Enhancing Literacy in First- and Second Language Learners with Behavioral and Learning Difficulties

Inaugural dissertation

for the attainment of the doctorate degree

of the Human Sciences Faculty University of Cologne in accordance with the doctorate regulations of

18.12.2018

submitted by

Anne Barwasser

from Monschau

July, 2021

Supervised by Ph.D., Professor Matthias Gruenke

Assessment

Primary Assessor: Matthias Gruenke, Ph.D., Professor, University of Cologne, Germany

Second Assessor: Kristie Asaro-Saddler, Ph.D., Professor, University at Albany, New York, US

Third Assessor: Conny Melzer, Ph.D., Professor, University of Leipzig, Germany

This dissertation was accepted by the Faculty of Human Sciences of the University of Cologne in July 2021

Date of the dissertation defense: 30.11.2021

Acknowledgement

Seeing a job as a passion rather than a duty is a wonderful gift in life. Many people supported and believed in me on the way to my promotion. I would like to thank all the people who have helped me walk this path with joy and confidence.

First, I would like to thank my supervisor, Prof. Dr. Matthias Gruenke, who believed in me from the very beginning, encouraged me in all matters, and gave me the opportunity to freely transform my enthusiasm into research. Without you, I would not be where I am right now. I would also like to thank Prof. Dr. Kristie Asaro-Saddler and Prof. Dr. Conny Melzer. Thank you for taking the time and effort in offering advice and reviewing and evaluating my research.

Further, thanks to my team for their support, good conversations, and assurance that we can rely on each other. Above all, I thank Dr. Susanne Hisgen for all her advice and emotional support in writing my dissertation. In addition, I would like to thank my colleague Prof. Dr. Karolina Urton, who gave me advice at the beginning of my journey. I've learned a lot from you, and I'm really grateful. Thank you for being such a good role model. I would also like to thank my colleague Prof. Dr. Mary Beth Calhoon, who motivated me throughout, showed me how to write good papers, and was always available with advice and support. The same goes for Prof. Dr. David Coker Jr., through whom I also learned to write good papers. Thank you both for believing in me and boosting my confidence.

I would also like to thank my husband, who always had my back and encouraged me at all times. Kamil, I know that it was not always easy, and I was often stressed out, but you were super patient and understanding. I cannot imagine a better partner in life. I would also like to mention my parents, my brother Simon, and my grandparents. You accompanied me every step of the way, and I always knew that you were there for me. Thank you so much for the tremendous support and encouragement. Without your patience, advice, and the knowledge that you are always behind me, I would not be the person I am now. Grandpa and grandma, you are not alive to see me make these steps, but I know how proud you would be. Finally, I thank my good friends Maureen Weaver and Cameron and Sophie Booth for their (English) advice and support.

Abstract

This dissertation is concerned with the improvement of literacy skills in primary and secondary school students with learning and behavioral difficulties—with German as a first language on the one hand and German and English as second languages on the other—displayed in two parts. This focus has been set against the background of an increasing number of students who perform with less literacy proficiency in first and second languages (L1 and L2) and the growing heterogeneity of students who face learning and behavioral difficulties that pose major challenges for teachers. Two specially designed and combined interventions, peer-tutorial Reading Racetracks (RT) and storytelling, will be presented that are characterized by great effectiveness in literacy and good usability for heterogeneous groups of students. For both parts, these interventions were evaluated in various constellations regarding their effectiveness for literacy. Group and single-case designs were used for this purpose and conducted in elementary, secondary, and special schools. Although the participants in the studies are very heterogeneous and certainly need further research, the results of all eight studies indicate that both types of interventions, RT in L1, and storytelling in L2, are highly effective in terms of literacy and thus provide an opportunity for students to improve significantly and regain enjoyment in learning. At the same time, the findings provide guidance for teachers on how to successfully design and implement interventions for L1 and L2 students with learning and behavioral difficulties across a wide range of grade levels.

Table of Contents

List of Abbreviations

ANOVA	Analysis of Variance
DCT	Dual Coding Theory
DRT	Dual Route Theory
EBD	Emotional Behavioral Disorder
GL2	German Second Language
GCs	Group Contingencies
IQ	Intelligence Quotient
ITRF-G	Integrated Teacher Report Form German
IRD	Improved Rate Difference
K	Kindergarten
L1	First Language
L2	Second Language
LD	Learning Disabilities
LQ	Reading Quotient
MBD	Mean Baseline Difference
NAP	Non-Overlap of All Pairs
OECD	Organisation for Economic Cooperation and Development
PA	Phonological Awareness
PAND	Percentage of All Non-Overlapping Data
PEM	Percentage Exceeding the Median
PISA	The Programme for International Student Assessment
PT	Peer-Tutoring
RR	Repeated Reading
RT	Reading Racetracks
SCR	Single-Case-Research
SELLMO	Scales for the assessment of learning and achievement motivation
SLRT	Salzburg Literacy Test
THEPHOBE	Test for the acquisition of phonological awareness and naming speed
WM	Working Memory
ZLT-II	Zurich Reading test II

1. Introduction

All students need effective literacy support to become literate adults. According to Bacon (1998), every human has the right to become a literate adult, which applies as much to second language (L2) acquisition as it does to an individual's first language (L1). As a subcomponent of literacy, the ability to read adequately is a foundational skill for literacy development (Oakes et al., 2010; Schaars et al., 2017). Being non-proficient in literacy enhances the risk of low income (Cree et al., 2012), can lead to difficulties in terms of emotional, economical, health, and social factors (Livingston et al., 2018), and impacts overall education as well as personal and social life matters (Brasseur-Hock et al., 2011; Chall & Jacobs, 2003; Macdonald et al., 2016).

To date, literacy proficiency has been decreasing among student populations in many countries such as Germany (Harju-Luukkainen et al., 2020). The gap between strong and low-achieving readers has widened over the years as a global educational challenge in both elementary and secondary school (Cirino et al., 2013; Harju-Luukkainen et al., 2020; McFarland et al., 2019; Mullis et al., 2017, National Center for Education Statistics, 2019; Strickland et al., 2013). The Programme for International Student Assessment (PISA) 2018 results reveal that only about 76% of 15-year-old students can minimally extract information from texts (i.e., reading level 2; Organisation for Economic Cooperation and Development [OECD], 2019). Also, in primary school, the number of students who achieve on the lowest proficiency level is increasing (Harju-Luukkainen et al., 2020). In addition, there is a steadily declining motivation to read among German readers (OECD, 2019).

Some students have a higher risk of non-proficiency in literacy than others. Significant groups are students learning the language of instruction as L2 (Wendt et al., 2016), students with behavioral (Carter et al., 2010; Garwood et al., 2017) and learning difficulties (Cirino et al., 2013; Solis et al., 2012). In Germany, about 30% of students have a migration background and are therefore likely to grow up multilingual with German as their L2 (Federal Government Commissioner for Migration, Refugees and Integration, 2014). These students are more likely to have significantly lower German literacy achievement, especially regarding reading and vocabulary, compared to their L1 peers (Gentry & Lindsey, 2008; Melby-Lervåg & Lervåg , 2014; Schleicher, 2019). Learning in an environment where the language of instruction is not the native language poses additional challenges for students (Becker-Mrotzek et al., 2012). This can also be applied to the English language, which is widely taught in German schools because English is of great importance (Crystal, 2015; Oxford, 2017). Challenges in becoming proficient in the English language can lead to long-lasting consequences (Molnár, 2008; Young et al., 2019).

Most non-proficient readers appear to have problems with lower-level reading skills such as reading fluency and word recognition in particular (De Jong et al., 2012; Gangl et al., 2018). These skills are foundational and must be fostered to counteract the risk that reading problems become entrenched and that adequate reading competence cannot be achieved as a result. Students with learning and behavioral difficulties can face severe challenges in reading (Fuchs et al., 2012; Garwood et al., 2017; Lane et al., 2008; Lerner & Johns, 2011; Solis et al., 2012), and their co-occurrence is undoubtable, which is quite discernable in literacy achievement (Arnold et al., 2005; Reid et al., 2004; Roberts et al., 2015).

In reference to Dewey (1916), the main task of education is to prepare students for active participation in a democratic society, and according to the Convention on the Rights of Persons with Disabilities, people with disabilities and accompanying special educational needs should have the opportunity to be educated in the general education system without exclusion because of their disability (United Nations, 2006, Article 24). Both references indicate that interventions should have the potential to consider the individuality of students to make education possible for all. In addition to the problems affecting primary and secondary education with regard to literacy and the declining motivation to learn, students with general learning difficulties and problem behaviors face even more daunting hurdles. The increasing heterogeneity in classrooms, students' learning difficulties, and behavioral problems pose challenges to teachers and practitioners, and support mechanisms must be designed that are adaptable to each student's unique need.

For effective support of literacy skills, teachers require adequate interventions and must consider that struggling students often tend to respond less to interventions (Peng & Fuchs, 2017). However, when suitable to the students, such interventions increase the probability that they will be improving their skills and will not develop serious difficulties (ibid.). Nevertheless, lower literacy abilities are no longer the focus from upper elementary on (e.g., Kent et al., 2017; Vaughn et al., 2003), even though the large number of struggling readers is widely known. Furthermore, literacy research on supervising secondary students and, more precisely, on student subgroups in primary and secondary education is still scarce (e.g., Ciullo et al., 2016). Although focusing on specific subgroups such as L2 learners and students with behavioral problems in addition to learning problems is so important, it appears that L1 research is still significantly more common than L2 research (Hall et al., 2019) and research on students with behavioral problems is declining (Garwood et al., 2020).

Building on this knowledge, this dissertation deals with effective literacy interventions that can be adapted to students' individual needs in Germany, focusing on German L1 learners in the first part, as well as German L2 and German students who are learning English as their L2 in the second part; both parts focus specifically on students with learning and behavioral difficulties. The overall aim of this work was to design interventions, test their effectiveness in

relation to the group of students selected, and gain new insights. The sharp increase in the number of less-proficient literate students (e.g., Harju-Luukkainen et al., 2020) illustrates, on one hand, that there is still not enough adequate support and, on the other hand, how important it is to provide all students with equal access to education through adequate literacy interventions. In addition, the dissertation addresses the conception of supports that are individually applicable to both children and adolescents with difficulties in learning and behavior with and without official diagnoses.

In the first part, an easy-to-use intervention combination of RT, peer tutoring (PT), and motivational elements was applied with regard to primary and secondary students with learning and behavioral difficulties who face challenges in reading. The interventions were conducted in small groups and whole classrooms. This combination of methods is based on general knowledge of research on effective components in the support of students with behavioral and/or learning difficulties. A final study takes the racetrack combination further by combining it with letter cluster training to examine the possible transfer effects on untrained items, with the intention of making the racetrack intervention even stronger. As it is of increasing importance to support L2 students, and since research in this field is not yet advanced enough (Hall et al., 2019), the second part focuses on a self-developed combined storytelling intervention and its effects on German and English as an L2, focusing on students with learning and behavioral difficulties. The conception of the storytelling intervention is adapted to findings from L2 research. As in the first part, this part also focuses on students with learning and behavioral difficulties who face hurdles in L2 acquisition. Finally, the second part is intended to illustrate that improvement of the L2 is also possible for students who have serious learning and behavioral difficulties. Both parts conclude with a discussion of the implications, and the whole dissertation offers a general conclusion and implications with regard to the findings from both parts

2. Reading Support for Students with Learning and Behavioral Difficulties

2.1 Word Reading

According to Frankel et al. (2016), literacy involves the process of reading, oral language, and writing. The first part of this dissertation focuses explicitly on reading. Due to the high number of students with difficulties in lower-level reading abilities and the resulting urgent need to foster and strengthen these abilities (De Jong et al., 2012; Gangl et al., 2018), the focus, more specifically, will be on word recognition.

Word recognition ability is a significant predictor of overall reading proficiency, including fluency and comprehension, at all grade levels - and students who still face hurdles at the word level in 3rd grade are at a higher risk for achievement failure in secondary school and dropping out (Brasseur-Hock et al., 2011; Kuhn et al., 2014; Mellard et al., 2010; Stanley et al., 2018; Toste et al., 2019; Zarić et al., 2021). A quarter of older elementary school students are unable to read words at their grade level, and an increasing number of adolescents are not reading age-appropriately (Brasseur-Hock et al., 2011; Cirino et al., 2013). Thus, reading at word level is an important aspect for continued research.

Reading fluency plays a significant role in appropriate word recognition and is thought to bridge the function between decoding and reading comprehension (LaBerge & Samuels, 1974; Pikulski & Chard, 2005;). It is defined as fast and accurate word reading, which in turn facilitates reading comprehension (LaBerge & Samuels, 1974). A lack of adequate reading fluency can lead to difficulties in text comprehension and general reading competency (Cromley & Azvedo, 2007; Little et al., 2017; Rasinski et al., 2017). A meta-analysis by García and Cain (2014) showed a correlation of r = 0.74 between lower-level reading skills and reading comprehension, underlining the importance in supporting basic reading skills. More precisely, word recognition is usually one main reason for inadequate reading fluency (Kuhn et al., 2014; Wolf & Katzir-Cohen, 2001) among primary and secondary school students (Boltzmann et al., 2017; Cirino et al., 2013; Eme et al., 2014; Paige et al., 2013; Yıldırım & Rasinski, 2014).

With respect to German reading, a longitudinal reading development study by Landerl and Wimmer (2008) showed that 70% of 115 German students who were dysfluent in 1st grade were still dysfluent in 8th grade, revealing the urgent need to counteract these manifestations and reduce the number of younger and older students with inadequate reading skills. Bar-Kochva et al. (2021) observed a general deficit of older German readers with difficulties in basic reading skills, including word reading, compared to normal readers. Based on this insight regarding older readers, which can be seen nationally and internationally, Cirino et al. (2013) suggested implementing basic reading interventions for older students instead of only focusing on primary school.

To sum up, reading words automatically can be a big challenge for students (e.g., Kuhn et al., 2014). According to the dual-route theory (DRT; Coltheart, 1993) and dual-route cascaded model (Coltheart et al., 2001), reading occurs through two routes: the direct route (directly accessing the mental lexicon and directly retrieving whole words to be read) and the indirect route (retrieving words through phoneme–grapheme correspondence). Both routes are important for word recognition and are both used by skilled readers (Ziegler et al., 2014). With regard to the ability to recall words directly from the mental lexicon, one often speaks of storing sight words that can be automatically recalled by learners (Ehri, 2005). Ehri's model of word reading (2005) consists of four stages: (1) pre-alphabetic (no alphabetic knowledge; print is

recognized due to memory; connections based on visual and context clues); (2) partial-alphabetic knowledge (beginning to connect spelling and pronunciation; readers are not yet able to fully decode and merely focus on initial and final letters); (3) full-alphabetic phase (the ability to fully decode words and understand the grapheme–phoneme correspondence [GPC]); and (4) consolidates alphabetic (connecting based on syllabic units). Building orthographic representations of word forms—storage of sight words—is fundamental for rapid word recognition because a lack can result in speed and accuracy challenges (Grainger et al., 2012; Moats & Tolman, 2019; Perfetti & Stafura, 2014). Orthographic processing means to retrieve entire words or letter combinations from the mental lexicon without making use of GPC (Ehri, 1997). Being proficient in word recognition has been shown to save cognitive resources that are needed for higher processing skills of reading such as reading comprehension (Kuhn & Stahl, 2003; O'Connor et al., 2007; Perfetti, 1985). Moreover, research demonstrates that orthographic knowledge contributes significantly to reading performance and is acquired through repeated print exposure (Holland et al., 2004). Problems in sight word reading are not only seen in non-transparent languages but also transparent languages like German (Ehri, 2005).

German is phonologically more transparent than, for example, the English language (Gangl et al., 2018) due to its mostly consistent letter–sound correspondence (Ziegler & Goswami, 2005; Rau et al., 2016). Wimmer et al. (2010) stated that both routes, lexical and sub-lexical, are inefficiently used among struggling German readers. They found under-activation in the brain regarding the lexical route - whole-word recognition - and the sub-lexical route–phonological processes. Gangl et al. (2018) conducted an eye-tracking study with dysfluent German 3rd and 4th graders. They found that dysfluent German readers rely on both sub-lexical and lexical reading routes. However, even though access to both routes can be restricted, according to Landerl and Wimmer (2008), word fluency is the only measure in word reading that differentiates between poor and good readers in more transparent orthographies, and reading development depends more on naming speed than on phonological awareness (PA; see also Joshi & McCardle, 2018; Landerl & Wimmer, 2000). Knoepke et al. (2014) showed that, with respect to German primary school students, orthographic decoding predicts sentence and text comprehension better than phonological recoding. A study by Zarić and Nagler (2021) examining poor readers among German primary school students and the impact of orthographic knowledge revealed that it contributes to word and sentence reading. In examining the topic, the urgent question arose as to where the relatively widespread reading problems come from.

Rapid Automatized Naming (RAN)

Based on Landerl and Wimmer (2008), one of the most important predictors in reading and especially in German reading is rapid automatized naming (RAN), which is the ability to quickly name shown items such as letter names and colors (Cohen et al., 2018). Cognitive processes that take place during rapid access to the mental lexicon, as well as word recognition, are comparable to the processes in RAN and thus make it a good predictor for reading ability (for meta-analysis of the German language, see Huschka et al., 2021; Landerl et al., 2019; Moll et al., 2009; Parrila et al., 2004) and oral reading fluency in particular (Conrad & Levy, 2007; Papadopoulos et al., 2016). Furthermore, a meta-analysis by Chen et al. (2021) found a moderate association between RAN and reading in transparent languages (r = .44; also see Araújo et al., 2015, r = 0.43), and, specifically, word-reading fluency has shown to be generally well predicted by RAN in a transparent language (e.g., Papadopoulos et al., 2016).

Phonological Awareness (PA)

PA describes the ability to identify and manipulate phonemes in spoken words (Ehri, 2005), which plays a primary role in word recognition (De Jong & Van der Leij, 2003; Karageorgos et al., 2020). However, according to a meta-analysis by Pfost (2015), about only 10-15% of the variance in later reading can be explained by PA in German reading. The results display a correlation of r = .33. A meta-analysis of less-transparent languages such as English found stronger correlations, resulting in the assumption that students learning more transparent languages have less challenges in phoneme–grapheme correspondence (Ziegler & Goswami, 2005). More specifically, Church et al. (2008) found that, with age, phonological brain regions are used less while reading, indicating the decreasing importance of PA with age. Fischer and Pfost (2015) stated that trainings in PA are more important in less-transparent languages than in German. In their review on the effects of PA support on reading for both less-proficient and stronger German readers, only small effect sizes were found. Even though PA seems to have a weaker impact on the German language than RAN, it is still an important component to consider (Schulte-Körne, 2011; Ziegler & Goswami, 2005).

Working Memory (WM)

Working memory (WM) can be seen as a cognitive system responsible for accessing the information needed for ongoing cognitive processes (Swanson, 2016; Wilhelm, 2013). Swanson and Stomel (2012) stated that learning difficulties can be caused by a malfunctioning WM, resulting in an inability to transform and store certain inputs. Problems in word reading and general academic performance can moreover derive from poor automatization due to less WM capacities (Gathercole et al., 2006; Pikulski & Chard, 2005). WM difficulties are highly related to difficulties in learning (De Weerdt et al., 2013; Moll et al., 2016; Sparks, 2021), and, in more detail, the ability to store and retain information is predictive of word reading

(Gathercole & Baddeley, 1993). A meta-analysis by Peng et al. (2018) displayed a small correlation among reading skills and WM (r = .29) and word reading and WM (r = .28). A meta-analysis by Ober et al. (2020) found a correlation between executive function and decoding of r = .32, which can be seen as moderate, specifically word reading r = .32 and non-word reading r = .33. The correlation between WM and word reading was r = .31 and non-word reading r = .34. No difference was found between age and grades. Even though these results do not show a strong correlation, WM needs to be considered as an influencing factor in reading.

Motivation

Despite WM, motivation is another main factor in learning (e.g., Sparks, 2021). Motivation can be defined as something that drives us to achieve a certain goal (Cerasoli et al., 2014). Even though cognitive variables also contribute to learning, it has been shown that motivation, after controlling for cognition, can influence learning success to a large extend (Gold, 2011; Lenhard & Lenhard, 2013; Taboada et al., 2009). In particular, repeated failure in school can lead to de-motivation (Aunola et al., 2002; Gruenke et al., 2017; Lenhard & Lenhard, 2013) and in turn, according to Guthrie and Wigfield (2000), students who are highly motivated are in general more active readers (e.g., Guthrie et al., 1999 r = .59 between reading and motivation in secondary students; Logan et al., 2011). Concerning reading it has been displayed that motivation to read is decreasing in Germany (OECD, 2019) indicating the urgent need to re-motivate students again. According to Becker et al. (2010) there is a relationship between reading competency and reading motivation and that a reduced motivation can lead to ineffective interventions derived from decreased effort and leading to reduced learning outcomes, especially for students with problem behavior (Nelson & Harwood, 2010; Scanlon et al., 2017). These are the reasons why motivation should always be taken into account when planning (school) interventions. Generally, research reveals that motivation declines with age and grades (Guthrie & Wigfield, 2000). Subsequently, a study by Gunobgunob-Mirasol (2020) on the attitudes of secondary school students towards reading found that reading attitudes decline with age and grade – making the important indication to go on foster reading in a joyful way. When interventions are specifically intended to focus motivation, students' reading development and motivation increases (e.g., Bates et al., 2016). As claimed by De Naeghel et al. (2012), encouraging students to read more belongs to a quality-criteria of education. After it became clear that reading proficiency can be affected by multiple factors and that basic skills are essential to reading proficiency, I asked myself which students are most affected by.

2.2 Learning and Behavioral Difficulties

Students with or at-risk for learning disabilities as well as students with behavioral difficulties face a higher risk of showing the most severe reading challenges (Fuchs et al.,

2012; Lane et al., 2008) largely impacting lower hierarchy reading skills such as word recognition and fluency (Cirino et a., 2013). Learning difficulties are the most common issues in education which often go in line with behavioral difficulties (Matson & Fodstad, 2010), and teachers face challenges in providing adequate support for these groups of students (Forlin & Chambers, 2011). Based on these findings, the emphasis in this dissertation in both parts is on these two student groups and on how to provide adequate literacy instruction where teachers urgently need help.

Learning Difficulties

There are several ways to describe students with learning challenges (e.g., learning disabilities [LD] or learning differences), but in general, the demarcation is very unclear, both nationally and internationally (Kraemer et al., 2021). Regarding the often-used term LD, Keogh (2020) stated there is still disagreement about the term. The studies conducted for this dissertation partly included students labeled as LD, specifically regarding the German definition of LD, which can be accompanied by a slightly reduced IQ in addition to academic failure and special needs (Gruenke & Cavendish, 2016; Hasselhorn & Gold, 2017). According to Gruenke and Morrisson-Cavendish (2016), LD refers to students who are unable to develop the knowledge, skills, and self-regulation necessary to meet the demands of the educational curriculum. Dombrowski et al. (2004) suggested that students classified as LD should demonstrate academic achievement deficits on standardized measures and impairments in academic performance, as evidenced by below-average grades (e.g., Sparks, 2016). However, completely independent of a diagnosis, all students with learning difficulties need support because they have an even greater risk of falling further behind in the education system (Grigorenko et al., 2020; Ritchie & Bates, 2013).

Due to the term inconsistencies and the fact that, independent from an official diagnosis, struggling students need deeper support, this dissertation uses the term *learning difficulties* as an umbrella term, according to the definitions provided by Gruenke and Morrisson-Cavendish (2016) and Dombrowski et al. (2004), for all students who face challenges with certain domains in school - more specifically difficulties in literacy in L1 and/or L2. The studies include students with and without a German diagnosis of LD—that is, all students who face severe challenges in literacy (Lenhard & Lenhard, 2013).

Many students with learning difficulties face challenges in reading and word-reading ability (Chard et al., 2002; Fletcher et al., 2007; Lerner & Johns, 2011; Solis et al., 2012; Strickland et al., 2013). Geary et al. (2020) found that memory contributes to learning difficulties, especially regarding reading difficulties when students need to store and retrieve specific information. Nonetheless, important to note is that up to 50% of students with LD and learning difficulties do not respond to interventions that are effective for students without (Al

Otaiba & Fuchs, 2006), which needs to be understood when planning interventions. However, with regard to intervention planning, it is also important to consider the group of students who have behavioral difficulties.

Behavior Difficulties

The same definitional conditions apply to the behavioral difficulties domain (Smith & Taylor, 2010). Thus, this dissertation uses the term *behavioral difficulties* to describe all students who struggle with behavioral patterns with or without official diagnosis of an emotional behavioral disorder (EBD). The participants labeled as having behavioral difficulties in the dissertation's studies do show externalizing behavior. Externalizing behavioral patterns are immensely challenging for teachers and for providing adequate support (Freiberg et al., 2009; Wills et al., 2010). According to Kauffman and Landrum (2009), externalizing problems mean behavior such as destruction, disruption, acting out, aggression, and impulsivity (see also Department for Education, 2012; Halonen et al., 2006). This explanation is in line with the German view on EBD, which is important to mention because the participants who are considered to have an EBD in the studies involved are classified according to the German definition and receive special needs support. All other subjects with problem behaviors were defined using results from the Integrated Teacher Report Form-German [ITRF-G], Volpe et al., 2018).

Students with or at risk for EBD are confronted with academic and behavioral challenges, which can lead to poor academic performances and social issues (Campbell et al., 2018; Garwood et al., 2017; Lane et al., 2008). Vaughn et al. (2002) found that students with EBD spend less time reading compared to their peers without behavioral challenges. It has been shown that students with problem behaviors, especially in reading, are at least one school year behind (Oakes et al., 2010; Vannest et al., 2009), whereas one main problem area is reading fluency (Hilsmier et al., 2016). Overall, challenging behavior is rising, especially among children (Carter et al., 2010). Nevertheless, a synthesis by Garwood et al. (2020) revealed there has been a decline in focus on interventions for students with EBD in the last 10 years. In general, when focusing on the outcomes among students with behavioral difficulties, most researchers have emphasized behavioral outcomes and not the improvement of academic achievement (Nelson et al., 2014).

How Both are Connected

Academic difficulties and problem behaviors co-occur, which is well seen in reading (Arnold et al., 2005; Reid et al., 2004; Roberts et al., 2020), and up to 50% of students with behavioral problems also showing severe learning difficulties (Al-Hendawi, 2012). In turn, students with learning difficulties show more behavioral difficulties than their normal developing

peers (Jordan et al., 2020). According to Gilmour et al. (2019), students with EBD and LD can fall behind their nondisabled peers with regard to reading by about grades 3 to 4. The authors advocated for the urgent implementation of appropriate interventions for these students. Becherer et al. (2020) revealed that problem behaviors have a negative impact on academic achievement in 5th to 9th graders, which corroborates Metsaepelto et al.'s (2015) findings. In general, students who face severe and repeated failures in school tend to avoid engagement in academic tasks (Neff et al., 2005). This non-engagement is common for students with behavioral and learning difficulties (Allday et al., 2011). However, engagement, in turn, is important with respect to enhancing academic outcome (Carini et al., 2006). Also, as Nelson et al. (2003) displayed in a meta-analysis, problem behaviors are predictive of the effectiveness of reading interventions.

Thus, many factors influence reading achievement and affect not only primary age students but also secondary age students. In addition, general learning difficulties and behavioral problems also influence reading and interventions' effectiveness (e.g., Nelson et al., 2003). Consideration should also be given to the fact that students who repeatedly fail in school may find themselves unwilling to participate in interventions, which leads to non-improvement in skills and in turn leads to an unwillingness to participate. Based on these findings, the question now arises as to what research suggests after taking all of these facts into account.

2.3 Effective Reading Support – What Research Suggests

As less-proficient readers' reading issues become more significant and skilled readers become even more skilled, referred to as the Matthew effect (Stanovich, 2009), it is of immense importance to provide less-skilled students appropriate interventions. According to a meta-analysis by Pfost et al. (2014), the Matthew effect is particularly true in word reading. At higher grade levels, students are required to use reading to comprehend information, making reading comprehension a key component of school success (Clemens et al., 2021; Steinle et al., 2021). However, the mastery of basic reading skills is necessary for students to even comprehend what they read (Cirino et al., 2013), and even though it is true that the number of readers who are less proficient and lack basic skills is rising (Gangl et al., 2018; McFarland et al., 2019; Mullis et al., 2017), little instruction in lower-level reading skills is provided later in elementary school (Kent et al., 2017). More specifically, word-reading instruction is typically absent from instruction after 2nd grade (Vaughn et al., 2003). Despite the needs of older primary school students, there is also a need to provide adequate interventions on basic skills for adolescent readers (Kamil et al., 2008). In general, researchers have observed at the primary level that instruction is rarely focused on explicit reading fluency, which, as a consequence, is not addressed at the secondary level (Ciullo et al., 2019; Wexler et al., 2018). This fact

underscores the importance of reading instruction at each grade level, depending on students' needs. In summary, research on secondary school students is not as advanced as research focusing on elementary students (Fien et al., 2018). Looking at specific student subgroups, literacy research is scarce, specifically in relation to secondary students with behavioral problems (Ciullo et al., 2016).

In their meta-analysis, Roberts et al. (2020) examined the effects of reading interventions for elementary students with reading and behavioral difficulties and found they can benefit from explicit reading instruction related to word reading in small groups, which, according to the authors, can be extended to students who struggle with reading and do not exhibit behavioral problems (see also Archer & Hughes, 2011). Generally, research has shown that explicitly smaller group interventions are effective for students with or at risk for LD and EBD (for a meta-analysis, see Hall & Burns, 2018; Roberts et al., 2020; Vaughn et al., 2003). A meta-analysis by Scammacca et al. (2015) focused on studies from 1980-2011 regarding the effects of reading interventions on reading outcomes in 4th to 12th graders and found the overall positive effects of reading interventions on all reading measures (d = 0.49), which is also true for L2 learners (for a meta-analysis, see Ludwig et al., 2019; d = 0.80 for reading fluency). All in all, however, interventions should be planned according to students' needs (Connor et al., 2014; Kamil et al., 2008), and it is important to keep in mind that research has revealed the challenge for an intervention to significantly influence the reading outcome in less-proficient readers, especially in older students (Solis et al., 2014). However, there is a notable dearth of reading research for secondary education compared to primary education even though figures have shown that both groups of students still face hurdles with lower-level reading skills (Cirino et al., 2013; Harju-Luukkainen et al., 2020; McFarland et al., 2019; Mullis et al., 2017, National Center for Education Statistics, 2019; Strickland et al., 2013). Moreover, many studies have been conducted with English-speaking students, but there is a definable lack of research in this area for German-speaking students.

These research suggestions posed additional questions about what the research specifically recommends in terms of effective reading interventions for struggling younger and older readers with basic reading, learning, and/or behavioral difficulties. This gave rise to the idea of how reading can be promoted at a lower hierarchical level.

Lexical and Sub-Lexical Training

As argued by Ehri (1999), reading instruction should first focus on the word level. In a meta-analysis of the effects of reading interventions on reading outcomes in 4th to 12th graders, Scammacca et al. (2007) found that, independent of grade, all students can benefit from word-level intervention, including those with learning difficulties (see Wanzek et al., 2010;

Torgesen et al., 2001; Toste et al., 2019). Further, Compton et al. (2005) investigated the effects of word training on reading comprehension in 3rd to 5th graders with below-average reading skills and found significant gains in reading comprehension (d = 1.15). Scammacca et al. (2015) stated that more research on word-learning interventions is needed, and Martin-Chang and Levy (2005) also pleaded for the implementation of isolated word-recognition training based on positive study results on word-reading trainings. Especially in Germany, studies on the effectiveness of whole-word training are rather a rarity.

Despite the whole-word training, according to Pikulski and Chard (2005), teaching multi-letter patterns helps students to transit from the full alphabetic to the consolidated alphabetic stage (e.g., Ehri, 2014). Also, Huemer et al. (2008) stated that interventions that go beyond whole-word training and use sub-lexical units might lead to greater transfer effects because single-word training effects are item specific. With regard to consonant clusters, researchers have shown that they enhance the reading of words, thus necessitating their use (Hintikka et al., 2008). Huemer et al. (2008) found the effects of sub-lexical training on trained and untrained words involve these clusters in German poor readers from the 2nd to 4th grade. However, Huemer et al. (2010) found that sub-lexical training had effects on trained words but not on untrained ones in Finnish poor readers from 4th to 6th grade (see also Grosche et al., 2013). These results display that there is still no agreement as to whether sub-lexical training clearly leads to transfer effects. In summary, word-reading training is effective in the context of reading. Likewise, training in letter clusters proves to be a good variant in promoting reading among struggling students with the additional chance of achieving transfer effects. Looking back at the fact that interventions for students with learning and behavioral difficulties should be designed in an appealing way (Bates et al., 2016; Nelson & Harwood, 2010; Scanlon et al., 2017), the question arises about how to realize a word and/or letter cluster training that is fun and promotes reading at the same time.

Repeated Reading Through Racetracks

One of the most efficient methods to enhance reading fluency is repeated reading (RR), meaning the repetition of words, passages, or text (Hintikka et al., 2008; Therrien, 2004). Its effectiveness has been demonstrated for students with and without learning and behavioral difficulties in primary and secondary education (Escarpio & Barbetta, 2016; Lee & Yoon, 2017; Musti-Rao et al., 2009; Stevens et al., 2017; Strickland et al., 2013; Wu et al., 2020) and in L2 (Gorsuch & Taguchi, 2008; Grabe, 2010; Jeon, 2009).

With respect to the dissertation's focus, RR will refer to repeated word reading. Automaticity in word recognition can be achieved by exposing students to words and patterns of words (Grabe 2010; Rasinski, 2014;). Repeatedly practicing words has been found effective specifically for poor readers (e.g., Berends & Reitsma 2006a; Martin-Chang & Levy 2005;

Steenbeek-Planting et al., 2012; Thaler et al., 2004; van Gorp et al., 2017). Martens and de Jong (2008) examined the effects of RR on reading speed in 4th- and 5th-grade dyslexic children, finding increased reading speed and accuracy (see also van Gorp, et al., 2014 for kindergarten children). However, RR interventions focusing on repeated word reading struggle to achieve transfer effects on reading fluency (Berends & Reitsma 2006a; Thaler et al., 2004). Berends and Reitsma (2006b) showed a positive effect of RR among Dutch children but stated that a transfer to untrained words was difficult (see also Kohnen et al., 2008). Van Gorp et al. (2017) investigated the effects of repeated word and pseudo-word reading in poor and strong 1st-grade readers. They found that both reader groups' accuracy and speed were enhanced in a transparent orthography. According to a research synthesis by Steinle et al. (2021), research on the effects of RR specifically in struggling adolescent readers is not yet advanced which is also true in German language. Thus, research needs to be extended in these specific areas.

To make reading interventions, as RR, a more enjoyable experience, especially for students with learning and behavioral difficulties, interventions should include game-like components. Games excite students at all levels and incorporate an element of fun that helps motivate students (Charlton et al., 2005). To teach word recognition properly while bearing possible motivational issues in mind, sight word training and RR can be embedded into a gamified method called Reading Racetracks (RT). RT is a simple method containing a playing field with empty squares, with the intention of putting certain content on these squares to be trained automatically. RT has been found effective for students with various special needs (Green et al., 2010; Hopewell et al., 2011) and for L2 learners (Gruenke & Barwasser, 2019; Sperling et al., 2019) regarding reading.

Direct and Explicit Instruction

For students with difficulties, direct instruction (DI) is an important part of being taught. Stockard et al. (2018) conducted a meta-analysis including over 300 studies from the last 50 years and found that DI is beneficial regardless of grade, age, or learning needs. However, according to Kuhn & Dean (2005), solely DI without practice might not be that effective and does not lead to long-term effects. According to Allington (2006), mastering reading skills requires DI combined with time for practice and feedback for students, especially for less proficient readers and those with behavioral challenges (Morgan et al., 2008). One method of DI is called explicit instruction, meaning content is conveyed not only directly but in small steps. Explicit instruction is especially beneficial for students with academic difficulties and learning disabilities (Archer & Hughes, 2011).

Therefore, one consideration was how to combine RT with direct explicit prior instruction of words and/or letter clusters, especially to give less proficient readers an orientation before doing the RT game. Other considerations included how to design the method so that it could be done with more students, and including a heterogeneous student body (stronger and less proficient students).

Peer-Tutoring

The rising heterogeneity of students makes it unavoidable to implement methods that can be used across a wide range of students (Carter et al., 2015; Scruggs et al., 2012). Peer tutoring (PT) can be defined as a student-mediated method in which two students or student groups work together on a specific task (Dufrene et al., 2010). PT has a positive impact on academic performance and reading (Dufrene et al., 2006; Gruenke, 2006; Hattie, 2008; Moeyaert et al., 2021; Slavin, 2011). It is effective for students with behavioral challenges (Bowman-Perrott et al., 2013, Tau-U = 0.76; Dunn et al., 2017; Ryan et al., 2004, d = 0.81 on reading; Sutherland & Snyoder, 2007), as well as students learning an L2 with and without learning difficulties (Cole, 2014 for meta-analysis [g = .486, p <.001]; Klingner et al., 2014; Sáenz et al., 2005). Moreover, PT also seems to positively influence behavior (Bowman-Perrott et al., 2014; Tau-U = 0.62 for behavior in pre-K–12). Ginsburg-Block et al. (2006) examined the effects of PT on behavior and found d = 0.45 for learning-related behavior and d = 0.28 for social competency. A review by Stenhoff and Lignugaris/Kraft (2007) revealed positive effects of PT in secondary education with individual student groups. In a research synthesis, Hudson et al. (2020) found that the most effective interventions on fluent word reading in primary school students were those implemented with a trained model as PT procedure.

Overall, giving feedback and correcting errors is beneficial in RR interventions (Kostewicz & Kubina, 2011; Lo et al., 2011). These practices can also be linked to motivation to work with peers as a social situation (Lee & Zentall, 2012). Rohrbeck et al. (2003) revealed that peer-assisted learning has an effect size of .59 for learning outcomes of elementary school students, with the greatest improvement found for the lowest achievers. Okilwa & Shelby (2010) conducted a literature synthesis and underlined the positive findings of PT independent of disability for 6- to 12-year-old students. Likewise, advantages have been shown for tutors in the PT procedure, which is also important to note (Leung, 2019). Nevertheless, tutors repotted to feel unsure about their role in PT (Vogel et al., 2007). In conclusion, according to Calhoon (2005), PT should be combined with reading instruction to improve reading fluency. Using PT alone without reading intervention would be difficult to achieve effects.

In reference to the sub-chapter on motivation and its enormous importance, I thought about how to further motivate students in addition to the playful character of the Racetracks.

The idea stems primarily from the fact that motivation is especially important for students with learning and behavioral difficulties, who face particular challenges to literacy.

Feedback and Positive Reinforcement

An effective intervention should not only target reading skills in isolation but also implement motivational components, especially when focusing on struggling students (e.g., Aro et al., 2018; Retelsdorf et al., 2014; van de Ven et al., 2017). Additional components such as feedback through self-graphing and positive reinforcement can lead to a further motivational boost (Bowman-Perrott et al., 2013; Leko, 2016; Mitchell, 2014). According to Hattie (2008), feedback is one of the most important components of learning. Troia (2009) says that it is important to ensure that students are guaranteed a sense of achievement, which also holds true for L2 learning (e.g., Bitchener, 2017; Ellis, 2012; Hansen & Liu, 2005; Li, 2010; Li et al., 2016; Lim & Rendandya, 2020; Dlaska & Krekeler, 2013).

Self-graphing is one component of self-monitoring. It describes a process where students graph their own learning progress and thus can track their learning development (Gunter et al., 2002). This can be effective in reading instruction in a student's L1 (McKenna & Bettini, 2018) and L2 (e.g., Albers & Hoffmann, 2012). McDaniel et al. (2013) also found positive effects of self-graphing on oral reading fluency in students with emotional and behavioral challenges. Schrauben and Witmer's (2019) review of feedback in reading programs revealed that task-specific feedback is especially effective, and that feedback regarding comparison of one's previous performances is more helpful than comparison with peer performances. Using self-graphing to track reading performance is beneficial to increase engagement and motivation (Menzies et al., 2009). Guzman et al.'s (2018) meta-analysis of self-monitoring on K–12 students' reading showed a large positive effect on reading performance (Tau-U = 0.79, $p < .001$). Sutherland and Snyder (2003) examined the effects of a combined peer-tutorial intervention including self-graphing and found positive effects on disruptive behavior and word-reading ability in middle school students with EBD (see Bruhn et al., 2015; Holifield et al., 2010).

Positive reinforcement can be combined with aspects like self-graphing procedure or group contingencies (GCs). Positive reinforcement means that a positive stimulus is given (reward prospect) with the aim of increasing the probability that a certain skill or behavior will be shown (Feist et al., 2017). In respect to reading, implementing positive reinforcement can be beneficial in one's L1 and L2 (Clay et al., 2019; Dolezal et al., 2007; Fromkin et al., 2018). Dolezal et al. (2007) examined a reinforcement package and its effects on reading among secondary students with challenges in reading, with large effects on reading and on-task behavior. Furthermore, GC procedure has been shown to be effective (e.g., Rohrbeck et al.,

2003). The procedures used in my dissertation's studies are interdependent GCs, meaning that all members need to behave according to a rule to achieve a goal (Tingstrom et al., 2006). Implementing GCs in an intervention was especially effective (Maggin et al., 2012; Maggin et al., 2017; Rohrbeck et al., 2003). GCs are commonly implemented in response to behavioral changes (Maggin et al., 2012). In my studies, GC was connected to academic outcome, not behavior. Little et al. (2015) elaborated on the effects of GC on general and academic behavior of children and found positive effects on both. Particularly in the area of PT (Fuchs et al., 2001), the use of the GC method has been shown to boost effectiveness in the area of academic achievement (Bowman-Perrot et al., 2013; Rohrbeck, et al., 2003).

2.4 Interim Conclusion

Students who experience repeated failure in academic subjects are at higher risk of avoiding any engagement in academic tasks (Neff et al., 2005) and up to 50% of students with LD do not react positively to interventions (Al Otaiba & Fuchs, 2006). Thus, it is necessary to make interventions an enjoyable experience, also for older students, as it is known that motivation increasingly declines with age (e.g., OECD, 2019). Furthermore, as Garwood et al. (2020) revealed, there is a declining focus on interventions for students with behavioral problems even though they are urgently needed. It is important to counteract this trend.

As RT has never been evaluated before as a PT method, and bearing in mind the positive effects of PT among students with and without learning and behavioral difficulties, the idea of adding PT to RT derived from the intention to create an even more powerful and adaptable intervention that fits the great heterogeneity of the student population. As rising individual differences among students make it a daily challenge for teachers to implement adequate support (Mullis et al., 2017), PT is one way to avoid this challenge. Adding motivational components to an intervention is beneficial to the focused student population, as noted by Guzman et al. (2018) and Bowman-Perrott et al. (2013). Thus, a combination of RT and PT embedded with self-graphing and positive reinforcement should produce exceptional outcomes in word reading across individual students—RT with PT should counteract the tendency in students with learning and behavioral difficulties to become disengaged or not respond to reading interventions (see Allday et al., 2011; Al Otaiba & Fuchs, 2006; Neff et al., 2005). In the first study, the emphasis was on older students with LD and EBD who severely struggled in word reading.

The following presentations of the studies are a short summary. Detailed information can be found in the attached articles. In addition, a more explicit discussion of the findings can be found at the end of the first part and the final discussion of this dissertation. Because of this, only brief conclusions are given below.

2.4.1 Summary Article 1 *(peer-reviewed)*

Barwasser, A., Urton, K., & Gruenke, 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.

Since it has been demonstrated that older students can also be affected by a lack of basic reading skills, and these students therefore might be confronted with enormous problems at school (Brasseur-Hock et al., 2011; Cirino et al., 2013), this study focuses on secondary school students. However, as stated before, basic reading skills are assumed to be mastered in secondary school and reading is used to acquire knowledge (Clemens et al., 2021; Kent et al., 2017; Steinle et al., 2021); thus, a student's reading urgently needs to be automatized. Evidence shows that students who have learning and behavioral difficulties in particular are faced with challenges in reading (Carter et al., 2010; Cirino et al., 2013; Garwood et al., 2017; Solis et al., 2012), and decreasing motivation to learn over time also plays a significant role here (OECD, 2019). The aim of this study was to evaluate the effects of a combined RT intervention designed for a PT setting with a self-graphing and GC procedure on the word reading of secondary students with LD and EBD. Additionally, we measured the variable again in follow-up measurements.

Methods

The final participants were German 5th- to 7th-grade students with LD (N = 8), four of whom had additional diagnoses of EBD with low reading achievement. The participants were selected based on a German reading screening (Wimmer & Mayring, 2014). To realize the PT learning aspect, strong readers were assigned as tutors to the participants with less proficient reading skills, who were defined as tutees (= participants). To evaluate the study objective, a multiple baseline design (Lane et al., 2017) was used, with participants divided into three groups with different baseline lengths. The groups were held three times a week for a total of eight weeks. After each baseline and intervention session, data was collected from participants as they were asked to read a pool of training words shown for one second each on PowerPoint slides (see Ehri, 2005). Each word read correctly within one second of appearance was counted as correct. For the baseline procedure, all tutees worked with their tutors on cognitive tasks for 15 minutes. For the intervention phase, the pairs of students played RT together while the tutors trained the tutees. This procedure was played out for 15 minutes. As a reward procedure, the students filled in a self-graphing sheet at any time after measurement according to their amount of correctly read words. Additionally, there was a group target of amounts that

needed to be achieved to get a reward as a group. Treatment fidelity was implemented, and social validity questionnaires were given to the students at the end of the intervention.

Results

For data analysis, the SCAN package for R was used. As overlap indices, I applied non-overlap of all pairs (NAP, Parker et al., 2011a), percentage exceeding the median (PEM, Ma, 2006), percentage of all non-overlap data (PAND, Parker et al., 2007) and Tau-U with possible A-phase trend correction (Parker et al., 2011b). Moreover, a level-2 regression analysis across all participants was conducted. Our results show improvement in words read correctly across all participants in a very short time. Two participants already started with higher values in the baseline phase. Accordingly, a ceiling effect was evident here. However, a slight improvement for these participants was also visible with stable follow-up data. Overall, the follow-up data was stable, with a slight drop in data for three participants. Moderate to very large overlap indices were found, and the regression analysis showed a statistically significant level ($p < .001$) and slope effect ($p < .05$), with an average increase of 0.60 words per intervention session. Social validity results displayed overall positive attitudes towards the intervention.

Conclusion

Overall, the intervention with the current components seems very effective in improving sight words in a very short time among secondary students with poor reading proficiency with learning and behavioral difficulties. No difference can be observed according to student characteristics such as EBD. All profited equally. Thus, this study adds a significant piece to the puzzle in reading research, especially with regard to secondary school students, among whom research is still less advanced (Ciullo et al., 2016; Fien et al., 2018). Furthermore, which is an important indication in reading research, is that all students were able to remember and recall the words they learned even after some time (Daniel et al., 2021). With respect to inclusion and implementation with the whole class, another question would be whether the RT can also be used in the classroom context in the sense of class-wide PT. PT has been shown to be an adaptable tool for individual students' needs, making it useful for whole classrooms with heterogeneous students. Especially in terms of learning inclusion, PT as a diversity tool can meet the academic and social needs of all students (Thomas & Loxley, 2007). However, teachers often do not feel well prepared for inclusive education and teaching a whole classroom, whereby proficiency gaps are widened (Yada & Savolainen, 2017). According to Peters et al. (2021), classroom instruction or a given classroom method needs to have the potential to be embedded differently for different students: in Germany, students with LD are mostly overlooked in the classroom even though they urgently need specific support. Thus,

another aim was to make the combined RT an even more universal tool by letting the whole class participate and evaluate the intervention with another age group.

2.4.2 Summary Article 2 *(peer-reviewed)*

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

Research has shown that the different aspects integrated into the combined RT intervention are also effective regarding primary school (e.g., Gruenke & Barwasser, 2019; Guzman et al., 2018; Holifield et al., 2010; Sperling et al., 2019). Thus, there is a need to look at primary school students' reactions to the intervention. Despite this, given the reality of inclusion in heterogeneous classrooms and the fact that teachers experience challenges to find methods that fit all students in a classroom (Yada & Savolainen, 2017), it is worth examining the RT intervention in a classroom setting involving all children at the same time. Furthermore, studies rarely provide follow-up data, which is urgently needed (Daniel et al., 2021). Therefore, we integrated two follow-ups into this study to estimate long-term effects.

Thus, the research aim was to evaluate the implementation of a PT–RT intervention with motivational components on the word-reading fluency of struggling 3rd-grade readers in an inclusive school setting. Moreover, assuming that the intervention effects would have to be measured again at a later time point, another important question was whether these effects would persist five and 10 weeks later.

Methods

For this purpose, we realized an experimental design with a control group and an experimental group using a preintervention, postintervention, and two follow-up measurements. The control group received a math racetrack intervention, and the experimental group received the RT intervention. Both interventions were the same except for the content that was presented. The interventions ran four times per week for 15 minutes, for a total of three weeks. To find the final sample (tutees; N = 44) eligible for the study, we conducted a German reading screening (Wimmer & Mayring, 2014). This screening was also used to identify who would be the tutors (stronger readers; reading quotient (LQ) >100) and who would be the tutees (less proficient readers (= participants); LQ <89). Sequentially, matching pairs of tandems were equally distributed in the control and experimental conditions. In the RT intervention, the tandems played the games in the same way as in the previous study. The same was true for the math racetrack condition, where math tasks were trained. To

realize a self-graphing procedure in a whole class intervention, a timer was set in both conditions for one minute after the official intervention has finished, and the tutees were asked to read as many words/solve as many math problems as possible. The results were entered in a self-graphing sheet for which rewards were given. At measurement times, the 30 training words were measured (in a random order for T1, T2, T3, and T4) using a PowerPoint with one word written per slide, and the children were asked to read the word within one second of its appearance on screen (Ehri, 2005). The amount of correctly read words functioned as the dependent variable. Social validity and treatment fidelity were additionally considered.

Results

We used a repeated-measures ANOVA in addition to a robust repeated-measures rmANOVA to validate the results. There was no difference between the groups at pretest (p = .94). Further, there was a significant interaction effect at the p <.001 level. The increasing results from pre- to posttest remained constant (p <.001, d = 5.77) at follow-up 1 (p <.001, d = 4.09). However, there was a slight decrease from pretest to follow-up 2 (p <.001, d = 3.59). We found a statistically significant main effect for group (p < .001, η^2 = .77). An additional rmANOVA also revealed equivalent results. Regarding social validity, all teachers rated the items as positive, which was also found in the RT students' answers.

Conclusion

The experimental group benefited significantly from the RT intervention compared to the control group, who were given math exercises but received the same procedure including racetracks, PT, and motivational components. More precisely, the results were stable across two follow-up measurements with only a slight decrease, which could be seen as proof of long-term effectiveness. Thus, combined RT also seem to work as whole-class intervention in primary school, making it even more important for teachers; this was in line with prior studies (e.g., Grabe 2010; Rasinki, 2014; Scammacca et al., 2007, 2015; Steenbeek-Planting et al., 2012; Torgesen et al., 2001) as well as with the first study of this dissertation.

Now that it was indicated via two studies that the RT in the applied combinatoric is well suited to promote whole-word retrieval among primary and secondary school students, a next step is to determine the extent to which the racetracks can be combined with sub-lexical training to achieve transfer effects to unknown words, making the intervention more powerful. It is difficult to see how extensively transfer effects can be created by this intervention, since it has already been shown that sight-word training has very specific effects only on those sight words that have been trained (Pikulski & Chard, 2005). Some studies show transfer effects on untrained words (e.g., Huemer et al., 2008) and other studies do not reveal transfer effects (Grosche et al., 2013; Huemer et al., 2010). One could work with letter clusters—for example, those that occur frequently in German—to create transfer to foreign words (and thus also

possibly to reading fluency) with these clusters, combining RT with direct reading instruction (Allington, 2006; Kuhn & Dean, 2005).

2.4.3 Summary Article 3 *(peer-reviewed)*

Barwasser, A., Hertel, S., & Grünke, M. (2021). The effect of a (sub)lexical intervention using peer racetracks on the reading performance of low-literacy students with behavior problems with and without learning disabilities. *Learning Disabilities. A Contemporary Journal, 19*(2), 143-159.

Referring to a possible improvement of RT by training letter clusters (e.g., Huemer et al., 2008) and the significance of direct, explicit instruction (Stockard et al., 2018), the next study focuses on a combination of a direct instructed letter-cluster training and subsequent RT to be automated in primary school children with behavioral difficulties with and without LD who have severe reading problems. Four of the final participants reported not having German as their L1. This also directs a focus to see whether our intervention might also be effective in L2 learning—a direct link to the second part of this dissertation. The research aim was to evaluate the combined intervention already used in the first two studies, while additionally implementing sub-lexical pattern training and documenting its effects on the acquisition of trained and untrained words.

Methods

The study was conducted at a German primary school. To begin, a reading screening (SLRT II; Moll & Landerl, 2010) was used to select 3rd and 4th graders suitable for the intervention regarding their reading skills. Moreover, the ITRF-G as a behavioral screening was used to determine problem behavior From the combination of problematic behavior and reading problems, five subjects were ultimately described in this paper. Four of the subjects had learned German as an L2, and one child spoke German as their L1. The initial sample size was 10 students; however, some participants were excluded from the study due to missing data. In addition, two of the students had an official LD diagnose.

A multiple baseline design (Lane et al., 2017) was applied across three groups. The whole procedure was held three times a week for 20 minutes over a period of 5 weeks in total. The dependent variables were 1) the number of correctly read training words, and 2) the number of correctly read untrained words, measured by using a PowerPoint presentation in which the training words and the untrained words were embedded on the slides in random order. Participants were asked to read the words within one second of them appearing on the screen (Ehri, 2005). As a baseline condition, students were engaged in cognitive tasks, after

which the dependent variables were measured. The following intervention consisted of two phases: 1) DI of so-called ”signal groups” (frequently occurring letter clusters in German taken from “*Blitzschnelle Worterkennung*” by Mayer, 2018), and 2) automated training of the signal groups using training words on the racetrack game, which was played in pairs as a PT activity – however this time, PT was realized through pairs of students with equal reading skills. The children were allowed to fill in two self-graphed sheets for trained and untrained words according to their amount of correctly read words after each measurement. Treatment fidelity and social validity were once again considered.

Results

The SCAN package was used for data analysis. The overlap indices NAP, PAND, and Tau-U were calculated. For training words, significant medium-to-large effects ($p < .01$-$p < .001$) were observed across all participants. As expected, untrained words had lower effects ranging from small to high, a few of which were not significant. With regard to untrained words, Ella profited least. Follow-up data was stable across the two variables. Plus, the social validity questionnaire showed that all children enjoyed the intervention and rated all items as positive. Furthermore, the children gave personal statements on the intervention.

Conclusion

There was a clear increase in training and transfer words for all children. However, since transfer effects are more difficult to achieve and more practice is used for achieving any transfer effect (e.g., Taguchi & Gorsuch, 2002), we found lower values for this variable. Ella was the student with the lowest scores on word and pseudo-word reading in pretesting, which might be an explanation for the lowest scores in transfer words. Slower successive increase of participants could be linked to still relying on the indirect route of reading. Less-proficient readers tend to read letter by letter instead storing them as orthographical representations (De Jong et al., 2012). Thus, they might need additional training in phonological processes and probably a higher degree of direct instruction. Nevertheless, in summary, it seems as though more intense training is needed to achieve transfer effects on unknown words. Some children already had high baseline values, which makes the data difficult to interpret. Also, the degree of problem behavior did not seem to affect the results negatively – further, no significant differences were found between the participants with an official diagnosis of LD compared with their peers without. More far-reaching, there did not appear to be a difference between L1 and L2 German speakers. However, this assumption needs to be taken with caution due to the small sample size. One must take into account that the children who dropped out of the study (due to COVID-19 situation) were all German L1 speakers; our intent was actually not to place an emphasis on L2 students. Nevertheless, now, that we have a majority of German L2 students in the sample, one must take a closer look at L2 speakers. Conducting a study with

mostly L2 students raised the question, “how can research support German L2 students with learning and behavioral difficulties?”

2.5 Conclusion

Taking all the studies and their results into account, I conclude that combining RT with PT character and motivational boosters can help students with learning and behavioral problems in primary and secondary education to read words automatically. Overall, the results of the three studies support previous studies on the positive influence of PT (e.g., Bowman-Perrot, 2009; Calhoon, 2005; Hudson et al., 2020; Moeyaert et al., 2021). Additionally, our results are in line with further research on racetracks (e.g., Erbey et al., 2011), isolated word reading training (e.g., Martin-Chang & Levy, 2005), and motivational components (e.g., Alberto & Troutman, 2008). Additionally, the third study supports previous research on the effects on untrained words (Huemer et al., 2008; Hintikka et al., 2008) and contradicts findings that did not report effects on untrained items (e.g., Grosche et al., 2013; Huemer et al., 2010).

Even though these indications are restricted to our samples, our results give important indications of what interventions for these student groups might look like. The combined RT method can also be used in a small group setting and in class settings, which is enormously important in terms of inclusion: teachers have classrooms full of individual students with different needs (Mullis et al., 2017). The letter cluster training showed that in addition to trained words, there are also effects on untrained words (although the effects were lesser than for trained words). Students who needed a longer time to achieve effects might have still relied on the non-lexical route of the dual-root theory, which is common for less proficient readers (De Jong et al., 2012) before they start to store clusters or whole words. Presumably these students would have needed more intensive direct instruction of the clusters and words (e.g., Stockard et al., 2018).

Most importantly, it is gratifying that all students appeared to have benefited, which, referring to Al Otaiba and Fuchs (2006) and Wills et al. (2010), is a complete success. These authors state that up to 50% of students with LD and learning difficulties do not respond to interventions that are effective for students without problems in learning, and students with externalizing behaviors present teachers with significant challenges in providing appropriate support. They commonly need positive experiences of literacy, since they seem to be the group of students who struggle most (Hollo et al., 2014). As Garwood et al. (2020) revealed, there has been a decline of focus on interventions for students with EBD in the last 10 years. According to Nelson et al. (2014), interventions for students with behavioral difficulties should also address academic outcomes instead of always focusing on behavioral improvement. These studies add another piece to the puzzle in literacy research—not only in terms of

learning difficulties, but also in terms of students with behavioral challenges. Moreover, follow-up data in Study 1 and Study 2 implicates the storage of learned words after the interventions have ended which is an important hint in intervention research (e.g., Daniel et al., 2021).

Our social validity questionnaires revealed that the participants had fun engaging in literacy intervention and hopefully will continue to be active in such interventions (Atkinson et al., 2002). Further, Mitchell and Sutherland (2020) stated that PT is an important method of social integration; students with special needs face greater social challenges (Krull et al., 2018). Considering that teachers face enormous challenges in providing adequate support for students with learning and/or behavioral issues (Forlin & Chambers, 2011; Wills et al., 2010), the racetracks seem to be an intervention that works for a wide range of students and can be used individually by teachers. The first study showed that RT not only works with primary school students (e.g., Grünke, 2019; Grünke & Barwasser, 2019) but also in higher grades. This is an important finding regarding the amount of struggling secondary school students experience with lower-level reading skills (Brasseur-Hock et al., 2011; Cirino et al., 2013). This study is in line with findings from Scammacca et al. (2007, 2015) that secondary school students can profit from word reading interventions. Also, because students with behavioral difficulties are less motivated to engage in reading activities due to negative experiences (Hollo et al., 2014), it is important to provide an appropriate intervention that can motivate these students. As seen in our results, students who faced behavioral challenges all profited from the support, and there seemed to be no difference between students with behavioral problems and those without behavioral problems. In Study 3, we observed no difference between students with and without official LD diagnoses, and there was no transparent difference between students with official EBD diagnoses/degree of externalizing behavioral problems and those without. However, completely independent of diagnosis, all subjects showed enormous problems in reading on previous screenings. These findings underline the urgent need to focus on all students with difficulties, not only on those with official diagnoses, as both groups need help in literacy (e.g., Ritchie & Bates, 2013).

One could assume that in Studies 1 and 3, the small group size also influenced the effect, as studies show that students with learning and behavioral difficulties benefit from smaller group interventions (Vaughn et al., 2003; Hall & Burns, 2018). However, Study 2 (the whole-classroom realization) yielded the greatest effect sizes across the experimental groups' participants. These results might indicate that the combined Racetracks can be effectively implemented beyond a small group setting, partially overriding the small-group effect. Of course, this conjecture would have to be investigated in more detail, as there is not yet sufficient evidence to support it.

In conclusion, as research has shown that word reading difficulties are predictive of poor future reading comprehension (e.g., Holland et al., 2004; Stanley et al., 2018; Toste et

al., 2019; Zarić et al., 2021), which can lead to severe difficulties in school throughout a student's career (Cirino et al., 2013), the presented studies try to keep students from facing later academic challenges. More precisely, referring to Landerl and Wimmer (2008), if German students show dysfluency in first grade, it is of immense importance to counteract this tendency in higher grades. Word recognition insufficiency is mostly seen in less proficient readers in primary and secondary school (Eme et al., 2014; Paige et al., 2013). Again, this stresses the importance of our studies, which can be implemented in both primary and secondary schools.

Limitations

With respect to limitations of the studies, one criticism is that the intervention consists of a multi components that has positive effects as a whole package, and it is not possible to figure out which component influenced the dependent variables to what extent. However, even though it should be important for research to separate the components from each other, this multicomponent intervention was created purposely based on prior research results on what an intervention needs to be successful in the student population we focused on (e.g., Van de Ven et al., 2017; Zentall & Lee, 2012). It is important to provide multicomponent interventions including motivational boosters (Aro et al., 2018). I agree with these opinions and would like to encourage the embedding of motivational aspects in all literacy interventions. Also referring to Calhoon (2005) that PT seems to need additional reading instruction since on its own reading was difficult to improve, the components' effectiveness can be estimated to a certain degree.

One other criticism could be that we evaluated the intervention on specific student groups, which might make it difficult to draw general conclusions. However, to counteract this assumption, it is necessary to emphasize that it was indeed our goal to evaluate the intervention for exactly our sample; as argued in the first part, this student group is more challenging to foster effectively (e.g., Forlin & Chambers, 2011). Nevertheless, the intervention needs more research on its effectiveness in students with different needs. Subsequently, the use of the German version of the Integrated Teacher Report Form (ITRF-G) to assess students with behavioral problems could be criticized for being a highly subjective assessment (Bennett et al., 1993). Nevertheless, Volpe et al. (2018) encouraged implementing this screening due to its good-to-excellent diagnostic accuracy.

Another limitation is the use of single-case research (SCR) designs and small-scale studies regarding generalization of results. Nevertheless, focusing on specific student sub-groups makes it more difficult to find a larger sample. Moreover, SCR is an important way to track students' learning progress and estimate when an intervention starts to work and how long students need to react (e.g., Lane et al., 2017). Additionally, the studies included a specific pool of words, especially regarding the first two studies, and did not measure

influences on general reading fluency. Thus, we do not know if the effects are restricted to the specific pool. Only the last study had a larger word pool of training words and considered transfer word effects while combining racetracks with sub-lexical training (see also Huemer et al., 2008). Although we did not measure transfer effects on general reading fluency, I refer to, e.g., Knoepke et al.'s (2014) statement that orthographic decoding (the focus of the studies) does have an impact on reading—even on reading comprehension (see also McArthur et al., 2015). Following this, the students in Study 1 were shown the same pool of words at the time of measurement. Even if the words appeared in a randomized order, a training effect cannot be completely ruled out by the measurements. In Study 3, the pool was much larger in comparison and the students did not get the same words displayed every time.

Another restriction is that it is difficult to draw conclusions from the third study, in which German L1 and German L2 students were involved—all benefitted, but no statements can be made regarding differences in L1s' and L2s' performances. As a last limitation, to date, it is not known if the combined RT are more effective than other reading methods, which would be an important aspect to consider in future studies.

Implications

Considering the good results as well as the limitations, I have been thinking about what further studies could look like and which areas I am still interested in. First, regarding the letter-cluster study, in general, it seems that students need much more time to transfer their knowledge on untrained items, which might suggest that the training needs to be longer and/or more intensive. Overall, more research is needed on untrained items, since studies revealed different outcomes (e.g., Huemer et al., 2008; Huemer et al., 2010). Of course, it would be of great interest to see if the intervention also works when components are omitted and which combination is ultimately the most effective. In addition, one would have to do this again separately for primary and secondary education and see how students with certain special educational needs react to different combinations. One realization could be an A-B-BC design evaluating the RT in the B phase and adding one of the other components (e.g., self-graphing) to the intervention in the BC phase to see whether there is a difference regarding the dependent variable (see Kazdin, 2011).

Another interest of mine is the implementation of direct behavior rating, as I am not only interested in how academic performances can be improved but how interventions affect behaviors (e.g., disruptive behavior) in addition to academic achievements (see Nelson et. al., 2014). The components used in the studies also showed positive effects on behavior in recent studies, e.g., Sheffield and Wallers (2010) found that a self-monitoring procedure reduced disruptive behavior among students with various disabilities (see also Bruhn et al., 2015). For PT, Bowman-Perrot et al. (2013) and Moeyaert et al. (2021) also found positive effects on

behavior. As it has been said that academic achievement and behavior correlate with each other (e.g., Arnold et al., 2005; Reid et al., 2004; Roberts et al., 2019), it is of great interest whether better reading performances would also lead to more appropriate behavior. Direct behavior rating would be the instrument of choice to get some answers. Regarding PT, in reference to Krull et al. (2018), it would be interesting to measure the degree to which students with special needs feel integrated after an intervention including PT. Plus, since the storytelling intervention is a group intervention, it might also be interesting to see if this affects social integration. Also of central interest are the opinions of the tutors. One study by Vogel et al. (2007) gave an insight into the fact that tutors were unsure how to deal with learning difficulties and how to build good relationships between tutors and tutees. Thus, from my point of view, tutor training is urgently needed to make the PT more transparent and teach the tutors how to be effective.

Referring to the transfer debate (e.g., Huemer et al., 2008), a further implication is to look at whether incorporating the learned words into a text helps students read the text faster, or to measure the effect on text reading fluency in general where trained words are not explicitly incorporated. Furthermore, with regard to letter-cluster training in Study 3, one could also try to promote at syllable or morpheme level (e.g., Beyersmann et al., 2019; Heikkilae et al., 2013; Mueller et al., 2020). A study by Mueller et al. (2020) found significant gains in German second graders in phonological recoding, word recognition and text-based reading following a word reading intervention with the focus on syllables.

Despite focusing on lower-level literacy skills, one other goal for the future would be to go beyond these lower-level skills and create other interventions on higher levels, such as reading comprehension for students who are proficient on lower levels but need help with more complex reading processes. Focusing on reading comprehension also means making sure that the sample is stable in lower-reading skills, such as word recognition. This is important because, e.g., a study by Mueller et al. (2015) has revealed that students with poor word recognition did not profit from a reading strategy intervention focusing on reading comprehension.

Also, L1 vocabulary has an impact on reading in L1 (e.g., Quinn et al., 2015). Although we did partly measure L1 vocabulary, it would be great to go into more detail and create studies focused on the relationship of L1 vocabulary and L1 literacy proficiency as already shown by, e.g., Dong et al. (2020). Another important implication is to measure RAN prior to conducting reading interventions to find a potential explanation for why some student might not benefit or might profit less. Since RAN is a significant factor in German and general reading (Chen et al., 2021; Kirby et al., 2010; Landerl & Wimmer, 2008; Swanson et al., 2013), a prior testing would give important indications. For example, one could implement the Test to Measure

Phonological Awareness and Naming Speed (THEPHOBE & Mayer, 2020) or the Züricher Lesetest II for 1st to 8th grade (ZLT-II, Petermann & Daseking, 2013).

Regarding pretest measurements, motivational screens can be used to estimate general motivation to learn or read. Also, the implementation of this questionnaire prior to and at the end of an intervention could give insights to any changes in motivation. As motivation acts as a good predictor for learning process, even after controlling for cognitive variables (Lenhard & Lenhard, 2013; Gold, 2011), and motivation is decreasing in German readers (OECD, 2019), embedding motivational screens in studies would establish important implications (e.g., Scales for Recording Learning and Achievement Motivation [SELLMO], Spinath et al., 2012). Referring to the criticism that we did not compare our method to other reading interventions, it is definitely a future goal to compare the method evaluated here with other possible (combined) methods to find out which method is best suited. Also, considering the indication that the combined RT works beyond small-group instruction, future studies can have a look at how students in small groups benefit compared to students in full classrooms.

A final implication concerns L2 learners, more specifically concerning German as an L2 and English as an L2, which are immensely important in Germany. There is an enormous number of students in Germany who learn German and English as an L2, and some of them have great difficulty, especially in L2 literacy (e.g., Wendt et al., 2016). These final considerations lead directly to the second part of my dissertation, which explicitly focuses on L2 learners and a specially designed intervention

3. Second Language Support for Students with Learning and Behavioral Difficulties

3.1 Second Language Learning (L2)

The global importance of learning an L2 is steadily increasing, and the number of L2 students is rising worldwide (Borodkin & Faust, 2014). In Germany, the growing number of multilingual students learning German as a second language (GL2) and English as a foreign language is rising (Schleicher, 2019; Wendt et al., 2016). Students with GL2 do not always reach an adequate level of German-language proficiency so that they can keep up in school, which can lead to increased learning difficulties affecting all school subjects (Jeuk, 2018; Schwab et al., 2013). Despite the challenge of acquiring adequate German literacy skills, these students are exposed to English, which is widely taught in German mainstream schools and can constitute further challenges (e.g., Molnár, 2008; Oxford, 2017). A lack of proficiency in English constitutes great problems for students regarding their future job chances and social involvement (Young et al., 2019).

The terms “L2” and “foreign language” refer to languages that are not the native language (L1). The difference, however, is that a second language is one that is acquired in a country where that language is spoken, while a foreign language is one that is not officially spoken in the country where it is learned (Dewaele, 2011; Jeuk, 2018; Oxford, 2017). Nevertheless, most research mainly uses the terms “second language” and “foreign language” identically, referring to both as L2 (Bisson et al., 2021; Channa et al., 2017). Based on this, the term L2, in this dissertation, includes any language being learned after the L1 (see also Lorette & Dewaele, 2015).

The acquisition of an additional language can be challenging due to the cognitive demands involved (Molnár, 2008). According to Cummins (2000), the L2 develops from the L1. This may mean that if students have not adequately developed their L1, these difficulties might also show up in the acquisition of an L2 (Sparks et al., 2008a; Sparks et al., 2008b). In their meta-analysis comparing L1 and L2 skills, Melby-Lervåg and Lervåg (2011) found a small correlation in oral language performance as well as medium to large correlation in phonological awareness and decoding (see also Bisson et al., 2021).

Students whose L1 is not the language used in school are challenged with becoming proficient in the language of school instruction while learning all of the curricular content, which places enormous hurdles in front of students attempting to meet curricular standards (August & Shanahan, 2006). Teachers are confronted with the challenging task to provide support to these students. In most European countries, about 10% of 15-year-old students do not speak the language of the country as their L1 (OECD, 2010). Students using an L2 also perform lower in L2 literacy than their peers using an L1, which is mostly based on differences in proficiency in the language of instruction (Wendt & Schwippert, 2017). A study by Schroeter and Schroeder (2017) revealed that German L1 and L2 secondary school students differ with regards to their linguistic skills (e.g., vocabulary) and text comprehension. The realization that learning an L2 can be quite complicated and, more importantly, the significance of understanding and using the language of instruction sufficiently led to the question of what components in the L2 are important to promote, particularly in literacy.

Vocabulary

According to Wilkins (1972), very little can be taught without grammar - nothing can be taught without vocabulary. To become proficient in a language, vocabulary is undoubtably one of the most important components to be fostered (Schmitt, 2008). Students with specific learning needs face particular challenges acquiring vocabulary (Gentry & Lindsey, 2008). Receptive and expressive vocabulary knowledge (Schmitt, 2014; Stuart, 2005) is important for reading comprehension—a significant aspect of becoming a literary adult (Holahan et al., 2018). A strong positive correlation between vocabulary knowledge and reading

comprehension was reported in a study by Stæhr (2008; r = .83) among English L2 learners with Danish (transparent orthography) as L1. Moreover, Stæhr (2009) showed a correlation of r = .70 between vocabulary knowledge and listening comprehension. To conclude, L2 vocabulary can predict growth in L2 reading comprehension over time (Lervåg & Aukrust, 2010; Schmitt et al., 2011) and limited amount of L2 input reduces the impact that the frequency and range of occurrence of words have on vocabulary learning in the L2 context. When an L2 is not constantly spoken at home, teachers may have the greatest influence on L2 vocabulary learning (Laufer, 2003). Sparks and Luebbers (2018) see an urgent need to focus more on vocabulary instruction in L2 learning (see also Harmon et al., 2009; Schmitt, 2008).

Reading

L2 reading difficulties are as heterogenous as L1 reading difficulties (Kato, 2018) and can be specifically challenging (Lee et al., 2020). Nevertheless, being able to read in an L2 is important as well (Anderson, 2015; Grabe, 2009). According to Melby-Lervåg & Lervåg (2014) students who start reading in their L2 often fall behind their L1 peers. In general, word reading in L2 is slower than reading words in L1 (Cop et al., 2015a), showing L2 readers to display greater amounts of fixations while reading (Whitford & Joanisse, 2018). In an eye-tracking study comparing L1 and L2 readers, Cop et al. (2015a) found that non-native speakers had more fixations and less word skipping in their L2 than L1. These findings indicated that more effort is required to read in an L2. Whitford and Joanisse (2018) found larger word frequency effects in L2 compared to L1 among bilingual children. Larger word frequency effects in L2 were also found for adults (e.g., Cop et al., 2015b). A main reason why L2 learners might struggle with reading is their lack of word recognition skills (Grabe & Stoller, 2011). A meta-analysis by Jeon and Yamashita (2014) found a correlation of L2 decoding (r = .56) and L2 vocabulary (r = .79) with L2 reading comprehension. Thus, both are strong predictors of L2 reading. The L1 also plays a specifically significant role in L2 reading (Akbari, 2017). For example, van Gelderen et al. (2007) found a significant impact of L1 reading on L2 reading for English adolescent learners with Dutch as their L1. PA also has an impact on reading acquisition, as mentioned in the first part. While PA has been found to be more prevalent in English than German reading, it is still of significance when acquiring German language skills (Schulte-Körne, 2011; Ziegler & Goswami, 2005). L1 and L2 readers are similar regarding RAN, working memory, and phonological awareness (Geva et al., 2000; Peng et al., 2018; Raudszus et al., 2018; Swanson, 2015).

Now that important components of L2 acquisition have been outlined, the following question again addresses L2 students with learning and behavioral difficulties, but this time with respect to L2 acquisition.

3.2 Language Learning with Behavioral and Learning Difficulties

Referring to the samples of the studies to be presented below, learning difficulties in the first three studies of the second part refer to the German language difficulties and not automatically to the participants' L1. The same goes partly for the English L2 studies. Here, students with an official LD diagnosis, which is referred to the German language students, display additional severe difficulties in learning English as an L2—all other participants with learning difficulties are labelled solely according to their English performance. However, all subjects in the studies included in the second part have enormous difficulties in L2 acquisition and thus, L2 learning difficulties. To date, it is not possible to say with certainty whether all probands also have learning difficulties in L1, even though research shows that there is a strong relationship and similar processes are involved (e.g., Abu Rabia & Siegel, 2002; Raudszus et al., 2018; Sparks et al., 2008a; Sparks et al., 2008b; Sparks et al., 2009; Swanson, 2015; von Hagen et al., 2021). Nevertheless, in reference to the dissertation's definition of learning difficulties, all students with serious learning hurdles are supported regardless of an official diagnosis or more specifically for the second part, regardless of additional known challenges in L1.

All in all, there is less research on L2 learning compared to L1, specifically with respect to LD (Hall et al., 2019). However, even though there is not yet enough evidence that students with LD automatically have problems in learning an L2, this fact is simply accepted in the field of education (Sparks, 2016). Sparks and Patton (2016) discussed that having severe difficulties in learning does not automatically mean that problems in the L2 must also be present. Research has shown that students with a diagnosis of LD do not necessarily face more academic difficulties in L2 courses than their peers without LD (Sparks, 2006; Sparks & Javorsky, 2000). According to von Hagen et al. (2021), even though less-proficient L1 students seem to have a higher risk of failing L2 acquisition courses (see also Maurer et al., 2021 for German students learning an L2), practitioners and researchers should not rely on the common belief that students with general learning difficulties are more likely to fail (Sparks, 2016).

Referring to general L2 difficulties, specifically, vocabulary acquisition seems to be a hurdle (Gentry & Lindsey, 2008). Vocabulary contributes significantly to the fact that L2 students display less proficient reading skills compared to their L1 peers (Geva & Farnia, 2012). This might be one explanation for overall L2 challenges. Generally, L2 students with learning difficulties show similar problems in WM and phonological processing to L1 students with learning barriers (e.g., Abu Rabia & Siegel, 2002). Further, common hurdles can also be seen in decoding skills as well as comprehension (Geva & Massey-Garrison, 2013). As Sparks (2009) claimed, teachers should offer appropriate methods for struggling L2 learners to be successful.

Continuing with behavioral difficulties, there seems to be an association between behavioral and language hurdles (Chow & Wehby, 2018; Hollo et al. 2014; Jansen et al., 2020). Participants with behavior challenges were labelled according to the ITRF-G results in the second part of this dissertation. One reason for the relationship between challenging behavior and language deficits might be that poorer language abilities are seen as a risk factor for the development of behavioral problems (Chow & Wehby, 2018; Hollo et al., 2014; Petersen & LeBeau, 2021). In their meta-analysis, Chow and Wehby (2018) found a significant negative relationship between language acquisition challenges and behavioral problems across age and time. Furthermore, Petersen and LeBeau (2021) stated that social and language skills and behavior difficulties are interrelated. They found that language ability might be correlated to later externalizing behavior problems with a modest effect size in teacher and parent behavior ratings. According to Keenan and Shaw (2003) challenging behavior can also result from being stressed due to language deficits. Moreover, it is rare to find adequate research focused specifically on language learning in students with learning and behavioral difficulties. In conclusion, research in L2 acquisition with regard to students who need additional support is expandable (e.g., Kormos, 2017). Unfortunately, the necessary attention has not been paid to this area so far. This fact and the previously explained facts finally led me to the question how effective L2 support can look like and what the research suggests.

3.3 Effective Language Learning - What Research Suggests

Even though it is assumed that L2 students with learning issues face the same hurdles as L1 students with learning difficulties (e.g., Linan-Thompson et al., 2016) one must take into account that interventions that are effective in the L1 can fail in the L2 (e.g., Moore & Klinger, 2014). Therefore, according to Moore and Klinger (2014), one should be cautious when choosing L2 interventions. However, some methods that are effective in the L1 have been found to have positive effects on L2 proficiency, but studies have rarely looked at specific sub-groups. The Matthew effect discussed earlier has also been demonstrated for language learning, such that the more advanced L1 abilities one has, the further L2 abilities will be developed (Bisson et al., 2021). Students who are learning German specifically as an L2 still lag behind their L1 peers, showing that educational institutions seem to struggle in providing adequate L2 education (Dixon et al., 2012) which is still true regarding the high proportion of L2 learners who currently struggle with L2 literacy (e.g., Harju-Luukkainen et al., 2020).

One of the most urgent demands in L2 support is the focus on vocabulary, especially in students at different language levels and with various learning difficulties (Filippini et al., 2012; Jitendra et al., 2004; Spies & Dema, 2014). Going a step further, phonics instruction is beneficially used in L2 for at-risk learners (Vadasy & Sanders, 2012) as, e.g., GPC (Hulme et al., 2012; Stuart, 2004) which is even more effective in L2 learning than in L1 (Ginns et al, 2019). Referring to motivation, researchers have found that correlations of motivation together

with aptitude in language proficiency show values mostly above r = 0.50 (Doernyei & Skehan, 2003). Considering the setting, for students with or at risk for LD, small groups are beneficial because more opportunities are provided for L2 practice (Baker & Kosty, 2012). Following von Hagen et al.'s (2021) argument that L1 reading skills can negatively affect L2 performance, the authors nevertheless argued that there is an urgent need for adequate methods for L2 learning as opposed to denying lower-performing students adequate opportunities to learn an L2. Besides the effective methods already explained in the first part of the paper in the context of L1 acquisition, I was also interested in findings related to L2 support.

Verbal and Nonverbal Input

Verbal and nonverbal input which is said to be immense important in L2 is discussed in the dual coding theory (DCT; Paivio, 2007). DCT is a cognition theory that can be applied to literacy learning. This theory posits that one can access memory verbally and nonverbally (Paivio, 2007). The verbal system is related to linguistic information (e.g., sound) and the nonverbal system to visual information (e.g., pictures; Paivio, 2014). Accessing both systems simultaneously is more beneficial than assessing only one system at a time, which is also seen in special education (Crosson et al., 2019; Kuder, 2017; Reed, 2010). Referring to vocabulary learning, words are better stored when taught through both systems (Paivio, 1991).

Concerning the nonverbal mode, according to Rieber (1994), using illustrations to learn words makes it easier to memorize them (see also Carpenter & Olson, 2012; Sadoski, 2005; Wright, 2010). Morett (2019) found that using images to teach vocabulary in an L2 is beneficial in adolescent learners. More precisely, implementing visual aids in L2 learning practice is seen as highly beneficial for students with and without LD (Richards-Tutor et al., 2016). Other nonverbal representations, such as gestures, have also been revealed to be effective in L2 word learning (Kelly et al., 2009; Macedonia et al., 2014; Morett, 2014; Porter, 2016). According to Klingner et al. (2014), reading instruction in L2 for students with and without academic challenges should entail audio and visual support (e.g., Pellicer-Sánchez et al., 2020).

Intentional and Incidental Learning

There are two main ways of learning a language: incidentally (learning as a by-product) and intentionally (being aware of the learning process) (Ellis et al., 2009; Webb & Nation, 2017). According to Peker et al. (2018), vocabulary is better remembered when incidentally learned through context compared to solely direct delivery of vocabulary through intentional learning (see also McKeown et al., 2018). However, other researchers state that direct intentional vocabulary instruction, for struggling L2 learners in particular, has been reported to be highly effective (e.g., August et al., 2009; Carlo et al., 2004; Lesaux et al., 2010; van Zeeland & Schmitt, 2013; Vaughn et al., 2009), which can be linked to secondary students (Lesaux et al., 2014) and younger students (Lugo-Neris, et al., 2010). In addition, the depth of vocabulary

knowledge can be enhanced when also learning additional facts about each word (Schmitt, 2014). A meta-analysis by Yousefi and Biria (2018) examining the effects of intentional L2 vocabulary instruction in language learning on L2 vocabulary and found effect size of d = .80. Intentional vocabulary instruction is the one practice that is most often used and established to be effective, especially for struggling students (Archer & Hughes, 2011; Hughes et al., 2017; Kuder et al., 2017; McLesky et al., 2017) to not only foster vocabulary acquisition but also L2 reading acquisition and PA (Gyovai et al., 2009). To conclude, intentional vocabulary instruction has been revealed to be more effective in L2 learning than incidental instruction, as found by a meta-analysis by Goo et al. (2015).

A meta-analysis by Maurulis and Neuman (2010) about the impact of vocabulary interventions on the language development of preschool and kindergarten children might give important insights into L2 learning. An overall effect size of *g* = 0.88 of vocabulary training on word learning was found. Moreover, a combination of incidental and intentional training led to a higher effect size (*g* = 1.21) than intentional (*g* = 1.11) and incidental (*g* = 0.62) training in isolation. Even though this study examined very young children in the L1, a combination could potentially be beneficial for older students concerning an L2.

Based on these findings, the question arose how DCT and incidental as well as intentional learning can be implemented in an intervention. Based on previous knowledge, it was assumed that a combination of intentional and incidental training as well as the implementation of verbal and nonverbal aspects is beneficial in the context of L2 with the group of students focused in this dissertation. Thus, the idea of creating a storytelling method arose.

Storytelling

Using stories in education has a long tradition. Particularly in L2 learning, stories are used to convey knowledge and enhance language proficiency (Li, 2007; Mol & Bus, 2011; Wells, 2009) and reading stories aloud to students has been proven beneficial in fostering language skills (Isbell et al., 2004; Mello, 2001). Moreover, listening to stories and talking about story content can enhance the motivation to read in L2 children (Cruz de Quiró et al., 2012). As a general statement, embedding stories into a learning environment can enhance literacy skills in students with profound and multiple learning difficulties (Fornefeld, 2013).

In L1 research, a study by Joffe et al. (2019) found positive effects on vocabulary for a combination of narrative intervention and vocabulary-instruction interventions in secondary students with language difficulties. Storytelling has also been found to be effective in German L1 primary school students (Suggate et al., 2013). Positive effects (specifically on vocabulary in young L1 children) were documented in a meta-analysis of interactive read-alouds by Mol et al. (2009), whereas expressive vocabulary (d = 0.62) and receptive vocabulary (d = 0.45) revealed small to moderate effects.

Regarding L2 acquisition, the use of stories in L2 learning is also important when learning vocabulary (Calderón et al., 2005). A study by Chlapana and Tafa (2014) examined the effects of story reading on L2 vocabulary in children 4–6 years old. The results showed the largest effects for the group in which stories were read aloud and target words were interactively discussed, compared to conditions with only brief or completely missing explanations of target words (see also Hickman et al., 2004). A meta-analysis by de Vos et al. (2018) found an effect size of g = 1.05 for spoken input on L2 word learning in kindergarten, primary, and secondary school students. Dickinson et al. (2018) found encouraging results of engaged book reading on vocabulary in preschoolers from various language backgrounds. Moreover, Calderón et al. (2005) stated that introducing new words before the story starts is easier for students, and it has been shown that exposure to printed words enhances L2 word learning (Chambrè et al., 2020). More far-reaching, a study by Kirsch (2016) observed positive effects of storytelling on vocabulary in English language learning for ages 10–11. Also, reading story while listening to it being read aloud has also been revealed to be effective for acquiring vocabulary and reading (Webb & Chang, 2012), compared to reading only or listening only (Chang & Millett, 2014; Webb & Chang, 2015). Reading while listening has been found to be effective for fluency, comprehension, and vocabulary in L2 learning (Chang & Millet, 2014; Webb & Chang, 2012). A meta-analysis by Fitton et al. (2018) discovered positive overall effects ($g = 0.28$, $p < .001$) for shared book reading (the reading of a book to a child by an adult using additional interactive techniques) on English L2 children's language outcomes. To end, combining GPC learning with book reading is advantageous (Shapiro & Solity, 2016) compared to phonics treatment without context (Yeung & Savage, 2020).

Storytelling developed for the studies in this dissertation has never been evaluated before, making it difficult to refer to proper studies and effects for comparison. It was oriented on the storytelling process by Roney (1996) that stories are told by a narrator communicating with the audience by using pronunciation, narrative structures, and mental imagery, and the audience responds to the story verbally and nonverbally. Additionally, storytelling was combined with practices suggested by research on effective language teaching. In general, storytelling seems to be described differently in each research study, since there is no universal definition. Nevertheless, some studies provide an indication of the effectiveness of different aspects used in the storytelling process developed within the framework of my qualification phase. Story interventions have been conducted in L1, mainly for kindergarten children, and in L2 for English language learners. Nevertheless, these studies suggest that storytelling might also work in GL2. However, to date, no study has been found based on storytelling in GL2; moreover, no study has used storytelling in a way similar to the process used in the presented studies of the dissertation.

3.4 Interim Conclusion

Richards-Tutor et al. (2016) conducted a research synthesis of the effects of literacy interventions for L2 learners from 2000–2012 for students with or at risk for LD. They concluded that studies with L2 learners with or those at risk for disabilities are rare compared to L1 studies. Also, they suggest that an intervention that works for some students might not work for others. Nevertheless, regarding literacy instruction, what works in an L1 can also work in an L2 (Ludwig et al., 2019; Schaars et al. 2019). However, interventions should be planned according to students' needs (Connor et al., 2014; Kamil et al., 2008), as stated in the first part of the dissertation for L1. As there is an urgent need for L2 interventions and a large research gap in adequate studies on learning and behavioral difficulties and L2 learning, the following studies are intended to contribute to closing this gap further. Based on the DCT, findings on intentional and incidental learning, and study results on reading aloud and storytelling, the specially designed storytelling method seems to be suitable to support students in L2 literacy. Through a combination of evaluative motivational components and explicit learning characteristics, the method also seems suitable for students with difficulties in learning and behavior. Based on this, in the following, I explore studies that refer to the use of the combined storytelling method in GL2 and English as L2 with heterogeneous student groups in primary and secondary education.

3.4.1 Summary Article 4 *(peer-reviewed)*

Barwasser, A., Urton, K., Knaak, T. & Grünke, M. (2021_accepted). *Intentional and incidental vocabulary acquisition through multi-component storytelling. The case of German L2 primary school students. Language Teaching Research.*

Vocabulary learning is one of the most important aspects of L2 acquisition and has strong influences on the development of adequate L2 literacy. At the same time, it is one of the biggest obstacles in L2 acquisition (Gentry & Lindsey, 2008; Holahan et al., 2018; Schmitt, 2008; Sparks & Luebbers, 2018). The combined storytelling intervention was first conducted in an elementary school with students in grades 2–4 with German as L2; these students struggled in German language use, and some had behavioral difficulties. We intensively trained vocabulary, which was measured receptively and expressively by the subjects, and asked for additional knowledge about the vocabulary that was not explicitly trained.

Methods

The participants were chosen based on their German L2 status and a German vocabulary test, for which scores <15th percentile were chosen to find the final sample. There were 9 final participants whose native languages were Turkish, Arabic, and Albanian. Three of

the participants showed externalizing behavioral problems. We used a multiple baseline design to estimate the intervention's effectiveness compared to the baseline condition (Lane et al., 2017). The children were divided into three groups, each having a different baseline length. The procedure was held three times a week over four weeks, with about 40 minutes for each session. A word pretest was implemented to establish the final words for training receptively and expressively. After each session, starting with the first baseline session, data was collected from each participant in terms of 1) words known receptively, 2) words known expressively, and 1) vocabulary facts. The measurements were based on the Peabody Pictures test (Dunn & Dunn, 2007); vocabulary facts were measured by asking specific questions about the words, and the children were asked to answer expressively. The baseline sessions were about cognitive tasks to be solved, and the intervention was divided into 2 stages: 1) direct instruction of 10 of the 20 training words while using flashcards with the word and a matching picture; 2) the storytelling itself in which words are embedded. The stories were adapted to the students' interests and told according to memory. When a training word appeared, the flashcards were taken, and the children were asked to remember what the word meant. In addition, treatment fidelity was implemented.

Results

The SCAN package by R was applied for data analysis. The NAP, PEM, and Tau-U to estimate the intervention's effectiveness were implemented. Further, we also conducted a piecewise regression analysis on level 2. For receptive vocabulary, steady baselines were visible and overlap indices showed moderate to large changes, which are statistically significant ($p < .05$-$p < .001$. The regression analysis yielded a significant level ($p < .001$), and slope effect ($p < .001$) was shown across all groups, with an average increase of 1.496 words per intervention session. For expressive vocabulary, overlap indices also revealed moderate to very large changes that can be seen as significant ($p < .05$–$p < .001$). Regression analysis showed the same results as for receptive vocabulary with an average increase of 1.688 words. Regarding vocabulary facts, overlap indices indicated moderate to very large changes ($p<.05$-$p<.001$), which are all significant. Regression analysis revealed the same results as the variables before and an average increase of 1.108 words per session.

Conclusion

As stated, vocabulary is very important, especially for reading (e.g., Holahan et al., 2018). Overall, for this study and these subjects, the receptive and expressive vocabulary that was explicitly trained were enhanced through the combined storytelling intervention, and the vocabulary facts that were mentioned in the stories without explicit training could also be retrieved. The results are promising, and all three dependent variables could be trained simultaneously with the intervention. There is no visible difference between the students

regarding their reaction. Also, problem behavior did not play a leading role here or in general L2 vocabulary knowledge. Unfortunately, L1 abilities were not measured. This may have determined explanations for certain results (e.g., Peng et al., 2018; Raudszus et al., 2018; Swanson, 2015). Also, no gender or age differences were visible. Thus, the results indicate that this intervention is suitable for a heterogeneous student population in L2 learning. The next question to be raised is whether word reading, which was the focus of the first part of the dissertation, can also be improved in L2 learning with storytelling in addition to vocabulary. It was of additional interest to find out how older students react to storytelling.

3.4.2 Summary Article 5 *(peer-reviewed)*

Barwasser, A., Lenz, B., & Grünke, M. (2021). A multimodal storytelling intervention for improving the reading and vocabulary skills of struggling German-as-a-Second Language adolescents with learning and behavioral problems. *Insights Into Learning Disabilities, 18*(1), 29–51.

Introducing the next study and looking back at the previous results, the storytelling intervention is effective in training German L2 vocabulary for younger students with and without behavioral or L2-learning difficulties. The next study examined the extent to which the storytelling intervention had a positive impact on word reading, in addition to vocabulary, in German L2 secondary school students with learning disabilities and behavioral challenges. The idea developed from the fact that L2 learners often have problems in reading and especially in word reading and showed more word fixations in eye tracking studies than L1 readers (e.g., Cop et al., 2015a; Melby-Lervåg & Lervåg, 2014; Whitford & Joanisse, 2018). Knowing that direct word recognition also plays a significant role in L2, word reading was evaluated as another dependent variable. Thus, the research aim was to evaluate if a combined storytelling intervention leads to an increase in expressive vocabulary and reading with respect to the described older student group.

Methods

The study used a multiple baseline design across participants (Lane et al., 2017) with three groups each, and with four students getting three different baseline lengths. The original participant group entailed nine students, whereas the current study only displayed four of them due to missing data points, which happened because of the COVID-19 pandemic. The four participants, two males and two females, were diagnosed with an LD, Turkish was their L1, and all scored poorly on the ITRF screening. Also, all subjects performed less-proficient in reading German words and German vocabulary. In addition, they were not able to read the chosen 30 words of training for the interventions. The words were crystallized through a word

pretest using PowerPoint with slides on which the words were written in a one-second rhythm (see Ehri, 2005). The study took place three times a week for 30 minutes over a four-week period. The dependent variable 1) “words read correctly” was assessed using the same procedure as in the pretest, but only containing the final words, which were randomly displayed to the students each time after the baseline and intervention sessions, asking them to read the words aloud within the rhythm. The other dependent variable, 2) “amounts of known words expressively” was assessed also using PowerPoint and embedding the matching pictures to the words on the slides. The students were asked to name the word. The first variable pool had 30 words and the second had 40. Both variables had the same words, but the expressive vocabulary had 10 more. The baseline was conducted using mathematics exercises and there were two phases for the interventions. For the sessions, self-written stories were used where the training words were embedded in. In each session, 10 of the pool were trained and directly instructed to the students with flashcards in phase 1 and trained in storytelling in phase 2. Storytelling consisted of the interventionist telling the story by heart while the students could also read the shown story—words embedded in the story were again trained with the flashcards. Treatment fidelity was additionally considered.

Results

Again, for the data analysis of this single case study, the SCAN package in R was used. The mean baseline difference (MBD; Campbell, 2003) was calculated by hand. Overlap indices PEM, NAP, and Tau-U were applied as well as a piecewise regression analysis level 2 across all students. Results showed a very rapid increase of both variables in all participants. Only one participant started with a high value average in the A phase but could display only an increase in the baseline phase, which can be visually seen through the MBD, as well as the overlapping indices. For expressive vocabulary, an increase of up to 1000% was found from baseline to intervention and moderate to large effects were seen across the participants, which are statistically significant ($p < .05$–$p < .001$). In the regression analysis, a significant level effect ($p < .001$) and a significant slope effect ($p < .05$) were shown. With regard to reading, the MBD was slightly lower compared to expressive vocabulary (up to 390%); however, overlapping indices showed a large change in all participants with statistical significance ($p < .05$). Again, a significant level effect was observed ($p < .001$) as well as a significant slope effect ($p < .001$) across all students.

Conclusion

As the results indicate, the intervention seemed to be effective on expressive vocabulary and sight word reading in GL2 secondary students with learning and behavioral difficulties. Thus, the results are in line with previous studies on fostering specific aspects in L2 learning (e.g., Maurulis & Neumann, 2010; Peters, 2014). Mattina showed slightly lower scores on

expressive vocabulary, which could be a result of her low performance on the general German vocabulary test (e.g., Alderson, 2005)). Unfortunately, due to the COVID-19 pandemic, social validity questionnaires could not be handed out and generally, data points are missing. Going a step further, it would be interesting to know how many variables storytelling could further affect positively simultaneously. In addition to reading and vocabulary, which can be successfully fostered through storytelling, looking at enhancing letter sound fluency would be beneficial. Even though, direct word recognition is more important in German (Landerl & Wimmer, 2008), the sub-lexical route does also play a role, especially for beginning readers (Karageorgos et al., 2020). Thus, it was assumed that it also might be beneficial in German L2, especially referring to the assumption that phonics training is more effective in L2 learners and for at-risk learners (Vadasy & Sanders, 2012)—more specifically, GPC training (Hulme et al., 2012; Stuart, 2004).

3.4.3 Summary Article 6 *(peer-reviewed)*

Barwasser, A., Bracht, J., & Grünke, M. (2021_accepted). A storytelling approach on vocabulary, reading and letter sound fluency of struggling first graders with German as Second Language with and without behavioral problems. *Frontiers in Psychology*

The following study is also based on the previous one and explores whether GPC also can be promoted, in addition to the effects on vocabulary and word reading. Even though automated word reading in German has been shown to be more important than PA (e.g., Landerl and Wimmer (2008), it is still recommended not to disregard it completely (e.g., Schulte-Körne, 2011), especially because phonological support in an L2 may be much more effective than in an L1 (e.g., Ginns et al., 2019). Thus, the research aim was to evaluate the multicomponent storytelling intervention's effects on expressive vocabulary, GPC, and reading in German L2 students from first grade with and without behavior problems.

Methods

A multiple baseline design was used to evaluate the intervention's effects (Lane et al., 2017) with three different group and starting the intervention on different days. The procedure took place three times a week for 25 minutes lasting, a total of 5 weeks. The participants (N=7) were selected based on a German vocabulary test, reading screening, and a screening for phonological skills showing below average performances. The children were aged 6–7 years and had German as their L2. The dependent variables were: 1) expressive vocabulary, 2) word reading, and 3) GPC. A total of 40 words of training were used and evaluated. The participants did not know the words before the intervention. For baseline conditions, the same content was used as for the previous studies as well as the storytelling intervention. The only difference

was that here the sounds were additionally trained. Treatment fidelity and social validity were implemented.

Results

The data was analyzed using the SCAN package in R. For further analysis beyond visual inspection and descriptive data, the overlap indices NAP, PEM, PAND, and Tau-U were used. In addition, a level 2 regression analysis was applied. The results for expressive vocabulary showed large significant changes for all participants ($p < .01$–$p < .001$). There was a slope effect for each group ($p < .05$–$p < .001$) and across groups ($p < .001$). The intervention also positively influenced letter sound fluency and revealed large effects for all children; only Tau-U showed a moderate effect for Niek. In the regression analysis, a slope effect was not reported. For the dependent variable reading, the results indicated moderate to very large changes with a significant slope effect for all but group 1 ($p < .05$). Social validity results implicated that the participants overall rated the intervention positively. Only two children rated "partly agree" in terms of reading, and one child rated "partly agree" with respect to self-graphing.

Conclusion

In summary, the combined storytelling intervention was successful in fostering the three evaluated variables simultaneously for the presented sample. Thus, results are consistent with previous findings of L2 support (e.g., Donegan & Wanzek; 2021; Maurulis & Neuman, 2010) and specifically with findings concerning GPC training in L2 (Shapiro & Solity, 2016; Yeung & Savage, 2020). Interestingly, problem behavior seems not to have had a negative impact on the intervention's effectiveness, as seen in part 1, even though students with problematic behavior might not react positively to a literacy intervention (Forlin & Chambers, 2011). However, as research has shown, vocabulary knowledge does play a role in word learning (e.g., Jeon & Yamashita, 2014). First-grade phonological skills could have played a role, and RAN – but RAN unfortunately was not measured before. Overall, the study showed that an L2 intervention that focuses on three variables and is adapted to students' everyday lives could indeed foster characteristics that make learning more difficult.

Now that the intervention has been evaluated several times in German as an L2, I wondered if the intervention was also suitable for other languages. In particular, the following study is on English as an L2 due to its global relevance (Young et al., 2019) and everyone should have the right to be taught in English. Another factor why English is the next focus is with respect to the dissertation's sample, as it is often argued that an additional language overburden students with an LD, and that it would be very difficult to teach English, especially to struggling students and students with an LD (Sparks, 2016; von Hagen et al., 2021).

Personally, I am resentful about this prevailing situation because I am convinced that, above all, it depends on how students are supported.

3.4.4 Summary Article 7 *(peer-reviewed)*

Knaak, T., Gruenke, M., & Barwasser, A. (2021). Enhancing vocabulary recognition in English foreign learners with and without learning disabilities: effects of a multi component storytelling intervention approach. *Learning Disabilities: A Contemporary Journal, 19*(1), 5–23.

Because it is extremely important for all students to learn the English language (Molnár, 2008; Oxford, 2017; Young et al., 2019) and since many assume that students with learning difficulties automatically have problems in L2 learning (Sparks, 2016), the next study aimed to look at how adolescent students with and without learning and/or behavioral difficulties benefit from English support through storytelling. In addition, we are filling in the large research gap of further knowledge regarding L2 learning and students with special needs. Thus, the research aimed to evaluate the multicomponent intervention using combined storytelling with group contingencies and self-graphing with regard to helping students with and without learning disabilities acquire English vocabulary.

Methods

The participants (N = 24) were secondary school students with and without LD who were chosen based on a researcher-developed English vocabulary test. The students were divided into six groups for baseline and intervention sessions. A multiple baseline design was applied across participants with different baseline length (Lane et al., 2017). The baseline and interventions sessions were held three times a week for 35 minutes over six weeks. Prior to the study, a vocabulary test of common English words was conducted to decide on the final words for the storytelling intervention. Out of this testing, we obtained 30 training words that functioned as dependent variables. The students received a sheet with English words on the left-hand side to translate into German on the right-hand side. For the baseline condition, students were engaged in cognitive tasks followed by a measurement of the dependent variable. For intervention, storytelling was introduced in the same two-step process as the prior studies. The 30 training words were taught using self-written stories. Additionally, self-graphing and group-contingency procedures were implemented at the start of the intervention phase. As in the baseline phase, the dependent variable was again measured after the intervention sessions. Additionally, the follow-up measurement was conducted. Treatment fidelity was also embedded.

Results

For data analysis, the SCAN package in R was applied to estimate the intervention's effectiveness. NAP, improve date Difference (IRD, Parker et al., 2011a) and Tau-U were applied. Moreover, a regression analysis was conducted both per group and across all groups. The results showed a sharp increase in the numbers of known words in each group, even though some variance was observed. Overlap indices indicated that the intervention was highly effective across groups (Tau-U: .96–1.00, p <.001; NAP: 1.00, p <.001; IRD: .96–1.00). The regression analysis also stressed this assumption by showing a statistically significant slope effect from phase A to B for all groups and across groups (p <.001) with an average increase of 2.26 known words per intervention session. Follow-up data was relatively stable.

Conclusion

Overall, the positive effects give some insight into the power of storytelling in English L2 learning for students with and without an LD. Group 5 and 6 showed a slower increase; however, this might be linked to the small number of intervention sessions. Above all, follow-up data indicated that the words could be remembered even after some time without intervention. One limitation might be that some participants were German L2 and thus, English L3, which could explain the varying data since the learning process gets more complicated with additional languages (e.g., Molnár, 2008). Further, we did not explicitly differentiate between students with and without an LD. Therefore, the next results (based on this study) emphasize the students with an LD and their reaction to the intervention.

3.4.5 Summary Article 8 *(peer-reviewed)*

Barwasser, A., Knaak, T., & Gruenke, M. (2020). The effects of a multicomponent storytelling intervention on the vocabulary recognition of struggling English as a foreign language learners with learning disabilities. *Insights into Learning Disabilities, 17*(1), 35–53.

Returning to Sparks (2016), students with LD do not automatically perform poorly in an L2—it probably depends a lot on how they are taught. Because of this, the next study data looks at the storytelling intervention's effects in English L2 on secondary students with an LD. Thus, the research aimed to evaluate the multicomponent intervention consisting of combined storytelling with group contingencies and self-graphing regarding helping students with LD acquire English vocabulary.

Methods

Out of a sample (*N* = 24), this paper focuses on four female secondary school students with LDs (age average 14.5 years). Also, the study was included based on a researcher-

developed English vocabulary test. A multiple baseline design across participants and groups was used (Lane et al., 2017). Jutta was in group 1, Lena and Pia in group 2, and Mia in group 3—all with a different baseline length from 4–6. To prevent repetition, I refer to the prior study where the measurement and baseline, as well as intervention procedure is described.

Results

As used in the previous studies, the SCAN package in R is also used here. We applied the MBD, and in addition NAP, percentage of nonoverlapping data (Parker et al., 2011), PEM, and IRD. Further, a regression analysis was used. All girls showed a great increase in the number of known English vocabulary words. Unfortunately, some measurements were missing due to absence of students in school. However, an increase was visible. Overlap indices revealed large effects for all participants, which are statistically significant ($p < .01$–$p < .001$). Regression analysis showed a significant slope effect for Jutta, Pia, and Mia. Lena, at least, displayed a significant level effect. The level 2 data indicates a highly significant slope effect across all girls with an average increase of 2.078 known words per interventions session. Follow-up data only showed a slight drop for Pia and Mia.

Conclusion

Overall, the storytelling intervention was effective with respect to students with LD, with relatively stable follow-up results. Although all baselines were relatively stable, there was an immense direct increase in the number of correctly known words for all students. Only one participant needed more sessions to reach the maximum word amounts. Nevertheless, there were large effects. The results indicated for English secondary school learners with an LD equal gains in vocabulary acquisition and is a first step to contradict the assumption that students with LD must exhibit problems in L2 acquisition (e.g., Sparks, 2016).

3.5 Conclusion

Looking at all five studies, both for the German and English L2, the results importantly indicate how students with particular difficulties in learning and behavior can be effectively supported in their L2 with a simple intervention in a very short time. Therefore, storytelling is a possible intervention in L2 acquisition, which is adaptable to individual students' needs.

Furthermore, the intervention fostered several important components in L2 (partly simultaneously). Because vocabulary knowledge is one of the most significant factors in language learning (Schmitt, 2008), it is very encouraging that receptive and expressive vocabulary can be enhanced through the combined storytelling intervention in English and German L2. In addition, since vocabulary contributes to reading, it is undoubtedly important in literacy (Stæhr, 2009). Further, L2 learners can struggle with word recognition (Grabe & Stoller, 2011). Two German L2 studies showed that the combined storytelling intervention could also

support automated sight word reading, thus contributing to a more rapid word recognition. This is specifically important since German is a transparent language in which naming speed (and thus, rapid word recognition) play important roles (Knoepke et al., 2014; Landerl & Wimmer, 2008). One study even fostered three different literacy components simultaneously in German L2: GPC, vocabulary, and word reading in young primary school students. Even though some say PA is not as important as rapid naming (Schulte-Körne, 2011; Ziegler & Goswami, 2005), I have decided to add this variable to the intervention with regard to the participants' age and the knowledge that PA does definitely influences younger learners (Knoepke et al., 2014). The study results are in line with reports that GPC trainings indeed are beneficial in L2 learning (Shapiro & Solity, 2016; Yeung & Savage, 2020). As shown Study 6, storytelling enhanced GPC in all participants, while expressive vocabulary and sight-word-reading also increased. Only two participants struggled with improving word reading. However, both had the lowest results on reading pretesting and one of the weakest results in vocabulary, which might have played a role here. As less-proficient readers rely on the non-lexical route (phonological recoding; e.g., Compton, 1997), they might have tried to read these words letter by letter instead of building an orthographic representation. This indicates that more sessions are probably necessary for these students and/or they might need more intensive training in PA, because phonological recoding is needed for building orthographic representations (e.g., Ehri & Robbins, 1992).

On a personal note, I am glad these results counteract the assumptions that struggling students automatically face difficulties in an L2 (Sparks, 2016; von Hagen et al., 2021). While they might face more challenges compared to students who are not academically challenged (e.g., von Hagen et al., 2021), they can also profit from L2 interventions, which is in line with Sparks (2016). Interestingly, according to the study by Maurer et al. (2021) examining the relation between German reading skills and English L2 outcome, it was concluded, that L2 difficulties were based on poor German reading overall – not based on being a German L2 learner. Thus, resulting again in the urgent need to focus on any student who struggle with literacy independent from any labelling.

In addition, with respect to behavioral problems, the studies that included students with behavioral difficulties revealed they reacted to the intervention the same as their peers without behavioral issues, even though behavioral difficulties are linked to language challenges (Chow & Wehby, 2018; Hollo et al., 2014). They might have benefitted from the small group size and/or storytelling enjoyable for them, motivating them to learn (Charlton et al., 2005; Roberts, 2020). Unfortunately, social validity questionnaires were only filled out for one study due to different reasons. Nevertheless, these students rated storytelling as very positive.

Overall, the results are in line with studies investigating intentional and incidental learning (Maurulis & Neuman, 2010; McKeown et al., 2018; Yousefi and Biria, 2018). However, the

studies discussed in this dissertation do not reveal to what extent intentional learning was more beneficial over incidental learning. Instead, the storytelling concept was partly based on Maurulis and Neuman's (2010) meta-analysis that states the significant advantage when combining intentional and incidental aspects. Moreover, in accordance with Kuder et al. (2017) and Paivio (2007), the positive effects of our combined storytelling could be from the verbal and nonverbal paths of learning a language according to the DCT (Paivio, 1991). Specifically, when accounting for WM, which plays an important role in learning, both parts of learning relieve the WM and the content is better stored (Paivio, 1991). Furthermore, storytelling as the main part also has a certain share in the effects, although to what extent is not proven. Returning to previous findings and studies on stories, storytelling and reading aloud, especially in the L2 context (Chlapana & Tafa, 2014; Hickman et al., 2004; Huang, 2006), the methodology was proportionally involved in the effects. Of course, combined storytelling cannot be compared with the existing storytelling methods because it is a combined intervention. Nevertheless, previous studies exploring stories have led to storytelling becoming the main method of the second part of this dissertation. Based on the results on a wide range of students and variables, there is strong indication that this combined intervention is exceedingly effective.

Limitations

Overall, there are some limitations I am aware of. First, as with the first part's studies, a larger sample size is needed to draw more general conclusions of the intervention's effectiveness or more single-case studies specifically with this intervention to call it evidence-based (e.g., see Tate et al., 2016). Furthermore, the studies were conducted with very heterogeneous students, which makes generalization, despite a small sample size, difficult. Another limitation in regard to the first part is the limited set of trained vocabulary. Unfortunately, no study was conducted in the second part of the dissertation tested for transfer effects, which is very important in research. Further, to date, we do not know if the students can use the vocabulary being taught in verbal communication. Regarding the language background, since we had students from diverse L1 backgrounds, we still do not know if the L1 played a role with respect to the intervention's effectiveness. Since L1 has an important impact in L2 learning (Maurer et al., 2021; van Gelderen et al., 2007), this could be assumed without knowing it for sure. As stated in the first part's limitations, which component influenced the dependent variables to what extent is unknown. However, again I refer to, e.g., Aro and colleagues (2018), who urgently called for using multicomponent interventions. Another limitation, which was also noted at the end of the first part, is that the intervention was never contrasted with another intervention. In summary, although the intervention is very effective, whether it is more effective than other interventions is unknown. Overall, the heterogeneous students in all studies should be stressed and there definitely needs to be more research on

the storytelling interventions' effectiveness for these sub-groups, precisely since we included participants who officially faced difficulties in L1 in addition to L2 and students from whom we did not know the L1 proficiency.

Implications

From the summary and the limitations, several implications can be derived, specifically for the second part of this work and especially for storytelling and L2 learning. Referring directly to the studies, it would be necessary to expand the intervention sessions for some students because the results indicated that specific students needed more automatization to achieve a gain in the dependent variables. One of my major interests is to figure out all participants' L1 language and their impacts on German and/or English learning because research suggest that decoding difficulties in L1 may result in challenges in learning an L2 appropriately (e.g., Helland & Kaasa, 2005; von Hagen et al., 2021). In reference to the criticism by Bialystok et al. (2010) that L2 studies rarely focus on L1 vocabulary, it would be beneficial to measure specific characteristics that have been evaluated in the L2 and in L1, (e.g., L1 reading) in future research because readers who are less proficient in their L2 reading can also face difficulties in L1 reading (e.g., Wendt et al., 2016) or indeed be good at L1 reading. Thus, the term "less-proficient readers" in general does actually not fit. They are just less-proficient readers in their L2. Nevertheless, the only way to figure this out is to implement measures that reveal their L1 skills. Also, the sixth study indicates that being able to read might be partly dependent on vocabulary knowledge. This is not too surprisingly because research has revealed the important relationship between vocabulary and reading (e.g., Jeon & Yamashita, 2014). Limbird (2007), who compared determinants of reading comprehension in L1 and L2 German-speaking elementary school students, suggested that the impact of vocabulary on reading was even more pronounced for L2 than for L1 speakers. Additionally, the degree of which the L2 is spoken at home or in the personal environment of the participants should be examined. Müller and Stanat (2006) concluded that whether students speak German at home affects their reading performance. Thus, the number of L2 speakers' use of German highly contributes to their reading literacy acquisition. Thus, future studies should measure vocabulary prior to the intervention's beginning as we did. Nevertheless, this indication led me to discover more about the relationship between vocabulary knowledge and reading achievement.

Another interesting area, especially in the context of storytelling, is students' listening abilities. In a study with Dutch elementary school students, Droop and Verhoeven (2003) demonstrated that listening skills not only strongly affected reading comprehension, but also that the effect was more pronounced in L2 compared with L1 speakers. Thus, in future studies I would also recommend considering L2 listening skills. Regarding German and English L2, PA does play a more leading role in English. Thus, when planning reading interventions in English L2, I would suggest also implementing PA components.

4. Overall Conclusion and Implications

4.1 Conclusion

This dissertation's overall intention was to evaluate self-created literacy interventions with regard to effectiveness in primary and secondary school students in L1 and L2 with learning and behavioral difficulties. In summary, the combined interventions of the first and second parts were an effective way to adequately support children and adolescents with learning and behavioral difficulties in literacy. Thus, it seems like these interventions can be implemented for a range of individual student groups. Referring back to how students with learning and behavioral difficulties are challenging to help improve literacy (e.g., Lerner & Johns, 2011; Solis et al., 2012; Solis et al., 2014; Wills et al., 2010), and that a high percentage of younger and older students who need special help, including those with language and behavioral issues, face enormous challenges in academic areas (Kultusministerkonferenz, 2018), this dissertation's results are encouraging and indicated that literacy acquisition might be dependent on how to foster students with special needs. With regard to our samples – the students learned certain content in a very short time when being taught in a joyful way. Also, considering students with learning and behavioral difficulties in terms of reading face the most challenges in lower-hierarchical reading skills (Cirino et al., 2013; Fuchs et al., 2012; Lane et al., 2008), the presented studies show that lower-level skills adequately can be fostered across all participants. Supporting a heterogeneous students' body effectively is mostly hampered by a lack of adequate methods, which can be adapted to individual students' needs (McLeskey & Waldron, 2011; Schmidt et al., 2002). This dissertation provides two possible combined interventions that address this very problem.

Further, this dissertation provides important implications on how both RT and storytelling can be promoted in terms of various literacy variable. The racetracks have also shown that they can be realized as PT interventions and combined with whole-word training in such a way that possible transfer effects can be achieved. These are fundamentally very simple—and important—interventions. The daily school routine often does not allow for embedding time-consuming support and needs easy-to-apply methods (Mitchell & Sutherland, 2020) such as RT and storytelling.

I am well aware of the individual studies' limitations, which were presented at the end of their respective parts; nevertheless, there is great potential for the interventions designed. The surveys of social validity, e.g., students' acceptance of an intervention in particular, have also shown once again that the participants like the interventions and that they have found pleasure in learning again. It should be emphasized, especially regarding storytelling, that the intervention can promote up to three variables simultaneously and thus shows great potential.

The combined racetracks with peer tutoring and motivational components are effective in sight word reading for both primary and secondary students, in small group settings, and in class settings. Transfer effects could be achieved via combining frequent letter clusters in German and racetracks as automation for primary students with L1 and L2. The racetracks lead to the second part of this dissertation, namely supporting students in L2 learning. Here, storytelling proves to be very effective with regard to memorizing vocabulary in secondary and primary students and adopting sight words as well as letter sound fluency. Overall, all questions of the summarized individual studies can be answered positively, though taking into account the limitations stated above. Besides the very good study results, I am particularly pleased with the students' feedback indicating that they felt they had improved in social validity. As explained above, when an individual improves, then their motivation increases and so does their willingness to learn (Guthrie et al., 1999; Nelson & Harwood, 2010; Scanlon et al., 2017).

This dissertation is an attempt to precisely break the vicious circle of rejecting literary interventions (Allday et al., 2011; Neff et al., 2005; Nelson & Harwood, 2010; Scanlon et al., 2017) because a person feels they are not good enough, but cannot get better if they do not practice. I also find the results in storytelling English for students with and without LD particularly great. Especially because those with LD have benefited just as much as the others, and this fortunately somewhat undermines the prevailing argument that teaching English would be too demanding for these students (Sparks, 2016). There are enormous difficulties in schools today in finding appropriate interventions for students with individual needs in L1 and L2 (Forlin & Chambers, 2011; Scheeler et al., 2009). My dissertation provides these interventions and ways to implement them that can be adapted to individual students' needs (Connor et al., 2014; Kamil et al., 2008). I wanted to contribute to making all students feel comfortable and enjoy learning, regardless of whether they have an official diagnosis. The limitations of the two parts also indicate how support could be improved and how the results should be interpreted in light of the literacy limitations.

Sparks (2016), Maurer et al. (2021) and von Hagen et al. (2021) again stress that it does not have to be about a specific label that students have received, but rather that overall students with less-proficient literacy skills face greater hurdles in L1 and L2 (see also Grigorenko et al., 2020; Ritchie & Bates, 2013). Because of this, it was absolutely the right decision to focus on all students who have difficulties in learning and behavior with and without official diagnosis (Oakes et al., 2010; Vannest et al., 2009). In sum, the studies' results are encouraging, and provide important information on implementing the interventions and add new research insights specifically regarding the to be focused participants and literacy instruction (e.g., Steinle et al., 2021). The results and the students' and teachers' feedback through the social validity questionnaires have proved that I am on the right track and that my thoughts on how to best combine an intervention for L1 and L2 were correct – of course

considering all limitations. The interventions followed the intention, according to De Naeghel et al. (2012), to comply with a quality criterion of education—namely to make literacy joyful again for the students.

4.2 Implications

Apart from the implications previously explained, the studies conducted have provided me further ideas about which aspects need to be looked at more closely, especially to improve the intervention and make it more adaptable to individual students. With regard to the rising importance of digital tools, the next step is to digitize the racetracks and storytelling. Thus, students would have access while at home and can learn whenever they want. Another further implication is to make both interventions, the combined Racetracks and combined storytelling, evidence-based for several dependent variables. For a treatment being labelled as evidence-based in single case research, studies are needed that are strong in methodology, and three different research teams across three different countries need to conduct them. In addition, a total of at least 20 cases is necessary (also see Horner et al., 2005). Thus, these combined interventions need more evaluation.

An additional interesting point to look at is based on Pugh et al. (2005), who found that native literacy instruction is beneficial for L1 and L2 literacy growth, which in turn supports the Cummins's (2000) hypothesis. Thus, it would be interesting to see whether an intervention in the students' L1 also has an impact on German/English L2. I would like to follow this up with more research on L2 acquisition, especially on the connections between L1 and L2. Plus, one really important point is that socio-economic backgrounds can explain poor L2 and L1 readers (Bradbury et al., 2015; Goldenberg et al., 2006). Thus, it would be interesting to also conduct more research to discover the relationship between socio-economic background and literacy, and how to increase these children's access to literature. I also would like to use racetracks more in L2 settings since one study in the first part showed that racetracks could also work for German L2's. In turn, I would like to embed storytelling in German L1 because research shows that listening to stories positively impacts L1 students (Joffe et al., 2019; Lenhart et al., 2018; Suggate et al., 2013) and that vocabulary knowledge is also crucial in L1 development (e.g., Droop & Verhoefen, 2003).

Further, we have several evidence-based methods in special education, but teachers rarely make use of them (Maheady et al., 2013). According to Cook and Odom (2013) practitioners struggle to implement interventions with fidelity. Scheeler et al. (2009) concluded that one reason might be because they were never taught how to implement these methods. Therefore, it is immensely important to teach these methods to the teachers in a comprehensible way and close this research-practice gap (see also Brock et al., 2017). There should be a continued focus on interventions that are easy to use yet effective, such as RT in

a PT setting or implementing storytelling, in addition to incorporating training for teachers on how to use these interventions as another important research aim.

Overall, there are many interesting areas in literacy promotion in both L1 and L2. In summary, the studies presented need to be improved, evaluated using more students, and integrated into schools where teachers should be trained to implement them with fidelity. Nevertheless, despite limitations and further implications, a lot can be taken away from the studies shown in order to be able to support students with learning and behavioral difficulties in L1 and L2 appropriately.

Bibliography

Abu Rabia, S., & Siegel, L. (2002). Reading, syntactic, orthographic, and working memory skills of bilingual Arab Canadian children. *Journal of Psycholinguistic Research*, *31*, 661–678.

Akbari, H., Ghonsooly, B., Ghazanfari, M., & Shahriari, H. (2017). Attitude toward reading: L1 or L2 or both. *SAGE Open*, *7*(3), 1–10.

Albers, C. A., & Hoffman, A. (2012). Using flashcard drill methods and self-graphing procedures to improve the reading performance of English language learners. *Journal of Applied School Psychology, 28,* 367–388.

Alberto, P., & Troutman, A. C. (2008). *Applied behavior analysis for teachers* (8th ed.). Merrill/Prentice-Hall.

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. Continuum.

Al-Hendawi, M. (2012). Academic engagement of students with emotional and behavioral disorders: Existing research, issues, and future directions. *Emotional and Behavioural Difficulties*, *17*(2), 125–141.

Allday, A. R., Bush, M., Ticknor, N., & Walker, L. (2011). Using teacher greetings to increase speed to task engagement. *Journal of Applied Behavior Analysis*, *44*(2), 393–396.

Allington, R. L. (2006). Fluency: Still waiting after all these years. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about reading fluency* (pp. 94–105). International Reading Association.

Al Otaiba, S. A., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities*, *39*(5), 414–431.

Anderson, N. J. (2015). Academic reading expectations and challenges. In N. Evans, N. Anderson, & W. Eggington (Eds.), *ESL readers and writers in higher education* (pp. 95–109). Routledge.

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, *107*(3), 868–883.

Archer, A. L., & Hughes, C. A. (2011). *Explicit instruction: Effective and efficient teaching*. The Guilford Press.

Arnold, E. M., Goldston, D.B., Walsh, A.K., Reboussin, B.A., Daniel, S.S., Hickman, E., & Wood, F.B. (2005). Severity of emotional and behavioral problems among poor and typical readers. *Journal of Abnormal Child Psychology, 33*(2), 205–17.

Aro, T., Viholainen, H., Koponen, T., Peura, P., Raeikkoenen, E., Salmi, P., & Aro, M. (2018). Can reading fluency and self-efficacy of reading fluency be enhanced with an intervention targeting the sources of self-efficacy? *Learning and Individual Differences, 67*(1), 53–66.

Atkinson, T. S., Wilhite, K. L., Frey, L. M., & Williams, S. C. (2002). Reading instruction for the struggling reader: Learning disabilities or emotional/behavioral disorders. *Preventing School Failure*, *46*(4), 158–162.

August, D. E., & Shanahan, T. E. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Lawrence Erlbaum Associates Publishers.

August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. J. (2009). The impact of an instructional intervention on the science and language learning of middle grade English language learners. *Journal of Research on Educational Effectiveness*, *2*(4), 345–376.

Aunola, K., Leskinen, E., Onatsu-Arvilommi, T., & Nurmi, J. E. (2002). Three methods for studying developmental change: A case of reading skills and self-concept. *British Journal of Educational Psychology*, *72*(3), 343–364.

Bacon, S. (1998). Conceptual knowledge, biliteracy, and the at-risk child. In M. Met (Ed.), *Critical issues in early second language learning* (pp. 313–318). Scott Foresman Addison Wesley.

Baker, D. L., & Kosty, D. (2012, February). *Effect of learning opportunities on the reading performance in Spanish of first grade English learners* [Paper presentation]. Pacific Coast Research Conference, Coronado, CA.

Banks, J., & McCoy, S. (2011). *A study on the prevalence of special educational needs*. National Council for Special Education. https://www.esri.ie/publications/a -study-on-the-prevalence-of-special-educational-needs

Bar-Kochva, I., Vágvoelgyi, R., Dresler, T., Nagengast, B., Schroeter, H., Schrader, J., & Nuerk, H.-C. (2021). Basic reading and reading-related language skills in adults with deficient reading comprehension who read a transparent orthography. *Reading and Writing*.

Bates, C. C., D'Agostino, J. V., Gambrell, L., & Xu, M. (2016). Reading recovery: Exploring the effects on first-graders' reading motivation and achievement. *Journal of Education for Students Placed at Risk*, *21*(1), 47–59.

Beauftragte der Bundesregierung fuer Migration, Fluechtlinge und Integration (2014, October). *10. Bericht ueber die Lage der Auslaenderinnen und Auslaender in Deutschland [*Federal Government Commissioner for Migration, Refugees and Integration*. 10th report on the situation of foreigners in Germany]*

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

Becker, M., McElvany, N., & Kortenbruck, M. (2010). Intrinsic and extrinsic reading motivation as predictors of reading: A longitudinal study. *Journal of Educational Psychology*, *102*(4), 773–785.

Becker-Mrotzek, M., Hentschel, B., Hippmann, K., & Linnemann, M. (2012, June). *Sprachfoerderung in deutschen Schulen – die Sicht der Lehrerinnen und Lehrer. Ergebnisse einer Umfrage unter Lehrerinnen und Lehrern [Language support in German schools - the teachers' view. Results of a survey among teachers]*. https://www.mercator-institut-sprachfoerderung.de/fileadmin/user_upload/Lehrerumfrage_Langfassung_final_30_05_03.pdf

Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology*, *85*, 347–356.

Berends, I., & Reitsma, P. (2006a). Addressing semantics promotes the development of reading fluency. *Applied Psycholinguistics*, *27*(2), 247–265.

Berends, I. E., & Reitsma, P. (2006b). Remediation of fluency: Word specific or generalized training. *Reading and Writing: An Interdisciplinary Journal*, *19*(2), 221–234.

Beyersmann, E., Grainger, J., & Castles, A. (2019). Embedded stems as a bootstrapping mechanism for morphological parsing during reading development. *Journal of Experimental Child Psychology, 182*, 196–210.

Bialystok, E., Luk, G., Peets, K.F. & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*(4), 525–531.

Bisson, M.-J., Kukona, A., & Lengeris, A. (2021). An ear and eye for language: Mechanisms underlying second language word learning. *Bilingualism: Language and Cognition*, *24*(3), 549–568.

Bitchener, J. (2017). To what extent has the published written CF research aided our understanding of its potential for L2 development? *ITL-International Journal of Applied Linguistics*, *167*(2), 111–131.

Boltzmann, M., Mohammadi, B., Samii, A., Muente, T. F., & Ruesseler, J. (2017). Structural changes in functionally illiterate adults after intensive training. *Neuroscience*, *344*, 229–242.

Borodkin, K., & Faust, M. (2014). Naming abilities in low-proficiency second language learners. *Journal of Learning Disabilities*, *47*(3), 237–253.

Bowman-Perrott, L. (2009). Classwide peer tutoring. An effective strategy for students with emotional and behavioral disorders. *Intervention in School and Clinic*, *44*(5), 259–267.

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.

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.

Bradbury, B., Corak, M., Waldfogel, J., & Washbrook, E. (2015). *Too many children left behind: the U.S. achievement gap in comparative perspective*. Russell Sage Foundation.

Brasseur-Hock, I. F., Hock, M. F., Kieffer, M. J., Biancarosa, G., & Deshler, D. D. (2011). Adolescent struggling readers in urban schools: Results of a latent class analysis. *Learning and Individual Differences*, *21*(4), 438–452.

Brock, M. E., Cannella-Malone, H. I., Seaman, R. L., Andzik, N. R., Schaefer, J. M., Page, E. J., Barczak, M.A., & Dueker, S. A. (2017). Findings across practitioner training studies in special education: A comprehensive review and meta-analysis. *Exceptional Children, 84*, 7–26.

Bruhn, A., Mcdaniel, S., & Kreigh, C. (2015). Self-monitoring interventions for students with behavior problems: A systematic review of current research. *Behavioral Disorders*, *40*, 102–121.

Calderón, M., August, D., Slavin, R., Duran, D., Madden, N., & Cheung, A. (2005). Bringing words to life in classrooms with English-language learners. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 115–136). Lawrence Erlbaum Associates Publishers.

Calhoon, M. B. (2005). Effects of a peer-mediated phonological skill and reading comprehension program on reading skill acquisition for middle school students with reading disabilities. *Journal of Learning Disabilities*, *38*(5), 424–433.

Campbell, J. M. (2003). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: A quantitative synthesis of single-subject research. *Research in Developmental Disabilities, 24*, 120–138.

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.

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education*, *47*(1), 1–32.

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., Lively, T. J., & White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, *39*(2), 188–215.

Carter, A., Wagmiller, R., Gray, S., McCarthy, K., Horwitz, S., & Briggs-Gowan, M. (2010). Prevalence of DSM-IV disorder in representative, healthy birth cohort at school entry: Sociodemographic risks and social adaptation. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(7), 686–689.

Carter, E. W., Moss, C. K., Asmus, J., Fesperman, E., Cooney, M., Brock, M. E., Lyons, G., Huber, H. B., & Vincent, L. B. (2015). Promoting inclusion, social connections, and learning through peer support arrangements. *Teaching Exceptional Children*, *48*(1), 9–18.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: a 40-year meta-analysis. *Psychological Bulletin*, *140*(4), 980–1008.

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 92–101.

Chall, J. S., & Jacobs, V. A. (2003). The classic study on poor children's fourth-grade slump. *American Educator*, *27*(1), 14–15, 44.

Chambrè, S. J., Ehri, L. C., & Ness, M. (2020). Phonological decoding enhances orthographic facilitation of vocabulary learning in first graders. *Reading and Writing*, *33*(5), 1133–1162.

Chang, A., & Millett, S. (2014). The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal*, *68*(1), 31–40.

Channa, L. A., Gilhooly, D., Lynn, C. A., M. S. A., & Soomro, N. H. (2017). Friend or foe? First language (MT) in second/foreign language (L2/FL) instruction & Vygotsky. *Journal of Language and Cultural Education, 5*(2), 25–39.

Chard, D. J., Vaughn, S., and Tyler, B. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *J. Learn. Disabil. 35,* 386–406.

Charlton, B., Williams, R. L., & Mclaughlin, T. F. (2005). Educational games: A technique to accelerate the acquisition of reading skills of children with learning disabilities. *International Journal of Special Education*, *20*(2), 66–72.

Chen, Y.-J. I., Thompson, C. G., Xu, Z., Irey, R. C., & Georgiou, G. K. (2021). Rapid automatized naming and spelling performance in alphabetic languages: A meta-analysis. *Reading and Writing*.

Chlapana, E., & Tafa, E. (2014). Effective practices to enhance immigrant kindergarteners' second language vocabulary learning through storybook reading. *Reading and Writing*, *27*, 1619–1640.

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.

Church, J. A., Coalson, R. S., Lugar, H. M., Petersen, S. E., & Schlaggar, B. L. (2008). A developmental fMRI study of reading and repetition reveals changes in phonological and visual mechanisms over age. *Cerebral Cortex*, *18*(9), 2054–2065.

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 and Writing*, *26*(7), 1059–1086.

Ciullo, S., Ely, E., McKenna, J. W., Alves, K. D., & Kennedy, M. J. (2019). Reading instruction for students with learning disabilities in grades 4 and 5: An observation study. *Learning Disability Quarterly*, *42*(2), 67–79.

Ciullo, S., Ortiz, M. B., Al Otaiba, S., & Lane, K. L. (2016). Advanced reading comprehension expectations in secondary school: considerations for students with emotional or behavior disorders. *Journal of Disability Policy Studies*, *27*(1), 54–64.

Clay, C. J., Bloom, S. E., Slocum, T. A., Samaha, A. L., & Callard, C. H. (2019). Language preference and reinforcing efficacy of praise in bilingual children with disabilities. *Journal of Applied Behaviors Analysis*, *53*(1), 536–544.

Clemens, N. H., Hsiao, Y.-Y., Lee, K., Martinez-Lincoln, A., Moore, C., Toste, J., & Simmons, L. (2021). The differential importance of component skills on reading comprehension test performance among struggling adolescent readers. *Journal of Learning Disabilities*, *54*(3), 155–169.

Cohen, M., Mahé, G., Laganaro, M., & Zesiger, P. (2018). Does the relation between rapid automatized naming and reading depend on age or on reading level? A behavioral and ERP study. *Frontiers in Human Neuroscience*, *12*, 73.

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., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589–608.

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.

Compton, D. L. (1997). Using a developmental model to access children's word recognition. *Intervention in School and Clinic*, *32*, 283–294.

Compton, D. L., Olinghouse, N. G., Elleman, A., Vining, J., Appleton, A. C., Vail, J., & Summers, M. (2005). Putting transfer back on trial: Modeling individual differences in the transfer of decoding-skill gains to other aspects of reading acquisition. *Journal of Educational Psychology*, *97*(1), 55–69.

Connor, C. M., Alberto P. A., Compton D. L., & O'Connor R. E. (2014). *Improving reading outcomes for students with or at risk for reading disabilities: A synthesis of the contributions from the Institute of Education Sciences Research Centers*. NCSER 2014-3000. National Center for Special Education Research.

Conrad, N. J., & Levy, B. A. (2007). Letter processing and the formation of memory representations in children with naming speed deficits. *Reading and Writing: An Interdisciplinary Journal*, *20*(3), 201–223.

Cook, B. G., & Odom, S. L. (2013). Evidencebased practices and implementation science in special education. *Exceptional Children, 79*, 135–144.

Cop, U., Drieghe, D., & Duyck, W. (2015a). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PloS One*, *10*(8), 1–38.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015b). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, *22*(5), 1216–1234.

Cree, A., Kay, A., & Steward, J. (2012). *The economic and social cost of illiteracy: A snapshot of illiteracy in a global context.* Final report from the World Literacy Foundation. http://hdl.voced.edu.au/10707/321997

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *99*(2), 311–325.

Crosson, A., McKeown, M., Robbins, K., & Brown, K. (2019). Key elements of robust vocabulary instruction for emergent bilingual adolescents. *Language, Speech, and Hearing Services in Schools*, *50*(4), 493–505.

Cruz de Quirós, A. M., Lara-Alecio, R., Tong, F., & Irby, B. J. (2012). The effect of a structured story reading intervention, story retelling and higher order thinking for English language and literacy acquisition. *Journal of Research in Reading*, *35*(1), 87–113.

Crystal, D. (2015). *English as a global language*. Cambridge University Press.

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Multilingual Matters Ltd.

Daniel J., Capin, P., & Steinle, P. (2021). A synthesis of the sustainability of remedial reading intervention effects for struggling adolescent readers. *Journal of Learning Disabilities*, *54*(3), 170–186.

De Jong, P. F. & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology, 95*, 22–40.

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.

De Naeghel, J., Van Keer, H., Vansteenkiste, M., & Rosseel, Y. (2012). The relation between elementary students' recreational and academic reading motivation, reading frequency, engagement, and comprehension: A self-determination theory perspective. *Journal of Educational Psychology*, *104(*4), 1006–1021.

Department for Education. (2012). *Pupil behaviour in schools in England: Educational standards and research division*. DfE Publications.

De Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhoefer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, *68*(4), 906–941.

Dewaele, J.-M. (2011). Reflections on the emotional and psychological aspects of foreign language learning and use. *Anglistik: International Journal of English Studies*, *22*(1), 23–42. http://eprints.bbk.ac.uk/5096/

De Weerdt, F., Desoete, A., & Roeyers, H. (2013). Working memory in children with reading disabilities and/or mathematical disabilities. *Journal of Learning Disabilities*, *46*(5), 461–472.

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. Macmillan.

Dickinson, D. K., Collins, M. F., Nesbitt, K., Spiewak Toub, T., Hassinger-Das, B., Burke Hadley, E., Hirsh-Pasek, K., & Michnick Golinkoff, R. (2018). Effects of teacher-delivered book reading and play on vocabulary learning and self-regulation among low-income preschool children. *Journal of Cognition and Development*, *20*(2), 136–164.

Dixon, L. Q., Zhao, J., Shin, J. Y., Wu, S., Su, J. H., Burgess-Brigham, R., Gezer, M. U., & Snow, C. (2012). What we know about second language acquisition: A synthesis from four perspectives. *Review of Educational Research*, *82*(1), 5–60.

Dlaska, A., & Krekeler, C. (2013). The short-term effects of individual corrective feedback on L2 pronunciation. *System*, *41*, 25–37.

Dolezal, D. N., Weber, K. P., Evavold, J. J., Wylie, J., & McLaughlin, T. F. (2007). The effects of a reinforcement package for on-task and reading behavior with at-risk and middle school students with disabilities. *Child & Family Behavior Therapy, 29*, 9–25.

Dombrowski, S. C., Kamphaus, R. W., & Reynolds, C. R. (2004). After the demise of the discrepancy: Proposed learning disabilities diagnostic criteria. *Professional Psychology: Research and Practice*, *35*(4), 364–372.

Dong, Y., Tang, Y., Chow, B.W.-Y., Wang, W., & Dong, W.-Y. (2020). Contribution of vocabulary knowledge to reading comprehension among Chinese students: A meta-analysis. *Frontiers in Psychology*, *11*, 525369.

Doernyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Blackwell.

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly*, *38*(1), 78–103.

Droop, M. & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly, 38*, 78–103.

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.

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.

Dunn, L., & Dunn, D. (2007). *Peabody picture vocabulary test* (4th ed.)*. PPVT-4.* Pearson.

Dunn, M. E., Shelnut, J., Ryan, J. B., & Katsiyannis, A. (2017). A systematic review of peer-mediated interventions on the academic achievement of students with emotional/behavioral disorders. *Education and Treatment of Children*, *40*(4), 497–524.

Ehri, L. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, *18*(1), 5–21.

Ehri, L. C. (1997). Sight word learning in normal readers and dyslexics. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 163–189). Erlbaum.

Ehri, L. C. (1999). Phases of development in learning to read words. In J. Oakhill & R. Beard (Eds.), *Reading development and the teaching of reading* (pp. 79–108). Blackwell.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188.

Ehri, L. C., & Robbins, C. (1992). Beginners need some decoding skill to read words by analogy. *Reading Research Quaterly*, *27*, 12–26.

Ellis, R. (2012). *Language teaching research and language pedagogy* (2nd ed.). Wiley-Blackwell, Inc.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching.* Multilingual Matters.

Eme, E., Lambert, E., & Alamargot, D. (2014). Word reading and word spelling in French adult literacy students: The relationship with oral language skills. *Journal of Research in Reading, 37*(3), 268–296.

Erbey, R., McLaughlin, T. F., Derby, M. K., and Everson, M. (2011). The effects of using flashcards with reading racetrack to teach letter sounds, sight words, and math facts

to elementary students with learning disabilities. *Int. Electron. J. Elementary Educ. 3,* 213–226.

Escarpio, R., & Barbetta, P. (2016). Comparison of repeated and non-repeated readings on the reading performances of students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, *24*(2), 111–124.

Feist, J., Feist, G. J., & Roberts, T.-A. (2017). *Theories of personality*. McGraw-Hill.

Fien, H., Anderson, D., Nelson, N. J., Kennedy, P., Baker, S. K., & Stoolmiller, M. (2018). Examining the impact and school-level predictors of impact variability of an 8th grade reading intervention. *Learning Disabilities Research and Practice*, *33*(1), 37–50.

Filippini, A., Gerber, M., & Leafstedt, J. (2012). A vocabulary-added reading intervention for English learners at-risk of reading difficulties. *International Journal of Special Education*, *27*(3), 14–26.

Fischer, M. Y., & Pfost, M. (2015). Wie effektiv sind Maßnahmen zur Foerderung der phonologischen Bewusstheit? Eine meta-analytische Untersuchung der Auswirkungen deutschsprachiger Trainingsprogramme auf den Schriftspracherwerb [How effective are trainings of phonological awareness? A meta-analysis of German language training programs and their effects on the acquisition of reading and spelling skills]. *Zeitschrift fuer Entwicklungspsychologie und Paedagogische Psychologie*, *47*(1), 35–51.

Fitton, L., McIlraith, A. L., & Wood, C. L. (2018). Shared book reading interventions with English learners. A meta-analysis. *Review of Educational Research*, *88*(5), 712–751.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities. From identification to intervention*. Guilford Press.

Forlin, C., & Chambers, D. (2011). Teacher preparation for inclusive education: Increasing knowledge but raising concerns. *Asia-Pacific Journal of Teacher Education*, *39*(1), 17–32.

Fornefeld, B. (2013). Storytelling with all our senses: Mehr-Sinn Geschichten. In N. Grove (Ed.), *Using storytelling to support children and adults with special needs* (pp. 78–85). Routledge.

Frankel, K. K., Becker, B. L. C., Rowe, M. W., & Pearson, P. D. (2016). From 'what is reading?' to what is literacy? *Journal of Education*, *196*(3), 7–17.

Freiberg, H. J., Huzinec, C. A., & Templeton, S. M. (2009). Classroom management: A pathway to student achievement. A study of fourteen inner-city elementary schools. *Elementary School Journal*, *110*(1), 63–80.

Fromkin, V., Rodman, R., Hyams, N., Amberber, M., Cox, F., & Thornton, R. (2018). *An introduction to language* (9th ed.). Cengage Learning.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children*, *78*(3), 263–279.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239–256.

Gangl, M., Moll, K., Jones, M. W., Banfi, C., Schulte-Koerne, G., & Landerl, K. (2018). Lexical reading in dysfluent readers of German. *Scientific Studies of Reading*, *22*(1), 24–40.

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: a meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, *84*(1), 74–111.

Garwood, J. D., Vernon-Feagans, L., & The Family Life Project Key Investigators. (2017). Classroom management affects literacy development of students with emotional and behavioral disorders. *Exceptional Children*, *83*(2), 123–142.

Garwood, J. D., Peltier, C., Sinclair, T., Eisel, H., McKenna, J. W., & Vannest, K. J. (2020). A quantitative synthesis of intervention research. *Behavioral Disorders. Flagship EBD Journals.*

Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education*, *8*(3), 259–272.

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A. M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, *93*(3), 265–281.

Geary, D. C., Hoard, M. K., Nugent, L., UEnal, Z. E., & Scofield, J. E. (2020). Comorbid learning difficulties in reading and mathematics: The role of intelligence and in-class attentive behavior. *Frontiers in Psychology*, *11*, 572099.

Gentry, J. E., & Lindsey, P. (2008). The impact of assistive technology on vocabulary acquisition of a middle school student with learning disabilities and limited English proficiency: A descriptive case study analysis. *Journal of the American Academy of Special Education Professionals*, 63–93. https://files.eric.ed.gov/fulltext/EJ1139312.pdf

Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL1. *Reading and Writing: An Interdisciplinary Journal*, *25*, 1819–1845.

Geva, E., & Massey-Garrison, A. A. (2013). Comparison of the language skills of ELLs and monolinguals who are poor decoders, poor comprehenders, or normal readers. *Journal of Learning Disabilities, 46*(5), 387–401.

Geva, E., Yaghoub-Zadeh, Z., & Schuster, B. (2000). Understanding individual differences in word recognition skills of ESL children. *Annals of Dyslexia*, *50*(1), 121–154.

Gilmour, A. F., Fuchs, D., & Wehby, J. H. (2019). Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities. *Exceptional Children*, *85*(3), 329–346.

Ginns, D. S., Joseph, L. M., Tanaka, M. L., & Xia, Q. (2019). Supplemental phonological awareness and phonics instruction for Spanish-speaking English Learners: Implications for school psychologists. *Contemporary School Psychology*, *23*(1), 101–111.

Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology*, *98*(4), 732–749.

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Oretega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443-483). Amsterdam: John Benjamins.

Gold, A. (2011). Lernschwierigkeiten. Ursachen, Diagnostik, Intervention [Learning difficulties. Causes, diagnostics, intervention]. In M. Hasselhorn, H. Heuer, & F. Roesler (Eds.), *Standards Psychologie* (pp. 7–322). Kohlhammer.

Goldenberg, C., Rueda, R. S., & August, D. (2006). *Sociocultural influences on the literacy development. Developing literacy in second-language learners: Report of the national literacy panel on language-minority children and youth*. Routledge.

Gorsuch, G., & Taguchi, E. (2008). Repeated reading for developing reading fluency and reading: The case of EFL learners in Vietnam. *System*, *36*(2), 253–278.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.

Grabe, W. (2010). Fluency in reading: Thirty-five years later. *Reading in a Foreign Language*, *22*(1), 71–83.

Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading* (2nd ed.). Pearson Education.

Grainger, J., Lété, B., Bertand, D., Dufau, S., & Ziegler, J. C. (2012). Evidence for multiple routes in learning to read. *Cognition*, *123*(2), 280–292.

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.

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), 31–51.

Grosche, M., Hintz, A., & Hoelz, A. (2013). Wortspezifische Lesetrainingseffekte bei gleichzeitig ausbleibendem Lerntransfer. Analyse eines computergestützten

Lesetrainings bei Foerderbedarf im Lernen. *Empirische Sonderpädagogik 5*(3), 222-236.

Gruenke, M. (2006). Zur Effektivitaet von Foerdermethoden bei Kindern und Jugendlichen mit Lernstoerungen: Eine Synopse vorliegender Metaanalysen [On the effectiveness of support methods for children and adolescents with learning disorders: A synopsis of existing meta-analyses]. *Kindheit und Entwicklung*, *15*(4), 239–254.

Gruenke, 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.

Gruenke, 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.

Gruenke, M., & Morrison-Cavendish, W. (2016). Learning disabilities around the globe: Making sense of the heterogeneity of the different viewpoints. *Learning Disabilities: A Contemporary Journal*, *14*(1), 1–8.

Gruenke, M., Sperling, M., & Burke, M. D. (2017). The impact of explicit timing, immediate feedback, and positive reinforcement on the writing outcomes of academically and behaviorally struggling fifth-grade students. *Insights into Learning Disabilities*, *14*(2), 135–153.

Gunobgunob-Mirasol, R. (2020). A survey of grade 5 and 6 pupils' attitudes toward reading. *Reading Psychology*, *41*(3), 229–239.

Gunter, P. L., Miller, K. A., Venn, M. L., Thomas, K., & House, S. (2002). Self-graphing to success: Computerized data management. *Teaching Exceptional Children*, *35*(2), 30–34.

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). Erlbaum.

Guthrie, J., Anderson, E., Alao, S., & Rinehart, J. (1999). Influences of concept-oriented reading instruction on strategy use and conceptual learning from text. *Elementary School Journal*, *99*(4), 343–66.

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), 160–168.

Gyovai, L., Cartledge, G., Kourea, L., Yurick, A., & Gibson, L. (2009). Early reading intervention: Responding to the learning needs of young at-risk English language learners. *Learning Disability Quarterly*, *32*(3), 143–162.

Hall, M., & Burns, M. K. (2018). Meta-analysis of targeted small-group reading interventions. *Journal of School Psychology*, *66*, 54–66.

Hall, C., Steinle, P. K., & Vaughn, S. (2019) Reading instruction for English Learners with learning disabilities. What do we already know, and what do we still need to learn? *New Directions for child and Adolescent Development*, *2019*(166), Special Issue, 145–189.

Halonen, A., Aunola, K., Ahonen, T., & Nurmi, J.-E. (2006). The role of learning to read in the development of problem behaviour: A cross-lagged longitudinal study. *British Journal of Educational Psychology*, *76*(3), 517–534.

Hansen, J. G., & Liu, J. (2005). Guiding principles for effective peer response. *ELT Journal*, *59*(1), 31–38.

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.

Harmon, J. M., Wood, K. D., & Kiser, K. (2009). Promoting vocabulary learning with the interactive word wall. *Middle School Journal*, *40*(3), 58–63.

Hasselhorn, M., & Gold, A. (2017). *Paedagogische Psychologie - Erfolgreiches Lernen und Lehren [Educational psychology - Successful learning and teaching]* (4th ed.). Kohlhammer.

Hattie, J. (2008). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Helland, T., & Kaasa, R. (2005). Dyslexia in English as a second language*. Dyslexia, 11*(1), 41–60.

Hickman, P., Pollard-Durodola, S., & Vaughn, S. (2004). Storybook reading: Improving vocabulary and comprehension for English language learners. *The Reading Teacher*, *57*(8), 720–730.

Hilsmier, A., Wehby, J., & Falk, K. (2016). Reading fluency interventions for middle school students with academic and behavioral disabilities. *Reading Improvement*, *53*(2), 53–64.

Hintikka, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency: Is it important to practice reading aloud and is generalization possible? *Annals of Dyslexia*, *58*(1), 59–79.

Holahan, J. M., Ferrer, E., Shaywitz, B. A., Rock, D. A., Kirsch, I. S., Yamamoto, K., Michaels, R., Marchione, K. E., & Shaywitz, S. E. (2018). Growth in reading comprehension and verbal ability from grades 1 through 9. *Journal of Psychoeducational Assessment*, *36*(4), 307–321.

Holifield, C., Goodman, J., Hazelkorn, M., & Heflin, L. J. (2010). Using self-monitoring to increase attending to task and academic accuracy in children with autism. *Focus on Autism and Other Developmental Disabilities*, *25*(4), 230–238.

Holland, J., McIntosh, D., & Huffman, L. (2004). The role of phonological awareness, rapid automatized naming, and orthographic processing in word reading. *Journal of Psychoeducational Assessment*, *22*(3), 233–260.

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.

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.

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. *Council for Exceptional Children*, *71*(2), 165–179.

Hudson, A., Koh, P. W., Moore, K. A., & Binks-Cantrell, E. (2020). Fluency interventions for elementary students with reading difficulties: A synthesis of research from 2000–2019. *Education Sciences*, *10*(3), 52.

Huemer, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency among poor readers of German: Many ways to the goal. *Annals of Dyslexia*, *58*(2), 115–137.

Huemer, S., Aro, M., Landerl, K., Lyytinen, H. (2010). Repeated reading of syllables among finnish-speaking children with poor reading skills. *Scientific Studies of Reading, 14*, 317-340.

Hughes, C. A., Morris, J. R., Therrien, W. J., & Benson, S. K. (2017). Explicit instruction: Historical and contemporary contexts. *Learning Disability Research & Practice*, *32*(3), 140–148.

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science*, *23*(6), 572–577.

Huschka, S. S., Georgiou, G. K., Brandenburg, J., Ehm, J.-H., & Hasselhorn, M. (2021). Examining the contribution of RAN components to reading fluency, reading comprehension, and spelling in German. *Reading and Writing*.

Isbell, R., Sobol, J., Lindauer, L., & Lowrance, A. (2004). The effects of storytelling and story reading on the oral language complexity and story comprehension of young children. *Early Childhood Education Journal*, *32*(3), 157–163.

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*.

Jeon, E. (2009). *Effects of repeated reading on L2 reading fluency and comprehension*. (Unpublished doctoral dissertation). Northern Arizona University.

Jeon, E. H., and Yamashita, J. (2014). L2 reading Comprehension and its Correlates: A Meta-Analysis. *Lang. Learn., 64*(1), 160–212.

Jeuk, S. (2018). *Deutsch als Zweitsprache in der Schule. Grundlagen, Diagnose, Foerderung [German as a second language in school. Basics, diagnosis, support]* (4th ed.). Kohlhammer.

Jitendra, A. K., Edwards, L. L., Sacks, G., & Jacobson, L. A. (2004). What research says about vocabulary instruction for students with learning disabilities. *Exceptional Children*, *70*(3), 299–322.

Joffe, V., Rixon, L., & Hulme, C. (2019). Improving storytelling and vocabulary in secondary school students with language disorder: a randomized controlled trial. *International Journal of Language and Communication Disorders*, *54*(4), 656–672.

Jordan, R. L. P., Fernandez, E. P., Costa, L. J. C., & Hooper, S. R. (2020). Internalizing and externalizing behaviors of children with writing disabilities. *Learning Disabilities Research and Practice*, *35*(2), 72–81.

Joshi, R. M., & McCardle, P. (2018). Models of reading in different orthographies: An introduction. *Journal of Learning Disabilities*, *51*(5), 419–421.

Kamil, M.L., Borman, G.D., Dole, J., Kral, C.C, Salinger, T., & Torgesen, J. (2008). Improving adolescent literacy: Effective classroom and intervention practices: A practice guide. & Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Karageorgos, P., Richter, T., Haffmans, M.-B., Schindler, J., & Naumann, J. (2020). The role of word-recognition accuracy in the development of word-recognition speed and reading comprehension in primary school: A longitudinal examination. *Cognitive Development*, *56*, 100949.

Kato, M. (2018). Exploring the transfer relationship of summarizing skills in L1 and L2. *English Language Teaching*, *11*(10), 75−87.

Kauffman, J. M., & Landrum, T. J. (2009). *Characteristics of emotional and behavioral disorders of children and youth* (9th ed.). Merrill/Pearson.

Kazdin, A. E. (2011). *Single-case research designs. Methods for clinical and applied settings*. Oxford University Press.

Keenan, K., & Shaw, D.S. (2003). Starting at the beginning: Exploring the etiology of antisocial behavior in the first years of life. In B.B., Lahey, T.E., Moffitt, A., Caspi (Eds.). The causes of conduct disorder and serious delinquency. NY: Guilford Press; p. 153-181.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*, 313–334.

Kent, S. C., Wanzek, J., & Al Otaiba, S. (2017). Reading instruction for fourth-grade struggling readers and the relation to student outcomes. *Reading and Writing Quarterly*, *33*(5), 395–411.

Keogh, B. K. (2020). Republication of future of the LD field: Research and practice. *Journal of Learning Disabilities*, *53*(1), 12–18.

Kirby, J. R., Georgiou, G. K., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly*, *45*(3), 341–362.

Kirsch, C. (2016). Using storytelling to teach vocabulary in language lessons: does it work? *The Language Learning Journal*, *44*(1), 33–51.

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., and Neeb, Y. (2014). Phonological recoding, orthographic decoding, and comprehension skills during reading acquisition. *Z. Erzieh. 17*, 447–471.

Kohnen, S., Nickels, L., Brunsdon, R. K., & Coltheart, M. (2008). Patterns of generalisation after treating sub-lexical spelling deficits in a child with mixed dysgraphia. *Journal of Research in Reading*, *31*(1), 157–177.

Kormos, J. (2017). *The second language learning processes of students with specific learning difficulties*. Routledge.

Kostewicz, D. E., & Kubina, R. M. (2011). Building science reading fluency for students with disabilities with repeated reading to a fluency criterion. *Learning Disabilities: A Multidisciplinary Journal*, *17*, 89–104.

Kraemer, S., Moeller, J., & Zimmermann, F. (2021). Inclusive education of students with general learning difficulties. A meta-analysis. *Review of Educational Research*, *91*(3), 432–478.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

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.

Kuder, S. J. (2017). Vocabulary instruction for secondary students with reading disabilities: An updated research review. *Learning Disability Quarterly*, *40*(3), 155–164.

Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about controlling variables? *Psychological Science, 16,* 866–870.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, *95*(1), 3–21.

Kuhn, M., Rasinki, T., & Zimmerman, B. (2014). Integrated fluency instruction: Three approaches for working with struggling readers*. International Electronic Journal of Elementary Education*, *7*(1), 71–82.

Kultusministerkonferenz. (2018). *Statistische Veroeffentlichungen der Kultusministerkonferenz, Dokumentation Nr. 214–Juni 2018. Sonderpaedagogische Foerderung in Schulen 2007 bis 2016 [Statistical publications of the standing conference of the Ministers of Education and Cultural Affairs of the Laender 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

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323.

Landerl, K., & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. *Applied Psycholinguistics*, *21*(2), 243–262.

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.

Landerl, K., Freudenthaler, H. H., Heene, M., de Jong, P. F., Desrochers, A., Manolitsis, G., Parrila, R., & Georgiou, G. K. (2019). Phonological awareness and rapid automatized naming as longitudinal predictors of reading in five alphabetic orthographies with varying degrees of consistency. *Scientific Studies of Reading*, *23*(3), 220–234.

Lane, J. D., Ledford, J. R., & Gast, D. L. (2017). Single-case experimental design: Current standards and applications in occupational therapy. *American Journal of Occupational Therapy*, *71*(2), 7102300010.

Lane, K. L., Barton-Arwood, S. M., Nelson, R. J., & Wehby, J. (2008). Academic performance of students with emotional and behavioral disorders served in a self-contained setting. *Journal of Behavioral Education*, *17*(1), 43–62.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? *Canadian Modern Language Review*, *59*(4), 565–585.

Lee, J., & Zentall, S. S. (2012). Reading motivational differences among groups: Reading disability (RD), attention deficit hyperactivity disorder (ADHD), RD+ADHD, and typical comparison. *Learning and Individual Differences*, *22*(6), 778–785.

Lee, J. A. C., Lee, S., Yusoff, N. F. M., Ong, P. H., Nordin, Z. S., & Winskel, H. (2020). An early reading assessment battery for multilingual learners in Malaysia. *Frontiers in Psychology*, *11*, 1700.

Lee, J., & Yoon, S. Y. (2017). The effects of repeated reading on reading fluency for students with reading disabilities: A meta-analysis. *Journal of Learning Disabilities, 50,* 213-224.

Leko, M. M. (2016). *Word study in the inclusive secondary classroom*. Teacher College Press.

Lenhard, W., & Lenhard, A. (2013). Learning difficulties. In L. Meyer (Ed.), *Oxford bibliographies in education* (p. 2). Oxford University Press.

Lenhart, J., Lenhard, W., Vaahtoranta, E., and Suggate, S. (2018). Incidental vocabulary acquisition from listening to stories: a comparison between readaloud and free storytelling approaches. *Educ. Psychol. 38*, 596–616.

Lerner, J., & Johns, B. (2011). *Learning disabilities and related mild disabilities: Characteristics, teaching strategies, and new directions* (11th ed.). Houghton Mifflin.

Lervåg, A., & Aukrust, V. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension between first and second language learners. *Journal of Child Psychology and Psychiatry*, *51*, 612–620.

Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for sixth graders in urban middle schools. *Reading Research Quarterly*, *45*(2), 196–228.

Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a randomized field trial. *American Educational Research Journal*, *51*(6), 1159–1194.

Leung, K. C. (2019). An updated meta-analysis on the effect of peer tutoring on tutors' achievement. *School Psychology International*, *40*, 200–214.

Li, S., Zhu, Y., & Ellis, R. (2016). The effects of the timing of corrective feedback on the acquisition of a new linguistic structure. *The Modern Language Journal*, *100*(1), 276–295.

Li, Y. (2007). How to learn English? (Translated title). *Common Wealth Magazine*, *2007 Special Issue on English Learning and Teaching*, 84–90.

Lim, S. C., & Renandya, W. A. (2020). Efficacy of written corrective feedback in writing instruction. A meta-analysis. *The Electronic Journal for English as a Second Language*, *24*(3), 1–26.

Limbird, C. K. (2007). *Phonological processing, verbal abilities, and second language literacy development among bilingual Turkish children in Germany* (Doctoral dissertation, Freie Universitaet Berlin, Germany).

Linan-Thompson, S.., Cavazos, H., McFarland, L., & Martinez, J. (2016). Language and Literacy Interventions for Els with LD. Two steps forward, one step back. In R. Schiff & R. Malatesha Joshi (Eds.), *Interventions in learning disabilities. A handbook on systematic training programs for individuals with learning disabilities* (pp. 191–205). Springer.

Little, C. W., Hart, S. A., Quinn, J. M., Tucker-Drob, E. M., Taylor, J., & Schatschneider, C. (2017). Exploring the co-development of reading fluency and reading comprehension: a twin study. *Child Development*, *88*(3), 934–945.

Little, S. G., Akin-Little, A., & O'Neill, K. (2015). Group contingency interventions with children—1980-2010: A meta-analysis. *Behavior Modification*, *39*, 322–341.

Livingston, E. M., Siegel, L. S., & Ribary, U. (2018). Developmental dyslexia: Emotional impact and consequences. *Australian Journal of Learning Difficulties*, *23*(2), 107–135.

Lo, Y., Cooke, N. L., & Starling, A. (2011). Using a repeated reading program to improve generalization of oral reading fluency. *Education and Treatment of Children*, *34*, 115–140.

Logan, S., Medford, E., & Hughes, N. (2011). The importance of intrinsic motivation for high and low ability readers' reading comprehension performance. *Learning and Individual Differences*, *21*, 124–128.

Lorette, P., & Dewaele, J.-M. (2015). Emotion recognition ability in English among L1 and LX users of English. *International Journal of Language and Culture*, *2*(1), 62–86.

Ludwig, C., Guo, K., & Georgiou, G. K. (2019). Are reading interventions for English language learners effective? A meta-analysis. *Journal of Learning Disabilities*, *52*(3), 220–231.

Lugo-Neris, M., Wood Jackson, C., & Goldstein, H. (2010). Facilitating vocabulary acquisition of young English language learners. *Language, Speech, and Hearing Services in Schools*, *41*(3), 314–327.

Ma, H. H. (2006). An alternative method for quantitative synthesis of single subject researches: Percentage of data points exceeding the median. *Behavior Therapy*, *30*, 598–617.

Macedonia, M. (2014). Bringing back the body into the mind: Gestures enhance word learning in foreign language. In G. M. Reevy & S. N. Bursten (Eds.), *Pedagogical psychology: Beyond the 21st century* (pp. 12–17). Frontiers Media SA.

Maggin, D., Johnson, A., Chafouleas, S., Ruberto, L., & Berggren, M. (2012). A systematic evidence review of school-based group contingency interventions for students with challenging behavior. *Journal of School Psychology*, *50*, 625–654.

Maggin, D. M., Pustejovsky, J. E., & Johnson, A. H. (2017). A meta-analysis of school-based group contingency interventions for students with challenging behavior: An update. *Remedial and Special Education, 38*(6), 353–370.

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.

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.

Martin-Chang, S. L., & Levy, B. A. (2005). Fluency transfer: Differential gains in reading speed and accuracy following isolated word and context training. *Reading and Writing*, *18*, 343–376.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300–335.

Matson, J. L., & Fodstad, J. C. (2010). Learning disabilities. In J. C. Thomas & M. Hersen (Eds.), *Handbook of clinical psychology competencies* (pp. 1301–1328). Springer.

Maurer, U., Jost, L.B., Pfenniger, S.E. et al. (2021). Effect of German reading skills and bilingualism on early learning of English as a foreign language in primary school children. Reading & Writing.

Mayer, A. (2018). *Blitzschnelle Worterkennung [Rapid word recognition]* (3rd edition). Borgmann Media.

Mayer, A. (2020). *Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit (TEPHOBE) [Test to assess phonological awareness and naming speed]*. Reinhardt.

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.

McDaniel, S. C., Jolivette, K., & Ennis, R. P. (2013). The effects of self-graphing on oral reading fluency for a student with E/BD within an alternative education school. *Journal of the American Academy of Special Education Professionals*, 69–82.

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.

McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., Diliberti, M., Forrest Cataldi, E., Bullock Mann, F., & Barmer, A. (2019). *The condition of education 2019*

(NCES 2019144). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

McKenna, J. W., & Bettini, E. (2018). Improving reading fluency skills for secondary students with emotional and behavioral disorders. *Beyond Behavior*, 27(2), 74–81.

McKeown, M. G., Crosson, A. C., Moore, D. W., & Beck, I. L. (2018). Word knowledge and comprehension effects of an academic vocabulary intervention for middle school students. *American Educational Research Journal*, *55*, 572–616.

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.

McLeskey, J., Barringer, M. -D., Billingsley, B., Brownell, M., Jackson, D., Kennedy, M., et al. (2017). High-leverage practices in special education. Arlington, VA: Council for Exceptional Children & CEEDAR Center.

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*, *34*(1), 114–135.

Melby-Lervåg, M. & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first-and second-language learners. *Psychological Bulletin, 140*(2), 409–433.

Mellard, D. F., Fall, E., & Woods, K. L. (2010). A path analysis of reading comprehension for adults with low literacy. *Journal of Learning Disabilities*, *43*(2), 154–165.

Mello, R. (2001). *Building bridges: How storytelling influences teacher and student relationships. In proceedings, storytelling in the Americas conference*. Brooks University Press.

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.

Merç, A., & Subaşı, G. (2015). Classroom Management Problems and Coping Strategies of Turkish Student EFL Teachers. *Turkish Online Journal of Qualitative Inquiry, 6*(1), 39-72.

Metsaepelto, R.-L., Pakarinen, E., Kiuru, N., Poikkeus, A.-M., Lerkkanen, M.-K., & Nurmi, J.-E. (2015). Developmental dynamics between children's externalizing problems, task-avoidant behavior, and academic performance in early school years: A 4-year follow-up. *Journal of Educational Psychology*, *107*(1), 246–257.

Mitchell, D. (2014). *What really works in special and inclusive education. Using evidencebased teaching strategies* (2nd ed.). Routledge.

Mitchell, D., & Sutherland, D. (2020). *What really works in special and inclusive education: Using evidence-based teaching strategies* (3rd ed.). Routledge.

Moats, L. C., & Tolman, C. A. (2019). *LETRS* (3rd ed.). Voyager Sopris Learning.

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.

Mol, S., Bus, A., & de Jong, M. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research, 79*, 979–1007.

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*(2), 267–296.

Moll, K., Fussenegger, B., Willburger, E., & Landerl, K. (2009). RAN is not a measure of orthographic processing. Evidence from the asymmetric German orthography. *Scientific Studies of Reading*, *13*(1), 1–25.

Moll, K., & Landerl, K. (2010). *SLRT-II: Lese-und Rechtschreibtest; Weiterentwicklung des Salzburger Lese-und Rechtschreibtests (SLRT) [Reading and spelling test; further development of the Salzburg reading and spelling test]*. Huber.

Moll, K., Goebel, S. M., Gooch, D., Landerl, K., & Snowling, M. J. (2016). Cognitive risk factors for specific learning disorder: Processing speed, temporal processing, and working memory. *Journal of Learning Disabilities*, *49*(3), 272–281.

Molnár, T. (2008). Second language versus third language vocabulary acquisition: A comparison of the English lexical competence of monolingual and bilingual students. *Toronto Working Papers in Linguistics (TWPL)*, *33*, 1–16.

Moore, B. A., & Klingner, J. K. (2014). Considering the needs of English language learner populations: An examination of the population validity of reading intervention research. *Journal of Learning Disabilities*, *47*(5), 391–408.

Morett, L. M. (2014). When hands speak louder than words: The role of gesture in the communication, encoding, and recall of words in a novel second language. *The Modern Language Journal*, *98*, 834–853.

Morett, L. M. (2019). The power of an image: images, not glosses, enhance learning of concrete L2 words in beginning learners. *J. Psycholinguist. Res., 48*(3), 643–664.

Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities*, *41*(5), 417–436.

Mueller, A. G., & Stanat, P. (2006). Schulischer Erfolg von Schuelerinnen und Schuelern mit Migrationshintergrund: Analysen zur Situation von Zuwanderern aus der ehemaligen Sowjetunion und aus der Tuerkei [School success of pupils with a migration

background: Analyses of the situation of immigrants from the former Soviet Union and Turkey]. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitaeten im Bildungswesen* [Disparities in education due to origin] (pp. 221–255). VS Verlag fuer Sozialwissenschaften.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/pirls2016/international-results/

Musti-Rao, S., Hawkins, R. O., & Barkley, E. A. (2009). Effects of repeated readings on the oral reading fluency of urban fourth-grade students. *Preventing School Failure*, *54*(1), 12–23.

National Center for Education Statistics. (2019). *The nation's report card: Trends in academic progress 2019*. National Center for Education Statistics.

Neff, K. D., Hsieh, Y. P., & Dejitterat, K. (2005). Self-compassion, achievement goals, and coping with academic failure. *Self and Identity*, *4*, 263–287.

Nelson, J., & Harwood, H. (2010). Learning disabilities and anxiety: A meta-analysis. *Journal of Learning Disabilities*, *44*, 3–17.

Nelson, J. R., Benner, G. J., & Boharty, J. (2014). Addressing the academic problems and challenges of students with emotional and behavioral disorders. In H. M. Walker & F. M. Gresham (Eds.), *Handbook of evidence-based practices for emotional and behavioral disorders: Applications in schools* (pp. 363–377). Guilford.

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.

Oakes, W. P., Mathur, S. R., & Lane, K. L. (2010). Reading interventions for students with challenging behavior: A focus on fluency. *Behavioral Disorders*, *35*(2), 120–139.

Ober, T. M., Brooks, P. J., Homer, B. D., & Rindskopf, D. (2020). Executive functions and decoding in children and adolescents: A meta-analytic investigation. *Educational Psychology Review*, *32*, 735–763.

O'Connor, R. E., White, A., & Swanson, H. L. (2007). Repeated reading versus continuous reading: influences on reading fluency and comprehension. *Exceptional Children*, *74*, 31–46.

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.

Organisation for Economic Co-operation and Development (OECD). (2010). *PISA 2009 results: What students know and can do student performance in reading, mathematics and science*. OECD Publishing.

Organisation for Economic Cooperation and Development (OECD). (2019). *PISA 2018 results (Volume I): What students know and can do*. OECD Publishing.

Oxford, R. L. (2017). Conditions for second language (L2) learning. In N. Van Deusen-Scholl & S. May (Eds.), *Second and foreign language education, encyclopedia of language and education* (pp. 27–41). Springer.

Paige, D. D., Magpuri-Lavell, T., Rasinski, T. V., & Smith, G. (2013). Interpreting the relationships among prosody, automaticity, accuracy, and silent reading comprehension in secondary students. *Journal of Literacy Research*, *46*(2), 123–156.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(3), 255–287.

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Erlbaum.

Paivio, A. (2014). Intelligence, dual coding theory, and the brain. *Intelligence*, *47*, 141–158.

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *34*(4), 388–399.

Papadopoulos, T. C., Spanoudis, G. C., & Georgiou, G. K. (2016). How is RAN related to reading fluency? A comprehensive examination of the prominent theoretical accounts. *Frontiers in Psychology, 7,* 1–15.

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

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.

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*, 284–299.

Parrila, R., Kirby, J. R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading*, *8*(1), 3–26.

Peker, H., Regalla, M., & Cox, D. T. (2018). Teaching and learning vocabulary in context. Examining engagement in three prekindergarten French classrooms. *Foreign Language Annals, 51*(2), 472–483.

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, Á. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: A eye-tracking study. *Studies in Second Language Acquisition*, *42*(3), 577–598.

Peng, P., & Fuchs, D. (2017). A randomized control trial of working memory training with and without strategy instruction: Effects on young children's working memory and comprehension. *Journal of Learning Disabilities*, *50*(1), 62–80.

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A metaanalysis on the relation between reading and working memory. *Psychological Bulletin*, *144*(1), 48–76.

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

Perfetti, C. A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37.

Petermann, F., & Daseking, M. (2013). Zuercher Lesetest – II. Weiterentwicklung des Zuercher Lesetests (ZLT) von Maria Linder und Hans Grissemann [*Zurich Reading Test - II. Further development of the Zurich Reading Test (ZLT) by Maria Linder and Hans Grissemann*] (4th, ed.) Hogrefe.

Peters, M. T., Foerster, N., Hebbecker, K., Forthmann, B., & Souvignier, E. (2021). *Effects of data-based decision-making on low-performing readers in general education classrooms: Cumulative evidence from six intervention studies*. https://osf.io/nj6tu/

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.

Pfost, M. (2015). Children's phonological awareness as a predictor of reading and spelling: A systematic review of longitudinal research in German-speaking countries. *Zeitschrift fuer Entwicklungspsychologie und paedagogische Psychologie*, *47*(3), 123–138.

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

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, *58*(6), 510–519.

Porter, A. (2016). A helping hand with language learning: Teaching French vocabulary with gesture. *Language Learning Journal*, *44*(2), 236–256.

Pugh, K. R., Sandak, R., Frost, S. J., Moore, D., & Mencl, W. E. (2005). Examining reading development and reading disability in English language learners: Potential contributions from functional neuroimaging. *Learning Disabilities Research and Practice*, *20*(1), 24–30.

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development*, *86*(1), 159–175.

Rasinski, T., & Young, C. (2017). Effective instruction for primary grade students who struggle with reading fluency. In M. Milton (Ed.), *Inclusive Principles and Practices in Literacy Education* (vol. 11, pp. 143–157). Emerald Publishing Limited.

Rau, A. K., Moll, K., Moeller, K., Huber, S., Snowling, M. J., & Landerl, K. (2016). Same same, but different: Word and sentence reading in German and English. *Scientific Studies of Reading*, *20*(3), 203–219.

Raudszus, H., Segers, E., & Verhoeven, L. (2018). Lexical quality and executive control predict children's first and second language reading comprehension. *Reading and Writing*, *31*(2), 405–424.

Reed, S. K. (2010). *Cognition: Theories and applications* (8th ed.). Wadsworth, Cengage Learning.

Reid, R., Gonzalez, J. E., Nordness, P. D., Trout, A., & Epstein, M. H. (2004). A meta-analysis of the academic status of students with emotional/behavioral disturbance. *The Journal of Special Education*, *38*, 130–144.

Retelsdorf, J., Koeller, O., & Moeller, J. (2014). Reading achievement and reading self-concept – Testing the reciprocal effects model. *Learning and Instruction, 29,* 21–30.

Rieber, L. P. (1994). *Computers, graphics, and learning*. Brown & Benchmark.

Richards-Tutor, C., Baker, D. L., Gersten, R., Baker, S. K., & Smith, J. M. (2016). The effectiveness of reading interventions for English learners: A research synthesis. *Exceptional Children, 82*, 144 –149.

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, *24*(7), 1301–1308.

Roberts, G. J., Solis, M., Ciullo, S., McKenna, J. W., & Vaughn, S. (2015). Reading interventions with behavioral and social skill outcomes: A synthesis of research. *Behavior Modification*, *39*, 8–42.

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*, 17–47.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A metaanalytic review. *Journal of Educational Psychology*, *95*, 240–257.

Roney, R. C. (1996). Storytelling in the classroom: Some theoretical thoughts. *Storytelling World*, *9*, 7–9.

Ryan, J. B., Reid, R., & Epstein, M. H. (2004). Peer-mediated intervention studies on academic achievement for students with EBD: A review. *Remedial and Special Education*, *25*(6), 330–341.

Sadoski, M. (2005). A Dual Coding View of Vocabulary Learning. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 21*(3), 221–238.

Sáenz, L. M., Fuchs, L. S., & Fuchs, D. (2005). Peer-assisted learning strategies for English language learners with learning disabilities. *Exceptional Children*, *71*(3), 231–247.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Reading interventions for adolescent struggling readers: A meta-analysis with implications for practice*. RMC Research Corporation, Center on Instruction.

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *Journal of Learning Disabilities*, *48*(4), 369–390.

Scanlon, D. M., Anderson, K. L., & Sweeney, J. M. (2017). *Early intervention for reading difficulties: The interactive strategies approach*. Guilford.

Schaars, M. M. H., Segers, E., & Verhoefen, L. (2017). Predicting the integrated development of word reading and spelling in the early primary grades. *Learning and Individual Differences*, *59*, 127–140.

Schaars, M. M. H., Segers, E., & Verhoeven, L. (2019). Cognitive and linguistic precursors of early first and second language reading development. *Learning and Individual Differences*, *72*, 1–14.

Schabmann, A., Schmidt, B. M., Klicpera, C., Gasteiger-Klicpera, B., & Klingebiel, K. (2009). Does systematic reading instruction impede prediction of reading a shallow orthography? *Psychology Science Quarterly*, *51*, 315–338.

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.

Schleicher, A. (2019). *PISA 2018. Insights and interpretations*. OECD. https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf

Schmidt, R. J., Rozendal, M. S., & Greenman, G. G. (2002). Reading instruction in the inclusion classroom. *Remedial and Special Education*, *23*(3), 130–140.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913–951.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, *95*, 26–43.

Schrauben, K. S., & Witmer, S. E. (2019). Feedback provided within structured reading programs: A systematic review. *Read. Writ. Q.,* 36, 1–18.

Schroeter, P., & Schroeder, S. (2017). The impact of L2 German on component processes of reading. *Journal of Research in* Reading, *40*(S1), 107–124.

Schulte-Koerne, G. (2011). Lese- und Rechtschreibstoerung im Schulalter - neuropsychologische Aspekte *[Reading and spelling disorder at school age - neuropsychological aspects]*. *Zeitschrift fuer Psychiatrie, Psychologie und Psychotherapie*, *59*(1), 47–55.

Schwab, S., Seifert, S., & Gasteiger-Klicpera, B. (2013). Reading intervention in second-grade children with poor reading abilities. *Procedia—Social and Behavioral Sciences*, *106*, 2205–2216.

Scruggs, T. E., Mastropieri, M., & Marshak, L. (2012). Peer-mediated instruction in inclusive secondary social studies learning: Direct and indirect learning effects. *Learning Disabilities Research & Practice*, *27*(1), 12–20.

Shapiro, L. R., & Solity, J. (2016). Differing effects of two synthetic phonics programmes on early reading development. *British Journal of Educational Psychology*, *86*, 182–203.

Sheffield, K. I. M., & Waller, R. J. (2010). A review of single-case studies utilizing selfmonitoring interventions to reduce problem classroom behaviors. *Beyond Behavior*, *19*(2), 7–13.

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

Smith, S. W., & Taylor, G. G. (2010). Education students with emotional and behavioral disorder. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (vol. 2, pp. 678–687). Elsevier.

Solis, M., Miciak, J., Vaughn, S., & Fletcher, J. M. (2014). Why intensive interventions matter: Longitudinal studies of adolescents with reading disabilities and poor reading comprehension. *Learning Disability Quarterly*, *37*, 218–229.

Solis, M. M., Ciullo, S., Vaughn, S., Pyle, N., Hassaram, B., & Leroux, A. (2012). Reading comprehension interventions for middle school students with learning disabilities: A synthesis of 30 years of research. *Journal of Learning Disabilities*, *45*, 327–340.

Sparks, R. (2006). Is there a "disability" for learning a foreign language? *Journal of Learning Disabilities*, *39*(6), 544–557.

Sparks, R. (2021). Identification and characteristics of strong, average, and weak foreign language readers. The simple view of reading model. *Modern Language Journal*.

Sparks, R., & Javorsky, J. (2000). Section 504 and the Americans with Disabilities Act: Accommodating the learning-disabled student in the foreign language curriculum (an update). *Foreign Language Annals*, *33*, 645–654.

Sparks, R., & Patton, J. (2016). Examining the simple view of reading model for United States high school Spanish students. *Hispania*, *99*, 17–33.

Sparks, R., Ganschow, L., & Patton, J. (2008a). L1 and L2 literacy, aptitude, and affective variables as discriminators among high- and low-achieving L2 learners with special needs. In J. Kormos & E. Kontra (Eds.), *Language learners with special needs: An international perspective* (pp. 11–35). Multilingual Matters.

Sparks, R., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term crosslinguistic transfer of skills from L1 to L2. *Language Learning, 59*(1), 203–243. https://doi.org/10.1111/j.1467-9922.2009.00504.x

Sparks, R. L. (2009). If you don't know where you're going, you'll wind up somewhere else: The case of foreign language learning disability. *Foreign Language Annals*, *42*, 7–26.

Sparks, R. L. (2016). Myths About Foreign Language Learning and Learning Disabilities. *Foreig Language Annals, 49*(2), 252-270.

Sparks, R. L., & Luebbers, J. (2018). How many U.S. high school students have a foreign language reading "disability"? Reading without meaning and the simple view. *Journal of Learning Disabilities*, *51*(2), 194–208.

Sparks, R. L., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2008b). Early first-language reading and spelling skills predict later second-language reading and spelling skills. *Journal of Educational Psychology, 100*(1), 162–174.

Sperling, M., Barwasser, A., and Gruenke, 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 Learn. Disabil. 16*, 79–90.

Spies, T. G., & Dema, A. A. (2014). Beyond word meaning: Vocabulary instruction for students with exceptional language needs. *Intervention in School and Clinic*, *49*, 271–280.

Spinath, B., Stiensmeier-Pelster, J., Schoene, C., & Dickhaeuser, O. (2012). *Skalen zur Erfassung der Lern- und Leistungsmotivation [Scales for assessing learning and achievement motivation]* (2nd ed.). Hogrefe.

Stæhr, L. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning*, *36*, 139–152.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, *31*(4), 577–607.

Stanley, C. T., Petscher, Y., & Catts, H. (2018). A longitudinal investigation of direct and indirect links between reading skills in kindergarten and reading comprehension in tenth grade. *Reading and Writing*, *31*(1), 133–153.

Stanovich, K. E. (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education*, *189*(1–2), 23–55.

Steenbeek-Planting, E. G., van Bon, W. H. J., & Schreuder, R. (2012). Improving word reading speed: individual differences interact with a training focus on successes or failures. *Reading and Writing*, *25*, 2061–2089.

Steinle, P. K., Stevens, E., & Vaughn, S. (2021). Fluency interventions for struggling readers in grades 6 to 12: A research synthesis. *Journal of Learning Disabilities*.

Stenhoff, D. M., & Lignugaris/Kraft, B. (2007). A review of the effects of peer tutoring on students with mild disabilities in secondary settings. *Exceptional Children*, *74*(1), 8–30.

Stevens, E. A., Walker, M. A., & Vaughn, S. (2017). The effects of reading fluency interventions on the reading fluency and reading comprehension performance of elementary students with learning disabilities: A synthesis of the research from 2001 to 2014. *Journal of Learning Disabilities*, *50*(5), 576–590.

Stockard, J., Wood, T. W., Coughlin, C. & Rasplica Khoury, C. (2018). The Effectiveness of Direct Instruction Curricula: A Meta-Analysis of a Half Century of Research. *Review of Educational Research, 88*, 479–507.

Strickland, W. D., Boon, R. T., & Spencer, V. G. (2013). The effects of repeated reading on the fluency and comprehension skills of elementary-age students with learning disabilities (LD), 2001–2011: A review of research and practice. *Learning Disabilities: A Contemporary Journal, 11*(1), 1–33.

Stuart, M. (2004). Getting ready for reading: A follow-up study of inner city second language learners at the end of key stage I. *British Journal of Educational Psychology*, *74*, 15–36.

Stuart, W. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33–52.

Suggate, S. P., Lenhard, W., Neudecker, E., & Schneider, W. (2013). Incidental vocabulary acquisition from stories: Second and fourth graders learn more from listening than reading. *First Language*, *33*(6), 551–571.

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

Swanson, H. L. (2016). Working memory and strategy instruction in children with learning disabilities. In R. Schiff & R. Malatesha Joshi (Ed.), *Interventions in learning disabilities. A handbook on systematic training programs for individuals with learning disabilities* (pp. 227–242). Springer.

Swanson, L., & Stomel, D. (2012). Learning disabilities and memory. In B. Wong & D. Butler (Eds.), *Learning about learning disabilities* (4th ed., pp. 27–58). Academic Press.

Swanson, H. L., Orosco, M. J., & Lussier, C. M. (2015). Growth in literacy, cognition, and working memory in English language learners. *Journal of Experimental Child Psychology*, *132*, 155–188.

Taboada, A., Tonks, S. M., Wigfield, A., & Guthrie, J. T. (2009). Effects of motivational and cognitive variables on reading comprehension. *Reading and Writing*, *22*(1), 85–106.

Tate, R. L., Perdices, M., Rosenkoetter, U., McDonald, S., Togher, L., & Vohra, S. (2016). The single-case reporting guideline in behavioural interventions (SCRIBE) 2016: Explanation and elaboration. *Archives of Scientific Psychology*, *4*(1), 10–31.

Thaler, V., Ebner, E. M., Wimmer, H., & Landerl, K. (2004). Training reading fluency in dysfluent readers with high reading accuracy: Word specific effects but low transfer to untrained words. *Annals of Dyslexia*, *54*, 89–113.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading. *Remedial and Special Education, 25,* 252–261.

Thomas, G., & Loxley, A. (2007) *Deconstructing special education and constructing inclusion*. Open University Press.

Tingstrom, D. H., Sterling-Turner, H. E., & Wilczynski, S. M. (2006). The good behavior game: 1969–2002. *Behavior Modification*, *30*, 225–253.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, *34*, 33–58.

Toste, J. R., Capin, P., Williams, K. J., Cho, E., &Vaughn, S. (2019). Replication of an experimental study investigating the efficacy of a multisyllabic word reading intervention with and without motivational beliefs training for struggling readers. *Journal of Learning Disabilities*, *52*(1), 45–58.

Troia, G. A. (2009). *Instruction and assessment for struggling writers.* Guilford.

United Nations. (2006). *Convention on the Rights of Persons with Disabilities*. http://www.refworld.org/docid/4680cd212.html

Vadasy, P. F., & Sanders, E. A. (2012). Two-year follow-up of a kindergarten phonics intervention for English learners and native English speakers: Contextualizing treatment impacts by classroom literacy instruction. *Journal of Educational Psychology*, *104*, 987–1005.

Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, *65*(2), 390–416.

van de Ven, M., de Leeuw, L., van Weerdenburg, M., & Steenbeek- Planting, E. G. (2017). Early reading intervention by means of a multicomponent reading game: Intervention affects reading game. *Journal of Computer Assisted Learning, 33*(4), 320–333.

van Gelderen, A., Schoonen, R., Stoel, R. D., de Glopper, K., & Hulstijn, J. (2007). Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology*, *99*(3), 477–491.

van Gorp, K., Segers, E., & Verhoeven, L. (2014). Repeated reading intervention effects in kindergartners with partial letter knowledge. *International Journal of Disability, Development & Education*, *61*(3), 225–239.

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.

van Hell, J. G., Bosman, A. M. T., & Bartelings, M. C. G. (2003). Visual dictation improves the spelling performance of three groups of Dutch students with spelling disabilities. *Learning Disability Quarterly*, *26*, 239–255.

Vannest, K. J., Temple-Harvey, K. K., & Mason, B. A. (2009). Adequate yearly progress for students with emotional and behavioral disorders through research-based practices. *Preventing School Failure*, *53*, 73–83.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening. *System, 41*, 609–624.

Vaughn, S., Levy, S., Coleman, M., & Bos, C. S. (2002). Reading instruction for students with LD and EBD: A synthesis of observation studies. *The Journal of Special Education*, *36*(1), 2–13.

Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Bryant, D. P., Dickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education*, *24*, 301–315.

Vaughn, S., Martinez, L. R., Linan-Thompson, S., Reutebuch, C. K., Carlson, C. D., & Francis, D. J. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English language learners: Findings from two experimental studies. *Journal of Research on Educational Effectiveness*, *2*(4), 297–324.

Vogel, G., Fresko, B., & Wertheim, C. (2007). Peer tutoring for college students with learning disabilities: perceptions of tutors and tutees. *Journal of Learning Disabilities*, *40*(6), 485–493.

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 off German schoolchildren. *Journal of School Psychology*, *66*, 25–40.

von Hagen, A., Kohnen, S., & Stadie, N. (2021). Foreign language attainment of children/adolescents with poor literacy skills: A systematic review and meta-analysis. *Educational Psychology Review*, *33*, 459–488.

Wanzek, J., Wexler, J., Vaughn, S., & Ciullo, S. (2010). Reading interventions for struggling readers in the upper elementary grades: A synthesis of 20 years of research. *Reading and Writing*, *23*, 889–912.

Webb, S., & Chang, A. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Journal*, *68*(3), 267–290.

Webb, S., & Chang, A. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*, 667–686.

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.

Wells, G. (2009). The social context of language and literacy development. In O. A. Barbarin & B. H. Wasik (Eds.), *Handbook of child development & early education* (pp. 271–302). The Guilford Press.

Wendt, H., & Schwippert, K. (2017). Lesekompetenzen von Schuelerinnen und Schuelern mit und ohne Migrationshintergrund [Reading competences of pupils with and without a migration background]. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, et al. (Eds.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich [Reading competences of primary school children in Germany in international comparison]* (pp. 219–234). Waxmann.

Wendt, H., Smith, D. S., & Bos, W. (2016). Germany. In I. V. S. Mullis, M. O. Martin, S. Goh, & K. Cotter (Eds.), *TIMSS 2015 encyclopedia: Education policy and curriculum in mathematics and science*. http://timssandpirls.bc.edu/timss2015/encyclopedia/

Wexler, J., Kearns, D. M., Lemons, C. J., Mitchell, M., Clancy, E., Davidson, K. A., & Wei, Y. (2018). Reading comprehension and co-teaching practices in middle school English language arts classrooms. *Exceptional Children*, *84*(4), 384–402.

Whitford, V., & Joanisse, M. F. (2018). Do eye movements reveal differences between monolingual and bilingual children's first-language and second-language reading? A focus on word frequency effects. *Journal of Experimental Child Psychology*, *173*, 318–337.

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 433.

Wilkins, D. A. (1972). *Linguistics in Language Teaching*. Hodder & Stoughton Educational.

Wills, H., Kamps, D., Abbott, M., Bannister, H., & Kaufman, J. (2010). Classroom observations and effects of reading interventions for students at risk for emotional and behavioral disorders. *Behavioral Disorders*, *35*, 103–119.

Wimmer, H., & Mayring, H. (2014). Salzburger Lese-Screening fuer die Schulstufen 2-9: SLS 2-9 [Salzburg Reading Screening for school levels 2-9: SLS 2-9]. Hogrefe.

Wimmer, H., Schurz, M., Sturm, D., Richlan, F., Klackl, J., Kronbichler, M., & Ladurner, G. (2010). A dual-route perspective on poor reading in a regular orthography: An fMRI study. *Cortex*, *46*, 1284–1298.

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading*, *5*, 211–239.

Wright, W. E. (2010). *Foundations for Teaching English Language Learners: Research, Theory, Policy, and Practice*. Caslon Publishing.).

Wu, S., Stratton, K. K., & Gadke, D. L. (2020). Maximizing repeated readings: The effects of a multicomponent reading fluency intervention for children with reading difficulties. *Contemporary School Psychology*, *24*(2), 217–227.

Yada, A., & Savolainen, H. (2017). Japanese in-service teachers' attitudes toward inclusive education and self-efficacy for inclusive practices. *Teaching and Teacher Education*, *64*, 222–229.

Yeung, S. S., & Savage, R. (2020). Teaching grapheme–phoneme correspondences using a direct mapping approach for at-risk second language learners: A randomized controlled trial. *Journal of Learning Disabilities*, *53*(2), 131–144.

Yildirim, K., & Rasinski, T. V. (2014). Reading fluency beyond English: Investigations into reading fluency in Turkish elementary students. *International Electronic Journal of Elementary Education*, *7*(1), 97–106.

Young, D., Schaefer, M. Y., & Lesley, J. (2019). Accommodating students with disabilities studying English as a foreign language. *Journal of Postsecondary Education and Disability*, *32*(3), 311–319.

Yousefi, M. H., & Biria, R. (2018). The effectiveness of L2 vocabulary instruction: A meta-analysis. *Asian-Pacific Journal of Second and Foreign Language Education*, *3*, 21.

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*.

Zarić, J., Hasselhorn, M., & Nagler, T. (2021). Orthographic knowledge predicts reading and spelling skills over and above general intelligence and phonological awareness. *European Journal of Psychology of Education*, *36*, 21–43.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29.

Ziegler, J. C., Bertrand, D., Lété, B., & Grainger, J. (2014). Orthographic and phonological contributions to reading development: Tracking developmental trajectories using masked priming. *Developmental Psychology*, *50*, 1026–1036.

Appendix A Article 1

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.

Abstract

Reading difficulties that are not addressed at the primary level continue to exist at the secondary level with serious consequences. Thus, it is important to provide struggling students with specific reading support. In particular, many students with learning disabilities (LD) and emotional behavioral disorders (EBD) demonstrate reading obstacles and are at risk for motivation loss. A multiple baseline design was used to evaluate the effects of a motivational reading racetrack as peer-tutoring on the word reading skills of secondary students with LD with and without EBD. The intervention was conducted through 4-5 baseline and 16-18 reading units three times a week for 15 minutes over 8 weeks. The results showed positive effects indicating a highly effective treatment. In addition, follow-up results were also promising. Our findings indicate that this multicomponent intervention has a positive effect on the word fluidity of low-achieving students in secondary education with LD and/or EBD.

Keywords: Reading fluency, reading racetracks, peer tutoring, learning and behavioral problems, multiple baseline design

Introduction

The Importance of Reading at the Secondary Level

Difficulties in reading at the secondary level are considered more serious than reading challenges at the primary level (Guerin & Murphy, 2015). Yet, the training of reading fluency is mainly carried out in the lower classes, as it is assumed that this is one of the tasks of primary school teachers (Rasinski et al., 2009). Thus, the promotion of reading at secondary level is often neglected (Edmonds et al., 2009). As a result, students with reading difficulties move further and further away from their typically performing peers, with the result that many fail to meet the requirements for each grade level. A recent edition of the Program for International Student Assessment (PISA) revealed that compared to the PISA survey in 2015, the reading performance of German youth had worsened (European Commission, 2018). Specifically, 21% performed below level 2 in reading which can be seen as high. Also, the survey showed that struggling German 15-year-olds did not enjoy reading as much as youth in other countries (Organisation for Economic Cooperation and Development [OECD], 2019). Acquiring the reading skills necessary to become successful far beyond school is a major challenge for many

students. Reading proficiency requires many complex steps. For example, lower-level processing skills such as decoding and reading fluency are necessary to advance towards higher-level skills such as reading comprehension (Chard et al., 2002; Kim et al., 2011).

Hurdles in Achieving Reading Proficiency

On the road to reading proficiency, fluency is extremely important as it functions as a bridge between decoding and understanding a certain text; thus, without fluency, working memory (WM) capacities are used to simply decode a text, leaving little effort left to spend on attention to content (Juffs & Harrington, 2011) and, consequently, poorer comprehension. While stronger readers do decoding and vocabulary retrieval automatically via long-term memory, weaker readers have to consume more WM resources to improve reading, and especially sight word reading (Peng et al., 2018; Sweller, 1994). A meta-analysis by Peng et al. (2018) found a moderate relation between WM and reading (r = .29). Specifically, WM and word recognition were more strongly related than WM and non-word reading, and WM was more related to word reading than sentence reading. Barriers in reading fluency arise primarily from poor automation of reading sight words, resulting in poor mastery of decoding skills (Ayala & O'Connor, 2013). Deficits in automation in word recognition, in turn, poses a tremendous challenge to reading comprehension (Perfetti & Stafura, 2014). Coltheart et al. (2001), in turn, proposed a “dual-route theory” with regard to reading acquisition consisting of a lexical route and a non-lexical route. Using the *lexical route* (orthographic decoding) words are accessed quickly, whereas the *non-lexical route* (phonological recoding) consists of decoding individual words to be read, making this a more arduous process. Students with hurdles in the area of learning tend to rely on the non-lexical route since they struggle with storing information properly and, as a result, experience challenges in retrieving information rapidly. But word recognition is needed in order to become a proficient reader and thus, needs to get early attention. For the German language, Knoepke et al. (2014) showed that skills on the lexical route predict text comprehension better than skills on the non-lexical route. This underlines the importance of promoting the lexical route. Moreover, for German, which tends to be one of the transparent orthographies, students with reading difficulties face hurdles especially in automated direct word recognition (Landerl & Wimmer, 2008).

Students who did not learn word recognition skills in the earlier grades will most likely have reading difficulties, not only in the higher grades but throughout adulthood as well (Leffingwell, 2016). Ehri (2005) developed a model that deals specifically with the lexical path and word recognition. This model consists of the following stages: pre-alphabetic, partial alphabetic, full alphabetic, and consolidated alphabetic phase, which describes the degree to which readers make memory connections between the written word and pronunciation. Automated consolidated words enable the reader to master reading by quickly and unconsciously retrieving a word from the mental lexicon via the lexical route (Ehri, 2005).

Students With Learning Disabilities and Emotional Behavioral Disorders

The majority of students with learning disabilities (LD) demonstrate hurdles in reading (Lerner & Johns, 2011), primarily reading fluency (Chard et al., 2002), due to challenges with processing information. Further, many students lose motivation to read and learn, and, understandably, get frustration (Martin et al., 2008). These factors may explain the PISA results with respect to reading motivation among German youth mentioned earlier.

Students with emotional behavioral disorders (EBD) present a growing challenge within the school setting (Forness et al., 2012). Problem behavior often has a negative effect on students' school careers (Chow & Wehby, 2018; Nelson et al., 2004), including a risk of kids dropping out of school (Bradley et al., 2008). Within the current context, students who face challenges with reading, spelling, and/or math often display inappropriate and aggressive behavior (Auerbach et al., 2008; Pierce et al., 2013). Additionally, it has been reported that students with behavioral issues have a higher risk of deficits in language compared to their peers without behavioral challenges, especially with respect to reading skills (Benner et al., 2002; Hilsmer et al., 2016; McCabe & Meller, 2004). A meta-analysis by Hollo et al. (2014) estimates that 81% of students with EBD have negative experiences with reading and writing that go unnoticed for a long time, as the main focus is on fostering appropriate behavior. Given the importance of reading literacy, the large number of underachieving secondary school students, and the high correlation between LD, EBD and inadequate reading proficiency and decreasing motivation, an intervention that addresses all of these is critically important.

Ways to Foster Reading Competency

Repeated reading and sight word training

In order to effectively combine the previous components and integrate them into a reading intervention, the method of repeated reading (RR) at the word level can be introduced as a core element. A synthesis by Stevens et al. (2017) revealed that RR interventions positively affected the reading fluency of students with LD. Moreover, small positive effects were also found with respect to comprehension. These findings concur with those of Chard et al. (2002) and support the use of drill-and-practice methods for automation. For example, in their study with sixth-grade students with LD and EBD Escarpio and Barbetta (2016) found that when the students read the same material repeatedly and got feedback from a tutor, they were able to read more words per minute and performed better on a reading comprehension test.

The addition of Ehri's model (2005) and the relevance of the lexical route helps make sight word training an effective option for improving reading proficiency. When teaching words, it is important to provide numerous opportunities to practice the specific words and give feedback. A meta-analysis by Scammacca et al. (2007) showed that older students with reading difficulties with and without LD (4th-12th graders) benefited from interventions that were focused on the word level. A follow-up meta-analysis by the authors (Scammacca et al.,

2015 reached a similar conclusion, showing the benefit of reading training at the word level. Thus, both studies showed that children can benefit from reading support up to grade 12, making it particularly relevant for secondary school readers who face severe failure in reading.

Reading racetracks

Repetitive sight-word reading can be embedded in a reading racetrack procedure. A racetrack consists of empty cells equipped with little flashcards including content such as phonemes, words, or mathematical exercises to be trained extensively (Erbey et al., 2011). While this procedure has been shown to be effective with second-language learners and students with diverse disabilities (Alexander et al., 2008; Gruenke, 2019; Gruenke & Barwasser, 2019; Hopewell et al., 2011; Sperling et al., 2019), to date it has not been investigated with secondary school students with LD with and without EBD.

Peer tutoring as a tool for inclusion

To make an intervention an inclusive tool, peer-tutorial learning can be added by having weaker and stronger children practice together. In general, peer-tutoring procedures are known for having a beneficial influence when being embedded in interventions (Mercer et al., 2011). These results are supported by the review of Stenhoff and Lignugaris/Kraft (2007) for secondary students in heterogeneous peer-tutoring settings. Okilwa & Shelby's (2010) literature synthesis points in the same direction by showing that peer tutoring effects academic achievement positive in a variety of subject areas for 6- to 12-year-olds regardless of their type of disability (learning disability, emotional or behavioral disability, and intellectual disability). This is also confirmed in the meta-analyses by Bowman-Perrott et al. (2013) and Moeyaert et al. (2021) for single case data, which show that peer-tutoring has an significant effect on both academic (see also McDuffie et al, 2009) and social-behavior outcomes. In the meta-analysis by Bowman-Perrott et. al. (2013) those with emotional and behavioral disorders benefitted most whereas Moeyaert et al. (2021) revealed a slightly larger effect on academic outcomes. With regard to reading skills a study by Calhoon (2005) found positive effects of peer tutoring with low-reading middle school students on phonological skills and reading comprehension, but not on reading fluency. However, it should be noted, that reading fluency was not taught directly, suggesting that peer tutoring might be effective on reading fluency as part of reading fluency training. The results regarding reading comprehension for secondary students with disabilities were also confirmed by a review of Alzahrani & Leko (2018). In general, peer-procedures seem to be beneficial in secondary special education (King-Sears, 2021). Considering students with reading and behavioral problems results show that when two students are working together in order to improve specific content, reading competency can be enhanced both for those with and without problem behavior (Bowman-Perrott et al., 2013).

The Advantages of Incorporating Motivational Components

Considering the findings of the PISA study in the context of motivation and the result that especially secondary school students with reading hurdles lose motivation and enjoyment in reading (OECD, 2019), there is an urgent need for motivational reinforcers to transform the reading experience into a more positive one for many students.

Group contingencies and self-graphing procedures

Elements such as group contingencies (GC, rewards dependent on group performance) and self-monitoring have also been demonstrated to be beneficial in the classroom as a means of supporting learning. The use of amplification systems (Bowman-Perrott et al., 2013) or, more specifically, GC procedures (Rohrbeck et al., 2003; Slavin, 1995) are particularly effective. In the implementation of tutorial learning, use of the GC procedures is a key success factor. Thus, research results confirm that procedures in which GCs are implemented, on average, achieve better results in terms of learning outcomes (Rohrbeck et al., 2003; Slavin, 1995) and improved social skills (Ginsburg-Block et al., 2006). Especially, interdependent group contingencies (IGC)procedures, in particular, have been found to be predictors of the success of peer-supported learning (Ginsburg-Block et al., 2006; Rohrbeck et al., 2003). Thus, studies using amplifiers had significantly greater effects on learning gains (i.e., Rohrbeck et al., 2003). Rohrbeck et al. (2003) published significant effects of using group reward contingencies in peer interventions (p<.05, g = 0.34 (with GC; g = 0.26 (without GC). Further, Popkin and Skinner (2003) showed that the use of specifically IGC has a positive effect on performance in different areas.

Self-monitoring procedures, which are related to self-regulation, have also proven to be beneficial for increasing performance. For example, Richards et al. (1976) found that students who monitored themselves in reading had stronger performance gains than students who received training without self-monitoring. More recently, a study by Stotz and colleagues (2008) showed positive effects on the number of total written words and number of correct word sequences with the implementation of a self-graphing procedure. Finally, Menzies et al. (2009) suggested that self-graphing particularly for reading performance can have positive effects on motivation and engagement.

Apart from the need for motivational elements, the demand for effective instructional methods that can be implemented with a heterogeneous learning group is increasing, especially due to the increasing heterogeneity of today's classrooms.

Research Aim

Given the increasing number of struggling secondary school readers with LD with and without EBD and the resulting risk of loss in motivation, an intervention that has a positive effect on both reading and motivation is essential. To make such an intervention applicable to inclusive classrooms, and therefore appropriate for students with varying abilities, the addition

of a peer-tutorial procedure would be helpful. To fill the research gap on the issue of sight-reading and secondary-level students, the present study investigated whether older students with challenges in learning and behavior could benefit from a combined racetrack intervention. Thus, the core research question of the study was as follows: Does an intervention consisting of peer-tutorial reading tracks with gamified components have a positive impact on the word recognition of struggling secondary school students with LD with or without EBD?

Materials and Methods

Participants and Setting

Participants were 16 students with LD and EBD in grades 5-7 attending a low social-economic German urban special needs school in North Rhine-Westphalia. First, consent forms were sent to the parents of prospective participants, and data were only collected on students whose parents had agreed to the survey. Subsequently, a German reading screening (Salzburger Reading Screening [SLS]; Wimmer & Mayring, 2014) was used in a first step in all classes (5-7, $N = 37$) to identify students with a reading quotient (RQ) below 89 as a cutoff for lower reading performance.

With regard to the intervention, which was to include peer tutoring with struggling and more advanced readers, the students with a lower RQ (<79) were selected as tutees and those with a higher RQ (>100) as tutors. In order to compile the reading pairs, the values of the reading screening were ranked, and the rank was divided in the middle. The student with the lowest score on the first half was paired (low RQ) with the student with the lowest score on the second half (high RQ) according to Fuchs et al. (1997). Care was taken to ensure that the students in the pairs understood each other well, based on advice from the teachers.

Overall, however, the reading performance of the participating classes was below an RQ of 95. Thus, the overall reading performance fell in the lower range. The reading screening resulted in 18 participants (9 tutors and 9 tutees). Only data on the tutees were collected because the tutors had to be able to read the words to be trained fluently on the racetrack in order to be eligible to participate. One participant was not included in the data analysis due to missing data; consequently, only data on eight tutees are shown in the following.

All eight tutees from whom data were collected had been diagnosed with LD. Four of the students we also diagnosed with an EBD. In Germany, the diagnosis of LD is determined contingent on repeated serious school failure in several subjects and EBD can be defined in Germany as getting special educational support with the focus on emotional and social development when a student cannot be adequately supported at school due to behavioral difficulties and his or her own development or that of his or her classmates is significantly disturbed or endangered. Both, students with LD and/or EBD receive special needs support in schools. All participants were native speakers of German.

Table 1

Characteristics of Participants

Participants	Gender	Age	IQ	Special Needs	Reading Proficiency (LQ)	German L1
John	male	11;7	70-85	LD/EBD	78	Yes
Timo	male	12;1	70-85	LD/EBD	<62	Yes
Emma	female	12;5	70-85	LD/EBD	64	Yes
Levin	male	11;7	70-85	LD	69	Yes
Ben	male	12;9	70-85	LD	<62	Yes
Sam	male	12;3	70-85	LD/EBD	<62	Yes
Seba	male	14;3	70-85	LD	68	Yes
Lauren	female	13;1	70-85	LD	<62	Yes

Note. Learning Disabilities (LD); Emotional Behavioral Disorder (EBD); First Language (L1); Reading Quotient (LQ; <89 underdeveloped reading; <79 weak reading)

Design

A multiple baseline design within an AB plan (Ledford & Gast, 2018) was implemented with a total of 24 planned measurement points and three different baseline lengths. The reason for using a multiple baseline design was the experimental control it provides by decreasing the probability of alternative explanations for intervention effects (Byiers et al., 2012). Each group was supervised by one female master's level student of special needs education and was taken out of the classroom for both the baseline phase and the intervention phase and supported in extra rooms. Data were collected after baseline sessions and after each intervention session. The students were randomly divided into three groups. The first small group had a baseline length of four sessions, the second group of five, and the third group of six, after which the intervention began directly for each group. Thus, John, Timo, and Emma started with a baseline length of four, Levin and Ben with a length of five, and the remaining three, Sam, Seba and Lauren, with six baseline sessions. In total, the groups were taken out three times a week over eight weeks Monday, Wednesday, and Friday always at the same time. The follow-up measurements took place four weeks after the end of the intervention, two weeks of which were Christmas holidays.

Dependent Variables and Data Collection Procedure

The measuring tool was a researcher-made PowerPoint presentation with a 30-slide word sequence, into which words that were to be read out for 1 second each were visibly inserted with one word per slide (Ehri, 2005). Data from each tutee were collected after each baseline and intervention session to evaluate the impact of the intervention and a possible correlation to increases in single word reading. The number of correct and incorrect words read out loud was recorded. A word was considered correctly read if the tutee read the word within the 1-second interval of its occurrence. A word was considered to be misread if the tutee either omitted part of the word, added something to the word, or read it incorrectly. If the student

corrected him/herself before the next word appeared, the word read aloud was recorded as correct. At no point during the word test did the students receive any help or feedback. The training words were shown in a different order at each measurement (and also at follow-up). The measurements were carried out by master's-level students in pairs to ensure impartiality, with an interrater reliability of 100%.

Material

The material consisted of a playing field in A3 size (11.7 x 16.5 inches), which was divided into 30 fields and embedded in a reading racetrack. Each team received a small wooden figure (i.e., a race car or an animal) and a dice. In addition, the students were given 30 white laminated flashcards in an envelope. Each flashcard contained a different word. In order to find the respective words, a PowerPoint presentation (the same procedure used for the measurement) with 120 words was used before the start of the study to increase the probability of students finding 30 words that were not stored as sight words. The 120 words were two to four syllables long and reflected the most frequently occurring words in the German language. The selection of words was taken from a list published by the University of Leipzig (https://wortschatz. uni-leipzig.de/de). Care was taken to ensure that the words were of similar difficulty and did not exceed two or three syllables in length. From these 120 words, the final 30 training words for the racetrack were selected with a mean word mid frequency of 60.08. A stopwatch was used to measure time and a training sheet was used as a line chart to record the individual results of each team, as part of the reward system. The training sheet was comprised of 12 lines, listed one below the other. Each row, in turn, had 30 blank boxes for the maximum possible number of correctly read words per measurement.

Procedures

Baseline

For the baseline condition, all students worked in their small groups and the assigned pairs in cognitive exercises, focusing mainly on sorting symbols into the correct order. The students were assigned as either tutee (low reading) and or tutor (more advanced reading). The length of the baseline condition was the same as the racetrack intervention in phase B (15 minutes). Subsequently, the measurement was performed individually for each subject. The groups were conducted at the same time in three different rooms.

Intervention

In phase B (reading racetracks), the participants practised the 30 words selected from the previous word-tests repeatedly in the same group as well as in tutor/tutee team as in the baseline condition.

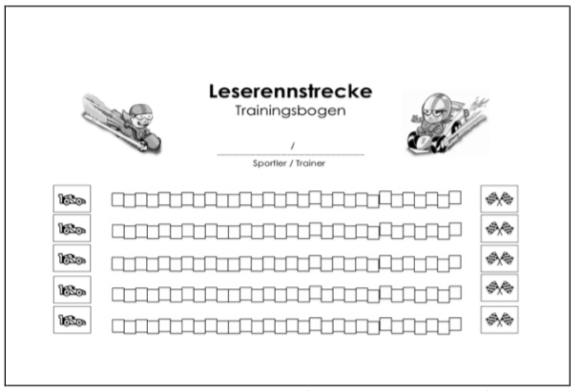
Prior to the start of the study, tutors were trained by the interventionists to provide feedback during the race game during a 1.5-hour training session. Tutors were given example situations, with the task being how they would respond as a tutor, and training on how an adequate tutor would respond. The tutors were then divided into tandems, with one of the tutors taking the role of the tutee and both playing the racetrack game as an example. The tutees in the study were not present for this training.

At the beginning of the intervention, the previously selected tandems, consisting of one tutee and one tutor, sat down at a table where the 30 index cards were placed on the board with the printed word facing down. In the first sessions, all tandems were given an intensive explanation of the racecourse procedure and the roles of tutor/tutee as coach/athlete. The tutors' role was to provide feedback and to correct if necessary and the tutee was asked to read around the racetrack. At the beginning, the tutees rolled the dice and moved their figure forward according to the number of points rolled. Then the card was turned over and the word printed on it was read aloud. Meanwhile, the tutors listened carefully, corrected, if there were no self-correction by the tutees within three seconds, and repeated the word again correctly. If the word was correct, the tutors praised the tutees and the tutees went on with their figure on the game board. During reading, the index cards remained on the table with the word facing up. When all the words were read, the deck was reshuffled and the game started again. After 10 minutes, a signal indicated that the game was over. Measurements were then taken for each tutee individually.

As a reward system, the children recorded the number of words read correctly by the tutees on a self-graphing sheet after each measurement in phase B to document their own learning progress. A line of 30 quarters represents one session and the quarters represent the number of words read correctly per session. Depending on the score achieved, there was a reward in the form of marbles. Tutees received one marble if they achieved the same number of words read correctly as last time, and two marbles if they improved. The marbles were kept in a container. A group target was set in terms of the number of marbles in the container, so that the whole group received a reward as a group contingency procedure. The reward system was intended to increase student motivation (Kim et al., 2011).

Figure 1

Self-Graphing Sheet



Note. Reading Racetracks (Leserennstrecke); Traininhg Sheet (Trainingsbogen); Athlet (Sportler)

Treatment Fidelity

To ensure treatment fidelity, a checklist was designed to be completed by the master's students after each session; in addition, for a third of the sessions, an external person filled out the questionnaire as well. The goal was to find out if the interventionists implemented the intervention as previously planned.

The checklist consisted of a table in which the subject codes were entered and whether they were present or not. Additional areas included "environment/framework conditions," "material," "course of support," "diagnostics and feedback," and "dealing with student behavior," each with several items to be answered on a 5-point Likert scale from 0 = "does not apply at all" to 4 = "applies completely." These areas were measured to ensure that the intervention was performed in exactly the same way in all three groups.

Before the study started, the first author gave a detailed briefing on the screening and conducting the baseline condition and intervention for two days in a row. In addition, a detailed guide was developed on how to conduct the study along with a time schedule. The first author was in regular weekly contact with the interventionists. The treatment fidelity agreement was 100%.

Social Validity

To measure social validity, after the study participating students were asked to rate the following eight items with the help of a self-designed questionnaire using a 5- point Likert scale ranging from 0 ("totally disagree") to 4 ("totally agree").

1) The racetrack helped me to read words correctly
2) I think the support also helps other students with reading difficulties
3) I understood the meaning of the intervention well

4) I learned a lot during the intervention
5) I enjoyed coming to the intervention
6) I would participate in the intervention again
7) The words were difficult
8) I enjoyed playing in pairs

Data analysis

Analyses were conducted using the SCAN package for R by Wilbert & Lueke (2019). First, a visual inspection is performed, and the descriptive data are presented. For a more in-depth analysis, overlap measures will be used to determine the effectiveness of the intervention, and a level 2 regression analysis will be conducted across all subjects, focusing on the slope, the increase from the A phase to the B phase, and the level effect of whether there is a direct increase in the onset of the intervention. Within the overlap measures, we use the Non-Overlap of All Pairs (NAP, Parker et al., 2011a), the Percentage Exceeding the Median (PEM, Ma, 2006), the Percentage of All Non-Overlap data (PAND; Parker et al., 2007), and the Tau-U derived from Kendall's rank correlation and Mann-Whitney U with possible A-phase trend correction (Parker et al., 2011b; A vs. B + TrendB - TrendA). The NAP is the percent improvement in data across phases, with 0-.65 indicating a weak effect, .66-.92 a moderate effect, and .93-1.0 a large effect. The PEM is the percentage of data points that exceed the median of the baseline. Less than .7 is a non-effective treatment, .7-.9 is a moderate effect, and above .9 is a large effect. PAND is the total number of data points that do not overlap between phases, with individual data points not biased by outliers. 50-70% is a weak effect, 70%-90% is a medium effect, and above 90% is a large effect. The Tau-U values can be divided into: up to 0.20 improvement can be considered as small change, 0.20 to 0.60 as moderate change, 0.60 to 0.80 as large change and above 0.80 as very large change.

Results

Visually, it was clear that two participants, John and Seba, started with higher values in the baseline, with Seba stabilizing at the end and a downward trend for John. Possible positive baseline trends can be seen for Emma and Lauren. All other baselines appear to be low and flat. In phase B, a rapid increase in the number of correctly read words can be seen for all students, with some even showing a ceiling effect. The follow-up data can be described as relatively stable, with all probands showing a slight decline in value, but still well above the values for phase A.

Table 2

Descriptive Data for Each Participant in Phases A, B, and E

	N (A)	N (B)	N (E)	M (A) (*SD*)	M (B) (*SD*)	Max (B)	M (E) (*SD*)
John	4	18	3	17.75 (3.30)	28.94 (1.68)	30.00	28.33(1.53)
Timo	4	18	3	5.00(1.41)	17.83(7.89)	27.00	18.67(0.58)
Emma	4	18	3	6.00 (3.46)	26.83(3.50)	30.00	25.67(1.53)
Levin	5	17	3	1.00(0.71)	14.12(5.27)	21.00	11.67(1.53)
Ben	5	17	3	2.00(1.22)	16.53(7.31)	26.00	15.33(1.15)
Sam	6	16	3	0.67(0.82)	9.00(4.77)	17.00	9.67(1.53)
Seba	6	16	3	9.80(2.17)	26.38(4.41)	30.00	26.33(1.53)
Lauren	6	16	3	1.67(1.63)	11.50(5.19)	18.00	12.33(1.15)

Note. Baseline (A); Intervention (B); Follow-up (E); Measurements (N); Mean (M); Standard Deviation (SD); Maximum Value (Max)

Overall, the average mean value in phases A, B., and E (follow-up) was 5.49, 18.90, and 18.50, respectively. This means that there was an overall increase of 1,790% from phase A to phase B. Three of the students achieved the maximum value of 30 during the intervention compared to the minimum value of 17 in phase B.

Table 3

Comparison of Overlap Indices for Number of Correct Recognized Words in Phases A and B

Participants	NAP	*p*	PAND	Tau-U	*p*
John	100	<.001	100	0.51	<.001
Timo	92.00	<.01	84.10	0.84	<.001
Emma	100	<.01	100	0.68	<.001
Levin	100	<.001	100	0.81	<.001
Ben	100	<.01	100	0.93	<.001
Sam	99.00	<.001	95.50	0.84	<.001
Seba	100	<.001	100	0.65	<.001
Lauren	93.00	<.001	90.90	0.80	<.001

Note. Non-Overlap of All Pairs (NAP); Percentage of All Non-Overlapping Data (PAND)

Regarding the NAP, all students achieved high values, ranging from 99.00-100.00, except for Timo, who reached a value of 92.00. These results can be interpreted as statistically significant either at the <.01 level or <.001 level.

For the PAND, a mean effect of 84.10 was found for Timo and a high effect size with values from 90.90-100.00 for the rest of the sample. Weighted Tau-U scores (A vs. B + trend B – trend A) showed a moderate effect for John (p<.001) and a large change for Lauren, (p<.001), Seba (p<.001) and Emma (p<.001). For Timo, Levin, and Sam a very large change was observed (p<.001). Furthermore, all results were statistically significant.

Table 4

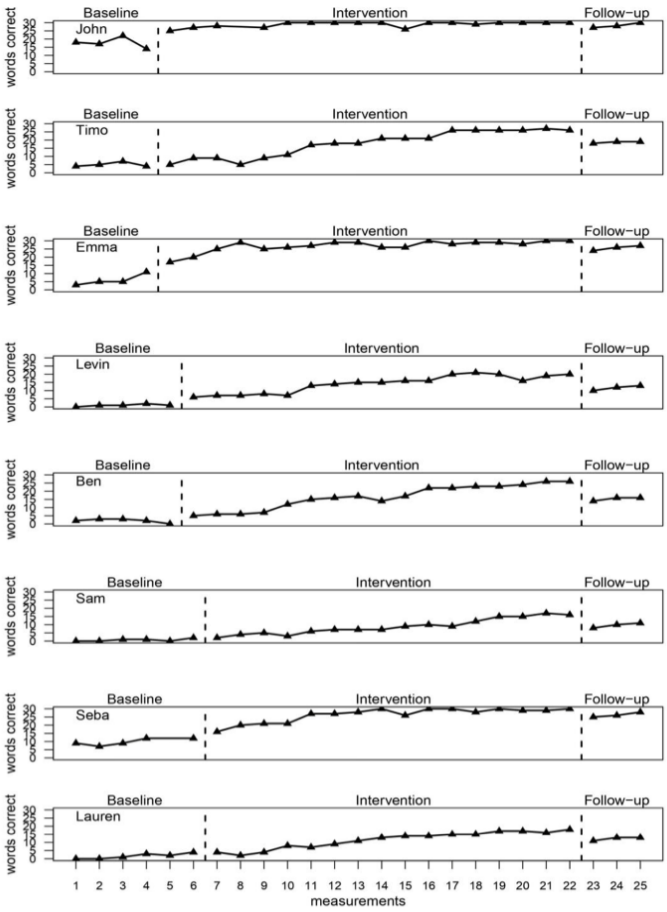
Regression Model for Number of Correct Recognized Words (Level-2 Analysis)

	B	SE	*t*	*p*
Intercept	4.811	2.516	1.912	.06
Trend	0.303	0.284	1.066	.29
LevelB	4.454	0.920	4.841	<.001
LevelE	4.524	5.663	0.799	.43
SlopeB	0.596	0.288	2.073	<.05
SlopeE	0.947	0.780	1.214	.23

The regression analysis across all participants at level 2 displayed a statistically significant level effect from phase A to phase B (p<.001). A statistically significant slope effect with an average increase of 0.60 words read correctly per session was found when comparing the two phases (p<.05). Furthermore, no statistically significant difference was found for the E phase compared to the B phase.

Figure 2

Amount of words read correctly for each participant



Social Validity

After the intervention, participating students were asked to complete the social validity questionnaire anonymously. Overall, they rated the intervention very positively on all issues. The highest score of *M* = 4.00 and an *SD* = 0, was given to items 2 (“I think the support also helps other students with reading difficulties”) and 8 (“I enjoyed playing in pairs”). This was immediately followed by items 1 (“the racetrack helped me to read words correctly”) and 5 (“I enjoyed coming to the intervention”) with a mean value of 3.88 and an *SD* = 0.33. Item 4 (“I learned a lot during the intervention”) received a mean score of 3.75 (*SD* = 0.43), item 3 (“I understood the meaning of the intervention well”), a mean value of 3.50 (*SD* = 0.25), and item 6 (“I would participate in the promotion again”), a mean value of 3.38 (*SD* = 0.48). Finally, responses to item 7 (“The words were difficult”) revealed that the training words were not too difficult for the students (*M* = 0.63, *SD* = 0).

Discussion

The objective of this study was to evaluate the effects of a peer-tutorial reading racetrack intervention on the word fluency of secondary students with LD and those with a co-morbidity of LD and EBD. In line with other research (e.g., Erbey et al., 2011; Green et al., 2010; Gruenke, 2019; Hopewell et al., 2011; Hyde et al., 2009), our results indicate that the reading racetrack intervention described in this paper was very effective in improving students’ ability to automate the reading of trained words. This also applies to the long-term effects. No significant decrease was evident here compared with the intervention effects for the group as a whole.

By applying the intervention at the secondary level and with students with LD, as well as students with LD and EBD, our study demonstrates that reading racetrack interventions can be used effectively with a heterogeneous student population. Further, while many previous studies have suggested that the intervention is effective in primary school (e.g., Gruenke, 2019), the present study provides evidence that secondary students can benefit from word-level reading interventions as already shown in the meta-analyses by Scammacca et al. (2007, 2015). According to the meta-analysis by Hollo et al. (2014), particularly students with EBP have had numerous negative experiences in reading and writing, so it is important to balance them with positive learning situations. That this is feasible with the intervention described here is clearly demonstrated by the students’ assessments of social validity – the students viewed the intervention as both helpful and motivating. Also, the results go in line with previous studies and meta-analysis on the effects of peer-tutoring regarding students with disabilities (Alzahrani & Leko; 2018; Bowman-Perrot, 2009; McDuffie et al., 2009; Moeyaert et al., 2021; Okilwa & Shelby, 2010; Stenhoff & Lignugaris/Kraft, 2007). Moreover, these findings follow on from

King-Sears (2021) that peer tutoring is generally well suited to secondary special education. Further, our study gives additional insights that reading fluency can be achieved through peer-tutoring when fluency is directly focused (see Calhoon, 2005).

Limitations and Recommendations for Further Research

The results of the present study must be interpreted with some reservations. Despite its encouraging results for secondary students, the study is subject to the same weaknesses as all single-case designs, including a lack of generalizability due to the small sample size, which affects the external validity of the study. However, this circumstance can be compensated for by the evidence of previous studies showing the effectiveness of reading racetracks for the training of sight words in German schools (e.g., Barwasser et al., 2021; Gruenke & Barwasser, 2019). Since the effectiveness of reading racetracks for students at higher grade levels has received little research attention so far, future studies should focus on this group of students in particular. Moreover, it has been shown that effect sizes with respect to peer-tutoring interventions are higher in quasi experimental designs and single group designs compared to randomized control trials indicating the fact that the stricter a research design is, the lower are the effect sizes (Zeneli et al., 2016). This is mainly due randomization of pre-tests which in turn control factors as e.g. maturation and history threats (e.g. Trochim, 2012). This fact should be considered when interpreting the effect sizes displayed for this study. Additionally, further studies based on a randomized experimental-control group design should attempt to replicate the present results.

A further limitation of the present study is that we did not include a differentiated analysis of the students according to those with LD only and those with a co-morbidity of LD and EBD. Therefore, it might be of interest for further research to investigate differential effects in relation to the particular special educational needs of students. This is especially true for the long-term effects of the intervention. Although there was no significant overall decrease in the effects over time, the visual inspection for the individual students indicates that for some students the competence level in the follow-up measurement decreased while it remained stable for others.

In addition to considering the tutees' perspective on social validity, the tutors' opinions also appear to be of central interest. The study by Vogel et al. (2007) gives an insight into the fact that the tutors were uncertain about dealing with learning difficulties and how to establish a good tutoring relationship, even though the interactions were rated as positive by both groups. Intensive training of the tutors could be considered here, which, in addition to teaching the content of the intervention, also clarifies the special support needs in learning of the tutees.

Finally, since the intervention consisted of several components (reading from the racetrack, motivational components peer tutoring), it is not possible to identify the specific

effects of each element of the intervention. Therefore, it remains to be investigated in future research to what extent each of the components adds to the overall effectiveness. In order to draw conclusions about the extent to which the effects of the present study can be attributed not only to the practice activity of reading words itself but also to the intervention implemented, it remains important to conduct randomized experimental control group designs in future studies.

Conclusion

In summary, our results confirm the effectiveness of a peer-tutorial reading racetrack intervention in promoting reading fluency for secondary students with LD and students with LD and EBD. Thus, the method has a wide range of application in terms of student age and special educational need. Given the small expenditure of materials and time makes this not only an effective but also an economic intervention for the classroom.

References

Alexander, C. T., McLaughlin, T. F., & Derby, K. M. (2008). The effects of reading racetracks on sight words across four elementary students with differing disabilities. *The Open Rehabilitation Journal*, *1*, 47–52. https://doi.org/10.2174/1874943700801010047

Alzahrani, T., & Leko, M. (2018). The effects of peer tutoring on the reading comprehension performance of secondary students with disabilities: A systematic review. *Reading & Writing Quarterly, 34*(1), 1-17. https://doi.org/10.1080/10573569.2017.1302372

Auerbach. J. G., Gross-Tsur, V., Manor, O., & Shalev, R. S. (2008). Emotional and behavioral characteristics over a six-year period in youths in persistent and nonpersistent dyscalculia. *Journal of Learning Disabilities*, *41*(3), 263–273. https://doi.org/10.1177/0022219408315637

Ayala, S. M., & O'Connor, R. (2013). The effects of video self-modeling on the decoding skills of children at risk for reading disabilities. *Learning Disabilities Research and Practice*, *28*(3), 142–154. https://doi.org/10.1111/ldrp.12012

Barwasser, A., Urton, K., Gruenke, M., Sperling, M., & Coker, D. L. (2021, accepted). Fostering word fluency of struggling third graders from Germany through motivational peer-tutorial reading racetracks. Reading & Writing.

Benner, G. J., Nelson, J. R., & Epstein, M. H. (2002). Language skills of children with EBD. A Literature Review. *Journal of Emotional and Behavioral Disorders*, *10*(1), 43–59. https://doi.org/10.1177/106342660201000105

Bowman-Perrott, L. (2009). Classwide Peer Tutoring. An effective strategy for students with emotional and behavioral disorders. *Intervention in School and Clinic, 44*(5), https://doi.org/10.1177/1053451208330898

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. https://doi.org/10.1080/02796015.2013.12087490

Bradley, R., Doolittle, J., & Bartolotta, R. (2008). Building on the data and adding to the discussion: The experiences and outcomes of students with emotional disturbance. *Journal of Behavioral Education*, *17*(1), 4–23. https://doi.org/10.1007/s10864-007-9058-6

Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single subject experimental design for evidence-based practice. *American Journal of Speech Language Pathology*, *21*(4), 397–414. https://doi.org/10.1044/1058-0360(2012/11-0036

Calhoon, M. B. (2005). Effects of a peer-mediated phonological skill and reading comprehension program on reading skill acquisition for middle school students with reading disabilities. *Journal of Learning Disabilities, 38*(5), 424-433. https://doi.org/10.1177/00222194050380050501

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

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

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

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Klingler Tackett, K., & Wick Schnakenberg, J. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research*, *79*(1), 262–300. https://doi.org/10.3102/0034654308325998

Ehri, L. (2005). Development of sight word reading: phases and findings. In M. J. Snowling & C. Hulme (Eds), *The Science of Reading: A Handbook* (pp. 135–154). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470757642.ch8

Erbey, R., McLaughlin, T. F., Derby, M. K., & Everson, M. (2011). The effects of using flashcards with reading racetrack to teach letter sounds, sight words, and math facts to elementary students with learning disabilities. *International Electronic Journal of Elementary Education*, *3*(3), 213–226.

Escarpio, R., & Barbetta, P. (2016). Comparison of repeated and non-repeated readings on the reading performances of students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, *24*(2), 111–124. https://doi.org/10.1177/1063426615574337

European Commission. (2018). *PISA 2018 and the EU. Striving for Social Fairness and Education*. Publications Office of the European Union.

Forness, S. R., Kim, J., & Walker, H. M. (2012). Prevalence of students with EBD: Impact on general education. *Beyond Behavior*, *21*(2), 3–10.

Fuchs, D., Fuchs, L. S., Mathes, P., & Simmons, D. (1997). Peer-assisted learning strategies: Making classroom more responsive to diversity. *American Educational Research Journal, 34(1)*, 174–206. https://doi.org/10.3102/00028312034001174

Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology*, *98*(4), 732–749. https://doi.org/10.1037/0022-0663.98.4.732

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.

Gruenke, 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

Gruenke, M., & Barwasser A. (2019). Enhancing sight-word fluency of second-language elementary students through reading racetracks. *International Journal of Technology and Inclusive Education (IJTIE)*, *8*(1), 1373–1378. https://doi.org/10.20533/ijtie.2047.0533.2019.0167

Guerin, A., & Murphy, B. (2015). Repeated reading as a method to improve reading fluency for struggling adolescent readers. *Journal of Adolescent & Adult Literacy*, *58*(7), 551–560. https://doi.org/10.1002/jaal.395

Hilsmer, A., Wehby, J., & Falk, K. (2016). Reading fluency interventions for middle school students with academic and behavioral disabilities. *Reading Improvement*, *53*(2), 53–64.

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

Hopewell, K., McLaughlin, T. F., & Derby, K. M. (2011). The effects of reading racetracks 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*(2), 693–710.

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*(1), 50–53. https://doi.org/10.2174/1874945300902010050

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*(2), 137–166. https://doi.org/10.1017/S0261444810000509

Kim, Y.-S., Wagner, R. K., & Foster, E. (2011). Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first-grade readers. *Scientific Studies of Reading*, *15*(4), 338–362. https://doi.org/10.1080/10888438.2010.493964

King-Sears, M. E. (2021). Secondary special education co-teachers in the United States and specialized reading instruction for adolescents with disabilities. *Educational Review*. https://doi.org/10.1080/00131911.2021.1872505

Knoepke, J., Richter, T., Isberner, M. B., Naumann, J., & Neeb, Y. (2014). Phonological recoding, orthographic decoding, and comprehension skills during reading acquisition. *Zeitschrift fuer Erziehungswissenschaft, 17*, 447–471. https://doi.org/10.1007/s11618-014-0560-z

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

Ledford, J. R., & Gast, D. L. (2018). *Single Case Research Methodology* (3rd ed.). Routledge. https://doi.org/10.4324/9781315150666

Leffingwell, T. J. (2016). The benefits of reading racetrack with flashcards, a smartboard and time delay usage in increasing sight word recognition and fluency with special education students. *Theses, Dissertations and Capstones*, 983. https://mds.marshall.edu/etd/983

Lerner J., & Johns, B. (2011). *Learning Disabilities and Related Mild Disabilities: Characteristics, Teaching Strategies, and New Directions* (11th ed.). Houghton Mifflin.

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

Martin, D., Martin, M., & Carvalho, K. (2008). Reading and learning-disabled children: Understanding the problem. *The Clearing House*, *81*(3), 113–118. https://doi.org/10.3200/TCHS.81.3.113-118

McCabe, P. C., & Meller, P. J. (2004). The relationship between language and social competence: How language impairment affects social growth. *Psychology in the Schools*, *41*(3), 313–321. https://doi.org/10.1002/pits.10161

McDuffie, K., Mastropieri, M. and Scruggs, T. 2009. Differential effects of peer tutoring in co-taught and non-co-taught classes: Results for content learning and student-teacher interactions. Exceptional Children, *75*(4), 493–510. https://doi.org/10.1177/001440290907500406

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), https://doi.org//doi/10.1177/0741932519855079

Nelson, J. R., Benner, G. J., Lane, K., & Smith, B. W. (2004). Academic achievement of K-12 students with emotional and behavioral problems. *Exceptional Children*, *71*(1), 59–73. https://doi.org/10.1177/001440290407100104

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 (ECD). (2019). *PISA 2018 results (Volume I): What students know and can do*. OECD Publishing. https://doi.org/10.1787/5f07c754-e

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

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

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*(1), 48–76. https://doi.org/10.1037/bul0000124

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

Pierce, M. E., Wechsler-Zimring, A., Noam, G., Wolf, M., & Tami Katzir, T. (2013). Behavioral problems and reading difficulties among language minority and monolingual urban elementary school students. *Reading Psychology*, *34*(2), 182–205. https://doi.org/10.1080/02702711.2011.626108

Popkin, J., & Skinner, C. H. (2003). Enhancing academic performance in a classroom serving students with serious emotional disturbance: Interdependent group contingencies with randomly selected components. *School Psychology Review*, *32*(2), 282–295. https://doi.org/10.1080/02796015.2003.12086199

Rasinski, T. V., Rikli, A., & Johnston, S. (2009). Reading fluency: More than automaticity? More than a concern for the primary grades? *Literacy Research and Instruction*, *48*(4), 350–361. https://doi.org/10.1080/19388070802468715

Richards, S. C., McReynolds, W. T., Holt, S., & Sexton, T. (1976). Effects of information feedback and self-administered consequences on self-monitoring study behavior. *Journal of Counseling Psychology*, *23*(4), 316–321. https://doi.org/10.1037/0022-0167.23.4.316

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), 240–257. https://doi.org/10.1037/0022-0663.95.2.240

Scammacca, N. K., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. (2007). *Reading interventions for adolescent struggling readers: A meta-analysis with implication for practice*. RMC Research Corporation, Center on Instruction.

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *Journal of Learning Disabilities*, *48*(4), 369–390. https://doi.org/10.1177/0022219413504995

Slavin, R. E. (1995). *Cooperative learning: Theory, research, and practice* (2nd ed.). Allyn & Bacon.

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.

Stenhoff, D. M., & Lignugaris/Kraft, B. (2007). A review of the effects of peer tutoring on students with mild disabilities in secondary settings. *Exceptional children, 74*(1), 8–30. https://doi.org/10.1177/001440290707400101

Stevens, E. A., Walker, M. A., & Vaughn, S. (2017). The effects of reading fluency interventions on the reading fluency and reading comprehension performance of elementary students with learning disabilities: A synthesis of the research from 2001 to 2014. *Journal of Learning Disabilities*, *50*(5), 576–590. https://doi.org/10.1177/0022219416638028

Stotz, K. E., Itoi, M., Konrad, M., & Alber-Morgan, S. (2008). Effects of self-graphing on written expression of fourth grade students with high-incidence disabilities. *Journal of Behavioral Education*, *17*, 172–186. https://doi.org/10.1007/s10864-007-9055-9

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Trochim, W. (2012). Design, the research method knowledge base. Retrieved from: http://www.socialresearchmethods.net/kb/design.php (March 9th, 2021)

Vogel, G., Fresko, B., & Wertheim, C. (2007). Peer tutoring for college students with learning disabilities: Perceptions of tutors and tutees. *Journal of learning disabilities,40*(6), 485–493. https://doi.org/10.1177/00222194070400060101

Wilbert, J., & Lueke, T. (2019). Package 'scan.' https://cran.r-project.org/web/packages/scan/scan. pdf

Wimmer, H., & Mayring, H. (2014). *Salzburger Lese-Screening fuer die Schulstufen 2-9: SLS 2-9* [Salzburg Reading Screening for school levels 2-9: SLS 2-9]. Hogrefe.

Zeneli, M., Thurston, A., & Roseth, C. (2016). The influence of experimental design on the magnitude of the effect size-peer tutoring for elementary, middle and high school settings: A meta-analysis. *International Journal of Educational Research, 76*, 211-223. https://doi.org/10.1016/j.ijer.2015.11.010

Appendix B Article 2

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

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 racetrack training with an integrated motivational system for the sight word fluency of German struggling elementary school students. The intervention comprised twelve 15-minute 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 η^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

Introduction

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 et al., 2001; Coltheart, 2005) 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; Roehrig et al., 2008; National Institute for Literacy, 2009; Nese et al., 2013; 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) (Zarić et al., 2020; Zarić & Nagler, 2021; Rothe et al., 2015;) 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., 2015). 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 = .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 (Rohrbeck et al., 2003; Bowman-Perrot et al., 2013; Pappas et al., 2010).

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 (Hirschkorn & 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 (Mercer et al., 2011; Spencer 2006; Dufrene et al., 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 = .77), as well as students with differing levels of social competencies (Ginsburg-Block et al., 2006; d =.28) and learning-related behaviors (Ginsburg-Block et al., 2006; d= .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; Gruenke, 2019; Hopewell et al., 2011; Hyde et al., 2009), 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 minutes, 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.

Table 1

Socio-Demographic Characteristics of Participants Comparing MR and RR

	Maths Racetracks (*N* =22)		Reading Racetracks (*N* =22)		*p*
	N	*M(SD)*	*N*	*M(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
	N		*N*		
Gender					.34
Male	9		6		
Female	13		16		
Special Needs					.55
LD	2		1		
Non LD	20		21		
German L2					.76
Yes	10		9		
No	12		13		

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

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* = .57 and *r* = .73. Additionally, the CFT-20-R is moderately correlated with school grades in mathematics (*r* = .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 12-15 are considered borderline and values between 16-40 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, in terms of gender, age, special needs, German as L2, cognitive abilities, reading proficiency and externalizing problem behavior were found (Table 1 & 2).

Table 2

Results of SDQ and Reading Test Comparing MR and RR

	Maths Racetracks (*N* =22)		Reading Racetracks (*N* =22)		*p*
	N	*M(SD)*	*N*	*M(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

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 1 s of the presentation. The one 1-s cycle was automatically preset, so the slides changed by themselves after 1 s. 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

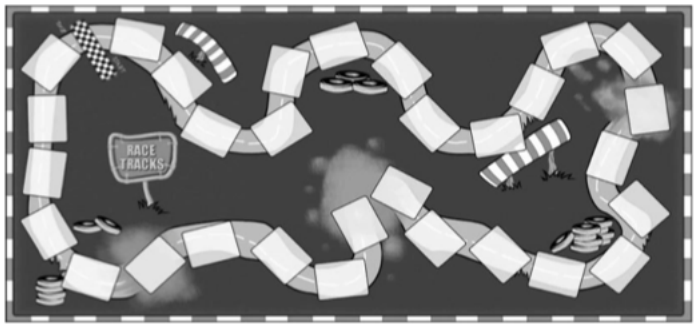
The intervention was conducted in the regular class setting. Both children of a pair (tandem) sat opposite each other and the racetrack game in the middle of them. At the start signal, the tutee rolled the dice and moved the piece to the appropriate square according to the number (see Figure 1). Flashcards 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 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 minutes. 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.

Figure 1

Self-made racetrack board



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 (Wolf, 1978; Briesch et al., 2013). 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) × 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 $p < .05$. 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* = .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.

Table 3

Descriptive Statistics for Words Read Correctly

Measures	Pretest	Posttest	Follow-up 1	Follow-up 2
	M (SD)	*M (SD)*	*M (SD)*	*M (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)

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

There was no statistically significant pretest difference between students in the reading racetracks and the math racetracks group for WRC $t(41.86) = .007$, $p = .94$, $d = 0.002$. The rmANOVA to test the interaction effect (Time × Group) showed a significant result, Greenhouse-Geisser $F(1.71, 71.73) = 133.60$, $p < .001$, *partial* $\eta^2 = .76$. The inner-subject main effect for time indicates that this factor has a statistically significant influence on WRC for the reading racetracks group, $F(1.62, 34.02) = 203.96$, $p < .001$, *partial* $\eta^2 = .91$. 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 (.86, $p = 1.0$, $d = 0.14$), 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* $\eta^2 = .16$. The pairwise comparisons assessing the influence of time revealed no significant differences. This refers to the comparison of the pretest with the posttest (.86, $p = .31$, $d = 0.39$) and with follow-up 1 (1.50, $p = .26$, $d = 0.46$) and follow-up 2 (2.10, $p = .15$, $d = 0.57$), as well as between the posttest and

follow-up 1 (.64, p = 1.0, d = 0.19) and follow-up 2 (1.23, p = .52, d = 0.32), and between follow-up 1 and follow-up 2 (.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* $\eta^2 = .77$. Although there was no group difference for the pretest $t(41.86) = .007$, $p = 94$, $d = 0.03$, 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 × Group ($F(3, 15.48) = 185.10$, $p < .001$, $\eta^2 = .97$), as well as the main effect for time ($F(3, 15.48) = 197.50$, $p < .001$, $\eta^2 = .97$) and group ($F(1, 19.90) = 160.48$, $p < .001$, $\eta^2 = .96$).

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.

Table 4

Social Validity Questionnaire Teachers

Items	*N*	*M (SD)*
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)

Note. 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)

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 5

Social Validity Questionnaire Students

Items	*N*	*M(SD)*
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)

Note. 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; Hopewell et al., 2011; Hyde et al., 2009, Gruenke, 2019), 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; McArthur et al., 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 (Gruenke & 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 (Laemsae et al., 2018; Amato-Zech et al., 2006; 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.

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., & Gruenke, 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*.

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.

Calabrèse, 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 & Visual Science, 57, 3192–3202. https://doi.org/10.1167/iovs.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. How can I transform teaching and student learning today.* 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.

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.

Dyslexia Action (2017). *Avoid dyslexia discrimination in the workplace.* https://dyslexiaaction.org.uk/contact-us/

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.). London, UK: 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.

Gruenke, 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

Gruenke, 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/ijtie.2047.0533.2019.0167

Guerin, A., & Murphy, B. (2015). Repeat reading as a method to improve reading fluency for struggling adolescent & adult literacy. *Journal of Adolescent & 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.

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

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. Europena policy perspectives and assessment strategies.* Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-38969-7

Hayes, C. (2016). The Effects of Sight Word Instruction on Students' Reading Abilities. *Education Masters*, Paper 327.

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

Hirschkorn, 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.

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 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 fuer 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 Veroeffentlichungen der Kultusministerkonferenz, Dokumentation Nr. 214 –Juni 2018. Sonderpaedagogische Foerderung in Schulen 2007 bis 2016 [Statistical Publications of the Standing Conference of the Ministers of Education and Cultural Affairs of the Laender 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.

Laemsae, J., Haemaelaeinen, R., Aro, M., Koskimaa, R., & AEyraemoe, 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). Bingley, UK: 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 lnowledge, *Sage Open, 9*(2). https://doi.org/10.1177/2158244019845182

Mayringer, H., & Wimmer, H. (2014). *SLS 2-9. Salzburger Lese-Screening fuer 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. *PeerJ, 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. korr. und erw. 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 & 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*, Article 5. 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 for Literacy (2009). *Developing early literacy: Report of the National Early Literacy Panel*. National Center for Family Literacy.

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*, 1–15. http://web.ebscohost.com.pluma.sjfc.edu

Pappas, D. N., Skinner, C. H., & Skinner, A. L. (2010). Supplementing accelerated reading with classwide 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-Koerne, 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/rrq.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.

Schroeder, S., Wuerzner, K.-M., Heister, J., Geyken, A., and Kliegl, R. (2015). childLex–Eine lexikalische Datenbank zur Schriftsprache fuer Kinder im Deutschen [childLex-A lexical database of written language for children in German]. Psychol. *Rundsch. 66*, 155–165. https://doi.org/10.1026/0033-3042/a0 00275

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., & Gruenke, 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

Table 6

Words of training

Laender *(countries)*	Anerkennung *(acknowledgement)*	Katastrophe *(catastrophe)*	Identitaet *(identity)*	rechtzeitig *(in time)*	zweifeln *(to doubt)*
zeichnen *(to draw)*	andererseits *(on the other hand)*	garantiert *(guaranteed)*	Langeweile *(boredom)*	jenseits *(beyond)*	selbstverstaendlich *(of course)*
akzeptiert *(accepted)*	Ereignisse *(events)*	anstellen *(to queue)*	puenktlich *(punctual)*	motiviert *(motivated)*	Taetigkeit *(activity)*
Aufklaerung *(reconnaissance)*	Gerechtigkeit *(justice)*	gezwungen *(forced)*	festgelegt *(set)*	keineswegs *(not at all)*	UEberzeugung *(conviction)*
erforderlich *(required)*	Auseinandersetzung *(examination)*	zuhoeren *(to listen)*	keinerlei *(none)*	Schwierigkeiten (difficulties)	nahezu *(almost)*

Appendix C Article 3

Barwasser, A., Hertel, S., & Grünke, M. (2021). A sub-lexical-patterns training with racetracks on trained and untrained words in low-literacy German students with behavioral problems with and without learning disabilities. *Learning Disabilities. A Contemporary Journal, 19*(2), 143-159.

Abstract

The ability to read adequately is one of the most important skills for students to achieve during their school career. Unfortunately, a large number of children do not have adequate reading skills and thus may face problems in many respects throughout their lives. Children with learning disabilities (LD) and behavioral problems are particularly prone to fall into this category and may additionally experience a general lack of motivation to learn. In the context of improving reading skills, fostering lexical and sublexical reading has proven successful. This single-case study (N = 5) investigated the effect of using a motivating intervention consisting of a (sub)lexical patterns training in combination with reading racetracks for the automation of common German sublexical patterns to improve the reading of trained and untrained words. The intervention, which was carried out three times a week for 20 minutes each over a period of five weeks, showed promising results, also in the follow-up measurements, including medium to strong effects on training words and small to strong effects on transfer words. These findings offer preliminary evidence of how to combine reading racetracks to create transfer effects in reading for low-achieving primary school students with severe learning and behavioral issues. *Keywords:* Peer Tutoring, Reading Racetracks; Lexical and Sublexical Reading; Learning

Introduction

The Significance of Reading Fluency and Sight Word Recognition

The ability to decode written and printed symbols is a key skill that has a decisive influence on almost all areas of life (Grigoryan, 2020; Macdonald et al., 2016). Regrettably, about 10-15% percent of those who have received at least basic education worldwide have major difficulties in understanding written texts (Dyslexia Action, 2017). In Germany, studies have found an upward trend of fourth graders with poor reading competency accompanied by declining reading motivation (Bos et al., 2017). In particular, many students with learning disabilities (LD) face challenges with regard to acquiring reading competency (Lerner & Johns, 2011; Solis et al., 2012), which can be partly linked to lower memory capacity, especially regarding word reading (Geary et al., 2020). In addition, there is a relationship between behavioral problems

and reading difficulties. That is, students who demonstrate behavioral challenges spend less time reading (Vaughn et al., 2002), and many perform at least a whole school year behind their peers (Oakes et al., 2010).

When reading fluently, an experienced reader retrieves words automatically from the mental lexicon (Morris & Perney, 2018; Young et al., 2020). The ability to read fluently is indispensable for reading comprehension and, thus, for overall reading proficiency (Ehri, 2005). According to the dual-route theory (DRT; Coltheart, 2005), the reading process happens through a lexical and a nonlexical route (sublexical). In the nonlexical route, graphemes are recorded and converted into phonemes (phonological recording), whereas in the lexical route, written words are mapped directly on to mental representations of word forms (orthographical decoding) (Coltheart et al., 2005). If students can process frequently occurring orthographic patterns of the German language, for example, this can guide direct word recognition of a large number of words (Mayer, 2018).

Experienced readers have built up these so-called “sight words,” which enables them to recognize a word from the mental lexicon within one second of its appearance (Ehri, 2005). Knoepke et al. (2014) demonstrated that readers of German (a transparent language; that is, there is a one-to-one relationship between meaning and form), orthographic decoding is a greater predictor of reading skills than phonological recoding (see also Tressoldi et al., 2007). According to Ehri (2014), being able to match sounds to orthographic patterns is critical for word recognition. Thus, nonlexical decoding skills also play a major role with respect to reading fluency (Harn et al., 2008). Therefore, combining both in an intervention seems to be a promising way to foster reading performance.

(Sub)Lexical Training and Repeated Reading

Reading instruction should start at the word level (Ehri, 2014). According to a meta-analysis by Scammacca et al. (2007), any student can profit from a word-level intervention. Further, word-recognition interventions have shown positive results (Martin-Chang & Levy, 2005; Scammacca et al., 2015). Going a step further, apart from whole-word training, Marinus et al. (2012) found that explicit training of sublexical patterns led to superior short- and long-term improvement in the rapid naming of trained and untrained patterns for primary students with poor reading skills. To become fluent readers, people need to read words or word patterns many times (Grabe, 2010); hence Repeated Reading (RR) has been found to be a promising tool for improving reading skills (Kostewicz et al., 2016; Lee & Yoon, 2017; Rasinski et al., 2016). The repetition of words and word patterns allows readers to automatically retrieve them from their mental lexicon (Zavala & Cuevas, 2019), thereby relieving the demands on working memory.

Reading Racetracks as a Realization of RR and (Sub)Lexical Training

Research has indicated that Reading Racetracks (RT) is an effective and enjoyable way to improve reading fluency through RR (Crowley et al., 2013). A game-based instructional method, RT uses a game board consisting of cells that contain items such as sight words. Numerous studies have shown the effectiveness of racetracks for both first-language (L1) and second-language (L2) students (Grünke, 2019; Grünke & Barwasser, 2019; Sperling et al., 2019); in addition, when combined with peer tutoring (PT) and motivational components, RT has been found effective for a wide range of students with and without LD (e.g., Barwasser et al., 2021a, 2021b). However, existing studies have not focused on possible transfer effects – solely on trained items.

Peer Tutoring as Inclusion Tool

Given the increase in inclusive educational programming, peer tutoring (PT), an evidence-based intervention whereby students are trained to provide instruction to their peers (Bertin & Narcy-Combes, 2007), is becoming a popular intervention choice. For example, PT has been found to be successful for promoting learning in collaborative and inclusive environments (Alzahrani & Leko, 2018). Of particular interest in the current context, multiple reading strategies have been effectively taught in peer-assisted learning arrangements (Sáenz et al., 2007; Van Keer & Vanderlinde, 2010). Overall, students who were integrated into a reading intervention that included peer tutoring performed better than students who only received teacher-centered reading instruction (Mueller et al., 2015). To our knowledge, RT have been only combined with PT in the studies by Barwasser et al. (2021a, 2021b).

Incorporation of Self-Graphing

Motivation also plays an important role in reading (Marinak & Gambrell, 2008). Students who lack motivation to engage in activities such as reading try to avoid them (Guthrie, 2000; Marinak & Gambrell, 2008); hence motivation must be given serious consideration in efforts to improve reading skills. Motivation techniques in reading interventions are particularly important for students who experience high levels of frustration and, therefore, are at risk of failing (Sideridis, 2002).

One possible technique is self-graphing, which is a specific type of self-evaluation wherein students write down their own performances, thus creating a visual representation of their performance over time (Gunter et al., 2002). In a meta-analysis on the effects of formative evaluation, Fuchs and Fuchs (1986) noted a significant increase in the academic achievement of students with disabilities through systematic formative evaluation procedures, especially when the data were presented graphically (see also Sutherland & Snyder, 2007).

Research Aim

Given that the gap between skilled and less skilled readers is widening in Germany, and indeed throughout the world, and that children with behavioral problems and LD face special challenges in his respect, it is of great importance to develop an intervention of a motivating and automating character and that can be used for a wide variety of students. Combining the aforementioned methodological aspects into one intervention, as done by Barwasser et al. (2021a, 2021b), the current study focused on the effects of a combined racetrack intervention on trained and, as a unique selling point, untrained words, to determine if the intervention is a universal tool that additionally can lead to transfer effects. Moreover, the social validity of the intervention – that is, how it is received by students – was examined.

The following research questions underlie the study: (a) Did the combined (sub)lexical patterns racetrack intervention with PT and self-graphing lead to an increase in the acquisition of trained and untrained (transfer) words? (b) Were the results stable at two-month follow-up? (c) How was the intervention received by the students?

Methods

Participants and Setting

The study was conducted in an urban elementary school in North Rhine-Westphalia, Germany. Students from third and fourth grade with low proficiency in word reading were targeted. Prior to the start of the study, consent to participate in the study was obtained from students' legal guardians.

To select the final participants, a multi-step procedure was used as follows. (a) A German reading screening that focuses on word reading (SLRT II; Moll & Landerl, 2010) was administered to the whole classes. The SLRT II consists of two 1-minute word and pseudo-word reading tests. The SLRT II's correlation with other reading screenings lies between .69 and .92, and its reliability is between .90 and .98. Students who achieved a percentile (PR) of <15 on both tests were eligible for the study (see Table 1). (b) A German vocabulary test (WS/ZF-R; Weiß, 2006) and a screening to assess externalizing problem behavior (Integrated Teacher Report Form [ITRF]; Volpe et al., 2018) were also used. The WS/ZF-R examines areas of crystallized ability (skills acquired through prior knowledge and experience); it contains a total of 30 items including words from semantic fields as well as abstract terms. The relevant task is to select from a series of five words the one that has the same or similar meaning as a given word. The WS/ZF-R has "good" to "very good" reliability (.87). The ITRF-G (short form) is a German translation of the American screening Integrated Teacher Report Form; it is considered by Volpe et al. (2018) as a universal and instructionally relevant behavioral screening to crystallize externalizing behavior. It consists of a total of 16 items, with eight items each assigned to the areas of "problems in learning behavior" and "oppositional

behavior;" the cutoff value for the total problem behavior is 13. The ITRF has a high internal consistency (total problem value: α = .91). In addition, a teacher questionnaire was distributed to assess further characteristics of the children.

The final sample consisted of 8 students, divided into three groups, each with different baseline lengths. The age range of students was between 8 and 10 years old. All children showed low proficiency in word reading and German vocabulary. Also, all children exhibited general problem behavior according to the ITRF, and two were diagnosed with LD. In Germany, LD means immense difficulties in more than one school subject mostly accompanied by a rather low IQ (Hasselhorn & Gold, 2017). Due to missing data (as a result of COVID-19 quarantine), this article focuses on the data of five students (*N* = 5). Reading pairs were formed based on teacher assessment and results of reading screening (that is the pairs of students had the same reading level).

Table 1

Demographic Characteristics of the Participants

Name	Gender	Age	Reading W/PW PR	Vocabulary PR	ITRF GP	SEN	German L2	Ethnicity
Latifa	female	10	7-8/4	10	21	LD	Yes	Lingala
Ella	female	9	<1/<1	2	18	LD	Yes	Greek
Lou	female	8	4-5/<2	0	15	/	No	German
Allai	female	8	1-2/14	2	14	/	Yes	Arabic
Jim	male	8	11-15/9-11	1	20	/	Yes	Thai

Note. Words (W); Pseudo Words (PW); Percentile (PR, cutoff 15); Integrated Teacher Report Form (ITRF); Main Problem Value (GP); Special Educational Needs (SEN); Second Language (L2); Learning Disabilities (LD).

Design

A multiple-baseline design across participants was applied to be able to exclude alternative explanations for the effectiveness of the intervention (Kazdin, 2010). The children were randomly assigned to the three groups; due to the COVID-19 pandemic, Group 3 and 4 could not be supported together. Group 1 had a total of five baseline measurements, Group 2, four, and Group 3, six (Kratochwill et al., 2013). In total there were 20 planned measurement points for baseline and intervention together. Before the baseline started, screenings were performed within 1.5 weeks. The baseline phase was then started three times a week, followed by the intervention phase, which also started three times a week over a period of five weeks. Data were collected after each baseline and each intervention session, as well as two months after the end of the intervention (follow-up). The groups were brought, one after the other, to rooms outside the students' regular classroom. As test leaders and interventionists, six

master's-level students of special needs education were employed. Four students supported the groups in pairs. The other two were responsible for the measurements to avoid bias.

Dependent Variables and Measurements

There were two dependent variables: (a) number of correctly read training words and (b) number of correctly read transfer words. A researcher-developed instrument was used, which consists of a PowerPoint presentation (PPT) in which a word to be read is shown on each slide. The slides are set to a 1-second rhythm (Ehri, 2005). After each baseline and each intervention session, the participants were assessed using this PPT. There were two pools of words, 70 training words and 70 transfer words, taken from "Rapid Word Recognition" (Mayer, 2018). Twenty words from each pool were used randomly in the PPT, which allowed the students to reach a maximum of 20 words for training words and 20 words for transfer words for each measurement. Care was taken to ensure that the 70 words were distributed equally among the measurement points and that the measurement points themselves did not differ significantly from each other with regard to word difficulty, as shown by a variance analysis using SPSS statistics. Afterwards the number of correctly read training and transfer words was noted.

Procedure

Baseline

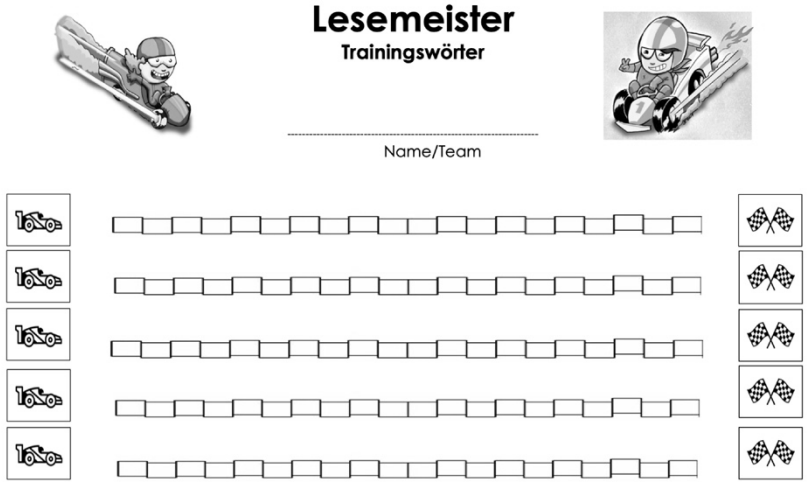
To engage the children in the baseline (Phase A), cognitive tasks were applied, consisting of logically continuing a certain sequence and crossing out symbols that do not fit into a row. These exercises were chosen since they do not foster reading, allowing the current state of the two dependent variables better to be estimated. Baseline was conducted in the same group as in the intervention. After 20 minutes, all participants were assessed independently with respect to the dependent variables.

Intervention

The intervention was made up of two stages. The first stage consisted of direct instruction of letter clusters and training words using large flashcards while the children sat in a semicircle in front of the interventionist for 10 minutes. For this stage, the materials of the German program *Blitzschnelle Worterkennung* (Rapid Word Recognition; Mayer, 2018) were used. This involved a pool of training words (70) and transfer words (70) as well as letter clusters. The letter clusters and matching training words were printed individually on 8.3 x 11.7-inch flashcards; the letter cluster was marked in green for the training words. The letter clusters were slowly presented, one by one, and then the appropriate training words were presented. The interventionist read all the words aloud, and the children were also asked to try to read the words and clusters.

The second stage consisted of the RT procedure where the pairs of students previously formed played a game for 10 minutes. A 11.7 × 16.5-inch game board with 14 empty cells was designed. In addition, the training words were printed individually on small flashcards that were distributed across the racetrack. The training words (always 14 out of the pool of 70) are placed upside down on each cell of the racetrack field, and the children take turns throwing a die. Child 1 throws a die and moves the game figure according to the number of eyes on the die and lifts the corresponding flashcard. The child tries to read the word and Child 2 corrects if possible. If the word was not read correctly, Child 2 reads the word and Child 1 repeats it. Now, both children think about which letter cluster is recognizable in the word. If the word is read correctly, students move right on while child 2 repeats the same process. The children collect the correctly read flashcards. Not correctly read words remain on the board. If both children cannot read the word correctly, the interventionists will help.

After each intervention session, all participants are assessed independently with respect to the dependent variables. To increase motivation, the children are asked to enter the number of correctly read training and transfer words in two self-graphing sheets after each measurement. Each sheet consists of several rows, one below the other (number of sessions), each with 20 boxes, representing the maximum number possible of correctly read words (see Figure 1).



Treatment Fidelity and Social Validity

Prior to the study, the interventionists and test leader were intensively trained in how to conduct the study. For treatment fidelity, a documentation sheet was designed to be filled out by the interventionists after each session and, additionally, by an external person for one third of the intervention time. The questionnaire contained a checklist consisting of the following items: (a) environment/external circumstances, (b) planning, (c) material, (d) course of the intervention, (e) diagnostics, and (f) feedback and dealing with students. Items were ranked on a 5-point Likert scale from 0 (not applicable at all) to 4 (completely applicable). At the end

of the checklist there is room for adding comments and/or remarks. The interrater reliability was 100%, both between the interventionists and between the interventionists and the external person across all groups.

To determine social validity, a questionnaire was designed consisting of 12 items to be rated by the participants on a 5-point Likert scale ranging from 0 = not true at all to 4 = completely true. *Item 1*: The racetrack game helped me to read words correctly; *Item 2*: I think that the support also helps other students with reading difficulties; *Item 3*: I have understood the purpose of the intervention well; *Item 4*: I learned a lot during the program; *Item 5*: I enjoyed coming to the program; *Item 6*: I enjoyed the support; *Item 7*: I would participate in the program again. *Item 8*: The letter clusters groups helped me to read better; *Item 9*: Drawing in the arcs was fun. *Item 10*: I would like to do something like that more often; *Item 11*: The words were difficult.

Item 12: I enjoyed playing in pairs.

Results

Descriptive Analysis

Training Words

The statistical program R was used throughout. Additionally, the mean baseline difference (MBD) was calculated by hand. Regarding the dependent variable training words, there was a visual increase from Phase A to Phase B for each participant; same pattern is reflected in the descriptive data (see Table 2). Latifa, Allai, and Jim reached the maximum value of 20 correctly read training words during Phase B, followed by Ella and Lou with 18, with Lou showing the greatest increase. Moreover, the follow-up data (Phase E) were stable across all participants. The mean values showed a slight increase compared to Phase B. The MBD from Phase A to B was as follows: Jim: 21,34 %, Allai: 29.30 %, and Latifa: 31.00%, with a greater increase for Ella: 91,41% and Lou: 200%. With respect to mastery, Latifa needed six sessions, Allai four sessions, and Jim three sessions. Ella reached her maximum value in Session 7 and Lou in Session 12.

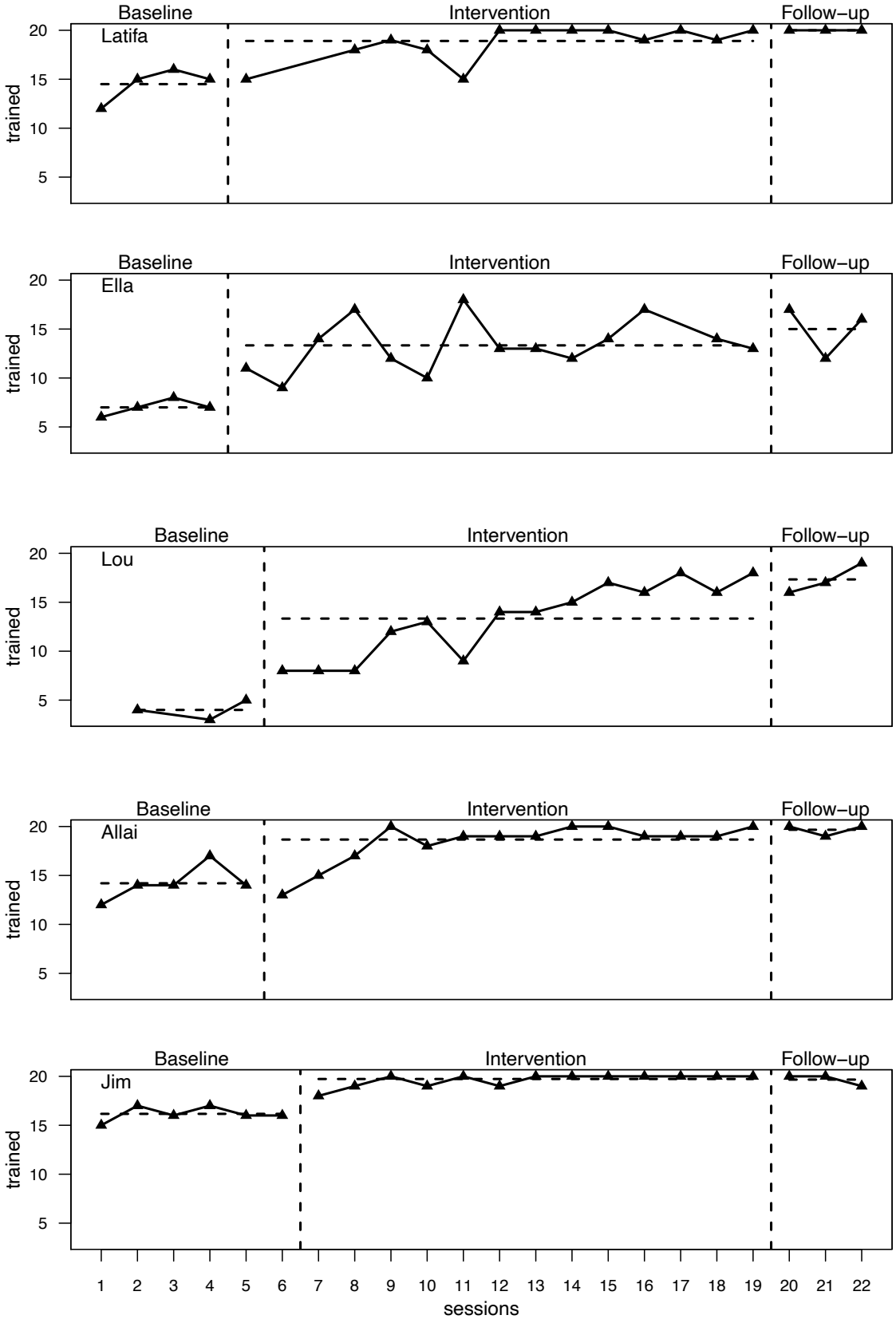


Table 2

Descriptive Data for Each Participant in A, B, and E Phase Training Words

	N (A)	*N* (B)	*N* (E)	MA (*SD*)	MB (*SD*)	MBD	ME(*SD*)	Max B
Latifa	4	15	3	14.50(21.73)	19.00(1.80)	31,00%	20.00(0.00)	20.00
Ella	4	15	3	7.00(0.82)	13.38(2.72)	91,14%	14.67(2.52)	18.00
Lou	5	15	3	4.00(1.00)	12.00(3.84)	200,00%	17.33(1.53)	18.00
Allai	5	14	3	14.20(1.79)	18.36(2.06)	29,30%	19.67(0.58)	20.00
Jim	6	13	3	16.17(0.75)	19.62(0.65)	21,34%	19.67(0.58)	20.00

Note. Measurements (*N*); Baseline (A); Intervention (B); Follow-Up (E); Mean (*M*); Standard Deviation (*SD*); Maximum (Max); Mean Baseline Difference (MBD).

Transfer Words

For the transfer words, there was a clear difference from the training words since visually all children increased more slowly in Phase B than with the training words (see Table 3). The descriptive data show that Latifa and Jim reached a maximum value (20 in Phase B, followed by Allai with 19, Ella with 13, and Lou with 12, again with the largest increase.

For the follow-up measurements, the data were relatively stable as for the training words. Overall, the mean values showed a slight increase from Phase B to Phase E. Regarding the MBD, Latifa and Jim showed the weakest increases of 25.81% and 26.25 % from Phase A to Phase B, respectively, followed by Ella (44.60%) and Allai (48.10%). Lou, again, had the greatest improvement, with 204.35 %. Regarding mastery, Latifa needed 12 sessions and Jim, 11. Ella's maximum value was reached in Session 4, Allai's in Session 9, and Lou's in Sessions 13.

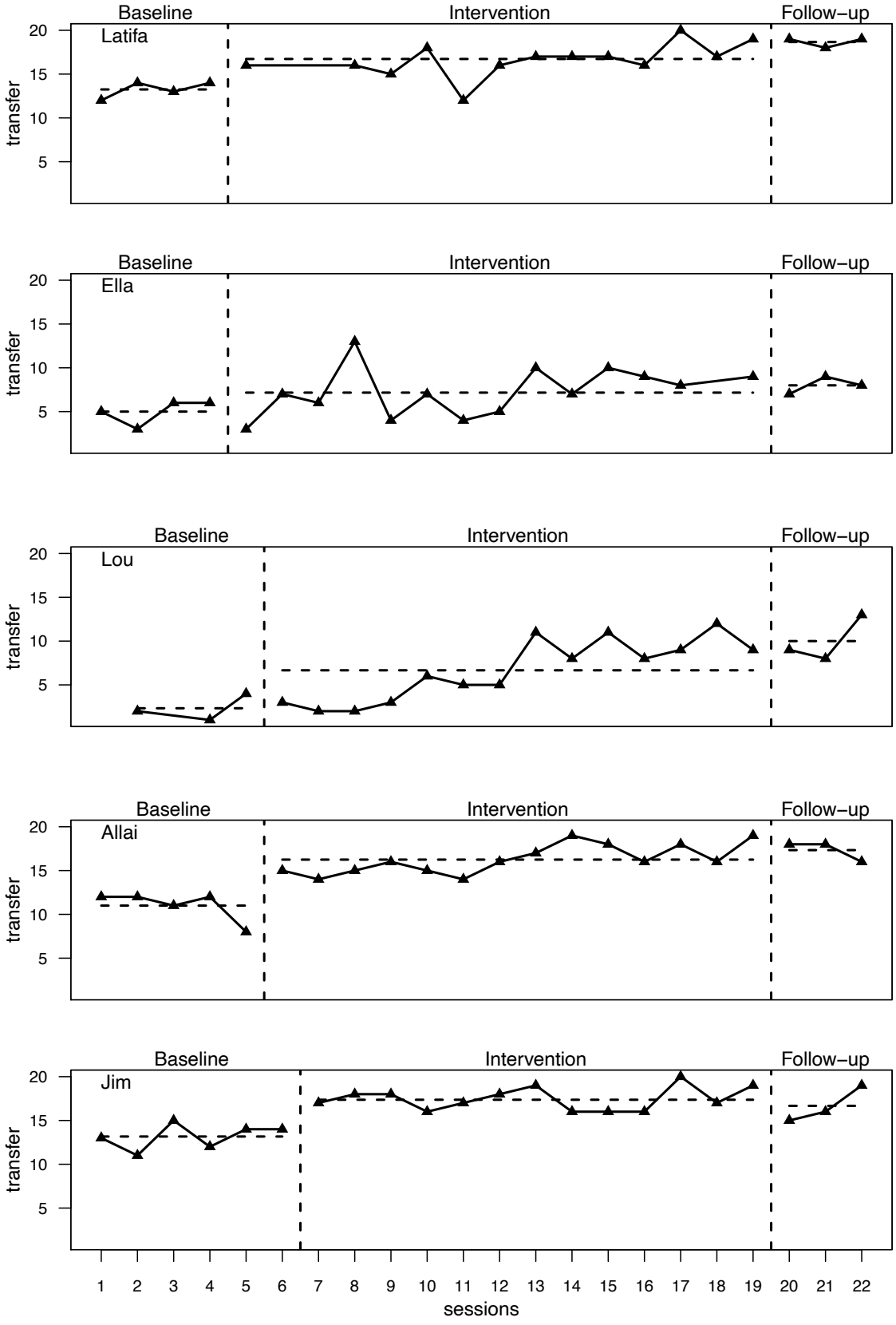


Table 3

Descriptive Data for Each Participant in A, B, and E Phase Transfer Words

	N (A)	*N* (B)	*N*(E)	MA *(SD)*	MB *(SD)*	MBD	ME*(SD)*	Max B
Latifa	4	15	3	13.25(0.96)	16.67(2.02)	25,81%	18.67(0.58)	20.00
Ella	4	14	3	5.00(1.41)	7.23(2.89)	44,60%	8.00(1.00)	13.00
Lou	5	13	3	2.33(1.29)	7.00(3.44)	204,35%	10.00(2.65)	12.00
Allai	5	14	3	11.00(2.83)	16.29(1.68)	48,10%	17.33(1.15)	19.00
Jim	6	13	3	13.83(2.14)	17.46(1.33)	26,25%	16.67(2.08)	20.00

Note. Measurements (*N*); Baseline (A); Intervention (B); Follow-Up (E), Mean (*M*); Standard Deviation (*SD*); Maximum (Max); Mean Baseline Difference (MBD).

Overlap Indices

Training Words

Overlap indices were used for further and more in-depth analysis. The Non-Overlap of All Pairs (NAP; Parker et al., 2011) and Percentage of All Non-Overlapping Data (PAND) Parker et al. (2007) were chosen as well as Tau-U due to possible A-Phase trends.

Regarding training words, Latifa (92.00; p<.01) and Allai (92.00, p<.01) achieved a medium effect for NAP while Ella (100.00, p<.001), Jim (100.00, p<.001), and Lou (100.00) displayed a strong effect. Latifa (76.47) and Allai (89.47) showed moderate effects for PAND while Ella (100.00), Jim (100.00), and Lou (100.00) displayed strong effects. The Tau-U showed a moderate effect for Latifa (0.43, p<.01), Ella 0.46, p<.01), and Allai (0.57, p<.001) and a large change for Jim (0.64, p<.001) and Lou (0.68, p<.001).

Transfer Words

For transfer words, the NAP value was 77.00 (p <.05) for Ella and 88.00 for Lou (p<.05), signaling a moderate change. Latifa (93.00, p<.01), Jim 100.00, p<.001), and Allai (100.00, p<.001) showed a high effect strength. PAND values resulted in a small effect for Ella (66.71) and a moderate effect for Lou (79.47) and Latifa (82.35). A highly effective treatment was reflected by Allai (100.00) and Jim (100.00). Taking into account a possible A-Phase trend, Tau-U values showed a moderate effect for Ella (0.43, <*.05*), Latifa (0.43, p<.05), Jim (0.47, p<.05), and Lou (0.54, p<.001) and a large change for Allai (0.76, p<.001).

Social Validity

The results of the social validity questionnaire indicated an overall very positive student attitude towards the intervention, with Items 5-8 (*Item 5*: I enjoyed coming to the program; *Item 6*: I enjoyed the support; *Item 7*: I would participate in the program again. *Item 8*: The letter clusters helped me to read better) rated highest with an overall score of 4 (completely correct). Further, the students rated the words as not too difficult overall. Finally, the children's self-

written comments show that all felt very well about the intervention, had fun, and would like to participate again (Latifa: “I would like to participate again in any case. It was a lot of fun.”; Ella: “*It was fun and I would like to participate again*.”; Lou: “*I had much fun. Thank you*.”; Allai: “*I miss you already. Thanks for helping us. Your game was a lot of fun.*”; Jim: “*I had much fun playing the racetrack game.*”).

Discussion

Main Findings

The aim of this study was to facilitate the reading of training and transfer words of children with severe reading difficulties with behavioral problems with and without LD using frequently occurring German letter clusters. Since it is of particular importance to create an intervention that can be used across a wide range of students, the present results are promising. Automatic decoding of words plays an immensely important role in reading (Burns, 2007; Knoepke et al., 2014), and given that the gap between less proficient and stronger readers is widening (Bos et al., 2017), it is necessary to find effective interventions. Above all, reading acquisition in primary school is critical (Musti-Rao et al., 2015) as a preventive measure. Overall, the variable training words showed moderate to strong effects, and also proved to be statistically significant. A main focus here was on Tau-U, which takes into account an A-Phase trend that was present purely visually in some participants’ performance. Three of the children reached the maximum value of 20 (mastery) in Phase B, closely followed by Ella and Lou with a value of 18. Unfortunately, Latifa, Allai, and Jim started with quite high values in Phase A; as a result, the maximum value of 20 in Phase B was quickly reached, and a ceiling effect was recognized for all three.

The follow-up data show stable values for all children. However, it should be noted that the pool of training words consisted of 70 words of which 20 were randomly selected for the measurement each time. Presumably, stronger effects would be achieved, if the number of training words were reduced. Nevertheless, the pool was large enough to avoid a single effect from the measurements.

For the variable transfer words, which is supposed to show whether the students could read untrained words through the cluster, as expected, lower effects were found, since it is more difficult to apply knowledge to unknown content than to retrieve known information. Here, the children started in Phase A with significantly lower values, whereby Latifa, Allai, and Jim again displayed the highest values. The increase in Phase B was visibly less steep than for the training words, but the Tau-U results showed moderate to strong effects that are statistically significant, except for Ella. Latifa and Ella seemed to have the most difficulty in applying their knowledge to unfamiliar words, presumably due to a lower degree of automation, especially

considering that they both have a LD. Children with an enormous lack in reading competency often rely on the non-lexical route of the DRT, thus, trying to read while synthesizing (De Jong et al., 2012). This may be the reason for the rather slow-successive increase in the data of Latifa, Ella, and Lou (all with enormous deficits in pseudo-word reading) and, above all, the reason for any variability in the data. Overall, however, it can be said that the intervention had a transfer effect, but probably not enough, especially for weaker students. Nevertheless, the results are consistent with those of Mayer (2008), who found a greater effect on training words than on transfer words.

According to Kern and Manz (2004), goals (e.g., of an intervention) are socially invalid if they do not serve clients (e.g., students). In the current context, the survey of social validity showed that the support and its goals not only added value for the students, but that they also had a lot of fun. Thus, the intervention is a tool that can be used to increase reading skills beyond training words and that is also fully accepted by the students and is associated with fun and enjoyment. Motivation plays an enormous role, especially for students with low reading skills, and the racetracks as used here are both effective and enjoyable.

Limitations

Despite the encouraging results, some limitations of the study must be mentioned. First and foremost, it is difficult to generalize the findings, since we are only dealing with a few children and whose characteristics are very similar. Second, restrictions due to the corona pandemic prevented students from different classes to be mixed. Thus, the children from third and fourth grade were separated. It would have been preferable to have some third and fourth graders in each group. However, fortunately, there was only a one-year grade difference; besides, there was a difference in reading ability between the third- and fourth-grade children. For example, some of the children scored higher numbers of correctly read words already in Phase A – especially Jim and Latifa – making it difficult to properly assess the effectiveness of the intervention because Phase B is capped with the number of words. This was particularly evident with Jim regarding the training words. However, in each case, a pool of 70 words was involved and 20 words were always randomly drawn – and not always the same 20 words. Jim achieved the maximum number almost every time in Phase B. Another limitation stems from the fact that the intervention consists of several method parts, and it is impossible to determine which part worked and how. However, since the intervention is very easy to use, it can be implemented well as a method package in schools if you know that the package works well. Furthermore, the children came from different language backgrounds and, as a result, their mother tongue may also have exerted an influence. Yet, Lou, the only native speaker of German, did not show any clear difference in the results compared to her classmates with

German as a second language. Moreover, only the first group included two students with LD. Here it would also make sense that at least one child with LD would be represented in each group. Nevertheless, there are no clear differences between the students with and without LD.

Recommendations for Future Research

In future studies, the method package should be evaluated on a larger sample to yield stronger validity. In addition, the individual method aspects could be evaluated separately in a group or single-case design (e.g., A-B-BC plan) to determine the effects of the different aspects. It would also be interesting to find out whether participants' linguistic background makes a difference – that is, whether there is a clear difference between native speakers of German and students with German First Language (L1). Since some children in the study had severe problems in reading, and thus had difficulty memorizing the words as whole words (De Jong et al., 2012), it would make sense to reduce the number of training clusters and words for those children while at the same time increasing the frequency of occurrence of the clusters and words.

Conclusion

Overall, this single-case study showed that reading racetracks may be combined in such a way that it has not only an effect on trained words but possibly also on untrained words. For this purpose, it is sufficient to use direct instruction in frequently occurring letter clusters before implementing the racetracks and to practice them automatically. Thus, the racetracks, which have already shown very positive effects on reading in many studies (e.g., Barwasser et al., 2021a; Barwasser et al., 2021b; Grünke, 2019; Sperling et al., 2019) may be described as an even more powerful tool if combined appropriately for struggling readers with behavioral problems, with and without learning disabilities.

References

Alzahrani, T., & Leko, M. (2018). The effects of peer tutoring on the reading comprehension performance of secondary students with disabilities: A systematic review. *Reading & Writing Quarterly*, *34*(1), 1–17. https://doi.org/10.1080/10573569.2017.1302372

Barwasser, A., Urton, K., & Grünke, M. (2021a). 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

Barwasser, A., Urton, K., Grünke, M., Sperling, M., & Coker, D. Jr. (2021b). Fostering word fluency of struggling third graders from Germany through motivational peer-tutorial reading racetracks. *Reading & Writing*, https://doi.org/10.1007/s11145-021-10172-3

Bertin, J. C., & Narcy-Combes, J. (2007). Monitoring the learner – who, why and what for? *Computer Assisted Language Learning*, *20*(5), 443–457. https://doi.org/10.1080/09588220701746021

Bos, W., Valtin, R., Hußmann, A., Wendt, H., & Goy, M. (2017). IGLU 2016: Wichtige Ergebnisse im Überblick [Important results at a glance]. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, & E.-M. Lankes et al. (Eds.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich [Reading competences of primary school children in Germany in international comparison]* (pp. 13–28). Waxmann.

Burns, M. K. (2007). Reading at the instructional level with children identified as learning disabled: Potential implications for response-to-intervention. School Psychology Quarterly, 22(3), 297-313.

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

Crowley, K., McLaughlin, T., & Kahn, R. (2013). Using direct instruction flashcards and reading racetracks to improve sight word recognition of two elementary students with autism. *Journal of Developmental and Physical Disabilities*, *25*(3), 297–311. https://doi.org/10.1007/s10882-012-9307-z

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

Dyslexia Action. (2017). *Avoid dyslexia discrimination in the workplace.* https://dyslexiaaction.org.uk

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

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, *18*(1), 5–21. https://doi.org/10.1080/10888438.2013.819356

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, *53*(3), 199–208. https://doi.org/10.1177/001440298605300301

Geary, D. C., Hoard, M. K., Nugent, L., Ünal, Z. E., & Scofield, J. E. (2020). Comorbid learning difficulties in reading and mathematics: The role of intelligence and in-class attentive behavior. *Frontiers in Psychology*, *11*, 572099. https://doi.org/10.3389/fpsyg.2020.572099

Grabe, W. (2010). Fluency in reading – Thirty-five years later. *Reading in a Foreign Language*, *22*(1), 71–83. http://nflrc.hawaii.edu/rfl

Grigoryan, A. (2020). Major stages of reading skills development. *Armenian Journal of Special Education*, *1*(1), 42–51. https://doi.org/10.24234/se.2020.1.1.71

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*, 1373–1378.

Gunter, P. L., Miller, K. A., Venn, M. L., Thomas, K., & House, S. (2002). Self-graphing to success: Computerized data management. *Teaching Exceptional Children*, *35*(2), 30–34. https://doi.org/10.1177/004005990203500204

Guthrie, J. T., Wigfield, A., & VonSecker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology*, *92*(2), 331–341. https://doi.org/10.1037/0022-0663.92.2.331

Harn, B. A., Stoolmiller, M., & Chard, D. J. (2008). Measuring the dimensions of alphabetic principle on the reading development of first graders: The role of automaticity and unitization. *Journal of Learning Disabilities*, *41*(2), 143–157. https://doi.org/10.1177/0022219407313585

Hasselhorn, M., & Gold, A. (2017). *Paedagogische Psychologie - Erfolgreiches Lernen und Lehren [Educational psychology - Successful learning and teaching]* (4th ed.). Kohlhammer.

Kazdin, A. E. (2010). *Single-case research designs: Methods for clinical applied settings*. Oxford University Press.

Kern, L., & Manz, P. (2004). A look at current validity issues of school-wide behavior support. *Behavioral Disorders*, *30*(1), 47–59. https://doi.org/10.1177/019874290403000102

Knoepke, J., Richter, T., Isberner, M., Naumann, J., & Neeb, Y. (2014). Phonological recoding, orthographic decoding, and comprehension skills during reading acquisition. *Zeitschrift*

für Erziehungswissenschaft, *17*(3), 447–471. https://doi.org/10.1007/s11618-014-0560-z

Kostewicz, D. E., Kubina, R. M., Selfridge, K. A., & Gallagher, D. L. (2016). A review of fixed fluency criteria in repeated reading studies. *Reading Improvement*, *53*(1), 23–33.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, *34*(1), 26–38. http://doi.org/10.1177/0741932512452794

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

Lerner, J., & Johns, B. (2011). *Learning disabilities and related mild disabilities: Characteristics, teaching strategies, and new directions* (11th ed.). Houghton Mifflin.

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.

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

Marinus, E., de Jong, P., & van der Leij, A. (2012). Increasing word-reading speed in poor readers: No additional benefits of explicit letter-cluster training. *Scientific Studies of Reading*, *16*(2), 166–185. https://doi.org/10.1080/10888438.2011.554471

Martin-Chang, S. L., & Levy, B. A. (2005). Fluency transfer: Differential gains in reading speed and accuracy following isolated word and context training. *Reading and Writing*, *18*, 343–376. https://doi.org/10.1007/s11145-005-0668-x

Mayer, A. (2008). *Phonologische Bewusstheit, Benennungsgeschwindigkeit und automatisierte Leseprozesse [Phonological awareness, naming speed and automated reading processes].* Shaker Verlag.

Mayer, A. (2018). *Blitzschnelle Worterkennung [Rapid word recognition]* (3rd ed.). Borgmann Media.

Moll, K., & Landerl, K. (2010). *SLRT-II: Lese-und Rechtschreibtest; Weiterentwicklung des Salzburger Lese-und Rechtschreibtests (SLRT) [Reading and spelling test; further development of the Salzburg reading and spelling test]*. Huber.

Morris, D., & Perney, J. (2018). Using a sight word measure to predict reading fluency problems in grades 1 to 3. *Reading & Writing Quarterly*, *34*(4), 1–11. https://doi.org/10.1080/10573569.2018.1446857

Mueller, B., Richter, T., Križan, A., Hecht, T., & Ennemoser, M. (2015). Word recognition skills moderate the effectiveness of reading strategy training in Grade 2. *Learning and Individual Differences*, *40*, 55–62. https://doi.org/10.1016/j.lindif.2015.05.001

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

Oakes, W. P., Mathur, S. R., & Lane, K. L. (2010). Reading interventions for students with challenging behavior: A focus on fluency. *Behavioral Disorders*, *35*(2), 120–139. https://doi.org/10.1177/019874291003500204

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping 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. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, *35*(4), 303–322. https://doi.org/10.1177/0145445511399147

Rasinski, T. V., Rupley, W. H., Paige, D. D., & Nichols, W. D. (2016). Alternative text types to improve reading fluency for competent to struggling readers. *International Journal of Instruction*, *9*(1), 163–178. https://doi.org/10.12973/iji.2016.9113a

Sáenz, L. M., McMaster, K. L., Fuchs, D., & Fuchs, L. S. (2007). Peer-assisted learning strategies in reading for students with different learning needs. *Journal of Cognitive Education and Psychology*, *6*(3), 395–410. https://doi.org/10.1891/194589507787382089

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Reading interventions for adolescent struggling readers: A meta-analysis with implications for practice*. RMC Research Corporation, Center on Instruction.

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *Journal of Learning Disabilities*, *48*(4), 369–390. https://doi.org/10.1177/0022219413504995

Sideridis, G. D. (2002). Goal importance and students at risk of having language difficulties: An under explored aspect of student motivation. *Journal of Learning Disabilities*, *35*(4), 343–356. https://doi.org/10.1177/00222194020350040501

Solis, M. M., Ciullo, S., Vaughn, S., Pyle, N., Hassaram, B., & Leroux, A. (2012). Reading comprehension interventions for middle school students with learning disabilities: A synthesis of 30 years of research. *Journal of Learning Disabilities*, *45*, 327–340. https://doi.org/10.1177/0022219411402691

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., & Snyder, A. (2007). Effects or 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. https://doi.org/10.1177/10634266070150020101

Tressoldi, P., Vio, C., & Iozzino, R. (2007). Efficacy of an intervention to improve fluency in children with developmental dyslexia in a regular orthography. *Journal of Learning Disabilities*, *40*(3), 203–209. https://doi.org/10.1177/00222194070400030201

van Keer, H., & Vanderlinde, R. (2010). The impact of cross-age peer tutoring on third and sixth graders' reading strategy awareness, reading strategy use, and reading comprehension. *Middle Grades Research Journal*, *5*(1), 33–45.

Vaughn, S., Levy, S., Coleman, M., & Bos, C. S. (2002). Reading instruction for students with LD and EBD: A synthesis of observation studies. *The Journal of Special Education*, *36*(1), 2–13. https://doi.org/10.1177/00224669020360010101

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 off German schoolchildren. *Journal of School Psychology*, *66*, 25–40. https://doi.org/10.1016/j.jsp.2017.11.003

Young, C., Rasinski, T., Paige, D., & Rupley, W. (2020). Defining fluency. Finding the missing pieces for reading fluency. *Literacy Today*, *37*(6), 32–33.

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.

Zavala, E., & Cuevas, J. (2019). Effects of repeated reading and rhyming poetry on reading fluency. *International Journal of Social Sciences & Educational Studies*, *6*(2), 64–87. https://doi.org/10.23918/ijsses.v6i2p64

Appendix D Article 4

Barwasser, A., Urton, K., Knaak, T. & Grünke, M. (2021). Intentional and incidental vocabulary acquisition in German L2 primary school students through multi-component storytelling. *Language Teaching Research.*

Abstract

The increasing number of students with German as a second language (GL2) poses a great challenge for schools. Previous studies show that especially young learners with a migration background are lagging behind in the acquisition of L2 literacy. Adequate vocabulary, including knowledge of word characteristics and word information, is essential for appropriate L2 language proficiency, and therefore both are crucial aspects of support that should be focused on. Thus, the aim of the study was to support vocabulary acquisition in students with GL2 regarding two aspects: (1) the intentional acquisition of expressive and receptive vocabulary, and (2) the incidental acquisition of vocabulary with respect to additional vocabulary facts that are not explicitly focused on. To this end, the effectiveness of a multi-component method consisting of storytelling and flashcards as well as motivational components was examined. A multiple baseline design was applied. The intervention was carried out with nine primary school students with GL2 in three small groups three times a week for four weeks. The results indicate that the training was very effective in increasing students' receptive and expressive vocabulary as well as additional vocabulary facts in just one month. The paper concludes with a critical discussion of the findings and provides practical suggestions for schools on how to adequately support and motivate students with GL2 in their vocabulary acquisition.

Keywords: Storytelling, Vocabulary, Vocabulary Facts, German Second Language, Motivation

Introduction

Migration and German Second Language Learning

Worldwide migration is increasing (International Organization for Migration, 2019). In Germany, where this study took place, the current number of the population with migrant background is 21.3 million (Federal Statistical Office, 2020). Due to this, multilingualism is very common in German schools and therefore teaching German as a Second Language (GL2) is a central task of the German education system. Here it is evident that students with GL2 lag far behind those with German as First Language (GL1) in terms of school success (e.g. Harju-Luukkainen et al., 2020; Kristen & Granato, 2007). This highlights the importance of "integrated

language education" in the classroom, whereby teachers ensure that students have sufficient opportunities to actively use and practice a language and receive individual support as needed (Gogolin et al., 2011; Ministry for School and Further Education NRW, 2016). However, if one asks teachers how competent they perceive themselves to be in this area, the results of a representative survey by Becker-Mrotzek et al. (2012) indicate that approximately 70% of all teachers educate students with a migration background, but only 30% feel well equipped for this task. This makes it necessary to provide teachers with adequate and effective methods that enable them to promote the language skills of their students in everyday school life.

Vocabulary as a main impact on Language Proficiency

A fundamental starting point for the development of students' linguistic competences is the development of vocabulary. According to Wilkins (1972) someone without grammatical knowledge can express very little, but without vocabulary this is not possible at all. Therefore, in language teaching and learning, it is central to focus on vocabulary (Amiryousefi & Ketabi, 2011; Clark & Paivio, 1991; Harmon et al., 2009; Schmitt, 2008), as the number of words that L2 learners know and are able to use correctly, has a major impact on their literacy skills (Laufer, 1992; Laufer & Nation 1995).

Vocabulary knowledge can be divided into receptive and expressive word retrieval (Nation, 2001; Schmitt, 2014). Receptive knowledge is the ability to recognize a word, but which cannot be actively produced (Webb, 2009; Nation, 1990). Receptive knowledge is considered to be acquired faster than expressive knowledge (Webb, 2008). The use of expressive vocabulary means that learners can actively apply words in a particular situation (Webb, 2008).

Regarding the importance of vocabulary for literacy skills, positive correlations are shown in several studies as in the one conducted by Stæhr (2009; r = .83) between receptive vocabulary knowledge and reading comprehension and for oral vocabulary knowledge on reading comprehension for L2 learners (Milton et al., 2010; r = .70). Furthermore, Suggate et al. (2013) revealed some evidence that vocabulary correlates with reading comprehension in 2nd and 4th graders. Also, a correlation has been found between listening comprehension and vocabulary knowledge (Stæhr, 2008; 2009). Considering vocabulary knowledge and writing performance, significant correlations are also found (Nation, 2001; Van Gelderen et al., 2004). Thus, the meta-analysis of Graham & Zohreh (2020) gives an indication that vocabulary knowledge explains about 31% of the variance in writing performance.

How to teach Vocabulary?

The foregoing illustrates that vocabulary knowledge can contribute significantly to language competence (Stæhr, 2008; Killic, 2019). Accordingly, it is central to address the question of which methods are suitable for effective vocabulary building in the context of second language acquisition. A useful theoretical background here is a multimodal approach (teaching a language in several ways). This will be explained in more detail in the following on the basis of the Dual Coding Theory (DCT, Paivio, 1991) and the Cognitive Theory of Multimedia Learning (CTMML, Mayer, 2005) in relation to second language acquisition.

Dual Coding Theory in L2 Learning

The DCT (Paivio, 1991) provides guidance in finding methods that can effectively promote language learning and vocabulary acquisition. It assumes that information presented to a person can be processed in two different systems: (1) verbally (e.g. through expressions) and (2) non-verbally (e.g. through images and gestures). Since the probability of retaining information is higher when the memory representation is based on both verbal and non-verbal coding, dual coding (DC) is of particular importance for information processing. Accordingly, it is favorable to teach vocabulary on the basis of the spoken and written words (verbally) as well as in parallel with a suitable picture (non-verbally). Due to the good representability of concrete words, DC possibly has a particularly positive effect here, since besides verbal information, non-verbal information can also be presented particularly well and subsequently stored in addition to verbal information. As will be discussed further below, DC can also be applied well in storytelling, as the narrative (verbal) can be enriched by means of pictures and gestures (non-verbal).

Cognitive Theory of Multimedia Learning in L2 Learning. CTMML (Mayer, 2005) is also based on a dual channel assumption seeing each individual as active knowledge constructors by connecting words and pictures meaningfully in order store them in the long-term memory. As well as the DCT the CTMML dual-channel assumption states that humans have two separate information processing channels the auditory/verbal and the visual/non-verbal. The verbal channel takes in information through hearing, while the visual channel processes facts through vision. Based on this theory, it can be assumed that L2 learning occurs more effectively through multimedia input (linking written and pictorial cues to visual as well as verbal systems) than when language is taught through a single-mode (Mayer, 2014; Ramizanali & Faez, 2019). Recent research supports both theories, DCT and CTMML, for vocabulary acquisition, as the combination of verbal and non-verbal aspects has proven to be beneficial for vocabulary teaching (Amiryousefi & Ketabi, 2011; Ramezanali et al., 2020; Uchihara et al., 2019).

Intentional and Incidental Vocabulary Learning and Teaching. Besides different modes of learning, learning and teaching of a language, especially vocabulary acquisition, can take place intentionally or incidentally (Graves, 2016). Incidental language learning means acquiring something without being aware of it (Zeeland & Schmitt, 2013; Webb & Nation, 2017), whereas intentional language learning is deliberate (Ellis et al., 2009). That vocabulary acquisition can occur effectively both intentionally and incidentally is shown by a number of studies. Thus, explicit vocabulary instruction proved to be effective for vocabulary acquisition as well as comprehension (e.g., National Institute of Child Health and Human Development (NICHD), 2000; Elleman et al., 2009; Scammacca et al., 2007). Considering the methods utilized, it appears that both nonverbal and verbal mnemonics (De Vos et al., 2018; Uchihara et al., 2019; Mizumoto & Takeuchi, 2009) and, in particular, the use of explicit instruction are effective for vocabulary acquisition (e.g., Ford-Connors & Paratore, 2015; Kamil et al., 2008; NICHD, 2000; Stahl & Fairbanks, 1986; Cervetti et al., 2016). This was also shown for linguistically diverse students (Lesaux et al., 2014). Furthermore, incidental learning is also central to language learning, as words learned through context are better remembered than when they are merely explicitly trained (Webb & Nation, 2017; Peker et al., 2018). In addition, contextual cues have a beneficial effect on incidental vocabulary learning (Teng, 2016; Webb, 2008). Thus, the use of both modalities is central as, according to Marulis & Neumann (2010), vocabulary learning is particularly effective when it is both, intentionally and incidentally, learned. Furthermore, since word knowledge has various levels, and the complete learning of a word is a multifaceted task it is assumed that additional information about the word is acquired during vocabulary learning, such as related ideas (Graves, 2016). This is of relevance as, in the sense of Tversky (1977), words do not represent isolated units, but rather are embedded in a complex network that connects previous experiences, associations with other words, concepts and ideas (Ford-Connors & Paratore, 2015; Nagy & Scott, 2000).

Motivation in L2 Learning

Another aspect of successful learning is motivation, which also affects L2 acquisition (Ellis, 1997; Doernyei, 2006; Grey et al., 2015). There is consensus that motivation is a major contributor to success here (see, e.g., Crookes & Schmidt, 1991; Doernyei, 2001; Doernyei & Skehan, 2003; García & de Caso, 2004; Gardner 1985; Horwitz et al., 1986; MacIntyre, 2002; Masgoret & Gardner, 2003; Saito et al., 2017; Sparks, 1995; Tremblay & Gardner, 1995). However, motivation is a dynamic parameter which varies depending on the task (Dornyei, 2020; Ellis, 2008). As such, in the context of facilitating learning, "the most pressing question related to motivation is not what motivation is but rather how it can be increased" Doernyei (2001, p. 51). According to Doernyei (2020), this needs to be clarified especially with regard to specific processes of L2 learning, such as vocabulary acquisition.

A Multi-Component L2 Intervention to Foster Vocabulary

Vocabulary Learning through Storytelling and Flashcards

Working with stories for improving language skills has a long tradition in early literacy (Snow, 1993; Wells, 2009). The development of phonemic awareness through reading stories at home with parents or in preschool education is seen as a prerequisite for later reading and writing abilities (Snow, 1993; Cochran-Smith, 1984). Stories therefore already play an important role in early language development and can also have an impact on later language learning as storytelling or more precisely hearing a story, has a significant impact on vocabulary development, especially when the story content is meaningful to the learners' interests (Mol & Bus, 2011; Kirsch, 2012). By presenting new vocabulary in a given story word knowledge is gradually expanded (Ellis & Brewster, 2002; Frishkoff et al., 2011; Landauer et al., 2011). In this process, each successive presentation creates links to other related content - in the sense of semantic networks (Borovsky et al., 2016; Ford-Connors & Paratore, 2015), which ultimately leads to the usage of the word productively in context (Durso & Shore, 1991). In particular, new vocabulary appears to be remembered best when it is taught in a meaningful context linked to learners' interests (Amiryousefi & Ketabi, 2011; Dong, 2013; Leons et al., 2009; Nation, 2015; Oxford & Scarcella, 1994). Storytelling can also be used in a systematic way to train vocabulary in students with L2. Positive effects were revealed here on students' reading comprehension and general language skills (Al-Mansour & Al-Shorman, 2011; Hemmati et al., 2015; Huang, 2006; Kim, 2010). High effects were found for both interventions when comparing storytelling with explicit word instruction and elaborative storytelling without explicit instruction, but these were more pronounced when the target words were explicitly instructed (Vaahtoranta et al., 2017). Suggate et al. (2013) found that German children in grades 2 and 4 showed less vocabulary growth when stories were read aloud compared to free storytelling.

With regard to the interpretation of the results of the reported studies, it must be critically noted that the term storytelling covers a wide variety of approaches. In the present study, storytelling is defined according to Roney (1996). In this sense, storytelling involves a narrator communicating with the audience by telling a story using pronunciation, narrative structures, and mental imagery, and the audience responds to the story verbally and non-verbally, with body language and mimic. In line with previous implementations in language teaching, this aims to increase students' vocabulary (vocabulary learning and comprehension) verbally and in combination with visual aids (e.g., Al-Mansour & Al-Shorman, 2011; Hemmati et al., 2015; Huang, 2006; Kim, 2010). Following the idea of the DCT and the CTMML, the present research targets to enhance students' intentional and incidental vocabulary learning by considering different modalities (verbal and non-verbal) in vocabulary learning through storytelling combined with flashcards. Thus, the story is read expressively by a narrator and accompanied

by gestures and the students can read along with the story themselves. Because pure storytelling is more of an incidental way of conveying words, illustrated flashcards, containing the pictures of the training words and the training words themselves, are appropriate to force intentional learning. As such, the use of flashcards in terms of DCT and CTMM also supports information processing and storage into long-term memory (Abbasian & Ghorbanpout, 2016; Oxford & Crookall, 1990; Thompson, 1987). Due to the direct instruction associated with the use of flashcards, the intervention is further expected to increase in effectiveness (Crowley et al., 2013; Rich et al., 2016; Schmitt, 2010; Thompson, 1987; Uchihara et al., 2019). The effectiveness of this combined storytelling method with flashcards for vocabulary training in English as L2 for students with learning disabilities was already demonstrated (Barwasser et al., 2020; Knaak et al., 2021).

Motivational Components

As mentioned before, motivation plays an important role in successful L2 acquisition (Ellis, 1997; Doernyei, 2006; Grey et al., 2015). Learners' engagement with lexical items can be boosted by using motivational components and it facilitates learning and cognition (Pintrich, 2003; Woodrow, 2017). In this regard, it is important to note that motivation is not a stable characteristically attribute of a learner but rather a dynamic process influenced by the way language is conveyed (Dornyei, 2020; Ellis, 2008). Therefore, the type of used tasks is significant. To ensure that each learner remains motivated during the task, motivational components can be embedded into a treatment (Cooper et al., 2007). Even though it can be assumed that storytelling alone, through the stories used, already increases students' motivation in L2 learning, the application of additional motivational components should ensure that the interest of all learners is maintained. According to the relevant research, methods of self-monitoring are an essential tool (Wells et al., 2017; Amato-Zech et al., 2006) to increase learners' self-regulation skills which positively influence academic success and self-motivation (Falkenberg & Barbetta, 2013; McDougall et al., 2012; Rafferty, 2010). Self-graphing as one method of self-monitoring procedures (Hirsch et al., 2013) allows students to make their own learning process transparent by filling in the results of a given task in a graph. By daily recording their results, students can follow their own improvements which seems to have a positive effect on school success, on task behaviour and motivation (Amato-Zech et al., 2006; Gunter et al., 2003).

As storytelling is furthermore not conducted in a one-to-one setting but with a group of students, methods that support working in group contexts could be helpful too. An effective tool for increasing motivation here is interdependent group contingencies (Bowman-Perrot et al., 2013; Cooper et al., 2007; Little et al., 2015; Popkin & Skinner, 2003). These methods can increase the on-task behaviour of students since they ensure that each member is responsible for the success of the entire group (Cooper et al., 2007; Ginsburg-Block et al., 2006; Wills et

al., 2016). In this case, each student's individual results are summed for the entire group to collectively earn an agreed-upon reward (see *Intervention*).

Research Aims

L2 research has led to various ways of acquiring a new language, but there is still too little research on effective vocabulary interventions in GL2 in Germany. Due to the increasing number of students with a migration background in recent years and the growing challenge for teachers to find effective and motivating support, it has become increasingly important to establish methods that can help diverse learners to build up a comprehensive vocabulary in German as a foundation for further successful learning at school.

The aim of the study at hand - based on the DCT and the CTMM as well as the findings on intentional and incidental learning and the importance of motivation in vocabulary acquisition - is to investigate the effects of a motivational two-stage storytelling intervention on the intentional acquisition of receptive and expressive vocabulary as well as the incidental acquisition of additional vocabulary facts in elementary school students with GL2. The research questions are as follows:

1) Does the multi-component storytelling intervention improve the German receptive vocabulary of struggling GL2 primary school students?

2) Does the multi-component storytelling intervention improve the German expressive vocabulary of struggling GL2 primary school students?

3) Does the multi-component storytelling intervention improve recall of additional vocabulary facts not intentional learned through the intervention of struggling GL2 primary school students?

Methods

Participants and Setting

The study was conducted at an inclusive primary school in North Rhine-Westphalia (Germany) among students in the 2nd, 3rd, and 4th grade. The subjects were selected in a multi-stage process: In a first step (1), the class teachers selected all children with GL2 as possible participants. All children learned German with the entry of kindergarten at age 3-4. As second step (2), the extent of German vocabulary was measured by a vocabulary test for the selected students, as GL2 does not necessarily lead to reduced vocabulary knowledge in German. All children with a percentile rank <15 were eligible for the study (N = 15). There were 7 children from 2nd grade, 5 from 3rd grade and 3 children from 4th grade. As third step (3) the German short version of the Integrated Teacher Report Form (ITRF-G; Volpe et al., 2018) to determine externalizing problem behavior was filled out by the class teachers. Since the storytelling intervention is a quite open method, we were interested in how students who face

behavioral problems react to the intervention since problem behavior is a challenge for teachers all around the world. As fourth step (4) a teacher questionnaire was used to collect general information about the subjects. The fifth and final step (5) was a word pre-test. This determined the final 20-word vocabulary for the intervention and ensured that none of the children had any of the words in their receptive or expressive vocabulary.

Informed consent was obtained from parents prior to data collection. Students' names were pseudonymized for privacy reasons. These are used in the following. Finally, of the 15 children, only *N* = 9 are included in the study, as school attendance was not possible for all the others due to the corona pandemic. The age range of the children was from 7.3 to 11.5 years. None of the children had a disability or received special needs support. Tugce, Mikail, and Cem showed an increased overall score (cutoff = 13) in the ITRF and thus indication of externalizing problem behavior. Tugce, Yusuf, Miray, Ferhat and Eyluel have Turkish as their L1, in the case of Elif and Mikail it is Arabic L1 and Cem and Defne speak Albanian as their L1 at home. For more information, see Table 1.

Table 1

Demographic Characteristics of the Participants

Name	Gender	Age	Vocabulary Test PR	ITRF GP	SN	GL2	L1
Tugce	female	7.3	2	18	0	Yes	Turkish
Elif	female	8.5	12	11	0	Yes	Arabic
Yusuf	male	10.5	14	10	0	Yes	Turkish
Miray	female	7.5	12	6	0	Yes	Turkish
Mikail	male	8.0	8	18	0	Yes	Arabic
Cem	male	11.5	7	15	0	Yes	Albania
Ferhat	male	9.3	3	3	0	Yes	Turkish
Eyluel	female	9.0	5	10	0	Yes	Turkish
Defne	female	10.0	10	9	0	Yes	Albania

Note. PR = percentile rank (<15 weak performance); ITRF = Integrated Teacher Report Form GP = main problem value (cutoff 13); SN = Special Needs; L1 = First Language; GL2 = German Second Language

Experimental Design

For the evaluation of the intervention, a multiple baseline design across participants was used (AB plan) with a total of twelve measurement points (Ledforst & Gast, 2014). The children were randomly divided into three groups. The baseline lengths varied between groups from four to six and the intervention sessions between eight to six time points. The baseline and intervention sessions were held three times a week over four weeks for 40 minutes. Three graduate university students of special needs education acted as interventionists and test leaders. Each interventionist was responsible for one of the three groups and their measurements. Before each baseline and intervention session, the participating students were

picked up from their classrooms. To allow for a disturbance-free implementation, the baseline and intervention session for each group took place in a separate room.

Measurements

Screenings

German Vocabulary Test. The vocabulary subtest of the German version of the Culture Fair Intelligence test (CFT; Weiß, 2006, cutoff percentile <15) was used to assess German word knowledge. It contains a total of 30 items with words from semantic fields and abstract terms. Students are shown one word and five other words in the margin. The task is to decide which of the five words best matches the one word shown. The test procedure is based on a representative sample of *N*= 2724 of all school types from the third to the 13th grade. In addition, a good reliability (*r* = .87) was examined (Weiß, 2006).

ITRF-G. The German short version of the ITRF-G consists of 16 items, of which eight items belong to 1) learning-related behavior (APD) and eight items belong to 2) oppositional/disruptive behavior (OD). The items are scored on a four-point Likert scale by the class teachers. Cutoffs for APD is ten, for OD is five, and for total problem score (GP) is 13.

Teacher Questionnaire. The teacher questionnaire consists of various questions about individual students such as age, date of birth, special needs support, first language, second language, language at home and whether the students receive additional help in German at school and/or outside of school.

Word pretest. Because the vocabulary in this study is based on the semantic field of animals, a total of 96 different animal words was selected. Each child was tested individually. The receptive testing was based on the *Peabody Pictures Test* in English (PPVT; Dunn & Dunn, 2007) and the expressive test was based on the *Expressive Vocabulary Test* (EVT; Williams, 2007). For the receptive part, pictures of animals were printed on little flashcards. Then, the child was asked to point to the corresponding picture card when an animal was named by the test leader. In the expressive part, animal pictures were presented by means of a PowerPoint Presentation, with one slide per animal. METACOM© symbols (Kitzinger, 2020) and online freely available images (www.pixabay.de) served as animal pictures. Based on this word pre-test, 20 animals were chosen which were neither stored receptively nor expressively across all children.

Data collection

In total there were three dependent variables. 1) receptive vocabulary, 2) expressive vocabulary and 3) vocabulary facts. The measurement was structured as follows: Three stacks of cards were placed on a table. 1) The first stack was for receptive vocabulary, 2) the second for expressive vocabulary and the 3) third for vocabulary facts. The first two stacks contained

20 cards per measurement (= all 20 training words). In order not to overload the students, the stack with vocabulary facts included only ten of the open-ended questions on the trained animal vocabulary per session. For receptive and expressive vocabulary, all training words were presented in random order, while for the vocabulary facts, always ten questions were randomly selected from a pool of 40 questions (there were two questions for each of the 20 animals). There was no time limit for answering.

Receptive vocabulary. To measure receptive vocabulary, the 20 words of training were written one on each card of the pile. The test leader uncovered the card and read the word aloud. Next to the stacks, all 20 animal pictures were spread out on the table. Each child should now point to the correct picture to the word read aloud. The correctly shown picture was considered to be the correct receptively known training word.

Expressive vocabulary. For expressive vocabulary, one animal pictures was printed on each card. The cards were shown to the participants one other the other and the child asked by the test leader to verbally express the fitting word. If a student was able to name the correct animal name, the word was considered expressively known.

Vocabulary Facts. For each of the 20 training words there were two questions of additional information that were told during the storytelling but not explicitly focused on. The cards for measurement had one question per card, e.g. “What animal builds a home out of up to 100 interconnected burrows?” and the kids were asked to actively name the fitting animal. For each measurement, always ten question cards were randomly shown to the students but making sure that all question cards were equally distributed across the measurement. For the measurement procedure, the interventionist turned the card over and read the question out loud. Each correct answer was considered correct explicitly vocabulary fact knowledge. The questions were conceived by the interventionists, paying attention to equal difficulty and the occurrence of the facts in the stories in equal parts. The stacks were processed one after the other. The test leader recorded student responses on a documentation sheet that was not visible to the participant.

Intervention Material

The three interventionists wrote the stories to be used for storytelling themselves (see Figure 1). There was a total of six to eight stories (dependent on baseline length). Each session had one story. In terms of content, the stories were built on each other.

Figure 1

Example of a Story Part

Die Reise zum Planeten Maximegalon Die erste Spur	Als die beiden Tiere schon eine Weile laufen, entdeckt der Hirsch auf dem Sandweg vor ihm Spuren. Die Spuren sehen aus wie zwei große Kreise. „Oh, wer ist hier gelaufen? Wo führt die Spur wohl hin?", fragt sich der Hirsch. Schnell werden die	anderen Tiere geholt, um die Spur gemeinsam zu verfolgen.

Note. Translation: Title: The journey to the planet Maximegalon. The first track; Text: When the two animals have been walking for a while, the deer discovers tracks on the sandy path in front of him. The tracks look like two big circles. "Oh, who has been walking here? I wonder where the trail leads?" the deer wonders. Quickly the other animals are fetched to follow the trail together.

Each story consisted of about 130 words, with particular attention to ensuring that all stories were of same difficulty and contained a similar sentence structure. Always 5 of the 20 training words were embedded in the stories (to avoid cognitive overload) and colored in red. The random assignment of words to stories was done via Excel, making sure that the words were equally distributed among the stories. The first stories served as an introduction to all 20 animal words (see Table 2), and all subsequent stories contained a random selection of five words, taking care to ensure that all words were used equally often at the end of the intervention. Together with the teachers it was decided on the vocabulary in the area of animals, since the children will deal with the topic of animals as part of their science lessons. In addition, the stories were adapted to the everyday interests of the students and contain characters with which the pupils can easily identify. The plot framework resembles an exciting story with several episodes. The stories were printed on 8.3 x 11.7-inch sheets with a maximum of 7 sentences on one page. These stories were stapled to a ring binder and placed in front of the children so that they could follow the story visually and acoustically.

Figure 2

Flashcard Example

Note. Hirsch (deer)

In addition, for Phase 1, 8.3 x 11.7-inch flashcards (see Figure 2) were used to introduce the words of training, consisting of one of the 20 selected words and a matching picture, printed on 8.3 x 11.7-inch sheets.

Table 2

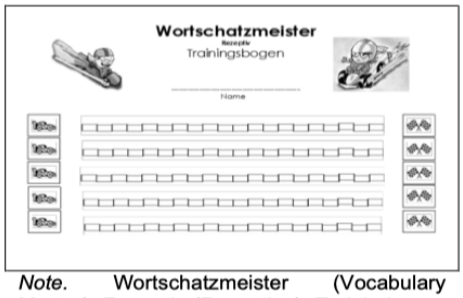
Words of Training

Frosch	Hirsch	Faultier	Wellensittich	Fuchs
(frog)	*(deer)*	*(sloth)*	*(budgie)*	*(fox)*
Biber	Ente	Meerschweinchen	Dachs	Garnele
(beaver)	*(duck)*	*(guinea pig)*	*(badger)*	*(prawn)*
Walross	Kamel	Moewe	Krake	Ameise
(walrus)	*(camel)*	*(seagull)*	*(octopus)*	*(ant)*
Fisch	Affe	Delfin	Hai	Eisbaer
(fish)	*(monkey)*	*(dolphin)*	*(shark)*	*(polar bear)*

The self-graphing sheets (see Figure 2) consisted of twelve rows one below the other and each row had 20 boxes corresponding to the maximum number of training words.

Figure 3

Self-Graphing Sheet



Note. Wortschatzmeister (Vocabulary Master); Rezeptiv (Receptive); Trainierbogen (Training Sheet)

Procedures

Baseline. As baseline condition, the children worked on cognitive tasks for 25 minutes. These cognitive tasks were explained to the children as puzzle tasks. They were given tasks that involved 1) recognizing which of the five symbols in a row did not match the others and 2) logically continuing a row with certain symbols by ticking the correct one from a selection of solutions. The baseline serves as a record of the actual state and is intended to collect the dependent variables without intervention. The cognitive load of the children through the tasks in the baseline leads to a more valid interpretation of the results, as pure attention effects can thus be mitigated. Students were already engaged at baseline for the exact same time period as the storytelling intervention. After each baseline the participants were tested in isolation with regard to the three dependent variables.

Intervention. The storytelling intervention consisted of two phases each session. Phase 1) included the introduction of always five words by direct instruction with the help of

the flashcards and from session two on the repetition of the last words for about ten minutes. Always five of the 20 training words were introduced in each session. The interventionist showed a flashcard with a picture and a matching word while hiding the written word asking if someone knows this animal and if anyone has ever seen it. This process was continued for each word. Phase 2) was realized through the method of storytelling for 15 minutes. A story was read aloud by the interventionists while using different gestures and expressions. The interventionist learned to tell the story by heard but since it has been shown that reading while listening is beneficial (Brown et al., 2008; Teng, 2018), the story with the marked training words was shown as well. If a word of the training appeared in the story, the telling process was interrupted, whereupon the corresponding flashcard, which was already used in stage one, was shown to the kids, and they were asked to name the correct animal. After this was completed, the telling process continued. The vocabulary facts which were also measured after each session were never explicitly taught during the intervention but just appeared in the story context without guiding the participants' focus on it. Thus, these facts were learned incidentally whereas receptive and expressive vocabulary were learned incidentally and intentionally.

Following each intervention, participants were tested in isolation on the three dependent variables.

For the reward system, two self-graphing sheets (see Figure 1) with 20 boxes for each session, one sheet for receptive and one sheet for expressive vocabulary for each participant was implemented. The small boxes per row were colored depending on the total number of correctly known receptive and expressive words achieved in each measurement. Thus, the children could track their learning progress after each session. Based on the colored boxes for each session, the reward system was implemented. In the group contingency procedure each group collected points, which were illustrated by marbles in a glass. Depending on each child's individual test results, this glass was filled with an appropriate number of marbles, and each child could place a maximum of four marbles in the glass per session. Points were determined as follows: Two points for a result that was better than the previous one, one point for keeping the result from the last time, and no points if the current result was lower than the previous one. The procedure was always carried out at the beginning of the next session for about five minutes, referring back to the last measurement results, so that all children could color the sheets at the same time. A group target of marble was set beforehand, so that each individual contributes to achieving the target in the sense of a group contingency procedure.

Treatment Integrity

The implementation quality of the intervention was ensured by a training of the interventionists by the first author prior the study. Interventionists were micro-instructed on how to conduct baseline sessions and intervention sessions as part of the storytelling intervention

in a 2-hour session. In addition, they were instructed in another 2-hour session on how to conduct pretesting and how to collect learning progress data in the multiple baseline design. Furthermore, testing the treatment integrity (TI) is a central aspect in order to make valid statements about the effectiveness of an intervention (Sanetti & Kratochwill, 2009). The TI checklist consisted of six scales: 1) environment/external circumstances (e.g. “Did the promotion take place in the familiar environment?”), 2) planning (“Was there a planning sheet for the current session?”), 3) material (“Was all material available?”), 4) course of the support (e.g. “Was the intervention implemented as planned?”), 5) diagnostics & feedback (e.g. “Was the learning progress of the students recorded?”) and 6) dealing with student behavior (e.g. "Did the students receive feedback?"). Scales oneto five could be answered with either yes or no. Scale six was scaled from 0 = not applicable at all to 4 = entirely applicable. The TI was filled in by the interventionists themselves after each session for each group and in addition an external person, who was also instructed in the intervention, filled in the TI sheet for each group for 1/3 of the interventions. Overall, the interrater reliability was 100%.

Data Analysis

For descriptive and quantitative analyses, the SCAN package for R Statistics by Wilbert and Lueke (2019) was used. Despite descriptive data, measures of non-overlap indices were calculated to assess the effectiveness of the intervention. First, visual inspection is resorted to, which is common in the context of single case research. In addition, descriptive data are reported, and for a more in-depth analysis, overlap measures are used to determine the effectiveness of the intervention, and a level 2 regression analysis is performed across all subjects, focusing on the slope, the increase from A phase to B phase, and the level effect, the direct increase at the onset of the intervention. With regard to the measures of overlap, the following measures were chosen: Non-overlap of All Pairs (NAP, Parker et al., 2011a), the Percentage Exceeding the Median (PEM, Ma, 2006) and Tau-U derived from the Mann-Whitney U and Kendall's rank correlation and, with A-phase trend correction (Parker et al., 2011b; A vs. B + TrendB - TrendA). The NAP is the percent improvement in data across all phases, with 0.00-65.00 indicating a weak effect, 66.00-92.00 indicating a moderate effect, and 93.00-100.00 indicating a large effect (Parker et al., 2011a). PEM is the percentage of data points exceeding the median baseline, where less than 70.00 is a non-effective treatment, 70.00-90.00 is a moderate effect, and above 90.00 is a large effect (Ma, 2006). Tau U values can be interpreted as follows: up to 0.20 improvement can be considered a small change, 0.20 to 0.60 a moderate change, 0.60 to 0.80 a large change, and above 0.80 a very large change (Parker et al., 2011b).

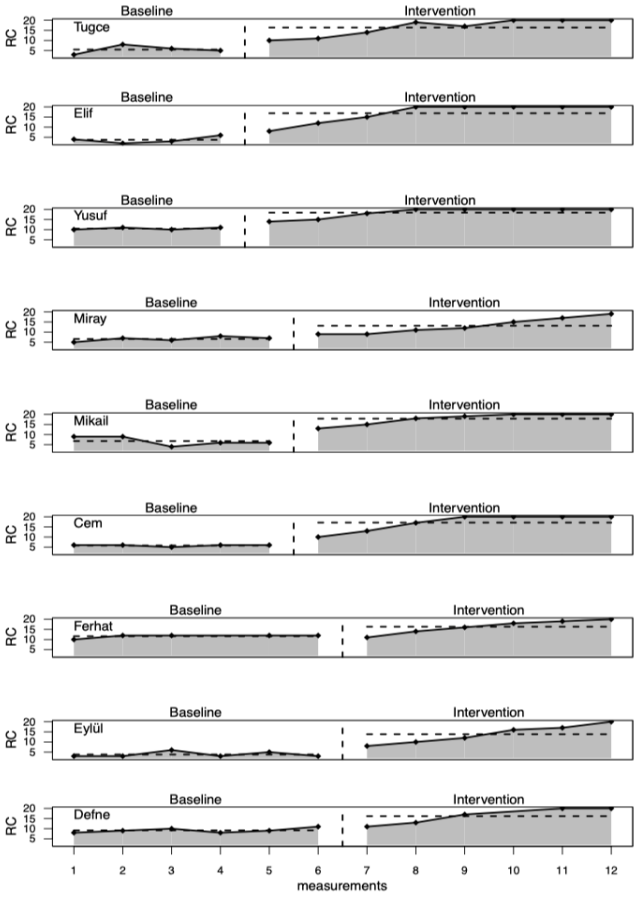
Results

Receptive Vocabulary

Regarding receptive vocabulary, a steady baseline can be established for all students (Figure 3). Moreover, it also shows that there is a significant increase in correctly known words in the B-phase and the highest number of correctly known words was quickly reached.

Figure 4

Receptive vocabulary (RC)



From the descriptive data, it can be seen that students can retrieve some words receptively already in the baseline. Nevertheless, it can be observed that in all participants there is an increase in the B-phase and almost all of them reach the maximum score of 20.00 in a very short time. Only Miray does not make it to 20.00, but she is close with 19.00.

Table 3

Descriptive Data for Each Participant in A and B Phase Regarding Receptive Vocabulary

	N (A)	*N (B)*	*M (A) (SD)*	*M (B) (SD)*	*Max (B)*
Tugce	4	8	5.50(2.08)	16.38(4.17)	20.00
Elif	4	8	3.75(1.71)	16.88(4.70)	20.00
Yusuf	4	8	10.50(0.58)	18.38(2.50)	20.00
Miray	5	7	6.60(1.14)	13.14(3.93)	19.00
Mikail	5	7	6.80(2.17)	17.86(2.79)	20.00
Cem	5	7	5.80(0.45)	17.14(4.10)	20.00
Ferhat	6	6	11.60(0.89)	16.33(3.39)	20.00
Eyluel	6	6	3.83(1.33)	13.83(4.58)	20.00
Defne	6	6	9.17(1.17)	16.20(4.09)	20.00

Note. Measurements (N); Mean (M); Standard Deviation (SD); Maximum Value (Max); A Phase (A); B Phase (B)

Almost all children except Ferhat (83.33) achieve a value of 100.00 for PEM. Statistically significant values (p<.05 to <.01) are found throughout the NAP. Expect Ferhat (87.00, p<.05) and Defne (98.00, p<.01) all students showed a maximum value of 100.00. The weighted Tau U results reveal a value of 0.47 (p<.05) for Defne and 0.50 (p<.05) for Ferhat, which can be considered a moderate effect. Elif and Yusuf have a value of 0.73 (p<.01), Miray (p<.01), Eyluel (p < .001) and Cem (p<.001) 0.76, a large change. Tugce achieves a value of 0.83 (p<.001) and Mikail of 0.86 (p<.001), which can be considered a very large change.

Table 4

Overlap Indices for Number of Correct Recognized Receptive Words

Participants	NAP	p	PEM	TauU	p
Tugce	100.00	<.01	100.00	0.83	<.001
Elif	100.00	<.01	100.00	0.73	<.01
Yusuf	100.00	<.01	100.00	0.73	<.01
Miray	100.00	<.01	100.00	0.76	<.01
Mikail	100.00	<.01	100.00	0.86	<.001
Cem	100.00	<.01	100.00	0.76	<.001
Ferhat	87.00	<.05	83.33	0.50	<.05
Eyluel	100.00	<.01	100.00	0.76	<.001
Defne	98.00	<.01	100.00	0.47	<.05

Note. Nonoverlap of All Pairs (NAP); Percentage Exceeding the Median (PEM)

The regression analysis reveals no trend in the A phase and a statistically significant level effect (p<.05) for group 1. Group 2 also shows no A phase trend. A statistically significant level effect (p<.01) and a significant slope effect (p<.001) with an increase of 1.681 scale points per intervention can be observed. No A-phase trend can be found for group 3. However, a statistically significant slope effect (p<.001) with an increasing value of 1.803 scale points on average per session can be identified. A statistically significant level effect (p<.001) and slope-effect (p<.001) is shown across all groups. Per intervention session, students increase by an average of 1.496 scale points. Furthermore, no A Phase trend can be identified.

Table 5

Regression Model Number of Correctly Known Receptive Vocabulary per Group and across Groups (Level 2 Analysis)

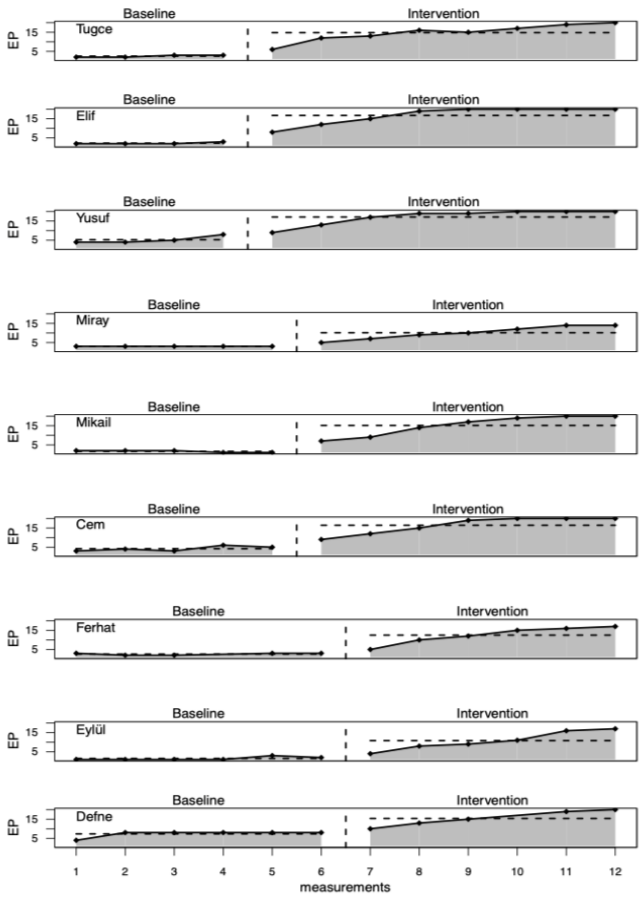
	B	*SE*	*t*	*p*
		Group 1		
Intercept	5.500	1.771	3.105	<.01
Trend	0.433	0.572	0.757	.46
Level	3.850	1.463	2.631	<.05
Slope	0.928	0.605	1.532	.14
		Group 2		
Intercept	6.800	1.324	5.134	<.001
Trend	-0.133	0.347	-0.384	.70
Level	3.724	1.258	2.960	<.01
Slope	1.681	0.404	4.159	<.001
		Group 3		
Intercept	7.315	1.557	4.697	<.001
Trend	0.230	0.220	1.047	.30
Level	-0.286	0.091	-0.262	.80
Slope	1.803	0.311	5.794	<.001
		Overall		
Intercept	6.829	0.835	8.177	<.001
Trend	0.086	0.189	0.452	.65
Level	2.716	0.731	3.716	<.001
Slope	1.496	0.220	6.794	<.001

Expressive Vocabulary

All baselines appear relatively flat. Only Yusuf possibly shows a positive trend, and Defne's values also increase in the baseline. All children show a very transparent increase in the B-phase, while some students even manage to reach the highest value immensely quickly. The intervention also seems to have a direct effect on all students, as an immediate increase is observed at the first intervention time.

Figure 5

Expressive vocabulary (EP)



The descriptive data support the previous estimates, since a distinct increase can be observed in the B-phase. Six students reach the maximum value of 20.00, Ferhat and Eyluel a value of 17.00 and Miray hit a value of 14.00.

Table 6

Descriptive Data for Each Participant in A and B Phase Regarding Expressive Vocabulary

	N (A)	*N (B)*	*M (A) (SD)*	*M (B) (SD)*	*Max (B)*
Tugce	4	8	2.50(0.58)	14.75(4.46)	20.00
Elif	4	8	2.25(0.50)	16.75(4.62)	20.00
Yusuf	4	8	5.25(1.89)	17.12(4.05)	20.00
Miray	5	7	3.00(0.00)	10.14(3.44)	14.00
Mikail	5	7	1.60(0.55)	15.14(5.34)	20.00
Cem	5	7	4.20(1.30)	16.43(4.50)	20.00
Ferhat	6	6	2.60(0.55)	12.50(4.51)	17.00
Eyluel	6	6	1.50(0.84)	10.83(4.96)	17.00
Defne	6	6	7.33(1.63)	15.40(4.16)	20.00

Note. Measurements (N); Mean (M); Standard Deviation (SD); Maximum Value (Max); A Phase (A); B Phase (B)

The maximum value of 100.00 is found for the NAP ($p<.01$), and PEM across all students. Regarding Weighted Tau U, Defne shows a small effect of 0.53 ($p<.05$). Ferhat (0.65,

p<.01), Eyluel (0.67, *p*<.01), Cem (0.73, *p*<.01), Elif (0.73, *p*<.01) and Yusuf (0.77, *p*<.001) achieve a large change. For Tugce (0.82, *p*<.001), Miray (0.83, *p*<.001) and Mikail, 0.92, *p*<.001 a very strong effect is manifested.

Table 7

Overlap Indices for Number of Correctly Known Expressive Vocabulary

Participant	NAP	p	PEM	TauU	p
Tugce	100.00	<.01	100.00	0.82	<.001
Elif	100.00	<.01	100.00	0.73	<.01
Yusuf	100.00	<.01	100.00	0.77	<.001
Miray	100.00	<.01	100.00	0.83	<.001
Mikail	100.00	<.01	100.00	0.92	<.001
Cem	100.00	<.01	100.00	0.73	<.01
Ferhat	100.00	<.01	100.00	0.65	<.01
Eyluel	100.00	<.01	100.00	0.67	<.01
Defne	100.00	<.01	100.00	0.53	<.05

Note. Nonoverlap of All Pairs (NAP); PEM: Percentage Exceeding the Median (PEM)

A statistically significant level effect (*p*<.001) and no A-phase trend was observed in group 1. Group 2 has a significant level effect (*p*<.05) as well as a slope effect (*p*<.001) with no A-phase trend. The participants improve by an average of 1.852 scale points per intervention. Group 3 shows a significant A-phase trend (*p*<.05) and a significant increase from A-phase to B-phase (*p*<.001) with 1.971. As for the results across all groups, the results reveal similar values (Level: *p*<.001, Slope: *p*<.001) with an increase of 1.688 with no overall A Phase trend.

Table 8

Regression Analysis for Number of Correctly Known Expressive Vocabulary per Group and across Groups (Level 2 Analysis)

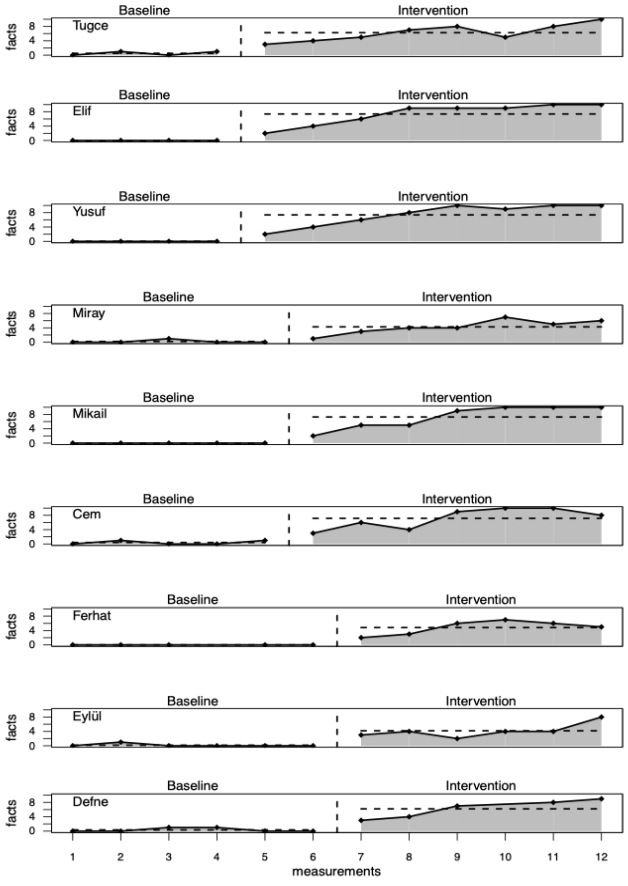
	B	*SE*	*t*	*p*
		Group 1		
Intercept	1.667	1.386	1.203	.24
Trend	0.667	0.456	1.435	.16
Level	4.643	1.188	3.908	<.001
Slope	0.940	0.492	1.913	.07
		Group 2		
Intercept	2.633	1.565	1.683	.10
Trend	0.100	0.366	0.274	.79
Level	2.962	1.325	2.235	<.05
Slope	1.852	0.426	4.350	<.001
		Group 3		
Intercept	2.671	1.498	1.783	.09
Trend	0.335	0.151	2.221	<.05
Level	0.255	0.749	0.341	.74
Slope	1.971	0.213	9.236	<.001
		Overall		
Intercept	2.819	0.825	3.418	<.01
Trend	0.205	0.171	1.199	.23
Level	3.005	0.657	4.575	<.001
Slope	1.688	0.199	8.493	<.001

Vocabulary Facts

For vocabulary facts, completely flat baselines can be found in all children and there is a clear increase in the B phase. Despite the rather short intervention times, students still manage to achieve the highest possible score in most cases.

Figure 6

Vocabulary Facts



The descriptive data shows a rapid increase in the B phase while 5 students reach the maximum value of 10.00, Defne hit 9.00, Eyluel 8.00 and Ferhat and Miray achieve a maximum value of 7.00.

Table 9

Descriptive Data for Each Participant in A and B Phase Correctly Known Vocabulary Facts

	N (A)	*N (B)*	*M (A) (SD)*	*M (B) (SD)*	*Max (B)*
Tugce	4	8	0.50(0.58)	6.25(2.38)	10.00
Elif	4	8	0.00(0.00)	7.38(3.02)	10.00
Yusuf	4	8	0.00(0.00)	7.38(3.07)	10.00
Miray	5	7	0.20(0.45)	4.29(1.98)	7.00
Mikail	5	7	0.00(0.00)	7.29(3.25)	10.00
Cem	5	7	0.40(0.55)	7.14(2.85)	10.00
Ferhat	6	6	0.00(0.00)	4.83(1.94)	7.00
Eyluel	6	6	0.17(0.41)	4.17(2.04)	8.00
Defne	6	6	0.33(0.52)	6.20(2.59)	9.00

Note. Measurements (N); Mean (M); Standard Deviation (SD); Maximum Value (Max); A Phase (A); B Phase (B)

Overall, overlap Indices up to 100.00 can be observed. Only Miray displays a NAP of 99.00 (p<.01) and Ferhat a moderate effect size of 0.55 (p<.05). Defne (0.61, p<.01), Cem (0.68, p<.01), Eyluel (0.71, p<.01), Miray (0.76, p<.001), Mikail and Tugce (0.79, p<.001) all achieve a large effect. Furthermore, Elif (0.84, p<.001) and Yusuf (0.90, p<.001) can claim a very strong effect.

Table 10

Overlap Indices for Number of Correctly Known Vocabulary Facts

Participant	NAP	p	PEM	TauU	p
Tugce	100.00	<.01	100.00	0.79	<.001
Elif	100.00	<.01	100.00	0.84	<.001
Yusuf	100.00	<.01	100.00	0.90	<.001
Miray	99.00	<.01	100.00	0.76	<.001
Mikail	100.00	<.01	100.00	0.79	<.001
Cem	100.00	<.01	100.00	0.68	<.01
Ferhat	100.00	<.01	100.00	0.55	<.05
Eyluel	100.00	<.01	100.00	0.71	<.01
Defne	100.00	<.01	100.00	0.61	<.01

Note. Nonoverlap of All Pairs (NAP); Percentage Exceeding the Median (PEM)

Group 1 shows both a significant level effect (p<.05) and slope effect (p<.01) with an increasing value of 0.973 scale points per session. Group 2 also shows a significant slope effect (p<.01) with an enhancement of 1.038. For group 3, a significant level (p<.05) and slope-effect (p<.001) can be stated with an average increase of 0.913. No overall A phase trend can be found in the groups. Across all groups, the analysis shows similar values with a significant level effect (p<.001) and slope effect (p<.001) and an average increase of 1.108. No trend in the A phase can be observed here either.

Table 11

Regression Model for Vocabulary Facts (Level 2 analysis)

	B	*SE*	*t*	*p*
		Group 1		
Intercept	0.000	0.804	0.000	1.00
Trend	0.067	0.294	-0.227	.82
Level	2.055	0.751	2.737	<.05
Slope	0.973	0.311	3.132	<.01
		Group 2		
Intercept	0.100	0.958	0.104	.92
Trend	0.033	0.258	0.129	.90
Level	1.686	0.934	1.806	.08
Slope	1.038	0.300	3.460	<.01
		Group 3		
Intercept	0.268	0.626	0.428	.67
Trend	-0.028	0.151	-0.187	.85
Level	1.854	0l.748	2.479	<.05
Slope	0.913	0.213	4.279	<.001
		Overall		
Intercept	0.406	0.413	0.981	.33
Trend	-0.070	0.112	-0.625	.53
Level	1.918	0.445	4.312	<.001
Slope	1.108	0.130	8.522	<.001

Discussion

In Germany, a considerable number of students belong to the first-generation whose families have a migration background. Previous studies indicate that these students are clearly disadvantaged in terms of academic success (e.g., Harju-Luukkainen et al., 2020; Kristen & Granato, 2007). Due to the limited language experience regarding the second language, they often have problems acquiring literacy skills (De Houwer, 2011; Hoff, 2009; Hoff et al., 2012). A basic requirement for this is an adequate vocabulary (Amiryousefi & Ketabi, 2011; Clark & Paivio, 1991). Accordingly, the aim of the present study was to promote receptive and expressive vocabulary and, indirectly, the associated vocabulary facts of each word, of primary school students with GL2. The multi-component intervention used includes the method of storytelling with integrated flashcards and was supplemented by aspects of increasing motivation.

The results of the study suggest that this method is an effective and inexpensive way of fostering receptive and expressive vocabulary among students. In addition to the results of the effect sizes, this is also evident from the fact that almost all students were able to achieve maximum receptive vocabulary scores after only six to eight intervention units (depending on the intervention group). With regard to the expressive vocabulary, this was obvious for two thirds of the students. The results thus confirm the findings of the study by Barwasser et al. (2020) and Knaak et al., (2021) on the effectiveness of a multi-component storytelling intervention for students with English as L2. Also, the findings go in line with the DCT and the CTMM (Paivio, 1991; Mayer, 2005; Amiryousefi & Ketabi, 2011; Ramezanali et al., 2020;

Uchihara et al., 2019) stating that a combination of verbal and non-verbal as well as visually and auditory aspects is beneficial in language learning. Additionally, this study can support previous findings that a combination of intentional and incidental vocabulary learning lead to high effect (Marulis & Neumann, 2010). In addition, it should also be emphasized that the present study was also able to show that the students were not only capable of building up the specifically trained receptive and expressive vocabulary in a very short time, but also that, without specific training, all students showed a significant increase in performance with regard to vocabulary facts of these trained words which were just embedded into the stories. Here, more than half of the students were able to achieve the maximum score when answering the questions on vocabulary facts. These findings go in line with incidental learning supporters (e.g., Webb & Nation, 2017; Peker et al., 2018).

Considering the characteristics of the individual students, no difference can be found between the participants across all three dependent variables. Thus, students from Arabic, Turkish and Albanian L1 backgrounds equally profited from the intervention. Also, even though some students showed problem behavior according to the ITRF screening, behavioral challenges did not have an impact on the results. Also, general L2 vocabulary knowledge did not seem to play a leading role. Plus, no gender or age differences can be found with respect to the results. Therefore, the results also provide a first indication that the method can be applied well for a heterogeneous student population. These results can be cautiously interpreted as an indication that the stories used in the intervention, in the sense of developing semantic networks, are capable of conveying further word knowledge to the students beyond the targeted development of receptive and expressive vocabulary.

Limitations and further research

Despite its encouraging results in developing vocabulary and vocabulary facts for students with GL2, this study is subject to the same weakness as all individual case designs. This concerns the lack of generalizability due to the small sample size, which reduces the external validity.

Here it should be noted in particular that the statements made can strictly speaking only refer to students with Turkish, Arabic and Albanian language background. Accordingly, it would be of interest in future studies to investigate the effectiveness of the method also with students with other native languages.

Furthermore, no conclusions can be drawn about the long-term nature of the effects. Following studies should investigate whether both, the receptive and expressive vocabulary, as well as the vocabulary facts can be retrieved in the long run. A further limitation is that the intervention consisted of several components (storytelling, flashcards, motivational components). It is therefore not possible to identify the specific impact of each component of

the intervention. It remains to be investigated in future research to what extent each of the components contributes to the overall effectiveness. Especially with regard to storytelling, it seems interesting to consider to what extent the method is able to teach students more complex knowledge content in terms of vocabulary facts. According to the DCT and the CTMM (Paivio, 1991; Mayer, 2005), the intervention contains a combination of verbal and non-verbal components. In this respect, too, no differentiated statement can be made about the effectiveness of the individual elements. In order to be able to develop effective interventions in the future, a differentiated analysis would be of interest here.

With regard to teaching a heterogeneous student population in inclusive education according to the broad concept of inclusion defined by Booth and Ainscow (2011), which includes not only students with different abilities but also those with different ethnic and national backgrounds, the results of our study, due to its implementation in small groups, can only give a limited indication of whether the support is suitable for use in inclusive education. Because of the importance of identifying evidence-based and economic methods that enable inclusive education for a diverse student body (Mitchell & Sutherland, 2020), further research should test the use of the method by teachers in inclusive classrooms. Moreover, since it has been shown that L1 vocabulary influences L2 vocabulary acquisition (e.g. Melby-Lervag et al., 2011; Maier et al., 2016) in future research it would be interesting to measure L1 vocabulary knowledge as well prior to the beginning of the intervention to have indications if students who also have insufficient L1 word knowledge might struggle more with L2 vocabulary interventions. Further, since all children could recall vocabulary facts which were not explicitly taught, next studies could compare storytelling with and without flashcards to see the amount of vocabulary growth also without explicit instruction. Also, since research found a correlation between reading ability and vocabulary growth (e.g. Suggate et al., 2013), it would be interesting to also measure L2 reading competency beforehand.

Practical Implications

The presented research provides valuable suggestions for supporting students with GL2. In addition to the effectiveness of the method, it is also shown that this method leads to a significant increase in competence of the students within a short time (six to eight units). This is particularly important since individual support for students often fails due to a lack of resources and adequate teaching methods (McLeskey & Waldron, 2011; Schmidt et al., 2002). The broad applicability of the method can be seen in the fact that the storytelling and flashcard methods used can be adapted to different teaching topics and age groups. While in the present study a training of animal names and characteristics was carried out, other topics are also conceivable, e.g. in the field of social sciences. The high flexibility of the method in terms of content is central, as it also allows the interest of the students to be taken into account. This is

an important indication for school practice, since motivational aspects play a decisive role in foreign language acquisition (Saito et al., 2017; Doernyei & Skehan, 2003; Klein, 1996).

Conclusion

Since vocabulary learning is of enormous significance for L2 language proficiency (Graham & Zohreh, 2020; Milton et al., 2010; Nation, 2001; Stæhr, 2008, 2009; Suggate et al., 2013; van Gelderen et al., 2004), it is important to implement interventions which positively influences even several aspects of vocabulary learning (receptive, expressive and additional vocabulary fact knowledge) simultaneously in an effective way. In summary, our results show that the multicomponent intervention used is an effective way to help students with GL2 to develop their vocabulary. In particular, the results show that the method was not only able to build up the vocabulary, but also to enhance vocabulary facts. Thus, the study provides a further contribution to the provision of evidence-based interventions for the fostering of students with GL2.

References

Abbasian, G.-R., & Ghorbanpour, E. (2016). The effect of flash card-based instruction on vocabulary learning by EFL learners. *International Journal of Humanities and Cultural Studies*, 1929–1938. https://www.ijhcs.com/index.php/ijhcs/article/view/2571/2398

Al-Mansour, S., & Al-Shorman, R. A. (2011). The effect of teachers' storytelling aloud on the reading comprehension of Saudi elementary stage students. *Journal of King Saud University. Language and Translation*, *23*(2), 69–76. https://doi.org/10.1016/j.jksult.2011.04.001

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

Amiryousefi, M., & Ketabi, S. (2011). Mnemonic instruction: A way to boost vocabulary learning and recall. *Journal of Language Teaching and Research*, *2*(1), 178–182. https://doi.org/10.4304/jltr.2.1.178-182

Barwasser, A., Knaak, T., & Gruenke, M. (2020). The effects of a multicomponent storytelling intervention on the vocabulary recognition of struggling English as a foreign language learners with learning disabilities. *Insight into Learning Disabilities*, *17*(1), 35–53.

Becker-Mrotzek, M., Hentschel, B., Hippmann, K., & Linnemann, M. (2012, June). *Sprachfoerderung in deutschen Schulen – die Sicht der Lehrerinnen und Lehrer.*

Ergebnisse einer Umfrage unter Lehrerinnen und Lehrern. [Language support in German schools - the teachers' view. Results of a survey among teachers.] https://www.mercator-institut-sprachfoerderung.de/fileadmin/user_upload/Lehrerumfrage_Langfassung_final_30_05_03.pdf

Booth, T., & Ainscow, M. (2011) *Index for inclusion developing learning and participation in schools*. Centre for Studies on Inclusive Education.

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Semantic structure in vocabulary knowledge interacts with lexical and sentence processing in infancy. *Child development*, *87*(6), 1893–1908. https://doi.org/10.1111/cdev.12554

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., Parker, R., & Anderson, C. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review*, *42*(1), 39–55. https://doi.org/10.1080/02796015.2013.12087490

Brown, R., Waring, R. & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. Reading in a Foreign Language, 20(2), 136–163.

Cervetti, G. N., Wright, T. S., & Hwang, H. (2016). Conceptual coherence, comprehension, and vocabulary acquisition: A knowledge effect? *Reading and Writing*, *29*, 761–779. https://doi.org/10.1007/s11145-016-9628-x

Clark, M. J., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. https://doi.org/10.1007/BF01320076

Cochran-Smith, M. (1984). *The Making of a Reader. Language and Learning for Human Service Professions Monograph Series*. Ablex Publishing Corporation.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Pearson Prentice.

Crowley, K., McLaughlin, T. F., & Kahn, R. (2013). Using direct instruction flashcards and reading racetracks to improve sight word recognition of two elementary students with autism. *Journal of Developmental and Physical Disabilities*, *25*, 297–311. https://doi.org/10.1007/s10882-012-9307-z

De Houwer, A. (2011). Language input environments and language development in bilingual acquisition. *Applied Linguistics Review*, *2*, 221–240.

De Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhoefer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, *68*(4), 906–941. https://doi.org/10.1111/lang.12296

Dong, Y. R. (2013). Powerful learning tools for ELLS: Using native language, familiar examples, and concept mapping to teach English language learners. *The Science Teacher*, *80*(4), 51–57.

Doernyei, Z. (2006). Individual differences in second language acquisition. *AILA Review*, *19*(1), 42–68. https://doi.org/10.1075/aila.19.05dor

Doernyei, Z., & Skehan, P. (2003) Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Oxford University Press. https://doi.org/10.1002/9780470756492.ch18

Dunn, L., & Dunn, D. (2007). *Peabody picture vocabulary test* (4th ed.). *PPVT-4.* Pearson. https://doi.org/10.1037/t15144-000

Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology: General*, *120*(2), 190–202. https://doi.org/10.1037/0096-3445.120.2.190

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children. A Meta-analysis. *Journal of Research on Educational Effectiveness*, *2*(1), 1–44. https://doi.org/10.1080/19345740802539200

Ellis, G., & Brewster, J. (2002). *Tell it again. The new storytelling handbook for primary teachers*. Penguin English.

Ellis, R. (1997). *Second language acquisition*. Oxford University Press.

Falkenberg, C. A. & Barbetta, P. M. (2013) 'The effects of a self-monitoring package on homework completion and accuracy of students with disabilities in an inclusive education classroom.' Journal of Behavioral Education, 22, pp. 190–210. https://doi.org/10.1007/s10864-013-9169-1

Federal Statistical Office (2020, December 07). *Population. Migration und integration. [Population. Migration and integration].*

https://www.destatis.de/EN/Themes/Society-Environment/Population/Migration-Integration/_node.html

Ford-Connors, E., & Paratore, J. R. (2015). Vocabulary instruction in fifth grade and beyond: Sources of word learning and productive contexts for development. *Review of Educational Research*, *85*(1), 50–91. https://doi.org/10.3102/0034654314540943

Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (2011). Predicting robust vocabulary growth from measures of incremental learning. *Scientific Studies of Reading*, *15*(1), 71–91. https://doi.org/10.1080/10888438.2011.539076

García, J. N., & de Caso, A. M. (2004). Effects of a motivational intervention towards writing of children with learning disabilities. *Learning Disabilities Quarterly*, *27*(3), 141–159. https://doi.org/10.2307/1593665

Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology*, *98*(4), 732–749. https://doi.org/10.1037/0022-0663.98.4.732

Gogolin, I., Dirim, İ., Klinger, T., Lange, I., Lengyel, D., Michel, U., Neumann, U., Reich, H. H., Roth, H.-J., & Schwippert, K. (2011). *Foerderung von Kindern und Jugendlichen mit Migrationshintergrund. Bilanz und Perspektiven eines Modellprogramms* (FoerMig ed.*). [Promotion of children and young people with a migration background. Balance and perspectives of a model programme.]* Waxmann.

Graham, K.M., & Eslami, Z. R. (2020). Does the simple view of writing explain L2 writing development? A meta-analysis. *Reading Psychology, 41*(5), 485–511.

Graves, M. F. (2016). *The vocabulary book: Learning and instruction* (2nd ed.). Teachers College Press.

Grey, S., Williams, J. N., & Rebuschat, P. (2015). Individual differences in incidental language learning: The role of learning styles, personality, and working memory. *Learning and Individual Differences*, *38*, 44–53. https://doi.org/10.1016/j.lindif.2015.01.019

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

Halle, T., Hair, E., Wandner, L., McNamara, M., & Chien, N. N. (2012). Predictors and outcomes of early versus later English language proficiency among English language learners. *Early Childhood Research Quarterly*, *27*(1), 1–20. https://doi.org/10.1016/j.ecresq.2011.07.004

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

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. https://doi.org/10.1007/978-3-030-38969-7

Harmon, J. M., Wood, K. D., & Kiser, K. (2009). Promoting vocabulary learning with the interactive word wall. *Middle School Journal*, *40*(3), 58–63. https://doi.org/10.1080/00940771.2009.11495588

Hemmati, F., Gholamrezapour, Z., & Hessamy, G. (2015). The effects of teachers' storytelling and reading story aloud on the listening comprehension of Iranian EFL learners. *Theory and Practice in Language Studies*, *5*(7), 1482–1488. https://doi.org/10.17507/tpls.0507.22

Hirsch, S. E., Ennis, R. P., & McDaniel, S. C. (2013). Student self-graphing as a strategy to increase teacher effectiveness and student motivation. *Beyond Behavior*, *22*(3), 31–39. https://doi.org/10.1177/107429561302200305

Hoff, E. (2009). *Language development*. Wadsworth.

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1–27. https://doi.org/10.1017/S0305000910000759

Horwitz, E. K., Horwitz, M. B., & Cope, J. A. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Huang, H. (2006). The effects of storytelling on EFL young learners' reading comprehension and word recall. *English Teaching & Learning*, *30*(3), 51–74.

International Organization for Migration (2019). *IOM's Global Migration Data Analysis Centre*. https://gmdac.iom.int/

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*(2), 137–166. https://doi.org/10.1017/S0261444810000509

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: IES practice*

guide (NCEE 2008-4027). National Center for Education Evaluation and Regional Assistance.

Kim, M. (2010). The effects of storytelling on adult English language learners. *Linguistic Research*, *27*(3), 447–473. https://doi.org/10.17250/khisli.27.3.201012.004

Kirsch, C. (2016). Using storytelling to teach vocabulary in language lessons: does it work? *The Language Learning Journal*, *44*(1), 33–51. https://doi.org/10.1080/09571736.2012.733404

Kitzinger, A. (2020, December 11). *METACOM.* https://www.metacom-symbole.de

Klein, W. (1996). Language Acquisition at different ages. In D. Magnusson (Ed.), *Individual Development over the Lifespan: Biological and Psychosocial Perspectives* (pp. 88–108). Cambridge University Press.

Knaak, T., Gruenke, M., & Barwasser, A. (2021)

Kristen, C., & Granato, N. (2007). *The education attainment of the second generation in Germany: Social origins and ethnic inequality* (IAB Discussion Paper No. 4/2007). http://doku.iab.de/discussionpapers/2007/dp0407.pdf

Landauer, T. K., Kireyev, K., & Panaccione, C. (2011). Word maturity: A new metric for word knowledge. *Scientific Studies of* Reading, *15*(1), 92–108. https://doi.org/10.1080/10888438.2011.536130

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 126–132). Macmillan. https://doi.org/10.1007/978-1-349-12396-4_12

Laufer, B., & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Ledford, J. R., & Gast, D. L. (Eds.). (2014). Single case research methodology: Applications in special education and behavioral sciences. Routledge. https://doi.org/10.4324/9780203521892

Leons, E., Herbert, C., & Gobbo, K. (2009). Students with learning disabilities and AD/HD in the foreign language classroom: Supporting students and instructors. *Foreign Language Annals*, *42*(1), 42–54. https://doi.org/10.1111/j.1944-9720.2009.01007.x

Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a

randomized field trial. *American Educational Research* Journal, *51*(6), 1159–1194. https://doi.org/10.3102/0002831214532165

Lin, P. M. S. (2014). Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. System, 42, 164e176. http://dx.doi.org/ 10.1016/j.system.2013.11.010.

Little, S. G., Akin-Little, A., & O'Neill, K. (2015). Group contingency interventions with children – 1980-2010: A meta-analysis. *Behavior Modification*, *39*(2), 332–341. https://doi.org/10.1177/0145445514554393

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

Maier, M. F., Bohlmann, N. L., & Palacios, N. A. (2016). Cross-language associations in the development of preschoolers' receptive and expressive vocabulary. Early Childhood Research Quarterly, 36, 49–63. https://d oi.org/10.1016/j.ecresq.2015.11.006

Mayer, R.E. Cognitive Theory of Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning*;

Mayer, R., Ed.; Cambridge University Press: Cambridge, UK, 2005; pp. 31–48.

McDougall, D., Morrison, C. & Awana, B. (2012) 'Students with disabilities use tactile cued self-monitoring to improve academic productivity during independent tasks.' Journal of Instructional Psychology, 30, pp. 119–130.

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*, 48–57. https://doi.org/10.1111/j.1540-5826.2010.00324.x

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, *80*(3), 300–335. https://doi.org/10.3102/0034654310377087

Melby-Lervag, M., & Lervag, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. Journal of Research in Reading, 34, 114–135. https://doi.org/10.1111/j.14679817.2010.01477.x

Meneses, A., Uccelli, P., Santelices, M. V., Ruiz, M., Acevedo, D., & Figueroa, J. (2018). Academic language as a predictor of reading comprehension in monolingual

Spanish speaking readers: Evidence from Chilean early adolescents. *Reading Research Quarterly*, *53*, 223–247. https://doi.org/10.1002/rrq.192

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. ChacónBeltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), Further insights into non-native vocabulary teaching and learning (pp. 83-98). Bristol, UK: Multilingual Matters.

Ministry for School and Further Education NRW (2016). *Integration durch Bildung. Maßnahmen und Initiativen des Ministeriums fuer Schule und Weiterbildung des Landes Nordrhein-Westfalen fuer zugewanderte Kinder und Jugendliche. [Integration through Education. Measures and initiatives of the Ministry for Schools and Further Education of the State of North Rhine-Westphalia for immigrant children and young people.]* https://www.schulentwicklung.nrw.de/q/upload/Schule_und_Zuwanderung/Pressemitteilung_MSW_24.05.2016.pdf

Ministry for School and Further Education NRW (2018). *Integration und Deutschfoerderung neu zugewanderter Schuelerinnen und Schueler. [Integration and German language support for newly immigrated pupils.]* https://www.schulministerium.nrw.de/themen/schulsystem/integration-durch-bildung/lernstandsberichte-fuer-neu-zugewanderte-schuelerinnen

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

Mizumoto, A., & Takeuchi, O. (2009). Examining the effectiveness of explicit instruction of vocabulary learning strategies with Japanese EFL university students. Language Teaching Research, 13(4), 425-449.

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*(2), 267–296. https://doi.org/10.1037/a0021890

Nagy, W. E., & Scott, J. A. (2000). Vocabulary processing. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (vol. 3, pp. 269–284). Erlbaum.

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Heinle & Heinle.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, *27*, 136–145.

National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel: Teaching children to read - An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* U.S. Goverment Printing Office.

Oxford, R. L., & Crookall, D. (1990). Vocabulary learning: A critical analysis of techniques. *TESL Canada Journal*, *7*(2), 9–30. https://doi.org/10.18806/tesl.v7i2.566

Oxford, R. L., & Scarcella, R. C. (1994). Second language vocabulary learning among adults: State of the art in vocabulary instruction. *System*, *22*, 231–43. https://doi.org/10.1016/0346-251X(94)90059-0

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

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

Peker, H., Regalla, M., & Cox, D. T. (2018). Teaching and learning vocabulary in context. Examining engagement in three prekindergarten French classrooms. *Foreign Language Annals*, *51*(2), 472–483. https://doi.org/10.1111/flan.12338

Pintrich, P. R. (2003). Motivation and classroom learning. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology. Educational psychology, Vol. 7* (pp. 103-122). Hoboken, NJ: Wiley.

Popkin, J., & Skinner, C. H. (2003). Enhancing academic performance in a classroom serving students with serious emotional disturbance: Interdependent group contingencies with randomly selected components. *School Psychology Review*, *32*(2), 282–295. https://doi.org/10.1080/02796015.2003.12086199

Rafferty, L. A. (2010) 'Step-by-step: Teaching students to self-monitor.' Teaching Exceptional Children, 43(2), pp. 50–58. https://doi.org/10.1177/004005991004300205

Ramezanali, N., & Faez, F. (2019). Vocabulary learning and retention through multimedia glossing. Language *Learning & Technology, 23*(2), 105–124. https://doi.org/10125/44685

Ramezanali, N., Uchihara, T., & Faez, F. (2020). Efficacy of multi glossing on second language vocabulary learning: A meta-analysis. *TESOL Quarterly.* https://doi.org/10.1002/tesq.579

Rich, T., Weber, K. P., McLaughlin, T. F., & Sells-Love, D. (2016). The effects of direct instruction flashcards to increase sight-word reading ability and cover, copy, and compare for spelling for a fifth-grade student with specific learning disabilities. *International Journal of Advanced Scientific Research*, *1*, 1–6.

Roney, R. C. (1996). Storytelling in the classroom: Some theoretical thoughts. *Storytelling World*, *9*, 7–9.

Saito, K., Dewaele, J. M., & Hanzawa, K. (2017) A longitudinal investigation of the relationship between motivation and late second language speech learning in classroom settings. *Language and Speech*, *60*(4), 614–632. https://doi.org/10.1177/0023830916687793

Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review*, *38*(4), 445–459.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Reading interventions for adolescent struggling readers: A meta-analysis with implications for practice*. RMC Research Corporation, Center on Instruction.

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. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Schmitt, N. (2010). *Vocabulary in language teaching.* Cambridge University Press.

Schmitt, N. (2014). Conceptual review article. Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913–951.

Siyanova-Chanturia, A., & Webb, S. (2016). Teaching vocabulary in the EFL context. In W. A. Renandya & H. P. Widodo (Eds.), *English language teaching today: Building a*

closer link between theory and practice (pp. 227–239). Springer International. https://doi.org/10.1007/978-3-319-38834-2_16

Snow, C. E. (1993). Families as Social Contexts for Literacy Development. *New Directions for Child Development, 61*, 11–24.

Sparks, R. (1995). Examining the linguistic coding differences hypothesis to explain individual differences in foreign language learning. *Annals of Dyslexia*, *45*, 187–214. https://doi.org/10.1007/BF02648218

Stæhr, L. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. Studies in Second Language Acquisition, 31, 577-607.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading, and writing. Language Learning Journal, 36(2), 139-152.

Stahr, L.S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, *31*, 577–607.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, *56*(1), 72–110. https://doi.org/10.3102/00346543056001072

Suggate, S.P., Lenhard, W., Neudecker, E., & Schneider, W. (2013). Incidental vocabulary acquisition from stories: Second and fourth graders learn more from listening than reading. First Language, 33(6), 551-571. https://doi.org/10.1177/0142723713503144

Teng, F. (2016). The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading. The Language Learning Journal. Electronic publication. doi: 10.1080/09571736.2016.1244217).

Teng, F. (2018). Incidental vocabulary acquisition from reading-only and multi-dimensional approach. Innovation in Language Learning and https://doi.org/10.1080/17501229.2016.1203328

Thompson, I. (1987). Memory in language learning. In A. Wenden & J. Rubin (Eds.), *Learner strategies in language learning* (pp. 15–30). Prentice-Hall.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. https://doi.org/10.1037/0033-295X.84.4.327

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis or correlation studies. *Language Learning*, *69*(3), 559–599. https://doi.org/10.1111/lang.12343

Vaatoranta, E., Suggate, S., Jachmann, C., Lenhart, J., & Lenhard, W. (2017). Can explaining less be more? Enhancing vocabulary through explicit versus elaborative storyrtelling. First Language, 1-20,

Van Gelderen, A:, Snellongs, P., & De Glopper, K. (2004). The effect of enhanced lexical retrieveal on second language writing. A classroom experiment. *Applied Psycholinguistics, 25*, 175-200.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. Language Learning, 61(1), 219e258. http://dx.doi.org/10.1111/j.1467-9922.2010.00593.x.

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 off German schoolchildren. *Journal of School Psychology*, *66*, 25–40. https://doi.org/10.1016/j.jsp.2017.11.003

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*, 232–245.

Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, *40*(3), 360–376. https://doi.org/10.1177/0033688209343854

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan, & J. C. Richards (Eds.), Language learning beyond the classroom (pp. 159e168). New York/London: Routledge.

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.

Wells, G. (2009). The Social Context of Language and Literacy Development. In: O. A. Barbarin & B. H. Wasik (eds.) Handbook od Child Development & Early Education (pp. 271–302). The Guilford Press.

Wells, J. C., Sheehey, P. H., & Sheehey, M. (2017). Using self-monitoring of performance with self-graphing to increase academic productivity in math. *Beyond Behavior*, *26*(2), 57–65. https://doi.org/10.1177/1074295617711207

Wilbert, J., & Lueke, T. (2019). Package 'scan.' https://cran.project.org/web/packages/scan/scan. pdf

Williams, K. T. (2007). *The expressive vocabulary test* (2nd ed.). AGS Publishing. https://doi.org/10.1037/t15094-000

Wilkins, D. A. (1972). Linguistics in Language Teaching. London: Edward Arnold.

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*, 58–76. https://doi.org/10.1177/0014402916658658

Woodrow, L. (2017) Motivation in Language Learning. In: Breeze R., Sancho Guinda C. (eds) Essential Competencies for English-medium University Teaching. Educational Linguistics, vol 27. Springer, Cham. https://link.springer.com/content/pdf/10.1007%2F978-3-319-40956-6_16.pdf

Appendix E Article 5

Barwasser, A., Lenz, B., & Grünke, M. (2021). A multimodal storytelling intervention for improving the reading and vocabulary skills of struggling German-as-a-Second Language adolescents with learning and behavioral problems. *Insights Into Learning Disabilities, 18*(1), 29–51.

Abstract

Students with learning and behavioral problems often face enormous barriers to successful school performance. When the language of instruction is not the students' first language and they are already in secondary school, this is compounded by additional challenges, especially with regard to vocabulary and reading – subjects that are fundamental to obtaining equal educational opportunities. In addition, difficulties in the language of instruction are often accompanied by declining student motivation. Direct word recognition and vocabulary acquisition are of great importance in language learning contexts. To address these areas, multimodal intervention approaches have proven useful. For this reason, a multimodal storytelling intervention was developed to help this group of students, in particular, to improve significantly in the aforementioned areas. Specifically, a multiple-baseline design across participants (N = 4) was applied to train the sight-word reading and vocabulary of adolescent students with learning and behavioral problems who spoke German as a second language. The training occurred three times a week over a period of five weeks. Results were promising for both variables in a short period of time, suggesting that the storytelling intervention in this form is an appropriate tool for teaching second-language German.

Keywords: storytelling, German as a second language, vocabulary, reading performance, learning disabilities, problem behavior, adolescents

Introduction

Second Language Acquisition and Immigration in Germany

Germany remains one of the most important destination countries in Europe for international migrants, with numbers steadily increasing from 2000 to 2019 (International Organization for Migration [IOM], 2019). Acquiring the national language is vital for to be able to fully participate in society. Unfortunately, a great proportion of children and adolescents encounter difficulties as they try to learn German to a native level, often associated with poor school performance and eventual dropout (Harju-Luukkainen et al., 2020; Isphording & Otten, 2014). Furthermore, young people with a migration background and insufficient German language skills experience unemployment and mental health issues more frequently than their

native peers, eventually resulting in lower income and higher levels of marginalization (Macdonald et al., 2016; Marksteiner et al., 2019).

Tragically, as the 2018 Programme for International Student Assessment results indicate, there is an increasingly large academic achievement gap between students with and without a migration background (Organisation for Economic Co-operation and Development [OECD], 2019) and, in general, multilingual students are often perceived as having major learning difficulties in school (Burr et al., 2015).

Challenges in Language Proficiency

Students learning a second language (L2) face the challenge of having to link their previously learned first language (L1) and the newly learned L2 (Isphording & Otten, 2014). Pae (2018) found that reading and writing skills in L1 and L2 are significantly correlated (see also Li et al., 2017; Lin, 2018). That is, if a student has difficulties in L1, it usually takes a great deal of effort to make progress in L2.

One of biggest obstacles in language acquisition is expanding vocabulary, a first step toward further progress in the new language (Adwani & Shrivastava, 2017). For example, a lack of word knowledge often results in an inability to build reading skills and, in turn, attain a productive level of writing (Hacking & Tschirner, 2017; Shin et al., 2019). Additionally, related to a lack of vocabulary is the capacity of working memory (Mackey & Sachs, 2012), which also plays a crucial role in reading development (Shin, 2020). Not surprisingly, all this combines to negatively affect students' motivation in school (Saito et al., 2017).

The challenges that almost everyone experiences when trying to acquire a foreign language are severely compounded for individuals who have learning problems (Haager & Osipova, 2017; Sarisahin, 2020), particularly deficits in working memory capacity and information-processing speed (Bishara & Kaplan, 2016). For them, trying to become familiar with a new language can be an overwhelming task (Sparks, 2012). The same goes for students with externalizing behavioral problems, due to their lack of sufficient self-control skills and motivation to learn (Maehler & Schuchardt, 2016).

The Importance of Vocabulary and Reading in a Second Language

Overall, linguistic ability is dependent on the development of vocabulary (Adwani & Shrivastava, 2017). Thus, while enormously challenging, expanding vocabulary plays a leading role in reading. In this regard, *word memory* refers to a mental lexicon that stores the information of acquired words at different levels, which can be divided into receptive and expressive vocabulary. *Receptive vocabulary* includes words that students can recognize but cannot actively use, whereas *expressive vocabulary* refers to words that students can actively use (Webb, 2005).

According to Coltheart et al. (2001), two different pathways can lead to word identification: a lexical and a non-lexical route. Through the *lexical route*, words are directly retrieved from the mental lexicon. In the *non-lexical route*, the letters of words are recoded into sounds via the application of grapheme–phoneme correspondence rules (Coltheart et al., 2001). The more words that are stored in the mental lexicon, the more working memory capacity is available to be devoted to reading comprehension (National Institute of Child Health and Human Development [NICHD], 2000).

With respect to the lexical route, Ehri (2005) named sight words, that is, words read within one second of their appearance, as a significant component in the reading process. Specifically, they activate the lexical pathway and thus lead to faster word recognition. Studies show that an increase in sight words in reading can lead to an increase in reading fluency (e.g., Scammacca et al., 2007) and, consequently, to improved reading comprehension (e.g., Perfetti & Stafura, 2014).

Adequate Second-Language Support

The dual-coding theory (Paivio, 1991) offers a potentially promising approach to teaching language skills. Students learn best when new information is processed in visual and verbal ways. Thus, explicit teaching of visual and verbal mnemonics has been found to have the greatest effect on the acquisition of new words (De Vos et al., 2018; Uchihara et al., 2019).

In this regard, new content can be connected to existing knowledge through explanations, pictures, and gestures to enhance memorization of words. For example, studies confirm that newly learned words are better retained when taught by linking visual and verbal modalities (e.g., Johnson & Mayer, 2009). This form of multimedia presentation has been found to be particularly effective for L2 acquisition (Bisson et al., 2013), as well as for improving the language development of students with LD (Brady et al., 2015). Furthermore, higher effects are achieved when the content taught in the L2 is embedded in meaning-bearing contexts (Peker et al., 2018).

Method of teaching is another important factor. For example, content can be taught in an implicit manner, without direct instruction, or explicitly using direct instruction (Ellis et al., 2009). However, promoting vocabulary is particularly effective when a combination of implicit and explicit instruction is used (Marulis & Neumann, 2010). Additionally, the effectiveness of reading on incidental learning of new words must be considered (Bisson et al., 2014). In a recent meta-analysis, Uchihara et al. (2019) found that the frequency of reading new terms is another important influencing variable in learning new words in a multimedia context. Furthermore, repetitive reading increases the reading fluency and, thus, the reading comprehension of students with and without learning problems (Lee & Yoon, 2017; Stevens et al., 2017).

Storytelling

The use of stories is a promising tool for teaching new content in the classroom (Landrum et al., 2019) as it combines many of the elements discussed above. Storytelling involves relating a narrative in an interactive way to engage the listeners in the process (Roney, 1996). To promote language development, learners should be provided with multimedia support at the three levels of reading, listening, and viewing (Feng & Webb, 2020; Mayer, 2005). Storytelling interventions seem to be useful for teaching general language skills and building students' listening and reading comprehension skills (Al-Mansour & Al-Shorman, 2011; Hemmati et al., 2015; Huang, 2006; Kim, 2010). Furthermore, this approach has proven to be effective for promoting vocabulary (Feng & Webb, 2020; Mello, 2001) and for building vocabulary skills in L2 English of students with severe learning difficulties (Barwasser et al., 2020; Knaak et al., 2021). Peters and Webb (2018) pointed to an increase in vocabulary among students through verbal delivery. New words are presented with the help of visualizations and the interactive reading aloud of a story.

New content in the L2 classroom should be explicitly taught (Hulstijn, 2005; Marulis & Neumann, 2010) because direct instruction methods are of great benefit particularly for at-risk students (Butler, 2020). In a meta-analysis, Stockhard et al. (2018) summarized strong effects of direct instruction on various participants. For example, using direct instruction, new words can be stored more effectively in students' mental lexicon (Uchihara et al., 2019). Various studies confirm these effects on students' sight vocabulary (e.g., Ehri, 2005; Fjortoft et al., 2014). Highly beneficial effects of direct instruction through the use of flashcards have also been demonstrated for students with learning problems (Fraher et al., 2019). For L2 learners, in particular, direct instruction provides another opportunity to store information from a story in the form of new words and to expand their vocabulary knowledge through context, which is necessary for improving their reading comprehension (Abbasian & Ghorbanpout, 2016; Liu, 2004).

Motivation and Second-Language Learning

Motivation plays a crucial role in learning in general (Stiensmeier-Pelster & Otterpohl, 2018). In language learning, for example, low motivation has been linked to decreasing performance (Grey et al., 2015). With regard to learning a new language, it is hypothesized that students' motivation arises in part from L1 learning and has further effects on L2 acquisition (Sparks, 2016). Therefore, it is important to integrate motivational components into instruction to support the language acquisition of students with risk factors (Anjomshoa & Sadighi, 2015).

Self-monitoring interventions provide a means of motivational support in this regard (Maag, 2019). One such method is self-graphing. In this strategy, students visualize their daily

progress in their own learning process (Hirsch et al., 2013). Various studies have shown the effectiveness of this method in terms of behavior and academic achievement, above all for students with emotional disorders (McDaniel et al., 2013; Sutherland & Snyder, 2007).

Research Questions

Against a background showing that students with German as an L2, in particular,

often have challenges performing well in school and that additional difficulties, such as learning and behavioral problems lead to increased difficulty, a storytelling intervention was developed that is theoretically and empirically oriented toward the L2 literature and addresses students' personal interests as a means of increasing motivation. The goal was to break the vicious cycle of underachievement and frustration and provide learners with a sense of accomplishment and a fair educational opportunity. Finally, given the limited amount of research on secondary-level students with low achievement in an L2 (Young-Scholten, 2015), the present study will help fill the research gap in this area.

Accordingly, the research questions were as follows:

1. Does a combined storytelling intervention lead to an increase in expressive vocabulary of struggling L2 German adolescents with learning and behavior problems?

2. Does a combined storytelling intervention lead to an increase in the number of sight words of struggling L2 German adolescents with learning and behavior problems?

Methods

Participants and Setting

The study was conducted at a school for students with special needs with a focus on learning disabilities (LD) in a large city in North Rhine-Westphalia, Germany. In Germany, an LD is defined as failure to develop the knowledge, skill, will, and self-regulation necessary to succeed in key subject areas. Unlike in the United States and other countries, in Germany, these problems can be accompanied by a moderately reduced IQ (70–85). As part of the definition of LD is also the requirement that a student's problems cannot be overcome without additional help (Gruenke & Cavendish, 2016). With regard to the participants in the current study, Grades 6–8 were targeted, specifically students with German as an L2. Accordingly, teachers made the first selection of potential subjects. Socioeconomic data on the participants were collected via a teacher questionnaire.

Before the data collection started, consent forms were sent to the parents of potential subjects. At the beginning of the study, 12 adolescents took part in a series of assessments: four students from Grade 6, four from Grade 7, and four from Grade 8. First, to determine the students' German vocabulary skills, the German vocabulary subtest as part of the Culture Fair

Intelligence Test (CFT; Weiß, 2006) was administered. The vocabulary test contains a total of 30 words, with each word having a total of four choices of words, one of which is the closest match to the one shown. The instrument is supposed to be a measure of an individual's range of vocabulary. The test procedure was based on a representative sample of all school types from Grade 3 to 13 (N = 2724), with a reliability of r = .87 (Weiß, 2006). All subjects who had a percentile rank below 15 were considered eligible for the study (N = 9).

To identify potential problem areas in reading, a German reading screening, the Salzburg Reading and Spelling Test (Moll & Landerl, 2010), was administered. The two subtests of word reading and pseudo-word reading were conducted within one minute. Norms were created based on data from 2,000 students from Grade 1 to 6. Reliability of the parallel test ranges from .90 to .98, and correlations with other reading tests range from .69 to .92.

In addition, the German translation of the Integrated System Teacher Rating Form (Volpe et al., 2018) was administered to assess externalizing behavior problems. Here, the shortened version consisting of 16 items was used due to time limitations. Behavior is rated in two dimensions (learning-related behavior and oppositional/disruptive behavior), with eight items each. In accordance with the information given by Volpe et al. (2018), students above a cutoff value of 13 are considered as having externalizing behavior problems.

To crystalize the selection of the final words for the intervention, which were to be words that the participants could neither read nor express actively, a word pretest was conducted. The word pool was taken from the Metacom symbols (Kitzinger, 2020), including words important in the students' everyday lives. A total of 140 words were selected, and the matching Metacom symbols were each put on a PowerPoint (PPT) slide. The students were asked if they knew the correct vocabulary term for the picture presented. The same 140 words were tested a second time on another day as written-out words on PPT slides in the context of reading. The PPT ran in a one-second cycle (see Ehri, 2005).

From these pretests, we obtained a pool of 40 words for expressive vocabulary and 30 words for reading, which were also included in the expressive pool. We chose 10 more words for the expressive vocabulary because the students performed better on the expressive pretest and we wanted to prevent ceiling effects.

Because the study was conducted during the COVID-19 pandemic, this paper focuses on four participants explicitly: Tana (male), Cem (male), Mattina (female), and Alen (male). The remaining students were dropped from the study due to missing data, mainly because of quarantine. Our sample included two participants in Grade 7 (Tana and Cem), one in Grade 6 (Mattina), and one in Grade 8 (Alen). All students spoke Turkish as their L1 and had started to learn German upon entry into kindergarten at the age of 3. Vocabulary knowledge, as well as reading ability, was severely low for all four participants.

Table 1

Characteristics of Participants

	Age	Gender	IQ	Special Needs	ITRF	Reading W	Reading PW	Vocabulary PR	L1
Tana	14;4	female	70-85	LD	18	<1	1-2	4	Turkish
Cem	14;7	male	70-85	LD	27	<1	1-2	5	Turkish
Mattina	13;6	female	70-85	LD	18	<1	1-2	2	Turkish
Alen	15;1	male	70-85	LD	14	3-4	7-10	4	Turkish

Note. LD = learning disability; ITRF = integrated teacher report form; L1 = first language; PW = pseudo word; PR = percentile; W = words.

Design

A multiple-baseline design across participants was chosen to be able to assess the individual learning progress of each student and to exclude alternative explanations for the effectiveness of the treatment as best as possible through the delayed start of the intervention (Ledford & Gast, 2018). The participants were divided into three groups of four children each, who started the intervention on different days after baseline. Group 1 was from Grade 6, Group 2 from Grade 7, and Group 3 from Grade 8.

Groups were randomly assigned to start the intervention. Group 1 had four planned baseline measurements, Group 2 had five baseline measurements, and Group 3 had six baseline measurements. A total of 18 measurement time points were planned. Due to the COVID-19 pandemic, the study had to be stopped after 11 measurement time points because the school had to close. The baseline measurements and the intervention took place three times a week within 30 minutes over a period of 4 weeks. Two well-instructed graduate students served as interventionists. They conducted the assessments and applied the training. The interventionists supervised all three groups together.

Dependent Variables and Measurement

There were two dependent variables, which we expected would be positively influenced by the intervention: expressive vocabulary and sight word reading. Both were tested using the word pool of the previously selected 40 and 30 words. The procedure was exactly the same as the word pretest. For expressive vocabulary, with 40 words, the appropriate pictures were each put on a PPT slide, and the youth were asked to name the correct expression for the picture. For sight word reading, with 30 words, the words were each written on a slide, and the students were asked to read them within one second of their occurrence (Ehri, 2005). The presentation in the context of reading was set at a one-second rate. First, the pool of expressive words was assessed, and then the pool of words for reading. The order of the

words was randomized for each measurement. Both interventionists performed the measurements, with each participant being assessed individually. The total number of words correctly identified (expressive vocabulary) and the number of words correctly read (sight vocabulary) were recorded.

Material

In the baseline phase, cognitive puzzles were used (taken from CFT-20-R, Weiß, 2006; German adapted version of the CFT) that the adolescents had to solve together. For the intervention phase, 8.3 x 11.7-inch cardboard flashcards were developed for Phase A, each showing the picture and the matching word together. For Phase B, short stories were written by the students themselves, adapted to the students' interests. The training words were embedded in each of the stories, with 10 per session drawn from the total pool of 40 and inserted into the stories in which the 30 reading words were included. The training words were marked in green in the text, while the rest of the text was black. Care was taken to ensure that the text was not too difficult and that it was of the same length across sessions. The substories built on each other. All stories were stretched in 8.3 x 11.7-inch on a ring binder for the students to follow the story. For the reward system, each student was given two self-graphing sheets with rows of boxes underneath each other, where the rows represented the number of sessions, and the 40/30 boxes per row represented the maximum number of correct words known or read.

Procedures

The baseline phase was used to measure the two dependent variables without the influence of a specific intervention directed at improving either expressive vocabulary or sight word reading. Students were instructed to solve cognitive puzzles together in 30 minutes. The cognitive puzzles contained, on the one hand, the logical continuation of a certain sequence and, on the other hand, the task of deciding which symbol, from five symbols shown, did not match the others. These task types were taken from the Basic Intelligence Test Scale 2 (CFT; Weiß, 2006). Just like the intervention, the baseline took place in the three small groups. After the 30 minutes were up, the participants were assessed individually with respect to the two dependent variables.

The three small groups remained intact throughout the intervention phase, so the storytelling intervention, like the baseline phase, also took place in small groups. The storytelling intervention was divided into two stages in each session. For each session, a short story was designed that included the 10 training words for that session. The reason for training only 10 out of 40 words per session was to avoid overtaxing the students. Stage 1 was the introduction and repetition of the expressive training words and reading of these words. In each session, 10 of the 40 words were directly taught, in random order. Care was taken to ensure

that students were exposed to the words in somewhat equal numbers. The participants were seated in a semicircle around the two interventionists, who introduced the words using the flashcards. To do so, they first covered up the word and referred to the picture, and then asked which of the students had an idea of what the word matching word would be. Then, the interventionists revealed the written word, read the word aloud and clearly, and then asked all students to read the word in chorus; individual students were also allowed to try. The aim was for students to store the image and form of the word in their mental lexicon. Additionally, the content of the stories from the previous session was repeated.

In Stage 2, after 10 minutes, the storytelling intervention was implemented. The interventionists recited the story as well as they could from memory according to the principles of storytelling, with lots of gestures and facial expressions. When a training word (marked in green) came up, the story was stopped, and the appropriate flashcard was consulted. The students were asked if they remembered the word and if anyone would like to read it. This continued for 15 minutes. Thus, the total length of each session was 25 minutes, always consisting of Stage 1 and Stage 2. Following each intervention session, students were again measured individually with respect to the two dependent variables.

Based on the results of each measurement, all participants plotted the correct number of words they knew on their two self-graphing sheets for expressive vocabulary, as well as the number of words read correctly. These steps was incorporated to show the students a learning progression to further motivate them.

Treatment Fidelity

The importance of treatment fidelity has gained enormous attention in recent years, especially with regard to better assessing treatment effects and the beneficial effects of an intervention (Nelson et al., 2012). To keep the implementation as similar as possible in all three small groups, which was additionally ensured by the fact that the same two interventionists instructed all groups, and to better estimate which outcome was influenced by the intervention, a guideline was created to which the interventionists had to strictly adhere. Additionally, a checklist was designed as part of the treatment fidelity to factually record the same implementation. The sheet was divided into the following areas: “Attendance,” “External Circumstances,” “Implementation of the Intervention,” “Student Behavior,” and “Feedback.” Both interventionists filled out this sheet after each session; also, one third of the way through the intervention, an external person came to observe the intervention in the three groups and to fill out the sheet. There was 100% agreement between the interventionists themselves, as well as between the interventionists and the external person.

Results

The statistical program R was used to analyze the data. Initially, the focus was on visual inspection and descriptive data, including the mean baseline difference (MBD) calculation (Campbell, 2003). For further analysis, overlap measures – the percentage exceeding the median (PEM; Ma, 2006), non-overlap of all pairs (NAP), and Tau-U (Parker et al., 2011), with additional correction for a possible baseline trend (A vs. B + Trend B + Trend A) – were also used. A Level 2 regression analysis was performed across all subjects with special attention to the trend as well as the level and slope effect from Phase A to Phase B (Huitema & McKean, 2000).

Expressive Vocabulary

For the first dependent variable, there was a clearly visible increase from Phase A to Phase B for Tana, Cem, and Alen, including direct increases when the intervention was applied. For Mattina, there was also a steady improvement, although it took her longer than the others. For Alen, a discrete trend tendency can be observed in Phase A. The descriptive data show the highest percentage increase for Mattina (1071%). Unlike the others, Mattina had a very flat baseline with lower values. Cem and Alen reached 35.00 words in Phase B, and Mattina had the lowest total, with 16.00.

Figure 1

Amount of Known Words Expressively

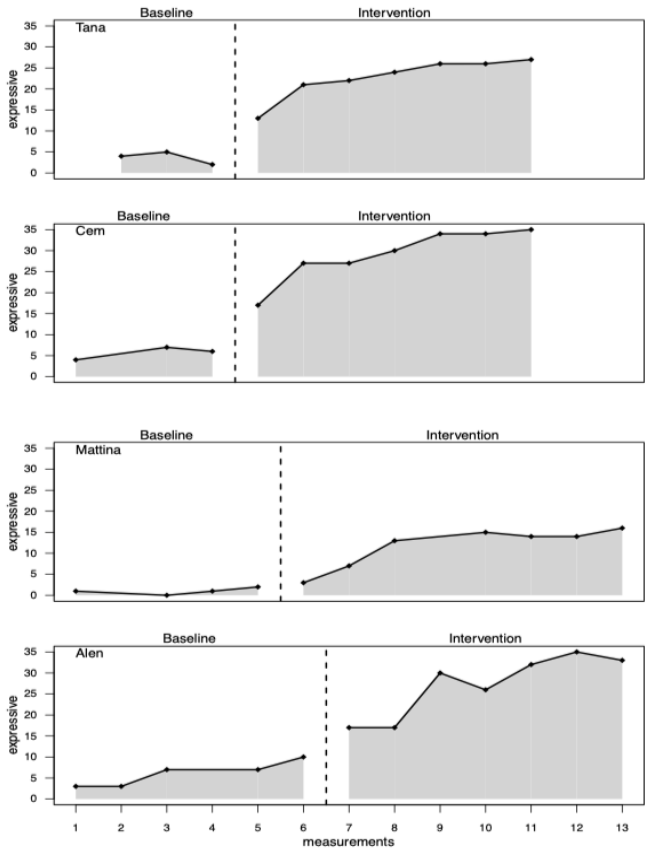


Table 2

Descriptive Data for Expressive Vocabulary

	N(A)	N(B)	*M*(A)*SD*	*M*(B) *SD*	Max A	Max B	MBD
Tana	4	7	3.67(1.53)	22.71(4.82)	5.00	27.00	518,80%
Cem	4	7	5.67(1.53)	29.14(6.31)	7.00	35.00	413,93%
Mattina	5	8	1.00(0.82)	11.71(4.82)	2.00	16.00	1071%
Alen	6	7	6.00(7.47)	27.14(7.47)	10.00	35.00	352,33%

Note. N = measurements; A = Phase A; B = Phase B; *M* = mean; *SD* = standard deviation; MBD = mean baseline difference.

The NAP and PEM show a value of 100.00 for all students, which can be described as a maximum value and indicates a strong effect. The Tau-U results show a moderate effect for Mattina ($p < .05$) and Alen ($p < .01$), and a large change for Tana ($p < .001$) and Cem ($p < .01$). In the regression analysis, a no statistically significant A-Phase trend was found throughout ($p = .26$). A significant level ($p < .001$) and slope ($p < .05$) effect can be reported, however, with the participants managing to improve expressively by an average of 1.410 words per intervention session.

Table 3

Overlap Indices for Expressive Vocabulary

	NAP	PEM	Tau-U	*p*
Tana	100.00	100.00	0.76	<.001
Cem	100.00	100.00	0.71	<.01
Mattina	100.00	100.00	0.53	<.05
Alen	100.00	100.00	0.55	<.01

Note. NAP = nonoverlap of all pairs; PEM = percentage exceeding the median.

Table 4

Regression Analysis for Expressive Vocabulary (Level 2)

	Phase B	*SE*	*t*	*p*
Intercept	2.048	3.534	0.579	.56
Trend	0.731	0.631	1.158	.26
Level	8.292	2.177	3.808	<.001
Slope	1.410	0.718	1.965	<.05

Reading

With regard to the visual vocabulary, strong increases from Phase A to Phase B were also found across all subjects, with Alen already showing relatively high values at baseline,

although a negative trend emerged. Here too level effects can be observed for all students, including Mattina, who showed a very steep increase in reading at the end of Phase B. Mattina also showed the strongest percentage increase from Phase A to Phase B. As expected, Alen showed the lowest increase, but the maximum value in Phase B was 30.00.

Figure 2

Amount of Words Read Correctly

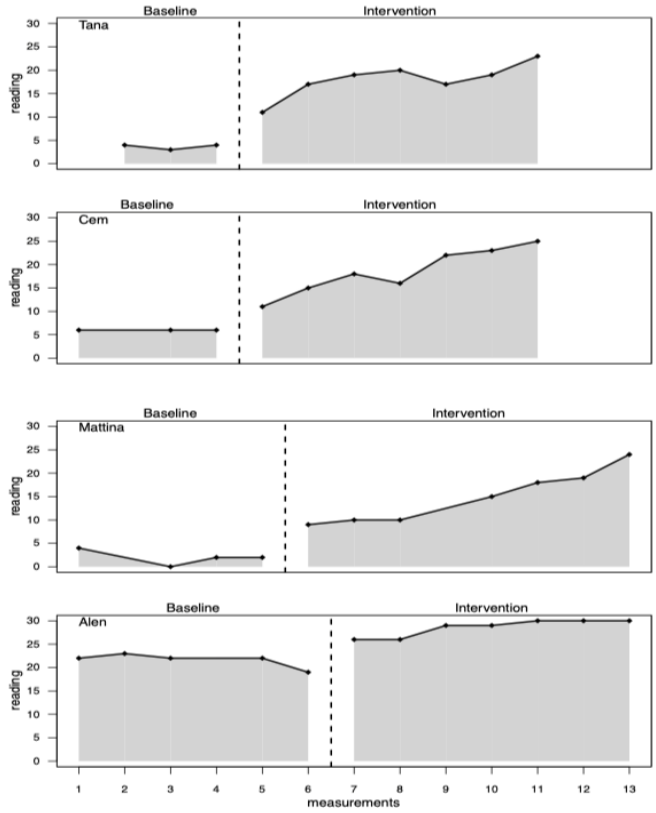


Table 5

Descriptive Data for Words Read Correctly

	N(A)	N(B)	*M*(A)*SD*	*M*(B) *SD*	Max A	Max B	MBD
Tana	4	7	3.67(0.58)	18.00(3.70)	4.00	23.00	390,46%
Cem	4	7	6.00(0.00))	18.57(5.00)	6.00	25.00	209,50%
Mattina	5	8	2.00(1.63)	15.00(5.66)	4.00	24.00	650%
Alen	6	7	21.60(1.52)	28.57(1.81)	23.00	30.00	32,27%

Note. N = measurements; A = Phase A; B = Phase B; *M* = mean; *SD* = standard deviation; MBD = mean baseline difference

All subjects showed a large change of 0.62–0.73 in Tau-U scores, with Tana and Mattina having the lowest scores. However, all values are considered statistically significant ($p < .01$). The analyses for the NAP and PEM showed the highest possible change across all students. The regression analysis displayed a statistically significant negative trend in Phase

A ($p < .05$), as well as a statistically significant level effect ($p < .001$) and slope effect ($p < .001$), with a beta coefficient of 2.454.

Table 6

Overlap Indices for Words Read Correctly

	NAP	PEM	Tau-U	*p*
Tana	100.00	100.00	0.62	<.01
Cem	100.00	100.00	0.73	<.01
Mattina	100.00	100.00	0.63	<.01
Alen	100.00	100.00	0.72	<.01

Note. NAP = nonoverlap of all pairs; PEM = percentage exceeding the median.

Table 7

Regression Analysis for Words Read Correctly (Level 2)

	Phase B	*SE*	*t*	*p*
Intercept	11.049	3.673	3.008	<.01
Trend	-0.790	0.365	1-2.167	<.05
Level	5.847	1.252	4.669	<.001
Slope	2.454	0.414	5.930	<.001

Discussion

Main Findings

German L2 students face tremendous obstacles as they try to acquire sufficient German vocabulary and reading skills to succeed academically. In cases where they also experience serious learning and behavior problems, the challenges often seem insurmountable. The purpose of this single-case study was to present a method using a self-designed multimodal storytelling intervention that can effectively train expressive vocabulary and sight-word reading simultaneously in a very short time for adolescents with severe learning and behavioral difficulties.

With regards to the first dependent variable, expressive vocabulary, three of the participants showed enormously strong increases after the start of the storytelling intervention. Mattina, who started with the lowest baseline data, also demonstrated an increase, although not as rapid as the others. She did, however, demonstrate the highest improvements in percentage terms under the MBD. Tana, Cem, and Alen showed a leveling effect, which means that they all seemed to respond immediately to the intervention. Tana and Cem benefited the most. Mattina and Alen showed moderate effects. Mattina reached the lowest percentile rank

in the German vocabulary screening, which might indicate that she had a harder time remembering words overall.

For the second dependent variable, sight-word reading, all students showed an increase from Phase A to Phase B. The increase was especially strong for Tana, Cem, and Mattina. Alen already had relatively high values in Phase A, which means that a rapid increase was not possible due to the ceiling effect. Accordingly, he also demonstrated the smallest percentage increase from Phase A to Phase B. Alen also seemed to show a negative trend at the baseline, which could indicate decreasing motivation because L2 learning is highly correlated with motivation (Grey et al., 2015). Because the variables were not addressed at the baseline, he might have been frustrated. Mattina showed the strongest increase with respect to MBD. In terms of effect sizes, Tana and Mattina benefited the least from the treatment. However, all effects can be considered strong.

Overall, the results align with findings from other studies on vocabulary learning (Barwasser et al., 2020; Bisson et al., 2013; Johnson & Mayer, 2009; Knaak et al., 2021; Peker et al., 2018; Peters, 2014) and memorization of sight vocabulary (Lee & Yoon, 2017; Stevens et al., 2017). The intervention seems to be suitable for both variables, with no clear difference in effectiveness between them. Overall, it is impressive that in such a short time (three times a week over five weeks), both variables could be effectively improved in students with several educational challenges.

Limitations

The promising findings of our research need to be considered alongside a number of limitations. First of all, we conducted a single-case study with only four participants, which makes generalization of the results difficult. However, the advantage of single-case studies is that they enable assessment of individuals and their responses to an intervention and, therefore, enable us to better adapt and optimize interventions.

Another issue is the limited number of measurement time points, mainly due to the COVID-19 pandemic. Nevertheless, even in a short period with only a few sessions, an increase in both dependent variables could be seen. That is, the intervention is time-efficient, which is very important for schools, as there is often not enough time outside of the standard curriculum to embed time-consuming interventions in the classroom.

Methodologically, the baseline with three measurement times could be criticized, because at least five measurement times per phase are recommended. According to Kratochwill et al. (2013), however, three measurements meet single-case design standards with reservations. In addition, it is often not possible to extend the baseline phase to the recommended length due to time constraints at schools.

Furthermore, we evaluated the benefits of a multicomponent intervention, and it is not possible to determine exactly which components had what effect; that is, it is not possible to say to what extent the flashcards alone, storytelling, and self-graphing contributed to the improvements in both dependent variables. However, the intention was not to determine the individual mechanisms of action, because it is a simple and easy-to-use intervention as a method package, which seems to work in this form.

Recommendations for Future Research

As further recommendations, the sample could be increased to include more students with specific characteristics to see how universally the storytelling intervention may be used. In addition, one could look at other dependent variables, such as writing and reading comprehension. Also, researchers could consider adopting a group design to increase the sample size and to compare storytelling with other interventions. Based on the experiences gained from the COVID-19 pandemic and the increasing importance of digitalization in schools, the storytelling intervention could be digitalized and made available to schools and students in the form of an app. In this way, teachers and students would also have access to the intervention if, for example, schools had to close and change to distance learning.

Conclusion

In summary, the simple storytelling method presented here can effectively help secondary-level students with German as an L2 who have learning and behavioral difficulties in expressive vocabulary and sight word reading in a very short time. Thus, the study contributes to the literature by showing that even severely struggling L2 students can learn a language if appropriate methods are used, including adapting them to the students' everyday lives and, thus, arouse their interest and motivation. Overall, this is an effective way to narrow the gap between groups of students differing in terms of school performance and provide all students with the opportunity for a fair and equal education.

References

Abbasian, G.-R., & Ghorbanpour, E. (2016). The effect of flash card-based instruction on vocabulary learning by EFL learners. *International Journal of Humanities and Cultural Studies,* 1929–1938.

Adwani, P., & Shrivastava, S. (2017). Analysis of factors affecting second language Acquisition. *International Journal of Social Sciences and Management*, *4*(3), 158–164. https://doi.org/10.3126/ijssm.v4i3.17247

Al-Mansour, S., & Al-Shorman, R. A. (2011). The effect of teachers' storytelling aloud on the reading comprehension of Saudi elementary stage students. *Journal of King Saud*

University. Language and Translation, *23*(2), 69–76. https://doi.org/10.1016/j.jksult.2011.04.001

Amiryousefi, M., & Ketabi, S. (2011). Mnemonic instruction: A way to boost vocabulary learning and recall. *Journal of Language Teaching and Research*, *2*(1), 178–182. https://doi.org/10.4304/jltr.2.1.178-182

Anjomshoa, L., & Sadighi, F. (2015). The importance of motivation in second language acquisition*. International Journal on Studies in English Language and Literature*, *3*(2), 126–137.

Barwasser, A., Knaak, T., & Gruenke, M. (2020). The effects of a multicomponent storytelling intervention on the vocabulary recognition of struggling English as a foreign language learners with learning disabilities. *Insight Into Learning Disabilities*, *17*(1), 35–53.

Bishara, S., & Kaplan, S. (2016). Executive functioning and figurative language comprehension in learning disabilities. *World Journal of Education*, *6*(2), 20–32. https://doi.org/10.5430/wje.v6n2p20

Bisson, M., van Heuven, W. B., Conklin, K., & Tunney, R. J. (2013). Incidental acquisition of foreign language vocabulary through brief multi-modal exposure. *PLoS ONE*, *8*(4), 1–7. https://doi.org/10.1371/journal.pone.0060912

Bisson, M. J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language Learning*, *64*(4), 855–877. https://doi.org/10.1111/lang.12085

Brady, N. C., Storkel, H. L., Bushnell, P., Barker, R. M., Saunders, K., Daniels, D., & Fleming, K. (2015). Investigating a multimodal intervention for children with limited expressive vocabularies associated with autism. *American Journal of Speech-Language Pathology*, *24*(3), 438–459. https://doi.org/10.1044/2015_AJSLP-14-0093

Burr, E., Haas, E., & Ferriere, K. (2015). *Identifying and supporting English learner students with learning disabilities: Key issues in the literature and state practice.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. https://files.eric.ed.gov/fulltext/ED558163.pdf

Butler, K. (2020). The value of direct instruction for at-risk-students. *Journal of Education and Development*, *4*(2), 10–16. https://doi.org/10.20849/jed.v4i2.741

Campbell, J. M. (2003). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: A quantitative synthesis of single-subject research. *Research in Developmental Disabilities*, *24*(2), 120–138. https://doi.org/10.1016/S08914222(03)00014-3

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

Ciullo, S., Ortiz, M. B., Al Otaiba, S., & Lane, K. L. (2016). Advanced reading comprehension expectations in secondary school: considerations for students with emotional or behavior disorders. *Journal of Disability Policy Studies*, *27*(1), 54–64. https://doi.org/10.1177/1044207315604365

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

De Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhoefer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, *68*(4), 906–941. https://doi.org/10.1111/lang.12296

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

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching.* Multilingual Matters. https://doi.org/10.21832/9781847691767-004

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening and viewing. Which mode of input is most effective? *Studies in Second Language Acquisition*, *42*(3), 499–523. https://doi.org/10.1017/S0272263119000494

Fjortoft, A., McLaughlin, T. F., Derby, K. M., Everson, M., & Johnson, K. (2014). The effects of two direct instruction teaching procedures to teach basic skills to two students with disabilities. *Multidisciplinary Journal of Education Psychology*, *4*(2), 1–32. https://doi.org/10.4471/remie.2014.09

Fraher, C., Jones, K., Caniglia, C., Crowell, G., Hastings, K., & Zumwalt, K. (2019). Effectiveness of direct instruction flashcards on sight word identification for a high school student with a specific learning Disability. *Insights Into Learning Disabilities*, *16*, 37–44.

Grey, S., Williams, J. N., & Rebuschat, P. (2015). Individual differences in incidental language learning: The role of learning styles, personality, and working memory. *Learning and Individual Differences*, *38*, 44–53. https://doi.org/10.1016/j.lindif.2015.01.019

Gruenke, M., & Cavendish, W. (2016). Learning disabilities around the globe. *Learning Disabilities: A Contemporary Journal, 14*(1), 1–8.

Haager, D., & Osipova, A. V. (2017). Enhancing academic instruction for adolescent English language learners with or at risk for learning disabilities. *Insights Into Learning Disabilities*, *14*(1), 7–26. https://files.eric.ed.gov/fulltext/EJ1165739.pdf

Hacking, J., & Tschirner, E. (2017). The contribution of vocabulary knowledge to reading proficiency: The case of college Russian. *Foreign Language Annals*, *50*(3), 500–518. https://doi.org/10.1111/FLAN.12282

Harju-Luukkainen, H., McElvany, N., & Stang, J. (2020). *Monitoring student achievement in the 21st century. European policy perspectives and assessment strategies*. Springer. https://doi.org/10.1007/978-3-030-38969-7

Hemmati, F., Gholamrezapour, Z., & Hessamy, G. (2015). The effects of teachers' storytelling and reading story aloud on the listening comprehension of Iranian EFL learners. *Theory and Practice in Language Studies*, *5*(7), 1482–1488. https://doi.org/10.17507/tpls.0507.22

Hirsch, S. E., Ennis, R. P., & McDaniel, S. C. (2013). Student self-graphing as a strategy to increase teacher effectiveness and student motivation. *Beyond Behavior*, *22*(3), 31–39. https://doi.org/10.1177/107429561302200305

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

Huang, H. (2006). The effects of storytelling on EFL young learners' reading comprehension and word recall. *English Teaching & Learning*, *30*, 51–74.

Huitema, B. E., & McKean, J. W. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement*, *60*(1), 38–58. https://doi.org/10.1177/00131640021970358

Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, *27*(2), 129–140. https://doi.org/10.1017/S0272263105050084

International Organization for Migration (IOM). (2019). *World migration report 2020*. https://publications.iom.int/system/files/pdf/wmr_2020.pdf https://doi.org/10.18356/b1710e30-en

Isphording, I., & Otten, S. (2014). Linguistic barriers in the destination language acquisition of immigrants. *Journal of Economic Behavior & Organization*, *105*, 30–50. https://doi.org/10.1016/j.jebo.2014.03.027

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, *101*(3), 621–629. https://doi.org/10.1037/a0015183

Kim, M. (2010). The effects of storytelling on adult English language learners*. Linguistic Research, 27*(3), 447–473. https://doi.org/10.17250/khisli.27.3.201012.004

Kitzinger, A. (2020, December 11). *METACOM.* https://www.metacom-symbole.de

Knaak, T., Gruenke, M., & Barwasser, A. (2021). Enhancing vocabulary recognition in English Foreign Learners with and without learning disability. The effects of a multi-component storytelling intervention approach. *Learning Disabilities: A Contemporary Journal, 19*(1), 1–18.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*(1), 26–38. http://doi.org/10.1177/0741932512452794

Landrum, R. E., Brakke, K., & McCarthy, M. A. (2019). The pedagogical power of storytelling. *Scholarship of Teaching and Learning in Psychology*, *5*(3), 247–253. https://doi.org/10.1037/stl0000152

Ledford, J. R., & Gast, D. L. (2018). *Single case research methodology: Applications in special education and behavioral sciences*. Routledge. https://doi. org/10.4324/9781315150666

Lee, J., & Yoon, S. Y. (2017). The effects of repeated reading on reading fluency for students with reading disabilities: A meta-analysis. *Journal of Learning Disabilities*, *50*(2), 213–224. https://doi.org/10.1177/0022219415605194

Li, S., Tao, S., Joshi, R. M., & Xu, Q. (2017). Second-language reading difficulties among native Chinese-speaking students learning to read English: The roles of native- and second-language skills. *Reading Research Quarterly*, *53*(4), 1–19. https://doi.org/10.1002/rrq.204

Lin, G. (2018). Modeling the relationship of metacognitive knowledge, L1 reading ability, L2 language proficiency and L2 reading. *Reading in a Foreign Language*, *30*(2), 209–231. http://nflrc.hawaii.edu/rfl

Liu, J. (2004). Effects of comic strips on L2 learners' reading comprehension. *TESOL Quarterly, 38*(2), 225–243. https://doi.org/10.2307/3588379

Lopes-Murphy, S. A. (2020). Contention between English as a second language and special education services for emergent bilinguals with disabilities. *Latin American Journal of Content & Language Integrated Learning*, *13*(1), 43–56. https://doi.org/10.5294/laclil.2020.13.1.3

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

Maag, J. W. (2019). Meta-analysis and quality of self-monitoring behavior interventions with children and adolescents. *International Journal of Education*, *11*(1), 40–67. https://doi.org/10.5296/ije.v11i1.14580

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. https://files.eric.ed.gov/fulltext/EJ1122799.pdf

Mackey, A., & Sachs, R. (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning*, *62*(3), 704–740. https://doi.org/10.1111/j.1467-9922.2011.00649.x

Maehler, C., & Schuchardt, K. (2016). Working memory in children with specific learning disorders and/or attention deficits. *Learning and Individual Differences*, *49*, 341–347. https://doi.org/10.1016/j.lindif.2016.05.007

Marksteiner, T., Janke, S., & Dickhaeuser, O. (2019). Effects of a brief psychological intervention on students' sense of belonging and educational outcomes: The role of students' migration and educational background. *Journal of School Psychology*, *75*, 41–57. https://doi.org/10.1016/j.jsp.2019.06.002

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, *80*(3), 300–335. https://doi.org/10.3102/0034654310377087

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.004

McDaniel, S. C., Jolivette, K., & Ennis, R. P. (2013). The effects of self-graphing on oral reading fluency for a student with E/BD within an alternative education school. *Journal of the American Academy of Special Education Professionals*, 69–82. https://files.eric.ed.gov/fulltext/EJ1135723.pdf

Mello, R. (2001). Building bridges: How storytelling influence teachers and student relationships. *The Journal of Higher Education*, *17*(6), 301–307. https://files.eric.ed.gov/fulltext/ED457088.pdf

Moll, K., & Landerl, K. (2010). *SLRT-II: Lese-und Rechtschreibtest; Weiterentwicklung des Salzburger Lese-und Rechtschreibtests (SLRT) [Reading and spelling test; further development of the Salzburg reading and spelling test]*. Huber.

National Institute for Literacy. (2009). *Developing early literacy: Report of the National Early Literacy Panel*. National Center for Family Literacy.

National Institute of Child Health and Human Development (NICHD). (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). U.S. Government Printing Office.

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research*, *39*(4), 374–396. https://doi.org/10.1007/s11414-012-9295-x

Organisation for Economic Co-operation and Development. (2019). *Programme for international students' assessment (PISA). Results from PISA 2018*. https://www.oecd.org/pisa/publications/PISA2018_CN_DEU.pdf

Pae, T. (2018). A simultaneous analysis of relations between L1 and L2 skills in reading and writing. *Reading Research Quarterly*, *54*(1), 109–124. https://doi.org/10.1002/rrq.216.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

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

Peker, H., Regalla, M., & Cox, D. T. (2018). Teaching and learning vocabulary in context. Examining engagement in three prekindergarten French classrooms. *Foreign Language Annals*, *51*(2), 472–483. https://doi.org/10.1111/flan.12338

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

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, *18*(1), 75–94. https://doi.org/10.1177/1362168813505384

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Roney, R. C. (1996). Storytelling in the classroom: Some theoretical thoughts. *Storytelling World*, *9*, 7–9.

Saito, K., Dewaele, J. M., & Hanzawa, K. (2017). A longitudinal investigation of the relationship between motivation and late second language speech learning in classroom settings. *Language and Speech*, *60*(4), 614–632. https://doi.org/10.1177/0023830916687793

Sarisahin, S. (2020). Reading comprehension strategies for students with learning disabilities who are emergent bilingual. *Intervention in School and Clinic*, *56*(1), 3–12. https://doi.org/10.1177/1053451220910731

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. RMC Research Corporation, Center on Instruction.

Schleppegrell, M., & O'Hallaron, C. (2011). Teaching academic language in L2 secondary settings. *Annual Review of Applied Linguistics*, *31*, 3–18. https://doi.org/10.1017/S0267190511000067

Shin, J. (2020). A meta-analysis of the relationship between working memory and second language reading comprehension: Does task type matter? *Applied Psycholinguistics*, *41*(4), 873–900. https://doi.org/10.1017/S0142716420000272

Shin, J., Dronjic, V., & Park, B. (2019). The interplay between working memory and background knowledge in L2. *TESOL Quarterly*, *53*(2), 320–347. https://doi.org/10.1002/tesq.482

Sparks, R. L. (2012). Individual differences in L2 learning and long-term L1-L2 relationships. *Language Learning*, *62*, 5–27. https://doi.org/10.1111/j.1467-9922.2012.00704.x

Sparks, R. L. (2016). Myths about foreign language learning and disabilities. *Foreign Language Annals*, *49*(2), 252–270. https://doi.org/10.1111/j.1944-9720.2009.01005.x

Stevens, E. A., Walker, M. A., & Vaughn, S. (2017). The effects of reading fluency interventions on the reading fluency and reading comprehension performance of elementary students with learning disabilities: A synthesis of the research from 2001 to 2014. *Journal of Learning Disabilities*, *50*(5), 576–590. https://doi.org/10.1177/0022219416638028

Stiensmeier-Pelster, J., & Otterpohl, N. (2018). Motivation at school and university. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and action.* Springer. https://doi.org/10.1007/978-3-319-65094-4_18

Stockhard, J., Wood, T. W., Coughlin, C., & Khoury, C. R. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research*, *88*(4), 479–507. https://doi.org/10.3102/0034654317751919

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. https://doi.org/10.1177/10634266070150020101

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis or correlation studies. *Language Learning*, *69*(3), 559–599. https://doi.org/10.1111/lang.12343

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

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33–52. https://doi.org/10.1017/S0272263105050023

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.

Young-Scholten, M. (2015). Who are adolescents and adults who develop literacy for the first time in an L2, and why are they of research interest? *Writing Systems Research*, *7*(1), 1–3. https://doi.org/10.1080/17586801.2015.998443

Appendix F Article 6

Barwasser, A., Bracht, J., & Grünke, M. (2021_accepted). A storytelling approach on vocabulary, reading and letter sound fluency of struggling first graders with German as Second Language with and without behavioral problems. *Frontiers in Psychology.*

Abstract

The number of students learning German as a Second Language (L2) is steadily increasing. Unfortunately, studies show that weak school performance affects a larger proportion of these students and additional behavioral problems can create even larger learning barriers. In order to master a language, the focus is not only on vocabulary, but also on reading, and studies show that multi-component intervention in reading and L2 acquisition is particularly promising. Therefore, this multiple baseline study focuses on a multi-component storytelling intervention on vocabulary, reading and letter sound fluency of low-achieving first graders with German as L2 with and without behavioral problems (N=7). The intervention was implemented 3 times a week over a 6-week period. Results show significant large to very large effects on vocabulary and moderate to large effects on letter sound fluency and reading, providing indication for the positive impact of storytelling on multiple aspects simultaneously for the focused sample.

Keywords: Storytelling, vocabulary, reading and letter sound fluency, German Second Language, behavior problems

Introduction

German as a Second Language

Education is largely dependent on language and in the German education system, the understanding and speaking of German at native language level is assumed. The entitlement to education thus demands that these students in particular have the opportunity to develop in the German education system (Becker-Mrotzek et al., 2012). According to the Federal Statistical Office, about 11% of the students at educational institutions have a migration background (Federal Statistical Office, 2020) and learning German as Second language (L2) (Aschenbrenner et al., 2016). The Programme for International Student Assessment (PISA) studies shows that students with a migration background perform significantly worse at school than students who learn German as their mother tongue (OECD, 2019) and it has been shown that a large proportion of fourth graders do not or only partly speak German at home (Hußmann et al., 2017). German L2 students struggle in schools leading to a challenge for the teachers

in designing appropriate lessons (Becker-Mrotzek et al., 2012) and a challenge for the students themselves with respect to educational opportunities.

Hurdles for Second Language Learners

To be proficient in language, various skills within language acquisition such as phonetics and literacy are needed (Aschenbrenner et al., 2016). But especially vocabulary learning is immensely important (Schmitt, 2011) and it is shown that particularly students with L2 experience severe failure in this area (Webb & Chang, 2012). In addition, letter sound fluency (LSF) is essential for language communication and acquiring the sound of individual letters presents a particular hurdle (Kim & Piper, 2019), and students who struggle with LSF are more likely to have difficulty in their reading skills later on as well (Piasta & Wagner, 2010). A reason for this might be that children fail to read because their overall L2 competence is not yet sufficient to read adequately (Wallace, 2014).

Also, it is widely known that a certain struggle in language development, as vocabulary, expression and reading, can be associated with problems in behavior (Peterson et al., 2013; Jansen et al., 2020). It has been reported that young children with language difficulties might develop problem behavior (Henrichs et al., 2013) which can get worse over time (Curtis et al., 2018). More specifically, deficits in language are connected to deficits in attention processing (Peterson et al., 2013) which can be linked to learning-related behavior (LRB). LRB, according to McClelland et al. (2006), includes abilities like staying focused, organizing school material and working on one's own. A meta-analysis by Chow and Wehby (2018) revealed a negative relationship between language deficits and problem behavior independent of age and time.

Important Language Components

Vocabulary is fundamental but challenging in a second language and influences all stages of acquisition (Ender, 2016). Vocabulary can be differentiated between expressive and receptive. Receptive vocabulary are words which can be recognized but not actively spoken whereas expressive vocabulary can be directly used (Schmitt, 2014). Significant correlations have been shown to exist between expressive vocabulary and reading ability in children from primary school (Wise et al., 2007). In general, it has been found that L2 vocabulary knowledge is linked to L2 reading comprehension (Lervåg & Aukrust, 2010). As in an L1, the automation of lower hierarchy processes, such as word recognition is fundamental for comprehension (Kramer & McLean, 2019). The Dual Route Model (DRM; Coltheart, 2005) describes two routes, the lexical and the non-lexical route, to show how readers read aloud. The lexical route refers to the mental lexicon where words can be automatically stored and retrieved (more important for irregular words: e.g., *"hoch" (high)* than for regular words: e.g., *"Sand" (sand)*). The non-lexical route goes through the grapheme-phoneme correspondence (e.g., important for non-words like "*brelo*" or "*blustof*"). In terms of direct word recognition, the direct route is

important, where sight words can be retrieved. Sight words are words that can be retrieved within one second of occurrence (Ehri, 2005). In addition to memorizing familiar words, Letter Sound Knowledge (LSK) also plays an important role in the non-lexical route of DRM because it enables readers to decode unfamiliar words (Ehri, 2002). Both approaches should be possible for a reader to build up adequate reading competence in a language. Clemens et al. (2017) found that LSF, a sub-component, was predictive of subsequent reading fluency with respect to kindergarten children. Through a mediation analysis of results from a large-scale intervention study (N=152), Hulme et al. (2012) showed that problems in LSK and phoneme awareness can cause difficulties in later word-reading-proficiency in 5-year-old children.

Fostering Second Language Acquisition

In order to counteract hurdles in second language acquisition and to offer L2 students an opportunity to acquire an L2 adequately, it is necessary provide effective support. The dual coding theory (DCT; Paivio, 2008) states that there is a verbal way and a non-verbal way (i.e., pictures) to store information underlining the importance of presenting new input verbally and non-verbally in a language, especially for L2 students (Huang et al., 2019). The verbal way is related to linguistic information (e.g., sound) and the non-verbal system is linked to visual information (e.g., pictures) (Paivio, 2007). According to Reed (2010) using both systems, maximizes the likelihood that information will be stored adequately.

Another way to train new content is either through explicit (intentional) training or implicit (incidental) training (Jin & Webb, 2020) – or a combination of both (Choo et al., 2012). intentional learning means that the learner is aware that he is learning something, and incidental learning means that the learner learns something like a by-product without being aware of it (Webb & Nation, 2017). In the case of incidental learning, it has been said that words are easier to acquire through repeated occurrence in context (Webb & Nation, 2017). Maurulis and Neuman (2010) conducted a meta-analysis about the impact of vocabulary interventions on the language development of pre-K and kindergarten children and found an overall effect size of g=0.88 of vocabulary training on word learning. Moreover, it was found that a combination of implicit training and explicit training lead to a higher effect size (g=1.21) than explicit (g=1.11) and implicit (g=0.62) in isolation. Hulme et al. (2012) found that teaching LSK and phoneme awareness explicitly in a reading and phonology intervention lead to an improvement of these two abilities.

It has been shown that multi component supports, including amongst other, phonics, vocabulary and fluency increases the probability of becoming a good reader (Foorman & Torgesen, 2001). A recently published literature review by Donegan and Wanzek (2021) showed that multi component reading interventions for elementary school with the highest effects incorporate instruction in decoding on the word level and in summary multicomponent

interventions are promising with regard to improve foundational reading skills and reading comprehension.

Storytelling

Listening to stories has been shown to influence language development on different areas positively in children (Isbell et al., 2004). Storytelling is a procedure where a teller tells a story in an authentic environment using gestures, vocalization and images to convey a certain message to the audience who are incorporated in the storytelling procedure (Mello, 2001). Storytelling has the ability to engage learners personally (Brewster et al., 2002), motivate learners, and spark interest in the subject matter (Wright, 2013). Using storytelling does have positive impacts on child's oral and written language development (Baker et al., 2013; Fien et al., 2011) and through the procedure of storytelling facts as well as vocabulary can be memorized better (Wajnryb, 2003). Lenhart et al. (2018) focused on the impact of story listening on vocabulary acquisition and found that vocabulary was acquired incidentally without any word explanation with a moderate effect (d=0.37) which was in turn not stable over time (age 3-6) concluding that using only incidental vocabulary training might not be sufficient enough. A meta-analysis by Mello (2001) indicates that using storytelling led to gains in vocabulary, fluency and writing skills, amongst other variables. Suggate et al. (2013) examined storytelling in 2nd and 4th grade German readers and revealed that more freely storytelling has more benefits than simply reading the story.

Read aloud has been shown to be effective for vocabulary, comprehension and narrative language in first graders (Baker et al., 2020) and for phonological awareness (Swanson et al., (2011). Since storytelling belongs rather to the implicit method, adding flashcards to storytelling in order to teach components explicitly would be, according to Maurulis and Neuman (2010), a further boost in effectiveness. Two additional studies by Barwasser et al. (2020) and Knaak et al. (2021) investigated a combined storytelling intervention consisting of implicit and explicit components on vocabulary acquisition in English language learning of students with and without learning disabilities showing that this combination is effective in the context of vocabulary acquisition. Barwasser et al. (2021) went a step further and examined the combined storytelling method in German Second Language learners from primary school on vocabulary and reading with overall positive effects.

Motivation and Self-Graphing

In second and foreign language acquisition, the ability to increase competence in a language often depends on how motivated a learner is (Ghenghesh, 2010). Adding motivational components to an intervention can be specifically successful (Leko, 2016; Bownman-Perrott et al., 2013). It has been shown that incorporating self-monitoring

procedures, such as self-graphing, the visualization of a student's own progress showing earlier scores and current scores (Stotz et al., 2008; Guzman et al., 2018; McKenna & Bettini, 2018), reading achievement can be improved for students with disabilities (Laurice & Eveleigh, 2011) and on task behavior as well as general academic productivity (DiGangi et al., 1991). Self-graphing can be realized by providing students with a graph overview where they can enter their scores after each measurement point in order to follow their own learning progress step by step. A meta-analysis by Guzman et al. (2018) revealed large effects of self-monitoring procedures on reading performance in K-12 students (*TauU* = 0.79, $p<.001$).

Research Questions

Based on the knowledge that there are a large number of low-performing German as a second language students in Germany, with both behavioral problems and motivation playing a significant role, a multi-component storytelling intervention was designed to simultaneously address three of important components in language learning: vocabulary, LSF, as well as sight word reading, and to investigate its effects on German L2 students with and without behavioral problems. In addition, we have implemented a social validity questionnaire in order to figure out the acceptance of the intervention rated by the participants. Assessing social validity is a necessity to crystallize the acceptance and usefulness of interventions (e.g., Briesch et al., 2013). Accordingly, the four research questions are:

1) Does a multi-component storytelling intervention lead to an increase in expressive vocabulary in German L2 students with and without behavior problems?

2) Does a multi-component storytelling intervention lead to an increase in LSF in German L2 students with and without behavior problems?

3) Does a multi component storytelling intervention lead to an increase in sight word reading in German L2 students with and without behavior problems?

4) How was the intervention evaluated by the participants in terms of social validity?

Material and Methods

Participants and Setting

The present study was conducted at an inclusive elementary school in a large city in North Rhine-Westphalia, Germany, targeting grade 1. To participate in the study, teachers of the respective classes were to identify all students who met the criterion 'German as a second language' (N=10). In addition, appropriate parental consent to participate in the study had to be obtained.

The teachers received a teacher questionnaire to obtain relevant information on the proposed students regarding socio-demographic characteristics.

Table 1

Characteristics of the participants

	Lio	Kim	Tila	Nele	Niek	Abden	Elif
Age	6;3	6;5	7;1	6;6	6;3	7;1	6;2
Grade	1	1	1	1	1	1	1
Gender	male	female	female	female	male	male	female
L1	Polish	Polish	Turkish	Chinese	Italian	Turkish	Turkish
LRB	12	4	13	3	14	10	6
Reading W (PR)	<7	<7	<2	9-13	7-8	16-17	5-11
Reading P (PR)	<2	<2	<4	24	6-8	19-23	8-10
BAKO (PR)							
Subtest PS	2	2	2	2	48	2	21
Subtest VS	42	6	19	53	6	6	61
Subtest RD	3	3	3	21	3	3	34
Subtest PI	8	8	8	74	8	8	21
Subtest SC	28	15	42	71	1	28	28
Subtest VD	57	7	23	57	23	23	57
Subtest WR	9	9	9	35	23	9	35
Total	7	0	2	39	2	1	31
Vocab (PR)	12	5	15	26	21	27	16

Note. Percentile (PR); Words (W); Pseudowords (P), Learning Related Behavior (LRB; cutoff 10); Native Language (L1); pseudoword segmentation (PS); vowel substitution (VS); residual word determination (RD); phoneme interchange (PI); sound categorization (SC); vowel length determination (VD); word reversal (WR); German Vocabulary Test (Vocab)

German Vocabulary Test

A vocabulary test (WS/ZF-R; Weiß, 2007) in the form of a group screening was used first to assess the students' verbal language skills. The WS/ZF-R measures colloquial vocabulary beyond the basic vocabulary of the German language and is used to determine the developmental level of verbal skills of students. The test sheet contains 30 multiple-choice items with five alternative answers each. Each task consists of a key word being given first. Subsequently, the respondents have to select the word from the five alternative answers that has a similar meaning as the given keyword. The reliability of the WS/ZF-R was assessed using the split-half method (N = 618), where values ranged from rtt = .79 to rtt = .90 with a mean value of rtt = .87. For the correlation with German grades (N = 689), the value was r = .48 (Weiß, 2007). The results are shown in percentiles (PR) and a PR under 15 means underdeveloped. For example, a percentile of 15 means 15 percent of the subjects in the norm

sample scored the same or fewer points. The participant with a PR of 15 therefore belongs to the 15 percent of the weakest in his age group.

Salzburg Reading and Spelling Test (SLRT II)

The Salzburg Reading and Spelling Test (SLRT II, Moll & Landerl, 2010) was used to assess reading ability at the word and pseudoword level. These two subtests each consisted of a one-minute reading fluency test by reading given words and pseudowords. The total time required is time-efficient at approximately five minutes. The parallel test reliability ranges from .90 to .98 and correlations with other German reading tests range from .69 to .92. All participants who had a percentile below 15 were selected for the study.

Test for phonological awareness (BAKO 1-4)

A test for phonological awareness for grades 1-4 was additionally used (BAKO 1-4, Stock et al., 2017). There are a total of 174 tasks divided into seven subtests: 1. pseudoword segmentation, 2. vowel substitution, 3. residual word determination, 4. phoneme interchange, 5. sound categorization, 6. vowel length determination, and 7. word reversal. The time required to complete the test is approximately 30 minutes. Norms are available for each grade level (N=876) and reliability shows that internal consistently varies by grade level (between $\alpha = .90$ and $\alpha = .92$, split-half reliability between $r = .90$ and $r = .94$). Criterion-related validity with reading or spelling performance measured by standardized tests varies by grade level between $r = .42$ and $r = .68$. (Stock et al., 2017). Results are again shown in PR.

Integrated Teacher Report Form (ITRF-G)

The Integrated Teacher Report Form (ITRF; Volpe et al., 2018) represents a multilevel screening procedure used to identify student behavior difficulties. In the present study, the ITRF-G short version is applied, which is the German translation of the English version. Using the test, evidence-based interventions can be individualized for all students with behavioral difficulties. In the research conducted, the screening is conducted by the classroom teachers as they are in the best position to assess the students' behavior. The teachers assess specific behaviors of the students on an assessment sheet and the items are created based on the factors "learning related behavior” and "oppositional/disruptive behavior". The ITRF-G is administered in a short version with 16 items, whereas the original version includes 47 items. The conducted short version has been positively evaluated and shows high internal consistency and sufficient test-retest reliability in terms of reliability and high external validity for all scales in terms of validity. The cutoff value for learning related behavior is 10 showing problems in this area (Volpe et al., 2018).

Word pretesting

To crystallize the final training words and to ensure that the words were not stored in either the expressive vocabulary or the mental lexicon for reading, words were auditioned prior to the study. Once for expressive vocabulary and once for reading. The pool of words (N=143) came from the Metacom Symbols (Kitzinger, 2020) and care was taken to ensure that words were taken which the children could use well in everyday life. These words were queried both expressively and in reading. For the reading test (day 1), the 143 words were integrated into a PowerPoint presentation so that one word was on each slide individually. The slides were scrolled in 1-second intervals, since according to Ehri (2005) a word is considered a sight word if it can be read within 1 second of its occurrence. Here, all words that could not be read were marked.

After a few days (day 2) the expressive test was performed with the exact same words. Here, the children were not shown the 143 words, but pictures matching the words. Here, too, there was a picture on a slide - there was no time limit. Now, for each picture, the children were asked what the word was called. All non-conscious words were marked and compared with the reading words. A total of 40 word-overlaps resulted for unknown expressive words and words not read correctly. The 40 training words in reading were the same as in vocabulary for the intervention and measurements later on. Thus, the children could neither read these words nor express them actively. The 40 training words, which were selected together with the teachers, had a mid-frequency of M= 10.5, meaning that the words appear 10.5 times per million words in a corpus (Brybaert et al., 2018). To estimate the frequency, we used the childlex database (Schroeder et al., 2015).

The students (N= 10) are divided into three groups. Group 1 had three children, group 2 had three children and group 3 had four children. All participants learned German with the entry of kindergarten at age 3-3;5. According to COVID 19 rules, groups were not allowed to be mixed across classrooms. Each group has a different baseline time and thus starts the intervention with a time delay. Three children are dropped from the data because they have too much missing data due to COVID 19 quarantine regulations. As a result, the finale sample for this paper is N=7.

Design

The present research utilized a multiple-baseline design across participants to examine the effects of the intervention. A single case analysis is often understood to be a study of one individual. However, a multiple-baseline design embeds subcases within an overall case. The introduction of the intervention is temporally staggered across the subjects. The goal of implementing a multiple-baseline design is to substantiate a cause-effects relationship by demonstrating that changes in the dependent variable only occur when the treatment is given

(Lane et al., 2017). First, a baseline of varying length is performed with 5-7 sessions. After each of these sessions, the dependent variables were collected. After completion of the baseline phase, the intervention starts in the following sessions. Data was also collected after each intervention session (e.g, baseline 1 – measurements; baseline 2 – measurements - …intervention 1 – measurements; intervention 2 – measurements; intervention 3 – measurements). Each group was randomly assigned to a specific baseline length resulting in group 1 = 5 baselines, group 2 = 6 baselines and group 3 = 7 baselines. The baseline and intervention sessions took place three times a week for 25 minutes, after which the children were measured individually for each of the three dependent variables. The entire period spanned 6 weeks and one week of diagnostic testing. Due to a previous school closure because of COVID 19, the study started later and comes to 18 measurement time points of originally planned 24. Two master's students for special needs education functioned as test leaders and interventionists. Both supported each group together.

Dependent variables and measurement

In total, there are three dependent variables: Expressive vocabulary, sight word reading, and LSF. The 40 training words were used for expressive vocabulary and for reading. For LSF, all letters from the German alphabet were measured.

1) **Expressive vocabulary**: The 40 training words were packed into a PowerPoint presentation in the form of pictures, with one picture per slide. For each picture, the child was asked if he knew the name of the word. The total number of correctly conscious words expressive was transferred to an Excel table per measurement point.
2) **LSF:** All letters of the German alphabet were mixed and written on two 8.3 x 11.7-inch sheets, so that a total of 104 letters could be seen. The child was now asked to pronounce as many sounds as possible correctly within 1 minute. A timer was set to 1 minute and the two test leaders listened attentively. The total number of correctly pronounced sounds was also entered in the Excel table for each measurement point.
3) **Reading:** The 40 training words written were embedded in a PowerPoint. Here, one word per slide was written down. The slides were separated by hashtags and were laid out in 1-second intervals (see Ehri 2005). Again, the total number of correctly read words was recorded in an excel table per measurement time point.

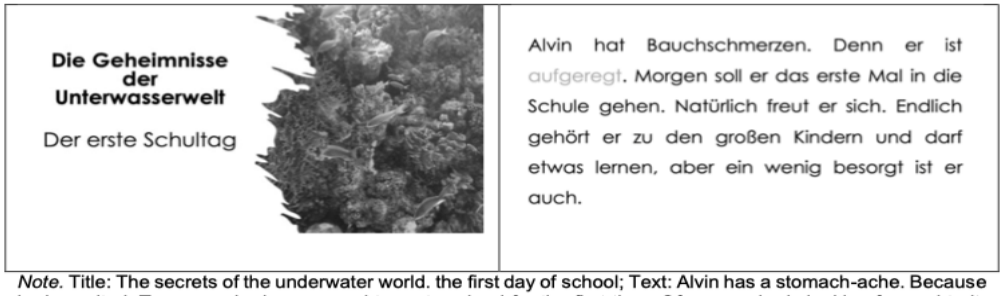
Intervention Material

For the direct instruction of the words and the sounds, a phonetic table and 8.3 x 11.7-inch flashcards with the letters on them and 8.3 x 11.7-inch flashcards with the picture and the matching word were used. For the storytelling intervention, short stories were required for each session. Before the study started, the master students talked to the children about their interests to determine the focus of the stories. In total, there was one full story with sub-

chapters per session. The stories (example Figure 1) were self-written with somewhat the same length and formatting. Additionally, care was taken to ensure that all words occurring were not too difficult. The training words were always embedded and from the pool of 40 words always 5 were taken into one story which appeared twice on one story. The words were randomly assigned to the stories, making sure that in the end the words occurred in equal proportions. The training words in the story were always highlighted in blue, while the rest of the font was black.

Figure 1

Example part of a story

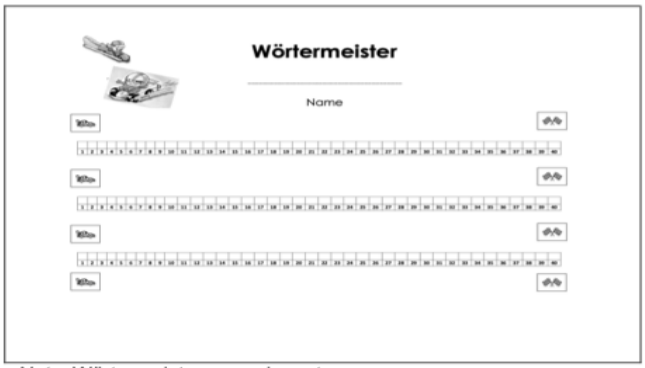


Note. Title: The secrets of the underwater world. the first day of school; Text: Alvin has a stomach-ache. Because he is excited. Tomorrow he is supposed to go to school for the first time. Of course, he is looking forward to it. Finally, he belongs to the big kids and is allowed to learn something, but he is also a little worried.

Regarding the motivational system, there were three self-graphing sheets for the children corresponding to the three dependent variables. Each sheet consisted of several rows one below the other, which were supposed to represent the sessions. The rows consisted of small boxes that were supposed to represent the number of words/sounds correctly known where the participants were asked to color the amount of correct known words/sounds after each measurement point.

Figure 2

Example of self-graphing sheet



Note. Wörtermeister = word master

Procedures

Baseline

The baseline (A phase) is used to record the actual state in a multiple baseline design. Before the storytelling intervention starts, all three groups go through a baseline phase of different lengths for the groups. The baseline activities must not have anything to do with reading, vocabulary or LSF, so that the dependent variables are not already promoted in the baseline. Thus, during baseline condition, games, puzzles and math problems are solved together in 25 minutes. These are simple tasks that do not explicitly promote vocabulary, reading or the LSF. Afterwards, the three dependent variables are measured for each child.

Storytelling

After the baseline (A) phases are all completed, the groups begin the intervention phase (B phase). The group constellations remain unchanged. Storytelling can be divided into two stages. In the 1st stage (10 minutes), the kids sit in a semi-circle around the interventionist who is firstly introducing the words to be learned directly to the participants. Both, the words and letters of the last story (despite session one), are repeated, and the words and letters of the current story are introduced through flashcards and a phonics table. In order not to overtax the children, only 10 of the 40 words are directly instructed per session. The interventionist holds up the flashcard with the word and the picture, covers the written word and asks the children, based on the picture, whether they know what it means. Then they talk about the word. Then the interventionist uncovers the written word and asks the children if anyone can read the word aloud. Then everyone reads together and then the interventionist reads the word again. After that, the interventionist lifts up the phonics picture. For each intervention session, 10 sounds were randomly selected to be trained. Using the phonics picture and the words, the interventionist asks, for example, for an "L": who knows how to pronounce that?" “And can you find the sound in one of our words?". The procedure lasts 10 minutes.

The 2nd stage (15 minutes) involves the process of storytelling. The stories were learned by heart by the interventionists and the text serves the children to follow the story and see the marked training words. Each story is told out loud to the students and if a training word is appearing in the story, the story is paused and the word as well as one sound is discussed using the appropriate flashcards (a word with a matching picture). After the storytelling, the three measurements are carried out with each child individually and feedback on the learning process follows on the self-graphing sheets. Each time after the measurement, each child enters the number of correct known items in two separate self- graphing sheets for the amount of correctly read words and correctly known word expressively.

Treatment Fidelity

In order to record treatment fidelity in the present study, the experimenters were first provided with a detailed script with steps to be followed. Additionally, the implementers were given a checklist to complete at the end of each intervention session without being aware of the intention of the sheet. This was used to reflect on compliance with what was outlined in the script. The checklist is divided into six sections: Environment/ external circumstances, planning, materials, procedure of support, diagnostics/ feedback, and handling student behavior during support using three response options ('yes'; 'no'; 'not applicable'). In addition, a free field was available to the investigators for comments on special features in the context of the support. The inter-rater reliability is 100%.

Social Validity

In order to measure the acceptance of the support by one of the students, a questionnaire was designed within the framework of social validity, which was handed out to the students at the end of the support. The interventionists were not present in order to avoid biased results and to obtain an honest opinion from the students. the questionnaire contains 9 items which should be rated on a scale from 0(=completely not agree) to 4 (= completely agree). The items were as follows: 1) Storytelling has helped me to be able to read words correctly; 2) *Storytelling helped me learn words and their meanings*; 3) *Storytelling helped me to pronounce sounds correctly*; 4) *I understood well the meaning of the promotion*; 5) *I have learned a lot during storytelling*; 6) *I gladly came to the intervention sessions*; 7) *The self-graphing sheets were fun*; 8) *The stories were great*; 9) *I would like to do more with stories in school*.

Data analysis

The entire data analysis was done using the statistics program "R" and the Scan Package for multiple baseline design analysis in order to estimate the intervention (B phase) effects compared to the baseline (A phase). The graphs (Figures 3-5) for each dependent variable serve for visual analysis. In addition, mean and median values of the two phases as well as the maximum values in phase A and phase B were determined and Mean Baseline Difference (MBDi). MBDi is a non-parametrical method which measures increase of a certain output from baseline (O'Brien & Repp, 1990). Further, overlap measures were used including the Non-Overlap of All Pairs (NAP, Parker et al., 2011a), the percentage exceeding the median (PEM, Ma, 2006), the Percentage of All Non-Overlapping Data (PAND; Parker et al., 2007), and finally the Tau-U additionally considering an A phase trend using the formular: A vs. B + TrendB − TrendA. TauU measures data non-overlap between phase A and phase B (Parker et al., 2011b).

The single-case reporting guidelines by Tate et al. (2016) suggest the use of inferential statistics to directly test for treatment effects. Even though there is still no universal gold standard for analyzing data from respective experiments, hierarchical piecewise regression modeling has become the most common tool for investigating the null hypothesis (Manolov, Amau, Solanas & Bono, 2010; Raudenbush & Bryk, 2002; Waddell, Nassar & Gustafson, 2011). In this approach, the data points during baseline of one individual are used to calculate a regression line and estimate the progression of the data during the intervention. Changes in level and/or slope across phases can then be tested for statistical significance (Level 1 analysis). Subsequently, data over several individuals can be accumulated to examine causal elements behind treatment effectiveness (Level 2 analysis). When regression modeling is used in group studies, each data point stems from a different individual. However, if this approach is applied in single case Level 1 research, the data points stem from one and the same person. One of the basic requirements for using parametric statistics (like regression analysis) is the independency of the distributed errors. There is no logical reason to assume that errors of different individuals are statistically associated. In contrast, the danger of autocorrelation in single case research is ever-present. For example, it is anything but unlikely that errors in observations that are close together in time are more similar than those that are more distant. The degree to which they correlate corresponds with the risk of incorrectly rejecting a true null hypothesis. To reduce the likelihood of mistakenly dismissing the absence of a given effect, we used a statistical package for R called SCAN (Wilbert, 2021) that controls for autocorrelation in single case data.

Results

Expressive Vocabulary

Overall, the visual baseline is very flat for all participants and there is a steady increase in the B phase. Tila (M=5.00), Abden (M=7.50) and Elif (M=8.83) start with slightly higher values in the A phase while Lio (M=0.00), Kim (M=0.50), Nele (M=1.80) and Niek (M=2.50) start very low. The highest mean value in the B phase is shown by Tila (M=31.42) and the lowest value is found in Niek (M=15.75). The highest increase is shown by Kim (3034%) and Lio (2469%9 and the lowest increase is shown by Baden (217.73%) and Elif (246.32%). Tila, Nele and Elif reach the maximum possible score of 40.00 in the B phase.

Figure 3

Amount of known expressive vocabulary

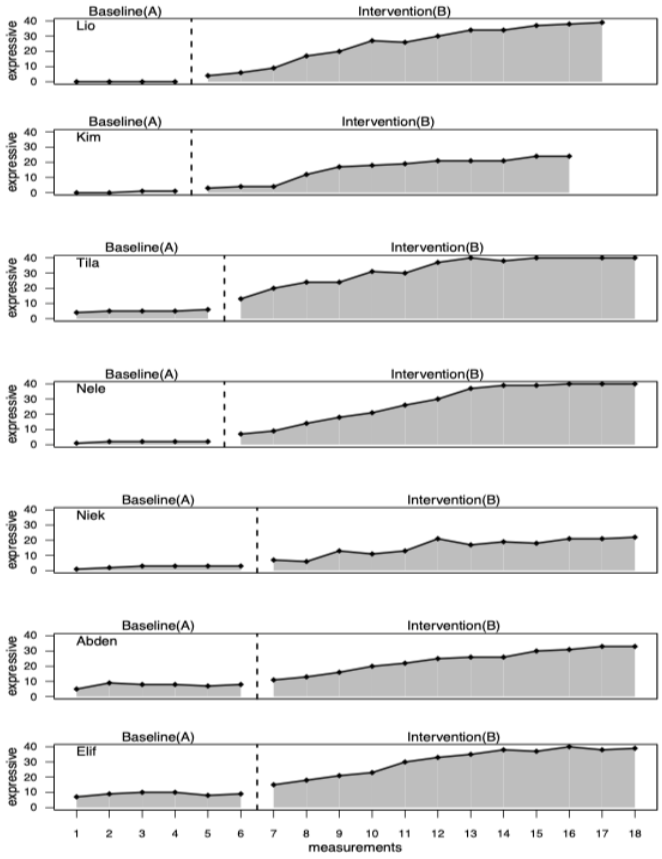


Table 2

Descriptive data for expressive vocabulary

Participants	N(A)	N(B)	M(A) SD	M(B) SD	MBDi	Md A	Md B	Max A	Max B
Lio	4	14	0.00(0.00)	24.69(12.41)	2469%	0.00	27.00	0.00	39.00
Kim	4	14	0.50(0.58)	15.67(7.91)	3034%	0.50	18.50	1.00	24.00
Tila	5	13	5.00(0.71)	31.42(9.30)	528,4%	5.00	34.00	6.00	40.00
Nele	5	13	1.80(0.45)	27.69(12.61)	1438,4%	2.00	30.00	2.00	40.00
Niek	6	12	2.50(0.84)	15.75(5.63)	527,6%	3.00	17.50	3.00	22.00
Abden	6	12	7.50(1.38)	23.83(7.57)	217,73%	8.00	25.50	9.00	33.00
Elif	6	12	8.83(1.17)	30.58(8.98)	246,32%	9.00	34.00	10.00	40.00

Note: measurements (N); Mean (M); Standard Deviation (SD); Mean Baseline Difference (MBDi); Median (Md); Maximum (Max); Minimum (min), A Phase (A), B Phase (B)

With regard to the overlap measures, the NAP shows the maximum value of 100.00 across all subjects (p<.001 - p <.01). The same picture can be seen for the PEM and the PAND. The Tau-U also shows statistically significant values (p<.001) which can be interpreted as a large change for Kim (0.69), Tila (0.70) and Niek (0.74) and as a very large change for Lio (0.83), Elif (0.84), Nele (0.88) and Abden (0.89).

Table 3

Overlap Indices for expressive vocabulary

Participant	NAP	p	PEM	PAND	TauU	p
Lio	100.00	<.01	100.00	100.00	0.83	<.001
Kim	100.00	<.01	100.00	100.00	0.69	<.001
Tila	100.00	<.001	100.00	100.00	0.70	<.001
Nele	100.00	<.001	100.00	100.00	0.88	<.001
Niek	100.00	<.001	100.00	100.00	0.74	<.001
Abden	100.00	<.001	100.00	100.00	0.89	<.001
Elif	100.00	<.001	100.00	100.00	0.84	<.001

Note. Non-overlapping of all pairs (NAP); Percentage of data points exceeding the median (PEM); The Percentage of All Non-Overlapping Data (PAND)

The results of the regression analysis show for group 1 a statistically significant slope effect from A phase to B phase (p<.05) with a beta coefficient of 2.464 and thus, an improvement by this value per intervention session. Group 2 shows a statistically significant level effect (p<.01) as well as a slope effect (p<.01) with an improvement of 2.379 per session. For group two a significant level effect (p<.05) and slope effect (p<.001) can also be seen with a beta coefficient of 1.668. As expected, a statistically significant level effect (p<.01) from the A phase to the B phase and a significant slope effect (p<.001) from the A phase to the B phase can be seen. The subjects managed to improve by 2.259 more expressive correctly conscious words per intervention session.

Table 4

Regression model for expressive vocabulary across all participants (Level 2- Analysis)

	B	SE	*t*	*p*
		Group 1		
Intercept	-0.250	3.805	-0.66	.95
Trend	0.200	1.170	0.171	.87
Level	1.697	2.693	0.630	.53
Slope	2.464	1.188	2.075	<.05
		Group 2		
Intercept	2.500	3.005	0.832	.41
Trend	0.300	0.791	0.379	.71
Level	7.231	2.437	2.966	<.01
Slope	2.379	0.814	2.924	<.01
		Group 3		
Intercept	5.311	3.326	1.597	.12
Trend	0.276	0.405	0.681	.50
Level	3.784	1.611	2.349	<.05
Slope	1.668	0.429	3.883	<.001
		Overall		
Intercept	3.456	2.229	1.550	.12
Trend	0.140	0.402	0.349	.73
Level	4.086	1.369	2.985	<.01
Slope	2.259	0.417	5.415	<.001

Letter Sound Fluency (LSF)

Visually, it can be said that the baselines here are not so flat compared to the expressive vocabulary and that positive trends can be partially assumed. Lio (M=21.75), Kim (M=31.00) and Tila (M=21.60) start relatively low and also show no trend tendency in the A phase. Nele (M=46.00), Niek (M=36.17), Abden (M=29.33) and Elif (M=37.83) start with slightly

higher values and show a positive trend tendency. Overall, however, there is also a clear increase for each test person in the B phase.

Figure 4

Letter Sound Fluency (LSF) in one minute

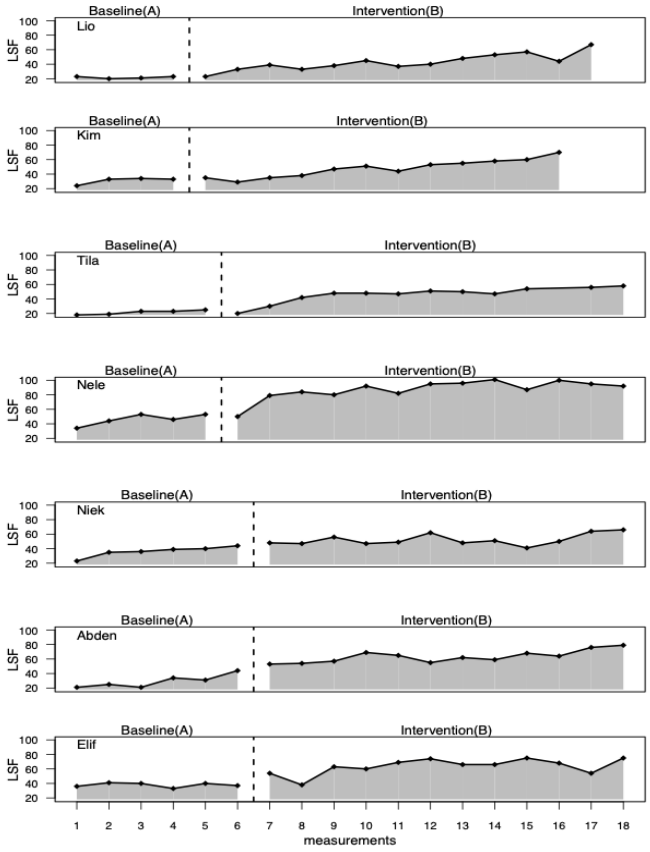


Table 5

Descriptive data for LSF

Participants	N(A)	N(B)	M(A) SD	M(B) SD	MBDi	Md A	Md B	Max A	Max B
Lio	4	14	21.75(1.50)	42.85(11.50)	97,01%	22.00	40.00	23.00	67.00
Kim	4	14	31.00(4.69)	47.92(12.16)	54,58%	33.00	49.00	34.00	70.00
Tila	5	13	21.60(2.97)	45.92(10.90)	112,59%	23.00	48.00	25.00	58.00
Nele	5	13	46.00(7.84)	87.50(13.35)	90,22%	46.00	92.00	53.00	101.00
Niek	6	12	36.17(7.19)	52.42(7.81)	44,93%	37.50	49.50	44.00	66.00
Abden	6	12	29.33(8.91)	63.41(8.46)	116,20%	28.00	63.00	44.00	79.00
Elif	6	12	37.83(3.06)	63.50(10.79)	67,86%	38.50	66.00	41.00	75.00

Note: measurements (N); Mean (M); Standard Deviation (SD); Mean Baseline Difference (MBDi); Median (Md); Maximum (Max); Minimum (min), A Phase (A), B Phase (B)

The overlap measures showed strong effects (94.00-100.00) for all children in the NAP, which were also statistically significant (p<.01-p<.001). The PEM shows a maximum value of 100.00 for Lio, Nele, Niek and Abden and a value of 91.67 for Kim, Tila and Elif. The PAND also shows that the intervention was highly effective for all subjects (91.18-100.00). The Tau-

U, taking into account a possible A phase trend, shows a moderate effect for Niek (0.52, *p*<.01), and a large change for the remaining children (0.62-0.69, *p*<.001).

Table 6

Overlap indices for LSF

Participant	NAP	p	PEM	PAND	TauU	p
Lio	98.00	<.01	100.00	91.18	0.69	<.001
Kim	94.00	<.01	91.67	93.75	0.64	<.001
Tila	95.00	<.01	91.67	95.00	0.62	<.001
Nele	97.00	<.001	100.00	96.92	0.64	<.001
Niek	99.00	<.001	100.00	98.61	0.52	<.01
Abden	100.00	<.001	100.00	100.00	0.67	<.001
Elif	96.00	<.01	91.67	95.83	0.64	<.001

Note. Non-overlapping of all pairs (NAP); Percentage of data points exceeding the median (PEM); The Percentage of All Non-Overlapping Data (PAND)

Regression analysis showed neither a significant level effect (*p* =.50) nor slope effect (*p*=.38) for group 1. The same can be said for group 2. Group 3, on the other hand, shows a statistically significant leveleffect from the A to the B phase (*p*<.05), but also a trend in the A phase (p<.05). Overall, there is a significant leveleffect (*p*<.05) and an A-phase trend (*p*<.01).

Table 7

Regression model for LSF across all participants (Level 2- Analysis)

	B	SE	*t*	*p*
		Group 1		
Intercept	22.750	4.911	4.633	<.001
Trend	1.450	1.529	0.949	.35
Level	-2.410	3.520	-0.685	.50
Slope	1.401	1.552	0.902	.38
		Group 2		
Intercept	25.100	14.713	1.706	.10
Trend	2.900	1.838	1.578	.13
Level	10.240	5.663	1.808	.08
Slope	-0.442	1.890	-0.234	.82
		Group 3		
Intercept	25.664	4.480	5.724	<.001
Trend	2.514	1.068	2.354	<.05
Level	9.532	4.244	2.246	<.05
Slope	-1.050	1.131	-0.928	.36
		Overall		
Intercept	23.614	5.182	4.557	<.001
Trend	2.666	0.807	3.304	<.01
Level	5.668	2.742	2.067	<.05
Slope	-0.470	0.838	-0.561	.58

Reading

Visual inspection shows enormously flat baselines with no positive trends. Significant increases in the B phases can only be found for five children. Lio and Kim initially show no improvement until the end, when there is a discrete increase. Kim (M=0.00) and Lio (M=0.25) start with the lowest values in the A phase and Niek (M=1.33) and Abden (M=0.83) with the highest values. the highest mean values in the B phase are shown by Nele (M=26.62) and Elif (M=20.25) and the lowest values by Lio (M=2.08) and Kim (M=2.67). The largest increase from A to B phase is shown by Nele (2562%) and Elif (2922%) and the least increase can be seen

in Kim (267%) and Lio (732%). Only Nele reaches the maximum value of 40.00 in the B phase. Lio and Kim show the lowest values with a maximum of 7.00 and 8.00.

Figure 5

Amount of correctly read sight words

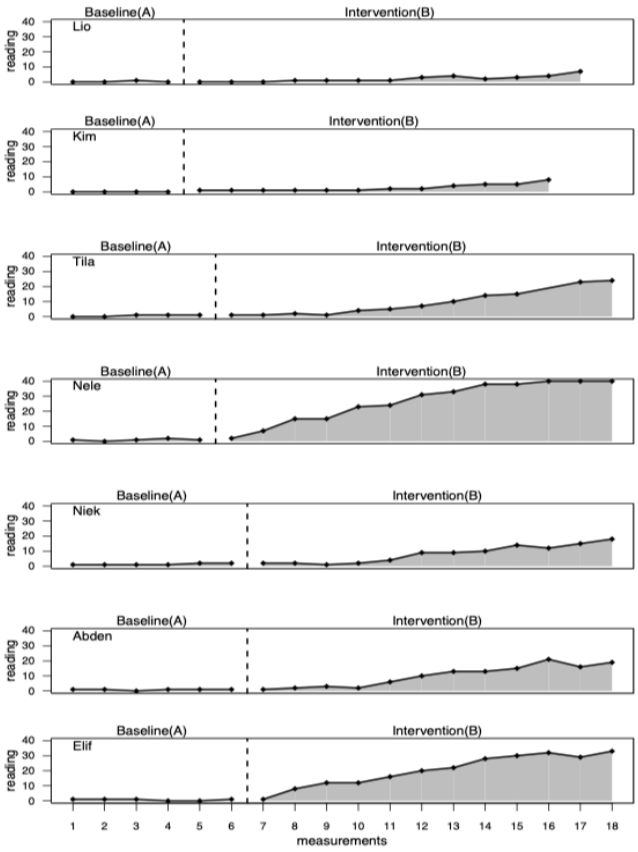


Table 8

Descriptive data for words read correctly

Participants	N(A)	N(B)	M(A) SD	M(B) SD	MBDi	Md A	Md B	Max A	Max B
Lio	4	14	0.25(0.50)	2.08(2.06)	732%	0.00	1.00	1.00	7.00
Kim	4	14	0.00(0.00)	2.67(2.31)	267%	0.00	1.50	0.00	8.00
Tila	5	13	0.60(0.55)	8.92(8.36)	1386,67%	1.00	6.00	1.00	24.00
Nele	5	13	1.00(0.71)	26.62(13.35)	2562%	1.00	9.00	2.00	40.00
Niek	6	12	1.33(0.52)	8.17(5.87)	514,29%	1.00	9.00	2.00	18.00
Abden	6	12	0.83(0.41)	10.08(7.10)	1114,46%	1.00	11.50	1.00	21.00
Elif	6	12	0.67(0.52)	20.25(10.49)	2922,39%	1.00	21.00	1.00	33.00

Note: measurements (N); Mean (M); Standard Deviation (SD); Mean Baseline Difference (MBDi); Median (Md); Maximum (Max); Minimum (min), A Phase (A), B Phase (B)

Further, the overlap measures for the NAP show a medium effect for Lio (82.00, p<.05), Niek (90.00, p<.01), and Tila (92.00, p<.01) and a strong effect for Abden (97.00, p<.001), Elif (97.00, p<.001), Nele (99.00, p<.011), and Kim (100.00, p<.01). The PAND testifies medium effects for all except Nele and Kim, who show strong effects. A similar picture emerges for the

PEM. The Tau-U shows a large change for Lio (0.61, *p*<.001), Kim (0.63, *p*<.001), Tila (0.69, *p*<.01) and Niek (0.69, *p*<.001). Abden (0.81, *p*<.001), Elif (0.87, *p*<.001) and Nele (0.88, *p*<.001) show a large to very large change.

Table 9

Overlap Indices for correctly words read correctly

Participant	NAP	p	PEM	PAND	TauU	p
Lio	82.00	<.05	76.92	70.59	0.61	<.001
Kim	100.00	<.01	100.00	100.00	0.63	<.001
Tila	92.00	<.01	75.00	82.35	0.69	<.001
Nele	99.00	<.001	100.00	94.44	0.88	<.001
Niek	90.00	<.01	91.67	75.00	0.69	<.001
Abden	97.00	<.001	91.67	83.33	0.81	<.001
Elif	97.00	<.001	91.67	86.11	0.87	<.001

Note. Non-overlapping of all pairs (NAP); Percentage of data points exceeding the median (PEM); The Percentage of All Non-Overlapping Data (PAND)

The results of the regression analysis at level 2 show no statistically significant level (*p*=.11) or slope effect (*p*=.18) for group 1. Group 2 shows a statistically significant slope from A to B phase (*p*<.05) with an increase of 2,503 correct words per intervention session. Group three shows a very similar picture (slope; B=2.502, *p*<.05). Overall, a significant slope effect can be observed with a beta coefficient of 1.224 (*p*<.05).

Table 10

Regression model for words read correctly across all participants (Level 2- Analysis)

	B	SE	*t*	*p*
		Group 1		
Intercept	0.000	0.892	0.000	1.00
Trend	0.050	0.323	0.155	.88
Level	-1.226	0.743	-1.650	.11
Slope	0.451	0.328	1.378	.18
		Group 2		
Intercept	0.050	5.937	0.008	.99
Trend	0.250	1.183	0.211	.83
Level	-1.226	0.743	-1.650	.54
Slope	2.503	1.217	2.057	<.05
		Group 3		
Intercept	0.778	2.819	0.276	.78
Trend	0.250	1.183	0.211	.93
Level	-1.836	2.012	-0.913	.54
Slope	2.503	1.217	2.057	<.05
		Overall		
Intercept	-1.603	2.911	-0.551	.58
Trend	0.608	0.558	1.091	.28
Level	-2.467	1.903	-1.296	.20
Slope	1.224	0.579	2.114	<.05

Social Validity

In terms of social validity, all participants have a very positive attitude towards the intervention overall. With regard to word reading, only Lio and Kim stated: "partly agree". Overall, "completely agree" dominates on all items. The children found that the storytelling helped them, they understood the meaning of the promotion and would like to do more storytelling in school. The students also liked the self-graphing. Only Niek rated: "partly agree".

Table 11

Results of social validity questionnaire

Items	Lio	Kim	Tila	Nele	Niek	Abden	Elif
Storytelling has helped me to be able to read words correctly	2	2	3	4	4	4	4
Storytelling helped me learn words and their meanings.	3	4	3	4	4	3	4
Storytelling helped me to pronounce sounds correctly.	4	4	4	3	4	4	4
I understood well the meaning of the promotion.	4	4	4	4	4	3	3
I have learned a lot during storytelling.	3	3	4	4	4	4	4
I gladly came to the intervention sessions.	4	4	4	4	4	4	4
The self-graphing sheets were fun.	3	3	4	4	2	4	4
The stories were great	4	4	3	4	4	4	3
I would like to do more with stories in school.	3	4	4	4	4	4	3

Note. 0= completely not agree; 1= not agree; 2= partly agree; 3= agree; 4= completely agree

Discussion

Main findings

The study presented was designed to estimate the effects of a storytelling intervention on the variables: vocabulary, LSF, and sight word reading in students with German as a second language with and without problem behavior. The background is the increasing number of students with GL2 and at the same time the increase of students with German as a second language and weak school performance especially in the area of reading. L2 students are educationally disadvantaged due to their deficits in the language. It is of particular importance to teach these students the language adequately in a motivating way.

Overall, the results are consistent with findings that have looked at multicomponent intervention (Donegan & Wanzek, 2021; Foorman & Torgesen, 2001) and the dual coding theory (Paivio, 2008) which states that using verbal and non-verbal system of process information is highly effective in order to finally store information. Moreover, the findings are also consistent with the meta-analysis by Maurulis and Neuman (2010) that conveying knowledge explicitly and implicitly in combination leads to the highest effects. Looking at the effectiveness on vocabulary acquisition, it can be seen that all subjects show an immense increase in the B phase, with all baselines being relatively flat. Niek, Kim and Tila show the weakest effects, although even these can be classified as large. Kim is by far the weakest in the vocabulary pretest with a PR of 5. For her, this may be due to the fact that she has great problems building vocabulary overall. In contrast, Tila and Niek perform better in the vocabulary pre-test, but unlike Kim, they have greater problems in learning-related behavior and the highest problem scores overall in the group. Particularly, problems in attention processing might be a reason here as describe in the literature (Peterson et al., 2013). Abden and Nele are among the strongest performers in terms of vocabulary, but both also show the

best results in the vocabulary pretest. It might be easier for them to learn new words if their overall vocabulary is already larger. While Abden has problems with learning-related behavior, which does not seem to play a major role here, Nele shows no problems in this regard. The results of vocabulary acquisition are consistent with the findings of Barwasser et al. (2020), Knaak et al. (2021) and Barwasser et al. (2021).

Furthermore, for the second dependent variable LSF, the baseline results are higher, i.e., some children have already had experience with German letter sounds, while others show a flat baseline with lower values. Niek is the weakest and Lio as well as Abden the strongest. Niek shows by far the weakest results in the pretest in the area of sound categorization, which could be a reason for his problems in the area of LSF. Overall, Abden is also one of the weakest students in the phonological awareness pretest but sound categorization is his best sub-category with a PR of 28. Like Abden, Lio also has problems in learning behavior which also does not seem to play a major role. However, overall results show that the intervention does have a positive impact on LSF which is an important indication since Hulme et al. (2012) has shown that problems in LSF is related to later word-reading difficulties which is referred to almost the same age as the participants of the current study.

With regard to sight word reading, the overall performance is weaker, especially for Lio and Kim. Except for Nele and Elif, the others seem to take longer to automate the words. One explanation for this could be that weaker readers often take the non-lexical route because they also have greater problems with the lexical route (De Jong et al., 2012). Thus, the children try to decode the words each time instead of storing them as a whole, for which the 1 second in the measurement is not sufficient. Thus, for these children it takes a longer time until they seem to change the route. Nele and Elif both have much higher scores in phonological awareness and also in pseudoword reading, which should make it easier for them to memorize the words as a whole more quickly, as they are better readers. In reading, they are among the strongest of the subjects in the pre-tests, along with Abden, who scores third best in effect sizes. Elif, like Nele, shows no problems in learning-related behavior. Lio and Kim are among the weakest subjects overall in terms of reading and phonological awareness. Perhaps the Polish L1 also plays a role because L1 background can influence L2 word recognition (Wang & Koda, 2007). According to Catts (1993) phonological awareness is more closely related to word recognition than measures of vocabulary in young 1st grade children with phonological difficulties and Lio as well as Kim perform poorly in both areas. Another explanation could be that Lio and Kim might have problems in rapid automatized naming, which is important with regard to naming speed and the retrieval of sight words from the mental lexicon, especially in the German language (Huschka et al., 2021; Landerl & Wimmer, 2008). Nevertheless, Abden, Niek and Tila also display severe problems in phonological awareness and need longer time to respond to the intervention in word recognition. Niek and Abden have better reading

performance in the pretest while Tila performs similarly weak in the pretest as Lio and Kim. So, what could be the reason? In the case of Tila, it could actually be the learning-related behavior that causes problems, or frustration, while in the case of Niek and Abden, the behavioral problems do not seem to have such an impact. One reason could be the overall better reading performance of the two students, which counteracts the problem behavior.

Overall, the intervention seems to work really well for one variable and well for the other two. Storytelling seems to also has an effect on the reading of sight words and goes partly in line by meta-analytic finding by Roberts et al. (2020) who were focused on foundational reading instructions for students with problem behavior in grades K-12 (g=0.86) as well as small group reading instruction for grade 1-4 (e.g., Scammacca et al. 2015). With regard to students who struggled with sight word reading, one can see that even with those with a slow increase, the increase seems to come after some time. Another assumption could be that the intervention should have been prolonged in order to achieve greater effects. Also, behavior might have played a role in some cases in combination with very low score in the pre-testings. Reflecting on the importance of motivation, especially in language acquisition, self-graphing probably contributed in part to the effects, as studies have pointed to the effectiveness of self-graphing in intervention and especially in reading interventions (Bowman-Perrott et al., 2013; Guzman et al., 2018; Leko, 2016; McKenna & Bettini, 2018; Stotz et al., 2008;). Especially regarding the social validity results where all children despite Niek, who seemed to be unsure, rated self-graphing as positive. Moreover, the results of the social validity questionnaire revealed that all participants rated the interventions as positive. With regard to reading, Lia and Kim gave worse scores than the others, but this is also understandable, since both could hardly benefit in sight word reading, also compared to the others.

It is also noticeable that the language background does not necessarily play a role. The Polish background is only noticeable when improving the visual vocabulary, but this does not necessarily mean anything. The sample is much too small to be able to make statements about this. Also, problem behavior did not seem to play a role across the board. This may be due to the fact that the children were taken out of the classroom and trained intensively in a small group. In general, small group interventions, especially with regard to reading have been shown in a meta-analysis by Hall & Burns (2018) to achieve a large effect size for elementary students (g = 0.64) (also see Nielsen & Friesen, 2012) which can be also referred to Roberts et al. (2020) who examined the effects for primary school students with behavioral problems in a meta-analysis.

Limitations

In addition to the promising results, there are some imitations: First, the intervention took place during the COVID-19 pandemic, where everyone in the school had to abide by

specific rules and it was generally unruly in the school. Groups were therefore not allowed to be mixed from different classes. With regard to reading, it can be seen that those with very weak performance at the phonological level also have greater problems storing the words as sight words. Here it would probably make sense to stay one level lower and train the LSF and other aspects of decoding more intensively. Furthermore, this is a multiple baseline study, which means that we focused on individual students, making it difficult to generalize the results. Nevertheless, the results give important indications with regard to the support of struggling students with GL2 with and without behavioral difficulties. The advantage of a multiple baseline study is that it allows us to see individual learning trajectories and to find out specifically how the intervention is received by different students.

Another limitation is that there is a certain probability that the children have also become better through the repeated measurements each time after the sessions. We have tried to counteract this by randomizing the order of the items in each test, but we cannot exclude it for sure. However, since there are no trends in the baselines where only testing was done, it could be argued that the influence of testing was not too great. A further minor limitation is the measurement time point of the first group in the baseline, since across the board at least 5 measurement time points are always recommended in each phase. after Kratochwill et al. (2013), however, at least three measurement time points are also sufficient to be able to make a statement. Due to time constraints, it was not possible to extend the baseline. And, as with all multi-component interventions, of course, one does not know which component worked for which parts. At the current time, it is not possible to say exactly to which parts the various components (such as self-graphing and implicit vs. explicit teaching) have had on the dependent variables. Since this intervention seems to work in this package, it is basically not the intention to examine the individual parts separately, as the package is very easy and straightforward to implement in the classroom.

Implications

A first goal would be to estimate the storytelling intervention on a larger sample and make generalized statements. Furthermore, the intervention would be compared to other interventions in order to see which support option seem to be most effective in the area of language acquisition. In the course of this, one could also look at whether the method also works with a whole class or if it is limited to small groups. In the context of digitalization and especially the current school closures worldwide due to the COVID-19 pandemic, which has once again shown how important digital learning is in schools, the storytelling intervention could be digitalized and made available via apps or web-based tools.

The intervention in its current form was rated very positively, which gives us an indication that despite the overall good effects on all three dependent variables, the

intervention is accepted across all participants. The further implication of this is to continue to conduct the social validity survey in future research to gain more insight into the overall acceptability of the intervention, which according to Briesch et al. (2013) is a necessity in intervention research. Last implications are the different languages and behavioral problems. It would be interesting to see whether the effects differ between children from different language backgrounds (Wang & Koda 2007). In addition, one could also record the abilities in the surveyed variables in the L1 in order to identify possible correlations here. Furthermore, the study looked at children with learning-related behavior problems. A continuation would be to see if the intervention would also help with students with disruptive behavior, which is a big challenge for teachers today (Rosenberg & Jackman, 2003). Also, measuring rapid automatized naming beforehand would be interesting since it is linked to rapid word retrieval and reading, particularly in the German language which is more transparent than e.g., English (Landerl & Wimmer, 2008).

Conclusion

It is enormously important to support struggling language learners in all components of a language in order to provide equal chances with respect to school and later job possibilities, especially to actively address the results of the PISA survey (OECD, 2019). Also, Morgan et al. (2008) showed that first graders with reading problems are more likely to show off task-behavior and general problem behavior in grade 3. Also considering the meta-analysis by Chow and Wehby (2018) on the negative relationship of language problems and behavioral difficulties, it is imperative to counteract this, particularly when students already display some kind of problem behavior. Also, one should consider the Matthew effect that stronger readers become stronger and weaker readers become weaker particularly in the first years of school because they start to dislike reading (Stanovich, 1986). Thus, early prevention in school failure is really important, specifically for students with GL2 and those with additional problem behavior who struggle with reading. This storytelling approach should give teachers, educators and researchers an indication of how an intervention in this area could look like which can train different areas of language at the same time and matches the concept of inclusion by Booth and Ainscow (2011) to integrate students with different competencies and characteristics as well as from different backgrounds.

References

Aschenbrenner, K.-H., Bauer, W., Gehring, S., Hazzouri, N., Holdorf, K., Jeuk, S., Junk-Deppenmeier, A., Maier, H., Mietzschke, G., Papesch, E., Stein, N., & Zillig, A. (2016). *Deutsch als Zweitsprache in der Grundschule* (2nd edition). Landesinstitut für Schulentwicklung (LS). https://www.ph-ludwigsburg.de/fileadmin/subsites/2b-spze-t-01/user_files/Service/Deutsch_als_Zweitsprache.pdf

Baker, S. K., Santoro, L. E., Chard, D. J., Fien, H., Park, Y., & Otterstedt, J. (2013). An evaluation of an explicit read aloud intervention taught in whole-classroom formats in first grade. *The Elementary School Journal*, *113*(3), 331-358.

Baker, D.L., Santoro, L., Biancarosa, G., Baker, S.K., Fien, H., & Otterstedt, J. (2020). Effects of a read aloud intervention on first grade student vocabulary, listening comprehension, and language proficiency. *Read and Writing, 33,* 2697–2724. https://doi.org/10.1007/s11145-020-10060-2

Barwasser, A., Knaak, T., & Grünke, M. (2020). The effects of a multicomponent storytelling intervention on the vocabulary recognition of struggling English as a foreign language learners with learning disabilities. *Insight into Learning Disabilities*, *17*(1), 35–53.

Barwasser, A., Lenz, B., & Grünke, M. (2021). A multimodal storytelling intervention for improving the reading and vocabulary skills of struggling German as a Second Language adolescents with learning and behavioral problems. *Insights Into Learning Disabilities, 18*(1), 29-51.

Becker-Mrotzek, M., Hentschel, B., Hippmann, K., & Linnemann, M. (2012). *Sprachförderung in deutschen Schulen – die Sicht der Lehrerinnen und Lehrer Ergebnisse einer Umfrage unter Lehrerinnen und Lehrern.* Mercator-Instituts für Sprachförderung und Deutsch als Zweitsprache, Universität zu Köln. https://www.mercator-institut-sprachfoerderung.de/fileadmin/user_upload/Lehrerumfrage_Langfassung_final_30_05_03.pdf

Booth, T., & Ainscow, M. (2011) *Index for inclusion developing learning and participation in schools*. Centre for Studies on Inclusive Education.

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, 39-55.

Brewster, J., Ellis, G., & Girard, D. (2002) The Primary English Teacher's Guide. New Edition. England: Pearson Education.

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.

Catts, H. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research, 36*, 948–958.

Choo, L. B., Lin, D. T., & Pandian, A. (2012). Language learning approaches: A review of research on explicit and implicit learning in vocabulary acquisition. *Procedia - Social and Behavioral Sciences*, *55*, 852–860.

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. http://dx.doi.org/10.1007/ s10648-016-9385-z

Clemens, N.H., Lai, M.H.C., Burke, M., & Wu, J.Y. (2017) Interrelations of growth in letter naming and sound fluency in kindergarten and implications for subsequent reading fluency. *School Psychology Review, 46,* 272-287.

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

Curtis, P. R., Frey, J. R., Watson, C. D., Hampton, L. H., & Roberts, M. Y. (2018). Language disorders and problem behaviors: A meta-analysis. *Pediatrics, 142*(2), e20173551. https://doi.org/10.1542/peds.2017-3551

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

DiGangi, S. A., Maag, J. W., & Rutherford, R. B. (1991). Self-graphing of on-task behavior: Enhancing the reactive effects of self-monitoring on on-task behavior and academic performance. *Learning Disability Quarterly, 14*(3), 221–230. https://doi.org/10.2307/1510851

Donegan, E. E., & Wanzek, J. (2021). Effects of reading interventions implemented for upper elementary struggling readers. A look at recent research. *Reading & Writing*, *35*(2), 62-71. https://doi.org/10.1007/s11145-021-10123-y

Ehri, L. C. (2002). Phases of acquisition in learning to read words and implications for teaching. *British Journal of Educational Psychology: Monograph Series, 1,* 7–28.

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

Ender, A. (2016). and explicit cognitive processes in incidental vocabulary acquisition. *Applied Linguistics*, *37*(4), 536-560. https://doi.org/10.1093/applin/amu051

Federal Statistical Office. (2020, October 20). *Ausländische Schüler/-innen nach Schularten*. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Schulen/Tabellen/allgemeinbildende-berufliche-auslaendische-schueler.html;jsessionid=2E832BA6153E455C10E6655AEF8A9F19.internet711?nn=209624

Fien, H., Santoro, L., Baker, S. K., Park, Y., Chard, D. J., Williams, S., & Haria, P. (2011). Enhancing teacher read alouds with small-group vocabulary instruction for students with low vocabulary in first-grade classrooms. *School Psychology Review*, *40*(2), 307-318.

Foorman, B. R., & Torgesen, J. (2001). Critical elements of class- room and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice*, *16*(4), 203–212. https://doi.org/10.1111/0938-8982.00020

Ghenghesh, P. (2010). The Motivation of L2 Learners: Does It Decrease with Age? *English Language Teaching, 3*(1), 128-141.

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), 160–168. https://doi.org/10.1037/spq0000199

Hall, M., & Burns, M. K. (2018). Meta-analysis of targeted small-group reading interventions. *J. Sch. Psychol. 66*, 54–66. https://doi.org/10.1016/j.jsp.2017.11.002

Henrichs, J., Rescorla, L., Donkersloot, C. et al. (2013). Early vocabulary delay and behavioral/emotional problems in early childhood: The generation R study. *Journal of Speech, Language, and Hearing Research, 56*, 553–566. https://doi.org/10.1044/1092-4388(2012/11-0169)

Huang, X., Kim, N., & Christianson, K. (2019). Gesture and vocabulary learning in a second language. *Language Learning*, *69*(1), 177–197. https://doi.org/10.1111/lang.12326

Hulme, C., Bowyer-Crane, C., Carroll, J. M.., Duff, F. H., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read. Combining intervention studies with mediation analyses. *Psychological Science, 23*(6), 572-577. https://doi.org/10.1177/0956797611435921

Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A. et al. (Eds.). IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich; Waxmann: Münster, Germany; New York, NY, USA, 2017.

Huschka, S.S., Georgiou, G.K., Brandenburg, J. *et al.* Examining the contribution of RAN components to reading fluency, reading comprehension, and spelling in German. *Read Writ* (2021). https://doi.org/10.1007/s11145-021-10145-6

Isbell, R., Sobol, J., Lindauer, L., & Lowrance, A. (2004). The effects of storytelling and story reading on the oral language complexity and story comprehension of young children. *Early Childhood Education Journal, 32*(3), 157-163. https://doi.org/10.1023/B:ECEJ.0000048967.94189.a3

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*, January. https://doi.org/10.1177/2396941519900076

Jin, Z., & Webb, S. (2020). Incidental vocabulary learning through listening to teacher talk. *The Modern Language Journal, 104*(3), 550-566. https://doi.org/10.1111/modl.12661

Kim, Y-S. G., & Piper, B. (2019). Cross-language transfer of reading skills: an empirical investigation of bidirectionality and the influence of instructional environments. *Reading and Writing*, *32*, 839–871. https://doi.org/10.1007/s11145-018-9889-7

Kitzinger, A. (2020). *METACOM. Symbolsystem zu Unterstützen Kommunikation*. Available: https://www.metacom-symbole.de [05.02.2021].

Knaak, T., Grünke, M., & Barwasser, A. (2021). Enhancing vocabulary recognition in English Foreign Learners with and without learning disability. The effects of a multi component storytelling intervention approach. *Learning Disabilities. A Contemporary Journal*, *19*(1).

Kramer, B., & McLean, S. (2019). L2 reading rate and word length. The necessity of character-based-measurement. *Reading in a Foreign Language, 31*(2), 201-225.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*(1), 26-38. http://doi.org/10.1177/0741932512452794

Lane, J. D., Ledford, J. R., & Gast, D. L. (2017). Single-case experimental design: Current standards and applications in occupational therapy. *American Journal of Occupational Therapy, 71*(2), 7102300010. https://doi.org/ 10.5014/ajot.2017.022210

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, 150–161.

Laurice, J. M., & Elisha, L. (2011). A review of the effects of self-monitoring on reading performance of students with disabilities. *Journal of Special Education, 45,* 43-53. https://doi.org/10.1177/0022466909349145

Leko, M. M. (2016). Word study in the inclusive secondary classroom. Teacher College Press.

Lenhart, J., Lenhard, W., Vaahtoranta, E., and Suggate, S. (2018). Incidental vocabulary acquisition from listening to stories: a comparison between readaloud and free storytelling approaches. *Educational Psychology, 38*(5), 596–616. https://doi.org/10.1080/01443410.2017.1363377

Lervåg, A., & Aukrust, V.G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry, 51*(5), 612–620. https://doi.org/10.1111/j.1469-7610.2009.02185.x.

Ma, H. H. (2006). An alternative method for quantitative synthesis of single subject researches: percentage of data points exceeding the median. Behav. Modif. 30, 598–617. doi: 10.1177/0145445504272974

Manolov, R., Arnau, J., Solanas, A., & Bon, R. (2010). Regression-based techniques for statistical decision making in single-case designs. *Psicothema, 22*(4), 1026–1032.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300–335. https://doi.org/10.3102/0034654310377087

Mayer, R. E. (2009). Multimedia learning (2nd ed.). New York, NY: Cambridge University Press.

McClelland, M., Acock, A., & Morrison, F. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*(4), 471–490. https://doi.org/10.1016/j.ecresq.2006.09.003

McKenna, J.W., & Bettini, E. (2018). Improving reading fluency skills for secondary students with emotional and behavioral disorders. *Beyond Behavior*, *27*(2), 74-81. doi:10.1177/1074295618779374

Mello, R. (2001). Building bridges: How storytelling influence teachers and student relationships. *The Journal of Higher Education, 17*(6), 301-307.

Moll, K., & Landerl, K. (2010). *SLRT-II: Lese-und Rechtschreibtest; Weiterentwicklung des Salzburger Lese-und Rechtschreibtests (SLRT) [Reading and spelling test; further development of the Salzburg reading and spelling test]*. Huber.

Morgan, P. L., Farkas, G., Tufis, P.A., & Sperling, R. A. (2015). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities, 41*(5), 417-436. https://doi.org/10.1177/0022219408321123

Nielsen, D. C., & Friesen, L. D. (2012). A study of the effectiveness of a small-group intervention on the vocabulary and narrative development of at-risk kindergarten children. *Reading Psychology, 33*, 269–299.

O'Brien, S., & Repp, A. C. (1990). Reinforcement-based reductive procedures: A review of 20 years of their use with persons with severe or profound retardation. *Journal of the Association for Persons With Severe Handicaps, 15*, 148–159.

OECD. (2019). Programme for International Student Assessment (PISA) results from PISA 2018. https://www.oecd.org/pisa/publications/PISA2018_CN_DEU.pdf

Paivio, A. (2008). The dual coding theory. In S. B. Neuman (Ed.), *Educating the other America* (pp. 227–242). Paul H. Brookes Publishing Company.

Paivio, A. (2007). Mind and its Evolution: A Dual Coding Theoretical Approach. Mahwah, NJ: Erlbaum.

Parker, R. I., Hagan-Burke, S., and Vannest, K. (2007). Percent of all nonoverlapping data (PAND): an alternative to PND. J. Spec. Educ. 40, 194–204. doi:10.1177/00224669070400040101

Parker, R. I., Vannest, K. J., and Davis, J. L. (2011a). Effect size in single case research: a review of nine non-overlap techniques. Behav. Modif. 35, 302–322. doi: 10.1177/0145445511399147

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

Peterson, I. T., Bates, J. E., D'Onofrio, B.M., Coyne, C.A., Lansford, J.E., Dodge, K.A., Pettit, G.S., & Van Hulle, C.A. (2013). *Journal of Abnormal Psychology, 122*(2), 542-557. https://doi.org/10.1037/a0031963

Raudenbush, S. W. Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Reed, S.K. (2010). Cognition: Theories and Applications (8th ed). Wadsworth, Cengage Learning: Bellmontt, CA, USA.

Roberts, G.J., Cho, E., Garwood, J.D. *et al.* (2020). Reading interventions for students with reading and behavioral difficulties: a meta-analysis and evaluation of co-occurring difficulties. *Educational Psychology Rev*iew, *32*, 17–47. https://doi.org/10.1007/s10648-019-09485-1

Rosenberg, M.S., & Jackman, L.A. (2003). Development, implementation, and sustainability of comprehensive school-wide behavior management systems. *Intervention in School and Clinic*, *39*(1), 10-21. https://doi.org/10.1177/10534512030390010201

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4-12: 1980-2011. *Journal of Learning Disabilities, 48*(4), 369–390. https://doi. org/10.1177/0022219413504995.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95,* 26-43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913–951. https://doi.org/10.1111/lang.12077

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., and Kliegl, R. (2015). childLex–Eine lexikalische Datenbank zur Schriftsprache für Kinder im Deutschen [childLex-A lexical database of written language for children in German]. Psychol. *Rundsch. 66*, 155–165. https://doi.org/10.1026/0033-3042/a0 00275

Stanovich, K. (1986). Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. *Reading Research Quarterly, 11*(4), 360-407

Stock, C., Marx, P., & Schneider, W. (2017). Basiskompetenzen für Lese-Rechtschreibleistungen. Din Test zur Erfassung der phonologischen Bewusstheit vom ersten bis vierten Grundschuljahr. Hogrefe: Göttingen

Stotz, K. E., Itoi, M., Konrad, M., & Alber-Morgan, S. R. (2008). Effects of self-graphing on written expression of fourth grade students with high-incidence disabilities. *Journal of Behavioral Education, 17*(2), 172-186. https://doi.org/10.1007/s10864-007-9055-9

Suggate, S. P., Lenhard, W., Neudecker, E., & Schneider, W. (2013). Incidental vocabulary acquisition from stories: Second and fourth graders learn more from listening than reading. *First Language*, *33*(6), 551–571. https://doi.org/10.1177/0142723713503144

Swanson, E., Wanzek, J., Petscher, Y., Vaughn, S., Heckert, J., Cavanaugh, C., et al. (2011). A synthesis of read-aloud interventions on early reading outcomes among preschool through third graders at risk for reading difficulties. *Journal of Learning Disabilities, 44*(3), 258–275. https://doi.org/10.1177/0022219410 378444.

Tate, R. L. et al. (2016). The Single-Case Reporting guideline In BEhavioural interventions (SCRIBE) 2016 statement. *Journal of School Psychology, 56*(1), 133-142.

Volpe, R. J., Casale, G., Mohiyeddini, C., Grosche, M., Hennemann, T., Briesch, A., & 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

Waddell, D. E., Nassar, S. L., & Gustafson, S. (2011). Single-case design in psychophysiological research: Part II - statistical analytic approaches. *Journal of Neurotherapy, 15*(2), 160-169.

Wanzek, J., Wexler, J., Vaughn, S., & Ciullo, S. (2010). Reading interventions for struggling readers in the upper elementary grades: A synthesis of 20 years of research. *Reading and Writing*, *23*(8), 889–912. https://doi.org/10.1007/s11145-009-9179-5

Wang, M., & Koda, K. (2007). Commonalities and differences in word identification skills among learners of English as a second language. *Language Learning, 55*(1), 201–222. https://doi.org/10.1111/j.0023-8333.2005.00290.x

Wallace, C. (2014). *The Cambridge guide to teaching English to speakers of other languages*. Cambridge, UK: Cambridge University Press.

Wajnryb, R. (2003). Stories. Narrative Activities for the Language Classroom. Cambridge: Cambridge University Press. http://assets.cambridge.org/97805210/01601/sample/978052 1001601ws.pdf.

Webb, S. A., & Chang, A. C-S. (2012). Second language vocabulary growth. *RELC Journal*, *43*(1), 113–126. https://doi.org/10.1177/0033688212439367

Webb, S., & Nation, P. (2017). How vocabulary is learned. Oxford: Oxford University Press.

Weiß, R. H. (2007). WS/ZF-R. Wortschatztest und Zahlenfolgentest – Revision. Ergänzungstest zum CFT 20-R. Manual. Göttingen: Hogrefe.

Wilbert, J. (2021). Analyzing single-case data with R and SCAN. Retrieved from https://jazznbass.github.io/scan-Book/.

Wise, J. C., Sevcik, R.A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. *Journal of Speech, Language, and Hearing Research, 50*(4), 1093–1109. https://doi.org/10.1044/1092-4388(2007/076)

Wright, A. (2013). RBT: Creating Stories With Children. Oxford University .

Appendix G Article 7

Knaak, T., Grünke, M., & Barwasser, A. (2021). Enhancing vocabulary recognition in English foreign learners with and without learning disabilities: effects of a multi component storytelling intervention approach. *Learning Disabilities: A Contemporary Journal, 19*(1), 5–23.

Abstract

The English language plays a major role around the world, making it important to learn English in order to participate and communicate in our globalized age. Adequate foreign language skills are important for everyday life and can even enhance performance in one's first language (L1). A growing number of very heterogeneous classrooms make it necessary to develop strategies that are beneficial for both high and low achievers. The purpose of this single-case study was to evaluate the effects of a multicomponent intervention consisting of storytelling, flashcards, and a reward procedure on 24 secondary-level students with and without learning disabilities. It appears that all participants benefited from the intervention and improved their sight word vocabulary knowledge. Students diagnosed with a learning disability showed the greatest improvements based on the results of visual analysis, effect size, and a piecewise regression analysis. In addition, follow-up data collected three weeks after the intervention showed that the effects maintained at a very high level.

Keywords: storytelling, flashcards, group rewards, vocabulary recognition, English as a foreign language

Introduction

Importance of English as a Foreign Language

Foreign language (L2) learning is an important aspect of education today. Almost two thirds of the world's population are able to speak two or more languages (Crystal, 2006). It is especially essential to acquire English, one of the most widely taught foreign languages worldwide (Crystal, 2003). Thus, English skills can enhance success in school and the chance of future employment (Reddy, 2016); in addition, it can increase social participation, since many aspects of society even in non-English speaking countries are influenced by English terms such as *social media, internet, news, advertisement, fashion*, and so on. In sum, to give everyone the opportunity to participate fully in a multilingual society, foreign language learning is an important aspect of education for every individual regardless of the degree of their academic abilities (United Nations, 2006).

For students whose literacy performance in their mother tongue lacks behind that of their peers, acquiring literacy in English as a foreign language is especially challenging (Fraser,

2007; Ganschow & Sparks, 2000; Kahn-Horwitz et al., 2006; Romonath et al., 2005). Nevertheless, numerous studies suggest that struggling children and youth can make great improvements in learning an L2 when provided with the right support (Sparks, 2006, 2016) while not being prevented from being successful in other subjects by learning a second language (Genesee, 2007; Kay-Raining Bird et al., 2012; Kohnert et al., 2005; Paradis, 2007). Further, L2 learning can increase motivation, participation and performance in students' first language (L1) due to positive cross-linguistic links between L1 and L2 (Erdos et al., 2014; Sparks, 2009; Sparks et al., 2008).

Due to the increasingly heterogeneous nature of classrooms today, teachers need methods that meet the needs of both high and low achievers when it comes to acquiring new language skills (Leons et al., 2009; McColl, 2005).

Vocabulary Acquisition

Overall performance in school depends heavily on the ability to retrieve specific information from long-term memory (Wolgemuth et al., 2008). Similarly, a main aspect of learning a foreign language involves the ability to recall vocabulary from long-term memory, the so-called *mental lexicon*. This area poses a particular challenge for many low achievers (Amiryousefi & Ketabi, 2011; Clark & Paivio, 1991; Simon, 2000). Therefore, sight word acquisition and repetitive practice are very important for relieving the demands on working memory and helping struggling students to store vocabulary and, as a result, lay the foundation for communicating in a foreign language (Coady & Huckin, 1997; Grabe, 2004; Morra & Camba, 2009; Schmitt, 2010). In particular, results of current research suggest that frequency of encounter (repetition) and additional support through, for example, visual or verbal mnemonics can facilitate vocabulary learning (Amiryousefi & Ketabi, 2011; Ramezanali et al., 2020; Uchihara et al., 2019).

According to the Dual Coding Theory (DCT) (Paivio, 1991), gives advice what effective instruction should focus on. This theory assumes that multiple modalities help students to memorize information more easily. For example, when information is presented in a verbal, visual, and gestural way, the likelihood of the information being remembered increases (Paivio & Lambert, 1981). In the case of vocabulary learning, the use of visual, verbal, and gestural connections could facilitate memorization, as the inclusion of multiple modalities seems to increase the chances of remembering new words (Paivio & Lambert, 1981).

Storytelling

Storytelling offers one way of meeting the aforementioned requirements of effective learning (verbalization, gestural support) and, therefore, can activate and motivate learners as teachers tell a story in a very active and communicative way (Roney, 1996). The story is presented visually so that learners can read along while listening. In storytelling, new words

can be conveyed while embedding them in a meaningful context, and can bring teacher and students into a communicative circle (Cameron, 2001; Ellis & Brewster, 2002; Roney, 1996). Through communicative interactions between teacher and students, motivation as well as language acquisition can be enhanced (Ellis & Brewster, 2002).

Furthermore, storytelling provides a relevant and motivating context that also appears to be effective when teaching vocabulary (Joshi, 2005). Thus, when new vocabulary is embedded into a meaningful context, it is more likely to be stored (Amiryousefi & Ketabi, 2011; Dong, 2013; Leons et al., 2009; Nation, 2015). If meaningful context is also highly connected to the learners' interests, it is beneficial for remembering vocabulary, especially over time (Oxford & Scarcella, 1994; Van, 2009).

To date, most studies that have examined storytelling in L2 learning have focused on improving reading comprehension and general language skills (Al-Mansour & Al-Shorman. 2011; Hemmati et al., 2015; Huang, 2006; Kim, 2010), with results suggesting great benefits in comprehension and general skills. However, no study has evaluated the effects of storytelling on vocabulary recognition of English L2 learners, nor have any studies focused on students in inclusive settings. Barwasser et al. (2020) have previously evaluated the use of storytelling with students with learning disabilities and found multicomponent storytelling as a beneficial method for this group of students.

Flashcards

According to the DCT, unknown words can be stored in the mental lexicon more efficiently when introduced with visual support (Schmitt, 2010; Thompson, 1987; Uchihara et al., 2019). By adding flashcards to a storytelling intervention, the requirement of the visual modality can be fulfilled. This method has proven to have a positive impact on sight word acquisition in different subjects (Crowley et al., 2013; Rich et al., 2016). Indeed, flashcards often present words to be trained supported by matching pictures, which facilitates learning, especially in English as L2 (Abbasian & Ghorbanpout, 2016; Oxford & Crookall, 1990; Thompson, 1987).

Motivational Components

To implement storytelling as well as flashcards in the vocabulary learning of English L2 learners with and without learning difficulties, it may be helpful to also include motivational components. Especially for struggling learners who most likely have experienced failure in their school career and therefore have developed foreign language learning anxiety (Horwitz, 2001), motivation is an essential component of learning (García & de Caso, 2004; Horwitz et al., 1986; Sparks, 1995). One way to motivate learners is to use a system of group rewards. Defined as a contingency in which receiving a reward is dependent on the behaviour of each member of

the group (Cooper et al., 2007), this strategy can help to improve students' on-task behavior, as every learner is responsible for the success of the whole group.

According to the relevant research, methods of self-monitoring are essential to ensure that such an intervention is successful. Thus, increasing learners' self-regulation skills can improve academic success, self-motivation, and self-efficacy (Falkenberg & Barbetta, 2013; McDougall et al., 2012; Rafferty, 2010).

Self-graphing as one component of self-monitoring has proven to be particularly practical in this context. As a method of visually recording students' own learning process (Hirsch et al., 2013), self-graphing lets students make their own progresses transparent by recording daily results in a graph, for example. Monitoring one's own learning curve has been found to have positive effects on-task behavior and academic success (Amato-Zech et al., 2006; Gunter et al., 2003; Legge et al., 2010).

Research Question

– The present study took place in Germany, a major non-English-speaking country. German teachers are predominantly faced with inclusive classrooms, as especially students with learning disabilities attend general education classes as part of the realization of the Convention on the Rights of Persons with Disabilities (European Agency, 2017). To adequately support every student in very heterogeneous classrooms, it is necessary to develop methods that teachers can implement rather easily and that are beneficial for students with different levels of ability.

– The aforementioned components of storytelling, flashcards, as well as motivational aspects, which all seem to have a positive impact on learning, and especially vocabulary learning, were combined in the study. The specific question underlying the study, therefore, was as follows: Can a multicomponent intervention consisting of storytelling, flashcards, group rewards, and self-graphing help students with and without learning disabilities acquire English vocabulary as sight words?

Methods

Participants and Setting

The study was carried out at a secondary school in a low socio-economic area in North Rhine-Westphalia. It was concentrated on two seven-grade classrooms, one of which was an inclusion class. All students of both classes were included in the selection process for the study (N = 33). Final participants were chosen based on the following criteria:

1) Low performance in English learning in general: the teachers gave information about the children's previous assessment performance in English, especially vocabulary knowledge and learning difficulties, on the basis of test scores in the classroom setting.

2) Low performance on an English vocabulary test: A researcher-developed vocabulary pretest was additionally used, consisting of 120 common English words. The exact structure of the test is described below under instrument since it has the same structure as the measurement for collecting data during baseline and intervention.

The words were selected from a list of the 1,000 most frequently used words in the English language (Education First, 2019). The vocabulary test was administered over three days (40 words each session) in order not to overwhelm the children. Students who translated 10 or fewer words correctly (German orthography was not taken into account) into German and were recommended by the teacher based on previous results were eligible for the study.

As a result of this process, a total number of 24 students, meeting the study criteria, participated in the study. Of these, three students were German second language learners from different language backgrounds and six had been diagnosed with a learning disability. In accordance with the common German definition, we considered students as having a learning disability if they had a disorder in one or more basic psychological processes that influences performance in listening, speaking, reading, writing, spelling, or mathematics (Turnbull et al., 2004) and were additionally diagnosed with an IQ between 70 and 85 leading to school failure and the need for special support. This definition differs from the one commonly used in North America, but it is compatible with the conception of a learning disability in other parts of the world, like the UK (Gruenke & Cavendish, 2016).

Through the aforementioned vocabulary test, 30 words were identified that were not stored as known words across the 24 participants for the storytelling intervention.

Materials

The stories told by three graduate university students who served as interventionists (prior to the study, the second author had briefed them on how to implement the treatment in three 45-minute sessions) were printed in big letters and presented in a ring binder that could be positioned so that students were able follow the story auditory and visually. The 12 stories for the intervention were authored by the interventionists (available from the first author upon request) who also conducted the intervention after being instructed by the authors of this study. They tried to keep the stories equal with regard to sentence structure and word difficulty. Every story consisted of 150 words and contained the same characters and was part of an overall framework plot. The stories dealt with topics related to the everyday reality of teenagers to arouse the participants' interest. Furthermore, each story contained 10 randomly assigned words out of the 30 words to be trained. To draw attention to these words, they were highlighted on the printed version. In addition, 30 flashcards, each presenting one of the 30 words and related pictures on a 8.3 x 11.7-inch sheet, were used. Finally, for the self-graphing procedure as well as the group reward, self-created graphs and tokens were used. The graphs consisted

of 12 successive rows (maximum number of intervention sessions), each with blank boxes for writing the translation of the 30 training words.

Design

In order estimate the effects of the intervention both process-wise and individually, a single-case with a multiple baseline design across participants (AB) was applied (Ledford & Gast, 2018). The study took place over a period of six weeks. Participants were randomly distributed over six groups of four students each, who received intervention together. Only the six students with learning disabilities were evenly assigned to the groups. The groups were randomly allocated to three intervention start times. Data was collected after baseline and intervention sessions, totaling 16 measurements. The groups started the intervention with a time delay as they had a minimum of four and a maximum of six sessions, whereby two groups always started the intervention at the same time. Consequently, the intervention took place 10-12 times. A follow-up measurement was conducted three weeks after the intervention with a two-week vacation in between to determine how well students maintained the vocabulary over time.

Measurement and Dependent Variable

The measurement consisted of a vocabulary test of the 30 words that were taught in the intervention step-by-step through storytelling. The measurement was conducted over a period of six weeks directly after the intervention sessions. Students had to complete the vocabulary test within 5 minutes. The measurement instrument was constructed the same way as the vocabulary pretest. The children received two sheets (8.5 x 11 inches) each with two columns. In the left-hand column the 30 English words were placed one below the other, and in the blank right-hand column the students were to write the correct German translation. As in the pretest, the German orthography was not evaluated. As long as it was clear that the students knew the German translation for the English word, the word counted as correct.

All three interventionists counted the correct words on the tests with 100% agreement. This was due to the fact that words that were counted as correct, for example synonyms, were determined by the authors beforehand. The order in which the 30 words were presented was randomized in each measurement. At the end of the intervention, participants were asked to complete a questionnaire for social validity. Specifically, students were asked to report whether they liked the storytelling method and would like to continue working with it, whether remembering new words through storytelling was easy, and whether they liked to work in a group. Finally, they were also asked to fill in their personal graph at the end of the sessions.

Procedures

During the baseline phase, students worked within their groups of four. For a period of 35 minutes, they worked on math and German worksheets that were randomly chosen for each

session. The vocabulary test, which served as the measurement, was completed directly after the working phase in each baseline session.

Within Phase B (intervention) storytelling as a multicomponent intervention was also implemented for a period of 35 minutes in each session. Groups were led by three master students who each were responsible for two storytelling groups. These were the same students who did the treatment fidelity checks and the data collection after each baseline and storytelling session. Implementation of the intervention was carried out as part of a two-step model, which included repetition of the 10 words from the previous session and introduction of 10 new words at the beginning as a pre-listening phase. This was realized with the help of flashcards and matching pictures to train meaning and pronunciation for 10 minutes.

The storytelling component followed as the second step of the intervention for 25 minutes. Each narrative contained 10 of the 30 unknown words, randomly chosen. Each word was used in at least three stories and appeared twice in the same story. As soon as one of the highlighted vocabulary was mentioned by the interventionist while telling the story, the corresponding flashcard with the word and the picture was brought out and discussed and repeated with the students. Every session was conducted according to this two-stage model. The first three intervention sessions were dedicated to mentioning all 30 words, 10 words at each meeting. Sessions 4–14 were focused on automation and repetition, with all words recurring randomly after being introduced the first time. Just as in the baseline condition, the participants were evaluated again with the 30-word vocabulary test after each session.

After each measurement of the B Phase, students completed the self-graphing sheet to see their own progress. They received one point for maintaining the level of known vocabulary and two points for improving their score. Additionally, the points of each member of the group was summed up to count the overall score of the whole group. The goal was to reach a previously defined number of known words to receive a group reward.

Treatment Fidelity

A manual was available for every interventionist with step-by-step explanations of how to implement the intervention in order to ensure identical treatment for each group. Additionally, the interventionists were asked to complete a treatment fidelity sheet after each session to make sure the standards of the study were constantly fulfilled. Furthermore, graduate students who were not involved in the treatment observed at least one third of all sessions and filled in a treatment fidelity sheet to make sure the intervention was implemented as planned. This checklist included 15 questions regarding the following general topics: environment (e.g., “Did the session take place without interruptions?”), material (e.g., “Was the material ready before the session started: stories, flashcards, ring binders, laptop?”), procedure (e.g., “Was the content of the previous lesson repeated before

the new session started?"), diagnostic/feedback (e.g., "Did the participants record their scores on the graph?"), and how the interventionists dealt with the students (e.g., "Was the attention of the participants drawn to the task?"). Most of the questions demanded a "yes/no" answer, but some had to be answered on a 5-point scale from "entirely true" to "does not apply at all." Review of the self-assessment scores and observation scores indicated that interventionists followed the criteria of the study at all times.

Results

Visual Analysis

Figure 1 shows an overall stable baseline with no trend. Directly after the intervention was implemented, all participants showed remarkable improvements. Each student improved recognition of correct training words, visible in a stable slope.

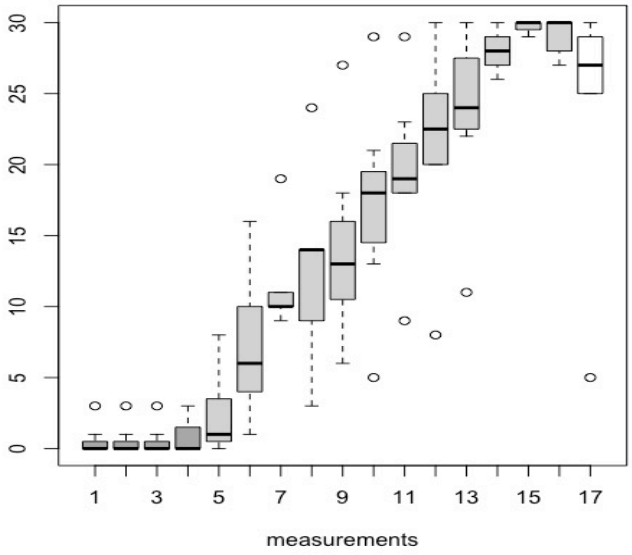


Figure 1. Number of correctly recognized words for phases A and B of groups 3 and 4 with baseline measurement times = 5.

As the intervention times increased, the variance in the results decreased, but there were some outliers below and above, indicating that children reacted differently to the intervention. However, groups 1 and 2 (see Figure 1) displayed less variance than the other four groups.

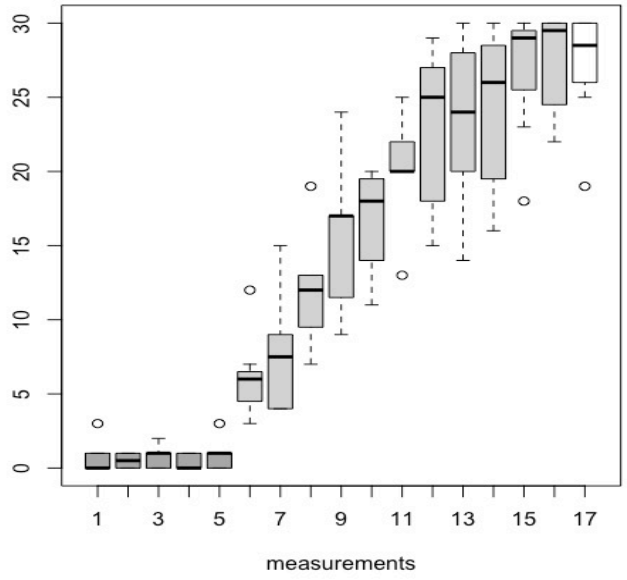


Figure 2. Number of correctly recognized words for phases A and B of groups 3 and 4 with baseline measurement times = 5.

Figure 2 also displays an overall stable baseline with no trend tendency. Again, immediately after the intervention, all participants improved. Specifically, all participants from groups 3 and 4 improved their recognition of vocabulary, as seen by a stable slope effect. In contrast, for groups 1 and 2, the variance of the results was greater per measurement time. However, there were fewer outliers. The greatest variance of outcomes was seen for measurements 12, 13, and 14.

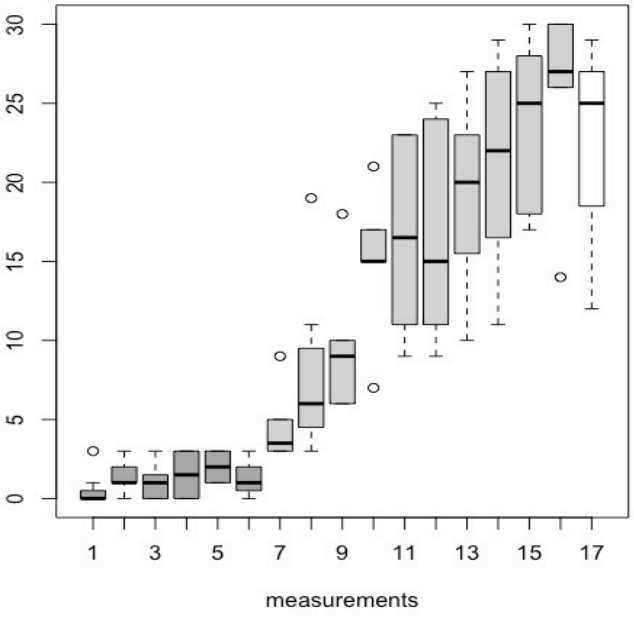


Figure 3. Number of correctly recognized words for phases A and B of groups 5 and 6 with baseline measurement times = 6.

Figure 3 shows minimal differences when comparing all three figures. The participants showed slower improvement but still a stable slope. Additionally, an enormous variance of results was recorded for groups 5 and 6, particularly for measurements 11, 12, and 14. Nevertheless, the members of these two groups also presented an improvement in their trained vocabulary.

In terms of the overall follow-up measurement, almost all participants in each of the groups remembered the same number of correctly recognized words as before the Easter break. Data was more stable for groups 1 and 2, along with groups 3 and 4. Groups 5 and 6 are characterized by rather varying results with outliers, but also a mean value that shows groups were able to maintain a good overall score.

Statistical Analysis

Descriptive statistics for the number of correctly recognized words for each group are depicted in Table 1. As illustrated, all groups showed great improvement when comparing Phases A and Phase B. Groups 3 and 4 showed the most significant improvement, with an average increase of 17.76 words, closely followed by groups 5 and 6, with an average increase to a total of 16.25 words in Phase B. However, groups 1 and 2 were not far behind and also displayed a strong improvement, with 14.73 words in the intervention phase.

Table 1

Descriptive Data

	n (A)	*n* (B)	*M* (A) *SD*	*M*(B) *SD*	Md (A) Min/Max	Md (B) Min/Max
Group 1 & 2	6	10	1.15 (1.19)	15.88 (8.80)	1 (0/3)	17.00 (10/24)
Group 3 & 4	5	11	0.71 (0.91)	18.47 (8.53)	0 (0/3)	20.00 (13.5/25)
Group 5 & 6	4	12	0.64 (1.09)	16.89 (10.12)	0 (0/3)	18 (5.5/29.00)

Note. M = Mean; Md = Median; SD = Standard deviation.

For further data analysis, overlap indices were calculated to estimate the effectiveness of the intervention. Specifically, Tau-U, which is a combination of the Mann-Whitney U-test and Kendall's rank correlation (Parker et al., 2011), non-overlap of all pairs (NAP), and improved rate difference (IRD) were calculated. Tau-U values ranged from .96 to 1.00 (p = <.001) across all groups, and there was no need for any baseline correction. These results cannot be

attributed to chance with a probability of less than 0.10%. Regarding NAP, all participants reached the maximal possible score of 1.00 (p <.001). Finally, IRD scores ranged from 0.96 to 1.00, indicating a high treatment effect.

Table 2

Effect Sizes for Number of Correctly Recognized Words

	TauU	*p*	NAP	*p*	IRD
Group 1 and 2	0.98	.00*	1.00	.00*	0.99
Group 3 and 4	1.00	.00*	1.00	.00*	1.00
Group 5 and 6	0.96	.00*	1.00	.00*	0.96

*significant at. The .001 level

In addition, a piecewise regression analysis was conducted across the three group constellations (level 1 analysis) and across all participants (level 2 analysis). All groups revealed a statistically significant slope effect, meaning that their performance increased gradually in the ability to recall the meaning of the English words presented to them. In addition, a significant level effect was found for groups 3 and 4. Concerning overall effects across all groups, a highly significant slope effect and a significant level effect emerged.

Table 3

Piecewise Regression Model for Number of Correctly Recognized Words (level 1 and 2)

	B	SE	*t*	*p*
		group 1 & group 2		
Intercept	0.716	1.652	0.434	.66
Trend	0.175	0.315	0.557	.60
Level	1.159	1.381	0.839	.40
Slope	2.186	0.350	6.251	.00**
		group 3 & group 4		
Intercept	0.894	1.425	0.627	.53
Trend	-0.009	0.340	-0.026	.99
Level	4.114	1.106	3.718	.00**
Slope	2.240	0.356	6.297	.00**
		group 5 & group 6		
Intercept	0.429	1.934	0.222	.83
Trend	0.086	0.508	-0.169	.87
Level	1.457	1.198	1.216	.23
Slope	2.291	0.519	4.417	.00**
		overall		
Intercept	0.751	0.939	0.800	.42
Trend	0.065	0.207	0.316	.75
Level	2.271	0.721	3.149	.002*
Slope	2.256	0.217	10.392	.00**

*significant at the .01 level. **significant at the .001 level.

Discussion

Main Findings

The purpose of this study was to examine the effects of a multicomponent intervention consisting of storytelling, flashcards, and motivational components on the vocabulary acquisition of 24 struggling secondary school students. Findings indicate that all participants benefited from the intervention and improved their sight word knowledge. In addition, follow-up data shows that almost all participants maintained their last score after three weeks, two weeks of which were Easter break. All effect sizes underline a high potency of this intervention. All participants who had been diagnosed with a learning disability were able to enhance their sight word knowledge as well. Moreover, they showed one of the greatest improvements during the intervention phase. This fact might not be too surprising, however, as low achievers tend to show greater gains than higher achievers, as phenomenon known as the “Robin Hood Effect” (Haefner et al., 2017).

Since the new words were introduced gradually, an immediate level effect was not expected. As shown in the visual analysis, improvement was visible after a couple of intervention sessions. Introducing the words in the stories step-by-step was intended to counteract an excessive cognitive demand, especially for struggling learners who face problems in mental retention and working memory processes. Groups 5 and 6 demonstrated slower improvement. This may be explained, on the one hand, as a result of having two pupils in these groups who performed lower than the others but still had great individual improvement, and, on the other hand, by only having 10 interventions while the other groups had 11 and 12. This finding leads to the assumption that 10 interventions might be insufficient to automatize all words as trained, or individual motivational problems might have played a role.

Limitations

The following limitations must be considered when interpreting the results of this study. First, the study was conducted with only 24 participants, making it difficult to generalize the findings. Moreover, our focus was English as a foreign language, so we cannot make general statements with respect to learning of other languages.

Additionally, some of our participants were German second language learners, for whom it might have been more difficult to remember and automatize the chosen trained sight words. Taking a look at the standard deviations and result variances in Phase B across all groups, it is obvious that some students demonstrated greater gain than others. One reason might be the fact that not all students were German L1. Even though all students showed improvement, it is important to determine the reasons for their different reactions to the intervention. Since we were implementing a multicomponent intervention, we do not know the

extent to which each component contributed to the results. Thus, a specific component may have a greater impact on the dependent variable than others.

This study was about the enhanced performance of sight words with regard to remembering the form and meaning of each word. We did not examine whether the chosen words could be read or written as well.

With regard to the measurement, it should be noted that it was carried out by the same persons who carried out the intervention. However, each survey was reviewed by all three master students with 100% agreement; thus, any interference by the master students with the results of the groups she supervised is highly unlikely.

Finally, as this was a pilot study to assess the effects of a self-developed intervention for the first time, the intervention was not compared to other interventions. Therefore, it is unknown whether it is better than other interventions that also focus on vocabulary learning.

Implications for Future Research

To generalize the results, future studies should include a larger sample size and examine whether storytelling can be used successfully with a whole class. In addition, the study should be repeated with students with different backgrounds, ages, and challenges, such as external behavioral problems. In future research, the impact on reading and writing skills could be studied to identify what areas storytelling can impact. Given that many students learn German as L2 and the fact that many students learn English as L3, the L1s of the subjects should be recorded in future research.

The intervention is both economical and quick and easy to implement. Design of the materials and implementation itself is simple and can be easily carried out by teachers. Furthermore, it can be adapted to different age groups by adjusting the themes of the stories and the respective training words to participants' everyday lives. Further, individual components can be exchanged and adapted to pupils with specific learning challenges

The results show that students with learning disabilities, even if only a handful in the current study, benefited from the intervention. This is of note given a recurring discussion about whether it is advisable to teach students with a learning disability a foreign language. Thus, there is an indication that students with learning difficulties also benefited from the intervention in terms of long-term effects. In a future study, a larger number of children with learning disabilities could be compared with students without learning disabilities in their response to the intervention.

An additional important area of further research would be to compare the storytelling intervention with another common English L2 intervention that also focuses on vocabulary training in an experimental group design.

Despite its limitations and need for further research, the present study provides important information about the effectiveness of this multicomponent intervention in expanding the vocabulary in English as a foreign language of struggling learners with and without learning disabilities in an inclusive environment. Above all, the simplicity of the intervention should lead to it being used to help children and young people learn English by creating a positive context and making students enjoy learning English through adaptation to the children's world and a high level of active foreign language learning.

Appendix H Article 8

Barwasser, A., Knaak, T., & Grünke, M. (2020). The effects of a multicomponent storytelling intervention on the vocabulary recognition of struggling English as a foreign language learners with learning disabilities. *Insights into Learning Disabilities, 17*(1), 35–53.

Abstract

In a globalized world, the ability to speak English has become increasingly significant. A large number of students who are learning English as Foreign Language face obstacles with acquiring and storing vocabulary. Especially children with learning disabilities are at risk because of a limited power of retention and motivational struggles. Having an adequate number of words stored in the mental lexicon facilitates the process of language learning and is essential in order to develop reading fluency. The purpose of this single case study was to evaluate the effectiveness of a two-stage storytelling model combined with flashcards and a rewarding procedure with respect to four struggling secondary school students with learning disabilities. All students showed tremendous increases in their performances with onset of the intervention. Visual analysis, effect sizes and a piecewise regression analysis indicate that the treatment was a complete success. In addition, a follow-up measurement underlines the long-lasting effectiveness extending beyond the intervention time.

Keywords: Storytelling, Motivation, Vocabulary Recognition, English Foreign Language, Learning Disabilities

Introduction

English is an important element of general education and overall success. Nowadays, there is almost no school curriculum in which English does not play a leading role (Ivancevic-Otanjac, 2016). Additionally, it influences the daily life of children and youth and thus, it is essential to learn not only with respect to the integration of pupils who need special support (Reddy, 2016).

Di Fino & Lomardino (2004) underline three global problem areas concerning foreign language learning which may result in failure: poor memorization skills, high level of anxiety and lexical grammar confusion. Especially the first has great influence on language learning due to a lack of remembering vocabulary with fatal repercussion. That, in turn, results in many L2 learners who read laboriously and far more slowly compared to their L1 (Fraser, 2007) and struggle with achieving an adequate reading fluency and language proficiency. Reasons for deficits concerning fluency derive from a lack of vocabulary stored in the mental lexicon.

Additionally, an appropriate set of known basic vocabulary is essential because by having a wide range stored in the mental lexicon, the working memory, with its limited capacities, is relieved, allowing it to pay more attention on higher level processes such as text comprehension and text writing in L2 (Grabe, 2004). When learning a new language, vocabulary knowledge is one of the central concerns. Without knowledge of words which can be automatically and quickly retrieved and recognized, learners cannot convey meaning and thus, communicative actions are restricted (Morra & Camba, 2009).

Learning disabilities (LD) affect vocabulary storage, working memory and information processing (Garcia & Tyler, 2010), and additionally, students with LD lack self-monitoring ability and learning strategies (Shippen, Houchins, Steventon, & Sartor, 2005). But, even though there is a certain struggle and a link between problems in L1 and L2 (i.e. Koda, 2005; Sparks, Ganschow & Patton, 2008; Abu-Rabia & Bluesetin-Danon, 2012; Sparks, 2012), Sparks (2009) states that students with LD should be enrolled in foreign language courses and being supported by appropriate teaching and instruction methods with the aim to become successful in these courses (i.e. through mnemonic devices, small class size and explicit linguistic teaching and reduced amount of content like successively introducing vocabulary) (Wight, 2015). Contrary to the fear of overtaxing learners with LD while introducing a new language, they are capable of learning a new language (Wydeell & Kondo, 2003; Lazda-Cazers & Thorson; 2008). They do not necessarily have more severe foreign language learning problems than their peers only because they are learning disabled (Sparks & Javorsky, 2000), it depends on how a language is conveyed. Learners with special needs can be successful in acquiring a foreign language when provided with adequate support (Sparks et al., 2008) by taking into account the main problem areas like memorization and information processing mentioned above. The way of how a new language is introduced is enormously important - it is all about modification of a language learning classroom (Lazda-Cazers & Thorson, 2008).

Paivio (1986) states that there are two existing ways of how students become better in learned material: verbal associations and visual imagery. The Dual Coding Theory (DCT; Paivio, 1986) differentiates between non-verbal and verbal mental systems in the context of processing imagery and linguistic information. If a person has stored a stimulus concept for a certain word as verbal and nonverbal concept, the person is able to retrieve both information for recalling - increasing the probability to remember a certain item. According to this theory, word concreteness and images should be emphasized in educational and cognitive tasks related to meaning (Clark & Paivio, 1991). Thus, providing students with verbal and nonverbal stimuli seems to be beneficial and (language) learning is more effectively when nonverbal and verbal modalities are combined (Peker, Regalla & Cox, 2018; Ginns, 2005; Mayer, 2009). In the context of vocabulary learning it is said that new words can be better learned when they

are combined with images (i.e. Mirhassani & Eghtesadei, 2007; O'Malley & Chamot, 1990). Students with learning disabilities profit from a multimodal instruction as well because they need more time to process language components and a deeper exposure to the language and the usage of visual tools support their language learning (Brady et al., 2015; Skinner & Smith, 2011). Furthermore, it leads to a faster incorporation of new facts and words into the long-term and short-term memory (Regalla & Peker, 2015).

In order to support vocabulary acquisition, words of training are commonly embedded into a context (Nation, 2015; Webb, 2008). Learning a new language is easier when content is put into a meaningful context because students are more likely to remember words and phrases within a story frame, especially over time (Van, 2009; Oxford & Scarcella, 1994). Atay & Ozbulgan (2007) showed that memory strategies like learning through context lead to an improvement of vocabulary knowledge. Moreover, there is a need for interesting topics, situated in meaningful contexts to trigger students' motivation (Grimm, Meyer & Volkmann, 2015). When unknown vocabulary is embedded into an interesting narrative, a foundation of being able to verbally utter these vocabulary and explain their meaning is easier. Also, an improvement of word knowledge can be observed though word meaning explanation and repeated reading (Webb, 2007). Studies show an improvement of gaining word meaning while using single reading and word meaning explanation (average of 15% gain; Senechal, 1997; Senechal & Cornell, 1993), repeated readings without word meaning explanation (average of 9% gain; Brabham & Lynch-Brown, 2002; Elley, 1989) and repeated reading and word meaning explanation (average of 26 %, Penno et al., 2002; Hargarve & Senechal, 2000).

Having in mind the main problems learners do face while learning a new language, it is necessary to develop interventions which combine the components of retention, the importance of context and the relevance of a multimodal instruction though nonverbal and verbal modalities.

The method of storytelling is a powerful tool in order to activate and motivate learners and moreover, to improve different kinds of skills while embedding teacher and students in a communicative circle (Roney, 1996; Ellis & Brewster, 2002; Cameron, 2001).

Storytelling can be defined as a process where a teller, by using mental imageries, vocalization and a narrative structure, conveys certain content through a story frame to audience (Roney, 1996). Through storytelling, interaction takes place and children adopt the teacher's pronunciation and thus, improve their sensitivity for rhythm, intonation and pronunciation (Ellis & Brewster, 2002).

There has not yet been research on the effects of storytelling on the vocabulary recognition of English L2 learners and moreover, no studies examining storytelling in the context of learning disabilities. Studies that are focusing storytelling in English as Foreign language mainly put an emphasis on improving general language skills and reading

comprehension (Al-Manosour & Al-Shorman. 2011; Hemmati et al., 2015; Kim, 2010; Huang, 2006), underlining storytelling to be a promising tool, especially with respect to the relevant role of context in language learning.

Another effective intervention in the field of word acquisition is said to be direct instruction flashcards which is a method to present and teach sight words. Proven effects can be found in the context of students with and without learning disabilities (Standish, McLaughlin & Neyman, 2012; Fraher, Jones, Caniglia, Crowell & Hastings, 2019; Brasch, Williams, & McLaughlin, 2008), in order to improve reading fluency (Kaufman, McLaughlin, Derby, & Waco, 2011) and sight words across different subjects (Crowley, McLaughlin, & Kahn, 2013; Rich, Weber, McLaughlin & Sells-Love, 2016; Erbey et al., 2011; Ruwe et al., 2011; Seines, McLaughlin, Derby & Weber, 2015) and in addition, positive effects can be found on sight words in English as Foreign Language L2 (Abbasian & Ghorbanpout, 2016). Through the usage of this flashcard procedure, automatization can be achieved and students who struggle with memorization and recall of certain information do benefit (Seine et al., 2015).

Since struggling learners fear to face repeated failure when learning a foreign language (Csizér et al. 2010) motivation is essential and adding motivational components to an intervention seems to be beneficial (Garcia & de Caso, 2004; Lepola, Saloner, Vauras, & Poskiparta, 2004). One realization of motivation is called “interdependent group contingencies” which are implemented into classrooms in order to improve learning outcome and behavior. It can be defined as a contingency in which the realization of getting a reward is dependent on the behavior of each group member (Cooper, Heron, & Heward, 2007). Students collecting points for meeting certain criteria which all together count for the whole team. Interventions which additionally introduced group contingencies showed great effects (Cooper et al., 2007; Popkin & Skinner, 2003; Little, Akin. Little & O’Neil, 2015). Furthermore, positive effects can be found on academic achievement (Rohrbeck, Ginsburg-Block, Fantuzzo & Miller, 2003; Bowman-Perrot et al., 2013; Pappas, Skinner & Skinner, 2010) and social behavior (Ginsburg-Block et al., 2006; Wills et al., 2016).

An additional tool which can be implemented in order to improve motivation is a self-graphing procedure which means the visual recording of the own progress. Self-graphing in order to make own progresses transparent has positive effects on “on task behavior” (Amato-Zech, Hoff & Doepke, 2006; Legge, deBar, Alber-Morgan, 2010) and enhances academic success (e.g., Gunter et al., 2003).

Vocabulary acquisition in a L2 needs context embedding as well as an adequate amount of repetition and direct instruction of words with respect to the long-term retention of these words. Additionally, adding motivational components and using verbal as well as nonverbal modalities are enormously important, especially regarding learners with learning

disabilities. Having in mind these facts, a multi-component intervention was realized in this study while uniting storytelling, DI flashcards as verbal and nonverbal instrument, self-graphing and group rewarding in order to foster vocabulary acquisition of secondary school students with learning disabilities in English as Foreign Language.

Methods

Participants and Setting

The participants of this study were 24 secondary school students of two seventh grades, one of which is an inclusive grade, in North Rhine-Westphalia, Germany. All participants were chosen based on three vocabulary pre-tests, each consisting of 40 most common words in English to estimate their amount of known words. The words were chosen from a list of the 1000 most common words in English (Education First, 2019) and from the suggestions of the teacher based on English school books and already taught vocabulary.

In addition, children were recommended by their classroom teachers on account of weak assessment performance in English, especially in the area of vocabularies. Six of the participants have been diagnosed with a learning disability, but only four will be focused in the present study due to missing data. Three master students of special needs education functioned as interventionists. All four participants were female and none of them has a migration background and thus, all three learned German as First Language.

Material

A binder was used to make the stories visible to all students. The story pages were put on the binder with an adequate size of words with the aim to follow the story not only auditory but visually as well. The stories were self-written and adapted to the students' everyday life interests. Out of the vocabulary pre-tests, 30 words were identified which were not stored as known words across all participants and were intensively taught during intervention period. Direct instruction flashcards, each consisting of one of the 30 chosen words and a matching picture, were printed on a DIN 4 sheet. Additionally, as motivational component, a self-graphing paper was designed to provide the possibility of documenting one's own learning progress. This paper displays 12 rows one below the other and each consisting of 30 empty boxes which represent the 30 words of training.

Design

A single-case with a multiple baseline design across participants (AB) (Ledford & Gast, 2018) was applied in order to estimate the effects of the intervention processually and individually for each participant. Over a period of six weeks data was collected each day directly after baseline activity and intervention. All 24 participants were randomly assigned to six groups consisting of four children who received storytelling intervention together, whereas

the students with learning disabilities, who will be specifically focused in this paper, were evenly distributed to the groups. All six groups were randomly assigned to the possible starts of intervention to increase internal validity (Dugard, File & Todman, 2012). The A phase had to consist of at least four and the B Phase of at least ten probes. Intervention took place three times a week at the same time. The different groups were taken out of regular lesson into separate rooms in order to reduce possible distractions.

Dependent variables and measurement

The number of correct recognized words functioned as the dependent variable. The participants were asked to fill in a vocabulary test consisting of the 30 chosen words to be trained in the storytelling intervention in 5 minutes. The students were asked to translate the English words in the left-hand side to German on the right-hand side whereas German spelling was not counted. It was important for the interventionist to see whether the words are recognized correctly or not. The amount of correct recognized words was counted.

Procedures

For baseline conditions all participants were working on exercise sheets for 20 minutes which had no connection with English language in order to estimate the current abilities of the students. The worksheets were randomly chosen and dealt with Maths and German exercises. Directly after the 20 minutes working phase, all participants were measured. For intervention condition, storytelling as a multi component intervention was implemented by three master students who each accompanied 2 storytelling groups. Out of a pool of the 30 words to be trained, 10 words were randomly assigned to each story and these ten words were intensively transferred by the interventionist. All participants were confronted with the 30 words at least three times during the entire intervention. The intervention was realized in a two- stage model consisting of a first pre-listening stage by repeating the ten words from the last story and introducing the ten new words with the help of flashcards and matching pictures to each word. The words were trained while thinking of the meaning and the correct pronunciation. On the second stage the storytelling procedure started.

The stories were always visible to the students and the words to be trained were marked with a green color. Any time a green word appeared while reading, the interventionist stopped the story, took the correct flashcards and repeated the word together with all participants. All stories were self-written adapted to the students' interest as a linear content with same characters solving certain problems. After each intervention all participants were measured again with the 30-word vocabulary test, whereas the order of the words of this test was random.

As motivational components, after any measurement, the participants were asked to fill in their own progress monitoring sheet. Each storytelling group collected points in order to achieve a reward at the end. The amounts of points resulted from the self-graphing paper. Students got

1 point for keeping the same score of vocabulary retention from last time and two points for getting better. These points were thrown by each participant in a receptacle visible to all group members. The goal was to hit a prior defined total number of words to get a group reward.

Treatment Fidelity

Treatment fidelity was realized by handing out a practical guidance that has to be followed by the interventionists. In addition, one third of interventions were observed by an unrelated person to make sure that the practical guidance was realized correctly by filling in a treatment fidelity sheet. Furthermore, each interventionist got an own treatment fidelity sheet to be filled out after each session by themselves. The three master students were in constant contact with the first author of this study.

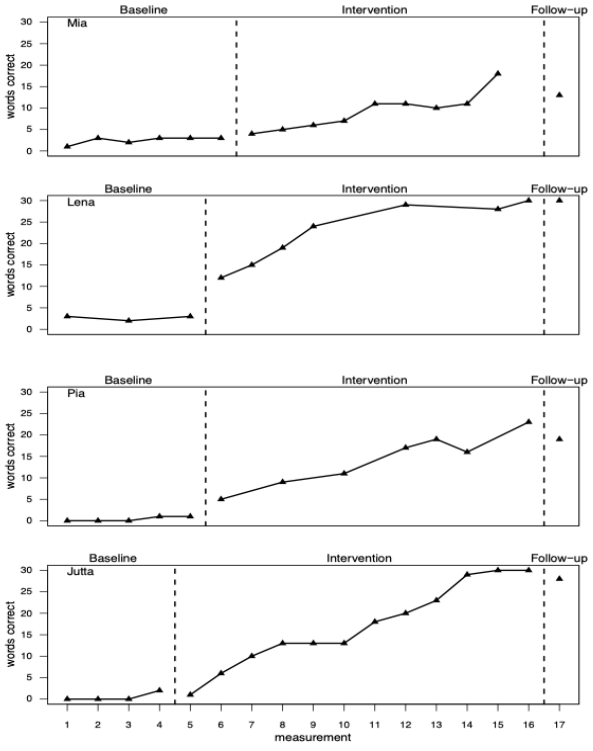
Results

Visual Analysis

Visual analysis was conducted with the help of the SCAN package for R by Wilbert (2018). As can be visually seen, all participants increased their vocabulary knowledge in the treatment phase. While having no baseline trends for all four participants, there scores improved visibly after the intervention was implemented. Pia and Lena had a direct level effect with the beginning of intervention whereas the other two had a slower increase of known words. Regarding the follow-up data, only minimal decrease of known words took place.

Figure 1

Amount of words known correctly for each participant in baseline, treatment condition and follow-up measurement.



Quantitative Analysis

Descriptive statistics for number of correct recognized words for each group are shown in table 1. All participants showed great improvement comparing A phase and B phase, whereas two participants reached the highest possible score.

Table 1

Descriptive data for each participant in A and B phase

	n (A)	*n* (B)	*M (A)* *SD*	*M(B)* *SD*	*Md* (A) *Min/Max*	*Md* (B) *Min/Max*
Mia	6	10	2.50(0.84)	9.22 (4.29)	3.00(1/3)	10.00(4/18)
Lena	5	11	2.67(0.58)	22.43 (7.18)	3.00(2/3)	24.00(12/30)
Pia	5	11	0.40(0.55)	14.29 (6.24)	0.00(0/1)	16.00(5/23)
Jutta	4	12	0.50(1.00)	17.17 (9.54)	0.00(0/2)	15.50(1/30)

Note. N = times of measurements, A = baseline, B= treatment phase, M= mean, Md = median, SD = standard deviation

The mean baseline difference (O'Brien & Repp, 1990) was applied showing an average increase of about 2400 % in the intervention phase compared to the baseline phase. Pia benefited the most while Mia benefitted the least. Even though official MBD value cutoffs do not exist, but 2400 % of performance increase can be seen as meaningful.

Additionally, non-overlap effect sizes were calculated, again using the SCAN package by Wilbert (2018) non-overlap of all pairs (NAP), percentage exceeding the median (PEM), percentage of non-overlapping data (PND) and the improve rate difference (IRD) (Parker, Vannest & Davis, 2011). All indices underline an improvement for all participants while Jutta shows the least improvement but still with considerable increase. The majority reached the maximum value of 100 % and NAP results indicate significant differences between phases A and B for all students.

Table 2

Effect Sizes for number of correct recognized words

	NAP	*p*	*PEM*	*PND*	*MBD*	*IRD*
Mia	100.00	.00**	100.00	100.00	269%	1.00
Lena	100.00	.01*	100.00	100.00	730%	1.00
Pia	100.00	.01*	100.00	100.00	5225%	1.00
Jutta	97.92	.01*	100.00	91.67	3646 %	91.67

Note. PND=percentage of non-overlapping data; PEM = Percentage exceeding the median; IRD=improved rate difference; MBD=mean baseline difference; NAP=non-overlap of all pairs. *significant at the .01 level ** significant at the .001 level; [11] very strong effects

Furthermore, we conducted a piecewise regression analysis as level 1 analysis (each student) and level 2 analysis (across all students). Trends in baseline cannot be found for any

participants but due to the short baseline phase trend information has to be interpreted carefully. Taking the visual inspection into account, no baseline trend is visible. Mia, Pia and Jutta showed a significant slope effect with an average of 1,5 scale points gained with each intervention point. For Lena, a significant level effect can be found. These results are in favor for a successively improvement of known words over time with the exception of Lena.

Table 3

Regression Model for number of correct recognized words (level 1 analysis)

	B	*SE*	*t*	*p*
		Mia		
Intercept	1.400	1.433	0.977	0.350
Trend	0.314	0.368	0.854	0.411
Level	-1.230	1.578	-0.779	0.452
Slope	1.119	0.418	2,676	0.022*
		Lena		
Intercept	2.667	3-543	0.753	0.480
Trend	0.000	1.037	0.000	1.000
Level	10.742	3.333	3.223	0.018*
Slope	1.662	1.081	1.538	0.175
		Pia		
Intercept	-0.500	1.485	-0.337	0.745
Trend	0.300	0.448	0.670	0.522
Level	2.463	1.603	1.537	0.163
Slope	1.422	0.477	2.978	0.018*
		Jutta		
Intercept	-1.000	2.096	-0.477	0.642
Trend	0.600	0.765	0.784	0.448
Level	-1.142	1.778	-0.643	0.533
Slope	2.001	0.779	2.570	0.025*

Note. * significant at the .05 level.

Level 2 analysis shows a significant slope effect across all participants at the .001 level, implying that all participants improved the amount of known English vocabulary with the amounts of interventions and supports the results of the level 1 analysis.

Table 4

Piecewise Regression Model for number of correct recognized words (level 2 analysis)

	B	*SE*	*t*	*p*
		Overall		
Intercept	2.315	2.239	1.034	0.307
Trend	-0.048	0.468	-0.102	0.920
Level	1.830	1.613	1.135	0.262
Slope	2.078	0.491	4.233	.00***

Note. *** significant at the .001 level

Discussion

Main Findings

This Study was conducted to evaluate the effects of a multicomponent intervention consisting of storytelling, DI flashcards as a verbal and nonverbal instrument, self-graphing and group rewarding on vocabulary acquisition in L2 of four secondary school students with learning disabilities.

Findings indicate high benefits for each of the four participants regarding sight word acquisition. The visual analysis shows continuous improvement during phase B. Individuals were able to improve their sight word knowledge which can be underlined by the results of the chosen effect sizes. Due to implementing new vocabulary step by step during the intervention, it was assumed that students would show a stable slope effect rather than direct quick improvements and level effects. The piecewise linear regression of each case and a hierarchical piecewise linear regression of all cases combined, confirm this assumption, as Mia, Pia and Jutta all showed high improvements after a couple of interventions whereas Lena was even able to increase correct words immediately after beginning of Phase B. Therefore, her results show a significant level effect. This indicates that she was able to recall almost all of the ten trained vocabulary from intervention one.

At the end of the entire intervention, Jutta and Lena were able to recall (almost) all 30 learned vocabulary. The follow up data which was collected after a two-week Easter break indicates a storage in the long term memory as all participants showed high memorization of the to be trained sight words. The intervention therefore seemed to help students with learning disabilities who usually struggled in learning new vocabulary in memorizing 30 new English vocabulary in a sustainable way.

Limitations

Nevertheless, the findings are subject to certain limitations. First, it would be impossible to draw conclusions how to teach students with learning disabilities to improve vocabulary knowledge as this study is only based on four cases and the outcomes are therefore of limited generalizability.

It was only tested if students would remember the meaning of the words. If they were also able to read, write or produce the words in their active language was not tested at all. Additionally, these four cases belonged to on particular age group and showed similar learning challenges. Learning disability in this study was defined as individuals whose skills are above those from intellectually disabled but still fail in all core subjects (Gruenke & Morrison Cavendish, 2016). This is the common definition in Germany and a number of other countries. Furthermore, our participants were only chosen by their previous school performance and an

English vocabulary test. To describe students more accurate, standardized tests would have been helpful.

With regard to the outcomes of this study we can only make vague statements about learning English vocabulary. If this method also helps German students to learn words from different languages or specific German words cannot be confirmed yet. Finally, we are not able to say which component of our multicomponent intervention contributed to the results to what extent.

Outlook

Despite the limitations of this study, the results are promising and give many options for future research. Future studies could include a larger number of students to increase the chance to generalize the results. This could also include to conduct the method of storytelling with larger groups or even whole classes to improve the benefit for teachers in everyday school situations. Evaluating this method with students of different ages could be interesting as well, as it is questionable if, for example students at the age of 15 or 16, would also benefit and be motivated by this kind of intervention.

As mentioned before we are unable to differentiate which component had the biggest impact to the results. Future research could try to figure that out by deleting certain components such as flashcards or motivational aspects when evaluating storytelling to see whether results are changing. Finally, one could examine if students are also able to write, read and produce words learned during storytelling.

References

Abbasian, G.-R., & Ghorbanpour, E. (2016). The Effect of Flash Card-based Instruction on Vocabulary Learning by EFL Learners. *International Journal of Humanities and Cultural Studies, Special March,* 1929–1938. Retrieved 3rd of February at https://www.ijhcs.com/index.php/ijhcs/article/view/2571/2398

Abu-Rabia, S., & Bluestein-Danon, D. (2012). A Study into the Results of an Intervention Program of Linguistic Skills in English (L2) and Its Effect on Hebrew (L1) among Poor Readers: An Examination of the Cognitive-Retroactive Transfer (CRT) Hypothesis. *Open Journal of Modern Linguistics, 2,* 131. https://doi.org/10.4236/ojml.2012.24017

Al-Mansour S., & Al-Shorman, R. A. (2011). The effect of teacher's storytelling aloud on the reading comprehension of Saudi elementary stage students. *Journal of King Saud University. Language and Translation, 23*, 69–76.

Amato-Zech, N.A., Hoff, K. E., & Doepke, K.J. (2006). Increasing on-task behavior in the classroom: Extension of self-monitoring strategiea. *Psychology in the Schools, 43*, 211–221

Atay, D. & Ozbulgan, C. (2007). Memory strategy instruction, contextual learning and ESP vocabulary recall. *English for Specific Purposes, 26*, 39–51.

Brady, N. C., Storkel, H. L., Bushnell, P., Barker, R. M., Saunders, K., Daniels, D., & Fleming, K. (2015). Investigating a multimodal intervention for children with limited expressive vocabularies associated with autism. *American Journal of Speech-Language Pathology, 24,* 438–459. https://doi.org/10.1044/2015_AJSLP-14-0093

Brasch, T. L., Williams, R. L., & McLaughlin, T. F. (2008). The effects of a direct instruction flashcard system on multiplication fact mastery by two high school students with ADHD and ODD. *Child & Family Behavior Therapy, 30*(1), 51– 59.

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,* 39–55.

Cameron, L. (2001). *Teaching Languages to Young Learners*. Cambridge: Cambridge University Press.

Clark, M. J., & Paivio, A. (1991). Dual Coding Theory and Education. *Educational Psychology Review, 3*(3), 149–210.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied Behavior Analysis (2nd Ed).* Upper Saddle River, NY: Pearson Prentice.

Crowley, K., McLaughlin, T. F., & Kahn, R. (2013). The effects of direct instruction flashcards, reading racetracks on sight word skills for two elementary students with autism. *Journal of Developmental and Physical Disabilities*, *26,* 297–316.

Csizér, Kata; Kormos, Judit & Sarkadi, Ágnes (2010). The Dynamics of Language Learning Attitudes and Motivation: Lessons From an Interview Study of Dyslexic Language Learners. *The Modern Language Journal, 94,* 470–487.

Di Fino, S.M., & Lomardino, L.J. (2004). Language learning disabilities. The ultimate foreign language challenge.Foreign Language Annals, 37(3), 390–400.

Dugard, P., File, P. & Todman, J. (2012). *Single-case and small-n experimental designs: A practical guide to randomization tests*. New York, NY: Routledge.

Education First (2019). 1000 most coomon words in English. Retrieved: 20.12.19 from: https://www.ef.com/wwen/english-resources/english-vocabulary/top-1000-words/

Ellis, G., & Brewster, J. (2002). Tell it Again. The New Storytelling Handbook for Primary Teachers. Harlow: Penguin English.

Erbey, R., McLaughlin, T. F., Derby, K. M., & Everson, M. (2011). The effects of using flashcards with reading racetrack to teach letter sounds, sight words, and math facts to elementary students with learning disabilities. *International Electronic Journal of Elementary Education, 3*(3), 213–226.

Fraher, C., Jones, K., Caniglia, C., Crowell, G., Hastings, K., & Zumwalt, K. (2019). Effectiveness of Direct Instruction Flashcards on Sight Word Identification for a High School Student With a Specific Learning Disability. *Insights into Learning Disabilities, 16* (1), 37–44

Fraser, C. A. (2007). Reading rate in L1 Mandarin and L2 English across five reading tasks. *The Modern Language Journal*, *91*, 372–394.

Garcia, S. B., & Tyler, B. (2010). Meeting the needs of English language learners with learning disabilities in the general curriculum. *Theory Into Practice*, *49*, 113–120.

García, J. N., & de Caso, A. M. (2004). Effects of a motivational intervention towards writing in children with learning disabilities. *Learning Disabilities Quarterly, 27*, 141–159.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15,* 313–331.

Ginsburg-Block, M. D., Rohrbeck, C. A. & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology, 98,* 732–749.

Grabe, W. (2004). Research on teaching reading. *ARAL*, *24*, 44–69.

Grimm, N., Meyer, M., & Volkmann, L. (2015). Teaching English. Tübingen: Narr.

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,* 28–31.

Hargrave, A. C. & Senechal, M. (2000). Book reading intervention with language-delayed preschool children: The benefits of regular reading and dialogic reading. *Journal of Child Language, 15.* 765–90.

Hemmati, F. (2015). The Effects of Teachers' Storytelling and Reading Story Aloud on the Listening Comprehension of Iranian EFL Learners. *Theory and Practice in language Studies, 7,* 1482–1488

Huang, H. (2006). The effects of storytelling on EFL young learners' reading comprehension and word recall. *English Teaching & Learning*, *30*(3), 51–74.

Ivančević-Otanjac, M. (2016). Students with language learning disabilities and difficulties in a foreign language classroom. *Special Education and Rehabilitation, 15*(4), 461–474. doi:10.5937/specedreh15-12071

Kaufman, L., McLaughlin, T. F., Derby, K. M., & Waco, T. (2011). Employing reading racetracks and DI flashcards with and without cover, copy, and compare and rewards to teach of sight words to three students with learning disabilities in reading. *Educational Research Quarterly*, *34,* 24–44.

Kim, M. (2010). The effects of storytelling on adult English language learners. *Linguistic Research, 27*(3), 447–473.

Koda, K. (2005). Insights into second language reading. Cambridge, ENG: Cambridge Press.

Lazda-Cazers, Rasma & Thorson, Helga (2008), Teaching Foreign Lan- guages to Students with Disabilities: Initiatives to Educate Faculty. In: Berberi, Tammy; Hamilton, Elizabeth C. & Sutherland, Ian M. (Hrsg.) (2008), *Worlds Apart? Disability and Foreign Language Learning*. New Haven & London: Yale University Press, 107–133.

Ledford, J. R., & Gast, D. L. (2018). Single case research methodology: Applications in special education and behavioral sciences. New York, NY: Routledge. https://doi.org/10.4324/9781315150666

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,* 43–52.

Lepola, J., Salonen, P., Vauras, M., CDATA[& Poskiparta, E. (2004). Understanding the development of subnormal performance in children from a motivational-interactionist perspective. *International Review of Research in Mental Retardation, 28,* 145–189.

Little, S. G., Akin-Little, A., &O'Neill, K. (2015).Group contingency interventions with children—1980-2010: A meta-analysis. *Behavior Modification*, *39,* 322–341.

Mayer, R. E. (2009). Multimedia learning (2nd ed.). New York: Cambridge University Press.

Mirhassani, S.A. & Eghtesadei, A.R. (2007). Improving vocabulary learning through mnemonic devices. *ILI Language Teaching Journal, 3*(2), 91–98.

Morra, S., & Camba, R. (2009). Vocabulary learning in primary school children: Working memory and long-term memory components. *Journal of Experimental Child Psychology, 104,*156–178

Nation, I. S. P., (2001) *Learning Vocabulary in Another Language*. Cambridge: CUP.

Nation, I. S. P. (2015). Principles guiding vocabulary learning through extensive reading. Reading in a Foreign Language, 27, 136–145.

O'Malley, M. & Chamot, A. (1990). Learning strategies in second language acquisition. Cambridge: Cambridge University Press.

Oxford, R. L. & Scarcella, R. C. (1994). Second language vocabulary learning among adults: State of the art in vocabulary instruction. *System*, *22*, 231–43.

Paivio, A (1986). Mental representations: a dual coding approach. Oxford. England: Oxford University Press.

Pappas, D. N., Skinner, C. H., & Skinner, A. L. (2010). Supplemental accelerated reading with classwide interdependent group-oriented contingencies. *Psychology in the Schools, 47*, 887–901.

Parker, R.I., Vannest, K.J., Davis, J.L., & Sauber, S.B. (2011). Combining non-overlap and trend for single-case research: Tau-U. *Behavior Therapy, 42*, 284–299.

Penno, J. F., Wilkinson, A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanaiton and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94,* 23-33.

Popkin, J., & Skinner, C. H. (2003). Enhancing academic performance in a classroom serving students with serious emotional disturbance: Interdependent group contingencies with randomly selected components. *School Psychology Review, 32*, 282-29.

Reddy, M. S. (2016). Importance of English Language in today's World. *International Journal of Academic Research, 4,* 179–184.

Regalla, M., & Peker, H. (2018). Yes they can. Demystifiying the effects of foreign language learning for students with special needs. Learning Language, 23, (1), 22-24

Roney, R. C. (1996). Storytelling in the classroom: Some theoretical thoughts. *Storytelling World, 9*, 7–9.

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,* 240–206.

Ruwe, K., McLaughlin, T. F., Derby, K. M., & Johnson, K. (2011). The multiple effects of direct instruction flashcards on sight word acquisition, passage reading, and errors for three middle school students with intellectual disabilities. *Journal of Developmental and Physical Disabilities, 23,* 241-255.

Seines, A., McLaughlin, T. F., Derby, K. M., & Weber, K. P. (2015). The effects of direct instruction flashcards on sight word skills of an elementary student with specific learning disability. *International Journal of Advances in Scientific Research, 1(3), 167–172.*

Senechal, M., & Cornell, E. H. (1993). Vocabulary acquisition thorugh shared reading experiences. *Reading Research Quarterly, 28(*4*),* 360-374.

Senechal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Child Language, 24,* 123-138.

Shippen, M. E., Houchins, D. E., Steventon, C., & Sartor, D. L. (2005). A comparison of twodirect instruction reading programs for urban middle school students. *Remedial and Special Education, 26,* 175–182.

Skinner, M., & Smith, A. (2011). Creating success for students with learning disabilities in postsecondary foreign language courses. *International Journal of Special Education,26,*42–57.

Sparks, R. L. (2012). Individual Differences in L2 Learning and Long-Term L1-L2 Relationships. *Language Learning, 62,* 5-27. https://doi.org/10.1111/j.1467-9922.2012.00704.x

Sparks, R.L., Ganschow, L., & Patton, J. (2008). L1 and L2 literacy, L1 and L2 aptitude, and L2 affective variables as discriminators among high- and low- Achieving, LD, and ADHD L2 learners In Language learners with special needs: An international perspective, ed. J.Kormos and E. H. Kontra, 11-35. Clevedon: Multilingual Matters.

Sparks, R., & Javorsky, J. (2000). Section 504 and the Americans with Disabilities Act: Accommodating the learning disabled student in the foreign language curriculum (an update). Foreign Language Annals, 33, 645–654. doi:10.1111/j.1944-9720.2000.tb00933.x

Sparks, R. L. (2009). If you don't know where you're going, you'll wind up somewhere else: The case of foreign language learning disability. *Foreign Language Annals, 42,* 7–26. doi:10.1111/j.1944-9720.2009.01005.x

Standish, K., McLaughlin, T. F., & Neyman, J. (2012). The effects of direct instruction flashcards and math racetrack on math facts with one elementary student. *Academic Research International, 2*(2), 241–246.

Van, T. (2009). The relevance of literary analysis to teaching literature in the EFL classroom. *English Teaching Forum 2-9,* 23–39.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. Applied Linguistics, 28, 46–65.

Webb, S. (2008). The effects of context on incidental vocabulary learning. Reading in a Foreign Language,20, 232–245.

Wight, S. M. S. (2015). Students with learning disabilities in the foreign language learning environment and the practice of exemption. *Foreign Language Annals, 48,* 39–55.

Wilbert, J. (2018). Package "Scan." Retrieved from: https://www.uni-potsdam.de/fileadmin01/projects/inklusion/scan/scan.pdf

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, doi 10.1177/0014402916658658

Wydell, T.N. & Kondo, T. (2003). Phonological deficit and the reliance on orthographic approximation for reading: a follow-up study on an English- Japanese bilingual with monolingual dyslexia. *Journal of Research in Reading, 26,* 33–48. doi:10.1111/1467-9817.261004

Appendix I Declaration of own Performance

Declaration of own performance

Article 1	
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.	
Research idea	Overall control
Current state of research	Overall control
Design/Measurements	Overall control
Study	Overall control
Data analysis	Overall control
Publication	With Co-authors

Article 2	
Barwasser, A., Urton, K., Grünke, M., Sperling, M., & Coker, Jr., D. L. (2021). Fostering word fluency of struggling third graders from Germany through motivational peer-tutorial reading racetracks. *Reading & Writing,*	
Research idea	Overall control
Current state of research	With Co-authors
Design/Measurements	Overall control
Study	Overall control
Data analysis	With Co-authors
Publication	With Co-authors

Article 3	
Barwasser, A., Hertel, S., & Gruenke, M. (2021). A sub-lexical-patterns training with racetracks on trained and untrained words in low-literacy German students with behavioral problems with and without learning disabilities. *Learning Disabilities. A Contemporary Journal, 19*(2), 143-159.	
Research idea	Overall control
Current state of research	Overall control
Design/Measurements	Overall control
Study	Overall control
Data analysis	Overall control
Publication	With Co-authors

Article 4	
Barwasser, A., Urton, K., Knaak, T. & Grünke, M. (2021). Intentional and incidental vocabulary acquisition in German L2 primary school students through multi-component storytelling. *Language Teaching Research.*	
Research idea	With Co-authors
Current state of research	Overall control
Design/Measurements	Overall control
Study	Overall control
Data analysis	Overall control
Publication	With Co-authors

Article 5	
Barwasser, A., Lenz, B., & Grünke, M. (2021). A multimodal storytelling intervention for improving the reading and vocabulary skills of struggling German-as-a-Second Language adolescents with learning and behavioral problems. *Insights Into Learning Disabilities, 18*(1), 29–51.	
Research idea	Overall control
Current state of research	Overall control
Design/Measurements	Overall control
Study	Overall control
Data analysis	Overall control
Publication	With Co-authors

Article 6	
Barwasser, A., Bracht, J., & Grünke, M. (2021_accepted). A storytelling approach on vocabulary, reading and letter sound fluency of struggling first graders with German as Second Language with and without behavioral problems. *Frontiers in Psychology*	
Research idea	Overall control
Current state of research	Overall control
Design/Measurements	Overall control
Study	Overall control
Data analysis	Overall control
Publication	With Co-Authors

Article 7	
Knaak, T., Gruenke, M., & Barwasser, A. (2021). Enhancing Vocabulary Recognition in English Foreign Learners With and Without Learning Disabilities: Effects of a Multi Component Storytelling Intervention Approach. *Learning Disabilities: A Contemporary Journal, 19*(1), 5–23.	
Research idea	With Co-authors
Current state of research	With Co-authors
Design/Measurements	With Co-authors
Study	With Co-authors
Data analysis	With Co-authors
Publication	With Co-authors

Article 8	
Barwasser, A., Knaak, T., & Gruenke, M. (2020). The effects of a multicomponent storytelling intervention on the vocabulary recognition of struggling English as a foreign language learners with learning disabilities. *Insights into Learning Disabilities, 17*(1), 35–53.	
Research idea	Overall control
Current state of research	With Co-authors
Design/Measurements	Overall control
Study	Overall control
Data analysis	Overall control
Publication	With Co-authors

Appendix J Declaration of Independence

Declaration of independence (according to §11 (1) 8)

“Ich versichere eidesstattlich, dass ich die von mir vorgelegte Dissertation selbständig und ohne unzulässige Hilfe angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit einschließlich Tabellen, Karten und Abbildungen, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keinem anderen Fachbereich zur Prüfung vorgelegen hat; dass sie noch nicht veröffentlicht worden ist, sowie dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Promotionsordnung ist mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Matthias Grünke betreut worden.”

"I affirm in lieu of an oath that I have prepared the dissertation submitted by me independently and without unauthorized assistance, that I have fully indicated the sources and aids used, and that I have identified in each case as borrowed those passages of the work, including tables, maps, and illustrations, which are taken from other works in wording or meaning; that this dissertation has not yet been submitted to any other department for examination; that it has not yet been published, and that I will not undertake such a publication before completion of the doctoral examination procedure. I am aware of the doctoral regulations. The dissertation I have submitted has been supervised by Prof. Dr. Matthias Gruenke."

Mönchengladbach, den 01.12.2021 Anne Barwasser

Appendix K List of Publications

*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*, 1373–1378.

*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.

*Hisgen, S., **Barwasser, A**., Wellmann, T., & Grünke, M. (2020). The effects of a multicomponent strategy instruction on the argumentative writing performance of low-achieving secondary students. *Learning Disabilities. A Contemporary Journal, 18*(1), 93–110.

Nobel, K., **Barwasser, A**., Melzer, C., & Grünke, M. (2020). Webbasiertes Distanzlernen. Erfahrungen im Rahmen einer Pilotstudie während der Covid-19-Schulschließungen zur digitalen Förderung der Aufsatzfähigkeiten von Schülerinnen und Schülern mit dem Förderschwerpunkt Lernen in der Sekundarstufe I. Zeitschrift für Heilpädagogik, 71. 465-479.

***Barwasser, A.,** Knaak, K., & Grünke, M. (2020). The effects of a multicomponent storytelling intervention on the vocabulary recognition of struggling English as a foreign language learners with learning disabilities. *Insights into Learning Disabilities, 17*(1), 35–53.

*Karnes, J., **Barwasser, A**., & Grünke, M. (2021). The effects of a math racetracks intervention on the single-digit multiplication facts fluency of four struggling elementary school students. *Insights into Learning Disabilities, 18*(1).

*Knaak, T., Grünke, M., & **Barwasser, A.** (2021). Enhancing vocabulary recognition in English foreign language learners with and without learning disabilities: effects of a multi component storytelling intervention approach. *Learning Disabilities. A Contemporary Journal, 19*(1).

***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

***Barwasser, A.,** Lenz, B., & Grünke, M. (2021). A multimodal storytelling intervention for improving the reading and vocabulary skills of struggling German-as-a-second-language adolescents with learning and behavioral problems. *Insights into Learning Disabilities, 18*(1), 29-51.

*Schröder, L., Urton, K., Vierbuchen, M., Schmitz, S., Knaak, T., Nobel, K., **Barwasser, A**., Grünke, M., Schulden, M., & Hillenbrand, C. (2021, under review). Digitale Förderung der Schreibkompetenz: Ein systematisches Literatur-Review. Empirische Sonderpädagogik

***Barwasser, A.,** Urton, K., Grünke, M., Sperling, M., & Coker, D. Jr. (2021). Fostering word fluency of struggling third graders from Germany through motivational peer-tutorial reading racetracks. *Reading & Writing*, https://doi.org/10.1007/s11145-021-10172-3

***Barwasser, A.,** Bracht, J., & Grünke, M. (2021). A storytelling approach on vocabulary, reading, and letter sound fluency of struggling first graders with German as second language with and without behavioral problems. *Frontiers in Psychology.*

***Barwasser, A.,** Urton, K., Knaak, K., & Grünke, M. (2021_accepted). Intentional and incidental vocabulary acquisition through multi-component storytelling. The case of German L2 primary school students. *Language Teaching Research.*

***Barwasser, A.,** Urton, K., Knaak, T., & Grünke, M., (2021_under review). A multi-faceted storytelling intervention on vocabulary of. German L2 secondary school students with learning disabilities with and without behavioral difficulties. *Learning Disabilities Quaterly.*

***Barwasser, A.,** Calhoon M.B., Knaak, K., & Grünke, M. (2021, under review). English foreign language vocabulary acquisition for adolescents with and without learning disabilities through a storytelling-program. *Learning Disabilities Research and Practice.*

***Barwasser, A.,** Dufrene, B., Schmitz, S., Nobel, K., Spilles, M., & Grünke, M. (2021, under review). Effects of a reading racetrack intervention and additional self-graphing on reading and behavior of struggling readers with emotional behavioral disorders and learning disabilities. *Journal of Behavioral Education.*

*Nobel, K., **Barwasser, A.,** Grünke, M., Asaro-Saddler, K., & Saddler, B. (2021_accepted). Using a Simultaneous Prompting Procedure to Improve the Qualitative Writing of Three Students with Learning Disabilities. *International Education Studies.*

***Barwasser, A.,** Hertel, S., & Grünke, M. (2021). The effect of a (sub)lexical intervention using peer racetracks on the reading performance of low-literacy students with behavior problems with and without learning disabilities. *Learning Disabilities. A Contemporary Journal, 19*(2), 143-159.

***Barwasser, A.,** Nobel, K., & Grünke, S. (2021_under review). Sub-lexical-units training on Reading of Students With Learning Disabilities, Behavioral Problems and German L2. *Learning Disabilities Research & Practice.*

***Barwasser, A.,** Hisgen, S., & Grünke, M. (2021_under review). Peer-racetracks on reading of L2 students with at-risk behavior. *Child Language Teaching and Therapy.*

***Barwasser, A.,** Nobel, K., & Grünke, M. (2021_under review). Peer-reading racetrack with and without self-graphing on reading and spelling of struggling 3rd grade readers with behavioral problems. *Theory and Research in Education.*

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

**peer-reviewed*