campbell, A. R., Bowman-Perrott, L., Burke, M. D., & Sallese, M. R. (2018). Reading, writing, math, and content-area interventions for improving behavioral and academic outcomes of students with emotional and behavioral disorders. *Learning Disabilities: A Contemporary Journal, 16*, 119–138.

Chard, D., Vaughn, S., & Tyler, B. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35*(5), 386–406. https://doi.org/10.1177/00222194020350050101

Chard, D. J., Ketterlin-Geller, L. R., Baker, S. K., Doabler, C., & Apichatabutra, C. (2009). Repeated reading interventions for students with learning disabilities: Status of the evidence. *Exceptional Children, 75*(3), 263–281. https://doi.org/10.1177/001440290907500301

Chow, J. C., & Wehby, J. H. (2018). Associations between language and problem behavior: A systematic review and correlational meta-analysis. *Educational Psychology Review, 30*(1), 61–82. https://doi.org/10.1007/s10648-016-9385-z

Cirino, P. T., Romain, M. A., Barth, A. E., Tolar, T. D., Fletcher, J. M., & Vaughn, S. (2013). Reading skill components and impairments in middle school struggling readers. *Reading & Writing, 26*(7), 1059–1086. https://doi.org/10.1007/s11145-012-9406-3

Cole, M. W. (2014). Speaking to read: Meta-analysis of peer-mediated learning for English language learners. *Journal of Literacy Research, 46*(3), 358–382.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Davenport, C. A., Konrad, M., & Alber-Morgan, S. R. (2019). Effects of reading racetracks on sight word acquisition for deaf kindergarteners. *The Journal of Deaf Studies and Deaf Education, 24*(2), 173–185. doi:10.1093/deafed/eny038

Dijkstra, T., & Van Heuven, W. J. (2002). Modeling bilingual word recognition: Past, present and future. *Bilingualism: Language and Cognition, 5*(3), 219–224.

Dufrene, B. A., Reisener, C. D., Olmi, D. J., Zoder-Martell, K., McNutt, M. R., & Horn, D. R. (2010). Peer tutoring for reading fluency as a feasible and effective alternative in response to intervention systems. *Journal of Behavioral Education, 19*(3), 239–256. https://doi.org/10.1007/s10864-010-9111-8

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

Forlin, C., & Chambers, D. (2011). Teacher preparation for inclusive education: Increasing knowledge but raising concerns. *Asia-Pacific Journal of Teacher Education, 39*(1), 1732.

Fraser, C. A. (2007). Reading rate in L1 Mandarin and L2 English across five reading tasks. *The Modern Language Journal, 91*(3), 372–394. https://doi.org/10.1111/j.1540-4781.2007.00587.x

Gnanadesikan, A. E. (2017). Towards a typology of phonemic scripts. *Writing Systems Research, 9*(1), 14-35.

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development, 90*(3), 985–992. DOI: 10.1111/cdev.13128

Grigorenko, E. L., Compton, D. L., Fuchs, L. S., Wagner, R. K., Willcutt, E. G., & Fletcher, J. M. (2020). Understanding, educating, and supporting children with specific learning disabilities: 50 years of science and practice. *American Psychologist, 75*(1), 37. https://doi.org/10.1037/amp0000452

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.

Guzman, G., Goldberg, T. S., & Swanson, H. L. (2018). A meta-analysis of self-monitoring on reading performance of K–12 students. *School Psychology Quarterly, 33*(1), 160.

Han, W. J. (2012). Bilingualism and academic achievement. *Child Development, 83*(1), 300–321. https://doi.org/10.1111/j.1467-8624.2011.01686.x

Harn, B., Fritz, R., & Berg, T. (2014). Effective literacy instruction in inclusive schools. In J. McLeskey, N. L. Waldron, F. Spooner, & B. Algozzine (Eds.), *Handbook of effective inclusive schools: Research and practice* (pp. 229–246). Routledge. https://doi.org/10.4324/9780203102930

Hollo, A., Wehby, J. H., & Oliver, R. M. (2014). Unidentified language deficits in children with emotional and behavioral disorders: A meta-analysis. *Exceptional Children, 80*(2), 169–186. https://doi.org/10.1177/001440291408000203

Jansen, R., Maljaars, J., Verhappen, A., Zink, I., Steyaert, J., & Noens, I. (2020). Problem behavior in young children referred with language difficulties: Relations to language and intentional communication. *Autism &*

Developmental Language Impairments, 5. https://doi.org/10.1177/2396941519900076

Kauffman, J. M. (2005). *Characteristics of children's behavior disorders* (7th ed.). Merrill.

Kaya, V. H., & Elster, D. (2018). German students' environmental literacy in science education based on PISA data. *Science Education International, 29*(2), 75–87. https://doi.org/10.33828/sei.v29.i2.2

Kempe, C., Gustafson, S., & Samuelsson, S. (2011). A longitudinal study of early reading difficulties and subsequent problem behaviors. *Scandinavian Journal of Psychology, 52*(3), 242–250. https://doi.org/10.1111/j.1467-9450.2011.00870.x

Kendeou, P., & O'Brien, E. J. (2018). *Reading comprehension theories: A view from the top down.* Routledge. doi:10.4324/9781315687384

Kieffer, M. J., Lesaux, N. K., & Snow, C. E. (2008). Promises and pitfalls: Implications of No Child Left Behind for identifying, assessing, and educating English language learners. In G. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, and school reform* (pp. 57–74). Corwin Press.

Kieffer, M. J., & Vukovic, R. K. (2013). Growth in reading-related skills of language minority learners and their classmates: More evidence for early identification and intervention. *Reading and Writing, 26*(7), 1159–1194. https://doi.org/10.1007/s11145-012-9410-7

Klingner, J. K., Boele, A., Linan-Thompson, S., & Rodriguez, D. (2014). Essential components of special education for English language learners with learning disabilities: Position statement of the Division for Learning Disabilities of the Council for Exceptional Children. *Learning Disabilities Research & Practice, 29*(3), 93–96.

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

Lafontaine, D., Dupont, V., & Schillings, P. (2018). Does reading literacy instruction vary according to language or culture? Similarities and differences in English-, German- and French-speaking education systems in PIRLS 2016. *IEA Compass: Briefs in Education, 4*, 1–11.

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

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology, 100*(1), 150–161. https://doi.org/10.1037/0022-0663.100.1.150

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

Lesaux, N. K., & Kieffer, M. J. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal, 47*(3), 596–632. https://doi.org/10.3102/0002831209355469

Lin, Y. C., Morgan, P. L., Hillemeier, M., Cook, M., Maczuga, S., & Farkas, G. (2013). Reading, mathematics, and behavioral difficulties interrelate: Evidence from a cross-lagged panel design and population-based sample of US upper elementary students. *Behavioral Disorders, 38*(4), 212–227. https://doi.org/10.1177/019874291303800404

Ma, H. H. (2006). An alternative method for quantitative synthesis of single subject researches: Percentage of data points exceeding the median. *Behavior Modification, 30*(5), 598–617. https://doi.org/10.1177/0145445504272974

Mahdavi, J. N., & Tensfeldt, L. (2013). Untangling reading comprehension strategy instruction: Assisting struggling readers in the primary grades. *Preventing School Failure: Alternative Education for Children and Youth, 57*(2), 77–92. https://doi.org/10.1080/1045988X.2012.668576

Marinak, B. A., & Gambrell, L. B. (2008). Intrinsic motivation and rewards: What sustains young children's engagement with text? *Literacy Research and Instruction, 47*(1), 9–26. https://doi.org/10.1080/19388070701749546

Martens, V.E.G., & de Jong, P. F. (2008). Effects of repeated reading on the length effect in word and pseudoword reading. *Journal of Research in Reading, 31*(1), 40–54. https://doi.org/10.1111/j.1467-9817.2007.00360.x

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. (2015). Sight word and phonics training in children with dyslexia. *Journal of Learning Disabilities, 48*(4), 391–407. https://doi.org/10.1177/0022219413504996

Menzies, H., Lane, K., & Lee, J. (2009). Self-monitoring strategies for use in the classroom: A promising practice to support productive behavior for students with emotional or behavioral disorders. *Beyond Behavior, 18*(2), 27–35.

Mercer, C. D., Mercer, A. R., & Pullen, P. C. (2011). *Teaching students with learning problems*. Pearson.

Moeyaert, M., Klingbeil, D. A., Rodabaugh, E., & Turan, M. (2021). Three-level meta-analysis of single-case data regarding the effects of peer tutoring on academic and social-behavioral outcomes for at-risk students and students with disabilities. *Remedial and Special Education, 42*(2), 94–106. https://doi.org/10.1177/0741932519855079

Morgan, D. L., & Morgan, R. K. (2009). *Single-case research methods for the behavioral and health sciences*. SAGE Publications. https://doi.org/10.4135/9781483329697

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. https://doi.org/10.1177/0741932514541485

Nelson, J. R., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy Interventions: A meta analytic review. *Learning Disabilities Research & Practice, 18*, 255–267.

Okilwa, N. S., & Shelby, L. (2010). The effects of peer tutoring on academic performance of students with disabilities in grades 6 through 12: A synthesis of the literature. *Remedial and Special Education, 31*(6), 450–463. https://doi.org/10.1177/0741932509355991

Organisation for Economic Cooperation and Development (OECD). (2010). *PISA 2009. Assessment Framework: Key competencies in reading, mathematics and science*. https://doi.org/10.1787/9789264062658-en

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percent of all nonoverlapping data (PAND): An alternative to PND. *The Journal of Special Education, 40*(4), 194–204. https://doi.org/10.1177/00224669070400040101

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011a). Effect size in single case research: A review of nine non-overlap techniques. *Behavior Modification, 35*(4), 302–322. https://doi.org/10.1177/0145445511399147

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011b). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy, 42*(2), 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Perfetti, C. A., & Dunlap, S. (2008). Learning to read: General principles and writing system variations. In *Learning to read across languages* (pp. 25–50). Routledge.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Petersen, I. T., & LeBeau, B. (2021). Language ability in the development of externalizing behavior problems in childhood. *Journal of Educational Psychology, 113*(1), 68–85. https://doi.org/10.1037/edu0000461

Qu, Q., Cui, Z., & Damian, M. F. (2018). Orthographic effects in second-language spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(8), 1325.

Ring, J. J., Barefoot, L. C., Avrit, K. J., Brown, S. A., & Black, J. L. (2013). Reading fluency instruction for students at risk for reading failure. *Remedial and Special Education, 34*(2), 102–112. https://doi.org/10.1177/0741932511435175

Roberts, G. J., Cho, E., Garwood, J. D., Goble, G. H., Robertson, T., & Hodges, A. (2020). Reading interventions for students with reading and behavioral difficulties: A meta-analysis and evaluation of co-occurring difficulties. *Educational Psychology Review, 32*(1), 17–47. https://doi.org/10.1007/s10648-019-09485-1

Rodriguez, D. D. L., Buetler, K. A., Eggenberger, N., Preisig, B. C., Schumacher, R., Laganaro, M., ... & Müri, R. M. (2016). The modulation of reading strategies by language opacity in early bilinguals: an eye movement study. *Bilingualism: Language and Cognition, 19*(3), 567–577.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal, 95*, 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*(3), 155–165. https://doi.org/10.1026/0033-3042/a000275

Schroeter, P., & Schroeder, S. (2017). The impact of L2 German on component processes of reading. *Journal of Research in Reading, 40*(1), 107–124. https://doi.org/10.1111/1467-9817.12078

Schroeter, P., & Schroeder, S. (2018). Differences in visual word recognition between L1 and L2 speakers: The impact of length, frequency, and orthographic neighborhood size in German children. *Studies in Second Language Acquisition, 40*(2), 319–339.

Shimono, T. R. (2018). L2 reading fluency progression using timed reading and repeated oral reading. *Reading in a Foreign Language, 30*(1), 152–179.

Shum, K.-K., Ho, C., Siegel, L., & Kit-fong Au, T. (2016). First-language longitudinal predictors of second-language literacy in young L2 learners. *Reading Research Quarterly, 51*(3), 323–344. https://doi.org/10.1002/rrq.139

Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education and Treatment of Children, 30*(4), 121–134. https://doi.org/10.1353/etc.2007.0033

Solis, M., Scammacca, N., Barth, A. E., & Roberts, G. J. (2017). Text-based vocabulary intervention training study: Supporting fourth graders with low reading comprehension and learning disabilities. *Learning Disabilities: A Contemporary Journal, 15*(1), 103–115.

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.

Sutherland, K. S., & Snyder, A. (2007). Effects of reciprocal peer tutoring and self-graphing on reading fluency and classroom behavior of middle school students with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders, 15*(2), 103–118.

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., et al. (2016). The single-case reporting guideline in behavioural interventions (SCRIBE) 2016 statement. *Journal of School Psychology 56*, 133–142. https://doi: 10.2522/ptj.2016.96.7.e1

Van Gorp, K., Segers, E., & Verhoeven, L. (2017). The role of feedback and differences between good and poor decoders in a repeated word reading paradigm in first grade. *Annals of Dyslexia, 67*, 1–25. https://doi.org/10.1007/s11881-016-0129-z

Volpe, R. J., Casale, G., Mohiyeddini, C., Grosche, M., Hennemann, T., Briesch, A. M., & Daniels, B. (2018). A universal behavioral screener linked to personalized classroom interventions: Psychometric characteristics in a large sample of German schoolchildren. *Journal of School Psychology, 66*, 25–40. https://doi.org/10.1016/j.jsp.2017.11.003

Wanzek, J., Al Otaiba, S., & Petscher, Y. (2014). Oral reading fluency development for children with emotional disturbance or learning disabilities. *Exceptional Children, 80*(2), 187–204. https://doi.org/10.1177/001440291408000204

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.

Wilbert, J., & Lueke, T. (2021). *Analyzing single-case data with R and SCAN.* Available at https://jazznbass.github.io/scan-Book/

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, 34*(8), 2031–2057.

Zigler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Zimmermann, L. M., Reed, D. K., & Aloe, A. M. (2021). A meta-analysis of non-repetitive reading fluency interventions for students with reading difficulties. *Remedial and Special Education, 42*(2), 78–93.