

# Effects of using curriculum-based measurement (CBM) for progress monitoring in reading and an additive reading instruction in second classes

Sven Anderson  · Jana Jungjohann  · Markus Gebhardt 

Published online: 17 January 2020  
© The Author(s) 2020

**Abstract** In this study, poor readers in second school year were selected from three schools ( $n=32$ ). Their reading skills were surveyed weekly using a CBM instrument over one school semester. Furthermore, they were supported by a five-week reading fluency instruction. The majority of students increased their weekly learning growth in regular teaching and in the phase with the additive instruction. The weekly learning growth was higher in the phase with the additive instruction in reading syllables ( $b=0.69$ ), reading words ( $b=0.44$ ) and reading comprehension at sentence level ( $b=0.45$ ) than without (reading syllables:  $b=0.49$ ; reading words:  $b=0.18$ ; reading comprehension:  $b=0.30$ ). Based on the results of this study, the benefit of CBM for adaptive reading instruction will be discussed.

**Keywords** Progress monitoring · Curriculum-based measurement (CBM) · Adaptive teaching · Reading fluency · Reading fluency instruction

---

S. Anderson (✉) · J. Jungjohann · Prof. Dr. M. Gebhardt  
Fachgebiet Entwicklung und Erforschung inklusiver Bildungsprozesse, Technische Universität  
Dortmund, Emil-Figge-Str. 50, 44227 Dortmund, Germany  
E-Mail: sven.anderson@tu-dortmund.de

J. Jungjohann  
E-Mail: jana.jungjohann@tu-dortmund.de

Prof. Dr. M. Gebhardt  
E-Mail: markus.gebhardt@tu-dortmund.de

## Effekte von Lernverlaufsdiagnostik und einer fünfwöchigen additiven Leseintervention in zweiten Klassen

**Zusammenfassung** In dieser Studie wurden leseschwache Schülerinnen und Schüler ( $n=32$ ) ausgewählt. Deren Lesekompetenzen wurden wöchentlich mit einem CBM-Instrument über ein Schulhalbjahr erhoben. Zusätzlich nahmen sie an einer fünfwöchigen Leseflüssigkeitsförderung teil. Die Mehrheit der Schülerinnen und Schüler konnte ihre Lesekompetenzen in der Phase ohne und mit der zusätzlichen Leseinstruktion steigern. Der wöchentliche Lernzuwachs war in der Phase mit der zusätzlichen Förderung im Silbenlesen ( $b=0,69$ ), im Wörterlesen ( $b=0,44$ ) und im sinnentnehmenden Lesen auf Satzebene ( $b=0,45$ ) höher als ohne die Förderung (Silbenlesen:  $b=0,49$ ; Wörterlesen  $b=0,18$ ; Leseverständnis:  $b=0,30$ ). Auf der Grundlage der Ergebnisse dieser Studie wird der Nutzen von CBM für den adaptiven Leseunterricht diskutiert.

**Schlüsselwörter** Lernverlaufsdiagnostik · Curriculumbasiertes Messen (CBM) · Adaptiver Unterricht · Leseflüssigkeit · Leseflüssigkeitsförderung

### 1 Introduction

Progress monitoring is a concept for identifying learning problems and for providing tailored instructions for each student. One type of progress monitoring is curriculum-based measurement (CBM; Fuchs 2017). CBM (German: Lernverlaufsdiagnostik; Klauer 2014) is a method of monitoring students' learning development over an extended period (e.g., a school semester or an entire school year). Students take short, standardized tests at regular intervals (e.g., weekly, monthly). Teachers can use the data for setting individualized learning goals and instruction planning (Stecker et al. 2005). The benefit of CBM for adaptive teaching in heterogeneous classes are currently being discussed with the goal of guaranteeing learning success for all students (Bremerich-Vos et al. 2017; Walter 2008). In adaptive teaching, teachers create learning opportunities to support students' individual learning needs. For this, teachers must have profound knowledge about learning development and effective instruction methods. Furthermore, teachers need ongoing information about individual students' learning progress (Vogt and Brühwiler 2014).

For reading, current results from PIRLS 2016 (Progress in International Reading Literacy Study; German: Internationale Grundschul-Lese-Untersuchung (IGLU); Bos et al. 2017) demonstrate the need for adaptive instructions and CBM in reading. At the end of class four, nearly 19% of primary school students do not have text-based reading comprehension skills. However, only 11% of primary school students receive additive reading support (ibid.). Without intensive reading support, low reading skills can cause major learning problems in secondary school (Wember and Greisbach 2018). Using progress monitoring with CBM instruments in primary school can support teachers to make individualized data-based decisions for adaptive teaching and offer poor reading students adequate learning opportunities (Schmidt and Liebers 2016). The aim of this study is to demonstrate a first verification of using

the CBM platform Levumi reading tests in a longitudinal section. For this purpose, reading performance of second-class students in regular teaching and in combination with an additive reading support will be surveyed over one school semester. The results provide information on the necessity of CBM for individualized opportunities in adaptive teaching of reading.

## 2 Reading development and reading instruction in primary school

In accordance with the educational standards in Germany, primary schools teach reading skills that enable children to read age-appropriate with comprehension (KMK 2005). For this, reading fluency has proven to be a central link between hierarchically low processes of visual word recognition and hierarchically high processes of comprehensive reading at sentence and text level (Samuels 2012). Reading fluency requires effective and efficient word recognition, which contributes to accurate, fast and expressive aloud reading. In German, most students have a high degree of reliability in word recognition at the end of primary school. Differences in word recognition separate good and poor readers (Knoepke et al. 2014). For quick and accurate word recognition, good readers use larger orthographic units like syllables or morphemes (Hasenäcker and Schroeder 2017). These students can then focus their cognitive resources on text comprehension. For phonological recoding, poor readers require more cognitive resources, which are then not available for comprehension (Perfetti 1985). For this group of students, reading fluency instruction can be a suitable way of improving reading skills (Gold et al. 2013). Students who have no or low text-based reading skills at the end of class four often lack the ability to read fluently until secondary school (Rosebrock et al. 2010).

In primary school, support programs in reading generally focus either on increasing reading fluency (Kuhn and Stahl 2003) or on strategies for comprehensive reading (Elleman et al. 2009). Reading fluency trainings have proven effective in increasing word recognition with a focus on syllables and morpheme structures (Müller et al. 2017; Ritter and Scheerer-Neumann 2009). As soon as word recognition routines have formed, reading aloud under repeated reading condition have shown to be particularly effective (Rosebrock and Nix 2006). For this, instructors often employ reading tandems that consist of a student with strong reading skills and one with poor reading skills. The stronger student tutors the poorer student. In the German-speaking countries, previous studies indicate the effectiveness of reading tandems in primary schools (Gold et al. 2013; Müller et al. 2013). For the improvement of reading comprehension, reading strategy knowledge is imparted (Munser-Kiefer 2012).

## 3 Curriculum-based measurement in reading

Since the 1980s, much research has been done in the field of CBM in the United States (Fuchs 2017). In Germany, CBM has been intensively discussed as well, especially in special education and educational psychology research (Klauer 2014). In

primary schools, reading curriculum-based measurement (R-CBM) and CBM maze are two types of CBM that are typically used to measure reading skills (Graney et al. 2010). With R-CBM, individual performance in oral fluency reading is measured by the number of correct read words over a period of one to five minutes using a word list or meaningful coherent text. R-CBM is typically used from first through to third school year. CBM maze was developed for screenings of students who can already read fluently and who focus on comprehensive reading. CBM maze is a group-based silent reading comprehension test that measures general reading comprehension. Each task contains single or multiple sentences in which a single word is deleted and must be replaced by the test-taker. For this purpose, the reader usually has a selection of several words available that are syntactically or semantically similar to the target word. Both CBM test methods provide valid measurement results of student achievement in reading (Ardoin et al. 2013). CBM maze is used from class three on. For both CBM tools, there are many studies assessing technical adequacy, which found strong alternative form reliability, moderate to strong criterion-related validity, and predictive validity (Graney et al. 2010; Shin et al. 2000).

Fuchs (2004) proposed research in CBM instruments in three stages. Stage one deals with the fulfilment of the quality criteria for status diagnostics. For stage two, longitudinal investigations test the usability as a CBM instrument. Stage three includes studies on implementation and professional handling in practice in order to test the benefits for the pedagogical practice. There is currently a lot of research at stage one. Research results for stages two and three are sparse. The utility of CBM depends on its successful implementation in school practice. Therefore, more research within stages two and three is needed (Voß and Gebhardt 2017). Furthermore, most CBM tests are based on the classical test theory (CTT). CBM tests should be assessed within the framework of item response theory (IRT) in order to make valid statements about dimensionality, sensitivity to change, and invariance of measurement. Because of the reliance on CTT when evaluating most CBM tests, these criteria are not yet sufficiently evaluated (Wilbert 2014).

To date, there are only a few CBM instruments in German-speaking countries that have been tested in longitudinal studies. For R-CBM with paper-and-pencil measures, Walter (2010) reported an average weekly learning growth in reading of 0.98 words per minute for second-class students, 1.12 words per minute for third-class students and 0.72 words per minute for fourth-class students. For CBM maze with paper-and-pencil measures, Walter (2011) observed an average weekly learning growth of 0.30 correctly assigned target words for second-class students, 0.44 for third-class students and 0.51 for fourth-class students. In a computer version, the average learning growth in the second class was 0.21 words, in the third class 0.32 words and in the fourth class 0.28 words (ibid.). In addition to the class level and the presentation form of the test, the timing of the measurements can also have an influence on the increase in learning. Christ et al. (2010) reported that learning growth rates are significantly higher in the fall than in the spring season, especially for second-class students. However, the results change depending on the difficulty of the test and the composition of the reference group. Therefore, similar CBM instruments can only be compared to a limited extent.

New research shows that CBM instruments can be administered as an online tool, which promises a higher usability in the field (Mühling et al. 2019). In Germany, there are multiple online instruments for CBM in the field of reading for which findings are available for stage one and two: the CBM online platform Levumi (Jungjohann et al. 2018a), the learning progress documentation Mecklenburg-Vorpommern within the framework of the Rügener Inclusion Model (Voß et al. 2016) as well as the CBM online platform quop (Förster and Souvignier 2011). So far, there are hardly any studies available on stage three for reading. Hebbeker and Souvignier (2018) have demonstrated positive overall effects of CBM instruments in third year reading instruction for quop over one school year. The use of additional feedback and support materials for teachers did not lead to any additional significant changes in reading performance. So far, few instruments offer support material and services in addition to testing. In a systematic review of CBM, Jungjohann et al. (2018c) report that within investigated instruments insufficient support was provided for teachers to link measurement results with concrete lesson planning. Overall, the supply of additional materials is still small. The use of CBM for adaptive teaching is particularly successful if additional material is available (Keuning et al. 2017).

## 4 Research questions

In heterogeneous classes, CBM can support adaptive teaching in a targeted manner if CBM instruments meet the quality criteria of status diagnostics (stage 1) and can sensitively measure changes in performance over time (stage 2). Based on this, we present two research questions in this study. First, we ask: (1) how do the Levumi reading tests correlate with a standardized reading comprehensive test at two measurement points? Next, we assess the longitudinal application of the reading tests of Levumi. Thus, our second question is: (2) can the Levumi reading tests measure learning development of students with low reading skills from second classes of primary schools over one school semester (14 measurement points in 20 weeks)? For this, we surveyed the weekly reading performance of students with low reading skills from second classes with Levumi reading tests. We measure the sensitivity to change of the web-based platform Levumi in regular lessons and in connection with additive reading instruction. The relationship between reading instruction and reading learning will be assessed and compared to an accompanying measurement phase in regular reading instruction.

## 5 Methods

### 5.1 Design and sampling

The present study took place over a total period of 20 weeks in the second half of the school year 2017/2018. During this period, the reading performance of all participating students was recorded at one pretest and one posttest. Each measurement used both the short version of the ELFE II instrument (Lenhard et al. 2017) and four

CBM reading tests in Levumi (Gebhardt et al. 2016). Six intervention groups were formed from the total sample using stratified random selection. The intervention groups consisted of the lowest performing readers and were formed on the basis of teacher surveys and ELFE II test results. In seven cases, extra students were identified as at-risk by their teachers and joined an intervention group. Each intervention group received Levumi tests for nine school weeks without additive support and then a five-week phase with an additive reading instruction. In a group setting with a maximum of six students, the students received the additive instruction as a reading fluency training twice a week (10 sessions of 45 min. each). The reading fluency training is based on the Levumi supporting materials (Jungjohann et al. 2017) and had a standardized procedure. It was designed from evidence-based methods of reading instruction. The training included improving reading accuracy and reading speed through exercises in syllable recognition and classification<sup>1</sup>. A high level of reading accuracy and reading speed are prerequisites for improving reading fluency (Fuchs et al. 2001). The training also included a reading aloud session in reading tandems. Over the entire study period, a trained person recorded the reading development of the intervention groups. Thus, for each intervention group, measurement results are available at 14 measurement points. After seven measurement points, all teachers simultaneously received an overview of the reading development of the students in the intervention groups recorded to date.

A total of  $N=146$  second-class students from three primary schools in North Rhine-Westphalia took part in the study. Of these students,  $n=114$  (thereof 51 female,  $M_{\text{age}}=7.98$ ;  $SD_{\text{age}}=0.55$ ; 98 students with immigration background) participated in the pre- and posttest with ELFE II and Levumi and were included in the calculations of this study. Each of the six intervention groups consisted of six students with low reading skills. In order to avoid teacher effects, we put the intervention groups together across classes. Participation in the study (i.e., doing the tests and participate the reading intervention) was voluntary. The parents were informed in advance in writing and asked for their consent. The data of  $n=4$  students in the intervention groups could not be included in the analyses due to the withdrawal of consent. In total, the learning development of  $n_{\text{intervention group of poor readers}}=32$  students (thereof 15 female,  $M_{\text{age}}=7.96$ ;  $SD_{\text{age}}=0.54$ ; 28 students with immigration background) from the six intervention groups could be analyzed for this study.

Based on the T-values of the overall test result with ELFE II, a comparison group with poor reading skills was selected from the overall sample. This group has the same number of students such as the intervention group ( $n_{\text{comparison group of poor readers}}=32$ ; thereof 12 female,  $M_{\text{age}}=8.11$ ;  $SD_{\text{age}}=0.64$ ; 29 students with immigration background). In an independent samples t-test, the mean differences between the students in the intervention group ( $M=33.53$ ;  $SD=4.86$ ) and the students in the comparison group ( $M=35.34$ ;  $SD=3.96$ ) were not statistically significant ( $t=1.64$ ;  $df=2$ ;  $p=0.11$ ). The remaining students mainly showed reading performance in the normal

---

<sup>1</sup> For example, the training included exercises for syllable synthesis and syllable segmentation of written words and pseudo words. The word material is based on levels 11 to 14 of the Kieler Leseaufbau (Dummer-Smoch and Hackethal 2016). The training materials will be published and will also be available free of charge on [www.levumi.de](http://www.levumi.de).

range in the ELFE II ( $M = 48.56$ ;  $SD = 5.37$ ). For this study, they were combined into the group of average readers ( $n_{\text{group of average readers}} = 50$ ; thereof 24 female;  $M_{\text{age}} = 7.89$ ;  $SD_{\text{age}} = 0.47$ ; 41 students with immigration background). Resulting, in this study, we divided the overall sample in three different groups of students: the intervention group of poor readers ( $n = 32$ ), the comparison group of poor readers ( $n = 32$ ) and the group of average readers ( $n = 50$ ).

## 5.2 Instruments

In this study, two instruments were used to measure the students' reading skills. The first was the paper and pencil short version of the reading comprehension test ELFE II (Lenhard et al. 2017). Second, multiple tests of the online platform Levumi ([www.levumi.de](http://www.levumi.de); Gebhardt et al. 2016) were used to measure reading fluency (R-CBM; read syllables: SiL-Levumi (Jungjohann et al. 2019); read words: WoL-Levumi; read pseudo words: Pseudo-WoL-Levumi) and reading comprehension at sentence level (CBM maze: SinnL-Levumi; Jungjohann and Gebhardt 2019). Previous research has tested the psychometric quality of the Levumi tests according to the IRT (Gebhardt et al. 2016; Jungjohann et al. 2018a, 2018b). The same difficulty level was used for all Levumi tests (Level 4). The Levumi tests are web-based and students do the tests on the screen. For each test, the platform randomly arrays the items to create an individual parallel test form for each student. The reading fluency tests (R-CBM) measure fluency by reading aloud individual syllables, words or pseudo words. Levumi tests are administered by a competent reader (e.g., research assistant, teacher). The test takes 60 seconds. Test-retest reliability of the syllable test is  $r = 0.85$  and within a period of eight weeks, the test is able to detect individual performance changes across multiple measurement points (Jungjohann et al. 2018b). Similar to the sentence level of the ELFE II, the students fill in missing words in a sentence for the Levumi reading comprehension test (CBM maze). The test takes 480 seconds<sup>2</sup>. In every task, the students choose from four options. The test can be administered across multiple devices (e.g., computers or tablets) simultaneously. The comprehension tests are also able to track significant performance changes over time within a period of three weeks (Jungjohann et al. 2018b).

## 6 Results

### 6.1 Research question 1: correlation between Levumi reading tests and standardized reading comprehension test ELFE II

Correlations were analyzed to show the relation between the Levumi reading tests and the standardized ELFE II reading comprehension test (see Table 1). Depending on the test and measurement point, there were positive correlations ( $p < 0.001$ ;  $df = 112$ ) between the raw values of the ELFE II and the raw values of the Levumi reading tests between  $r = 0.52$  and  $r = 0.81$ . In the pretest, there were positive correlations

<sup>2</sup> The test was shortened to 300 seconds under consistent reliability for the 2018/2019 school year.

**Table 1** Pearson correlation between Levumi reading tests and ELFE II reading comprehension test

Test	ELFE II word level Pre	ELFE II sentence level Pre	ELFE II total score Pre	ELFE II word level Post	ELFE II sentence level Post	ELFE II total score Post
SiL- Levumi Pre	0.65***	0.70***	0.71***	0.58***	0.65***	0.63***
SiL- Levumi Post	–	–	–	0.57***	0.59***	0.61***
WoL- Levumi Pre	0.73***	0.73***	0.77***	0.67***	0.72***	0.72***
WoL- Levumi Post	–	–	–	0.59***	0.65***	0.65***
Pseudo- WoL- Levumi Pre	0.60***	0.62***	0.64***	0.56***	0.63***	0.61***
Pseudo- WoL- Levumi Post	–	–	–	0.53***	0.59***	0.59***
SinnL- Levumi Pre	0.67***	0.81***	0.75***	0.62***	0.73***	0.69***
SinnL- Levumi Post	–	–	–	0.52***	0.64***	0.59***

*Note.* The focus of the study is on the Levumi reading tests. Therefore, the correlations between the pre-test results of ELFE II and the post-test results of the Levumi reading tests are not presented

\*\*\* $p < 0.001$

between  $r = 0.60$  and  $r = 0.81$ . In the posttest, there were positive correlations between  $r = 0.52$  and  $r = 0.65$ .

A comparison between the results of the pretest with Levumi and the results of the posttest with total score in ELFE II shows medium positive correlations between  $r = 0.61$  and  $r = 0.73$ . Similar results were found in the context of Levumi and the ELFE subtest at word and at sentence level.

## 6.2 Research question 2: measurement in reading performance over 14 measurement points

In order to make statements of the development of reading over the second half of the school year, the sums of the correctly solved items of the three groups were analyzed. Table 2 shows the mean values and standard deviations of the correctly solved items of the Levumi reading tests for pretest and posttest, separately for the intervention group of poor readers, the comparison group of poor readers and the group of average readers. In the posttest, the mean values in all reading tests are



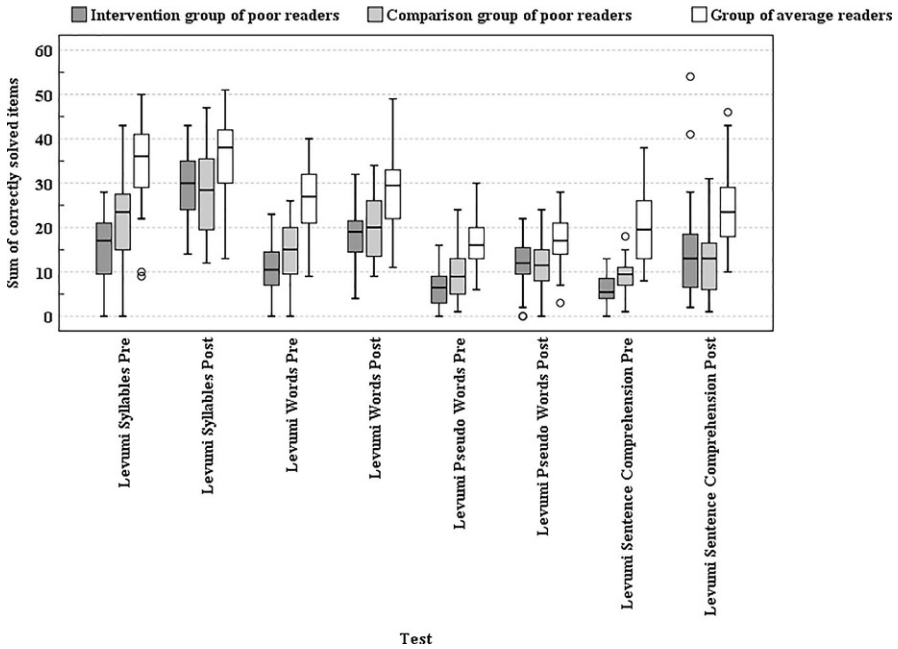
**Table 2** Mean values and standard deviations representing the Levumi reading tests SiL-Levumi, WoL-Levumi, Pseudo-WoL-Levumi, SinnL-Levumi and the total scores of the ELFE II for the intervention group, the comparison group and the group of average readers

Levumi tests										ELFE II test	
Group	Syllables		Words		Pseudo	Pseudo	Sentence	Sentence	Total	Total	
	Pre	Post	Pre	Post	Words Pre	Words Post	Compr. Pre	Compr. Post	raw score Pre	raw score Post	
Intervention group of poor readers ( $n = 32$ )	14.72 (7.38)	29.75 (7.42)	10.25 (5.47)	18.19 (7.07)	6.41 (4.48)	11.84 (6.02)	6.00 (3.63)	14.53 (10.91)	20.19 (7.34)	28.59 (8.68)	
Comparison group of poor readers ( $n = 32$ )	21.81 (10.51)	27.84 (9.47)	14.69 (7.32)	20.03 (7.25)	9.90 (5.86)	11.84 (5.57)	9.44 (3.66)	12.85 (7.40)	22.56 (6.26)	31.72 (10.23)	
Group of average readers ( $n = 50$ )	34.34 (8.77)	37.16 (8.41)	26.84 (7.19)	28.22 (8.41)	16.74 (5.55)	17.42 (5.63)	20.05 (8.92)	24.64 (9.29)	41.75 (8.86)	51.42 (9.11)	

higher than in the pretest. In the SiL-Levumi, all groups achieved the highest scores in both pretest and posttest. In contrast, in the Pseudo-WoL-Levumi test all groups achieved the lowest values in pretest and posttest.

The groups also differ within the individual reading tests. The group of average readers achieved the highest average values in all reading tests in the pretest and the posttest. The readers in the intervention group have achieved the lowest values in all reading tests in the pretesting. In posttest, the readers in the intervention group were able to achieve higher mean values than the poor readers in the comparison group in the SiL-Levumi and in the SinnL-Levumi.

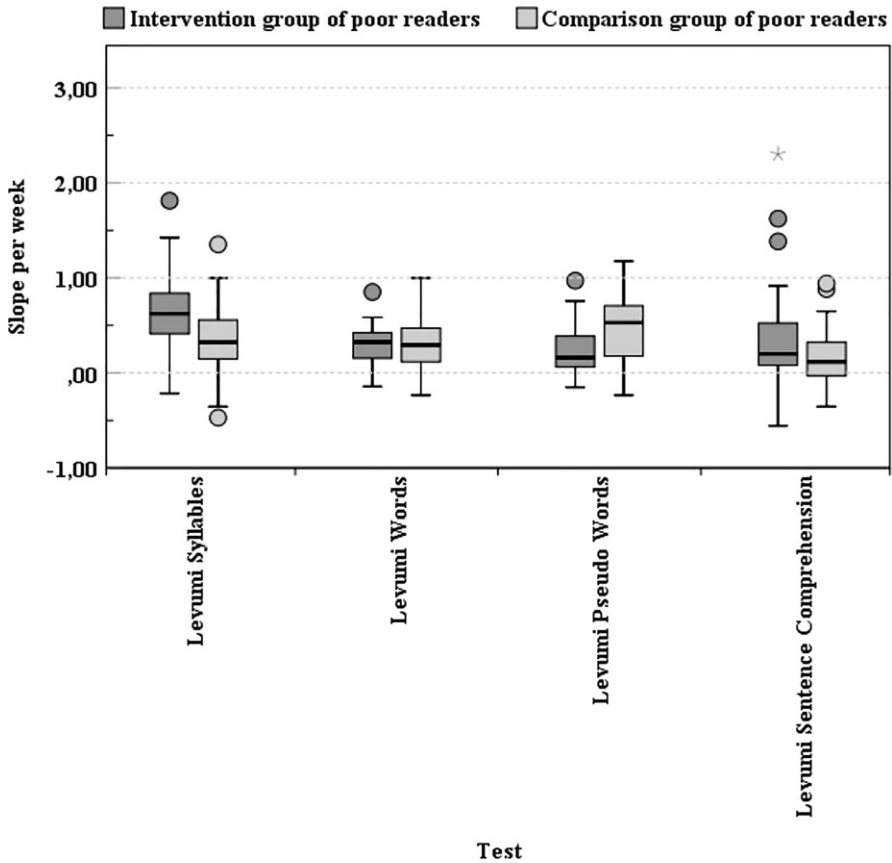
Both, the students in the intervention group and in the comparison group have a significant learning growth in reading with great effects over a school half-year. The average readers were able to significantly improve their reading performance in the SiL-Levumi and the SinnL-Levumi but not in the WoL-Levumi and the Pseudo-WoL-Levumi. For the SiL-Levumi, repeated-measures ANOVA showed a significant learning growth for all three groups (intervention group,  $F = 138.27$ ;  $df 1, p < 0.001$ ,  $\eta_p^2 = 0.817$ ; comparison group,  $F = 24.94$ ;  $df 1, p < 0.001$ ,  $\eta_p^2 = 0.446$ ; group of average readers,  $F = 10.84$ ,  $df 1, p = 0.002$ ,  $\eta_p^2 = 0.181$ ). For the WoL-Levumi there was a significant learning growth for the intervention group ( $F = 72.52$ ,  $df 1, p < 0.001$ ,  $\eta_p^2 = 0.701$ ) and comparison group ( $F = 37.51$ ,  $df 1, p < 0.001$ ,  $\eta_p^2 = 0.548$ ). The group of average readers did not significantly improve their reading performance in word reading ( $F = 3.20$ ,  $df 1, p = 0.08$ ). For the Pseudo-WoL-Levumi both groups, the intervention ( $F = 43.27$ ,  $df 1, p < 0.001$ ,  $\eta_p^2 = 0.583$ ) and the comparison group ( $F = 9.11$ ,  $df 1, p = 0.03$ ,  $\eta_p^2 = 0.102$ ) also improved significantly, while the group of average readers ( $F = 1.74$ ,  $df 1, p = 0.19$ ) did not. In SinnL-Levumi, all three groups have significantly improved their reading skills (intervention group,  $F = 22.24$ ,  $df 1, p < 0.001$ ,  $\eta_p^2 = .418$ , comparison group,



**Fig. 1** Distribution of the sum values of the Levumi reading tests SiL-Levumi, WoL-Levumi, Pseudo-WoL-Levumi and SinnL-Levumi for the intervention group of poor readers, the comparison group of poor readers and the group of average readers over the second half of the school year 2017/2018 ( $n=114$ )

$F=24.31$ ,  $df 1$ ,  $p<0.001$ ,  $\eta_p^2 = 0.263$  ; group of average readers,  $F=16.38$ ,  $df 1$ ,  $p<0.001$ ,  $\eta_p^2 = 0.285$  ). There are also big differences in reading skills within the individual groups. Fig. 1 shows the distribution of the sum values of the Levumi reading tests in the pretest and posttest divided into the intervention group, the comparison group and the average readers group. The median for all tests and groups is higher in the posttest than in the pretest. In syllable reading, the performance of the intervention group from pretest to posttest has improved to a greater extent than that of the comparison group. This result also applies to the Pseudo-WoL-Levumi. In WoL-Levumi and SinnL-Levumi, the median of the intervention group has come very close to the median of the comparison group over the same period.

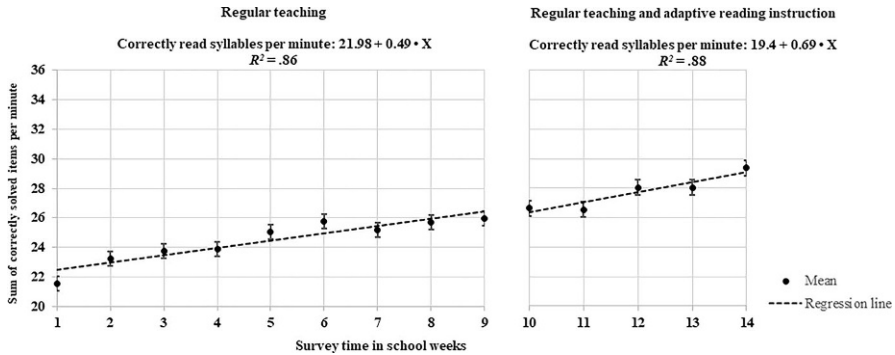
For the intervention group and the comparison group, the distribution of individual learning growth per week provides a first indication of the additional utility of additive reading instruction. Fig. 2 shows the distribution of individual learning growth for the Levumi reading tests for these two groups. In both groups, there is a broad trend in weekly learning growth. Moreover, in both groups, there was a small proportion of students without learning growth. In the intervention group, 31 students had a positive learning growth in the SiL-Levumi (30 in the WoL-Levumi, 30 in the Pseudo-WoL-Levumi and 25 in the SinnL-Levumi). In contrast, 27 students in the comparison group had a positive learning growth in the SiL-Levumi (28 in the WoL-Levumi, 28 in the Pseudo-WoL-Levumi and 21 in the SinnL-Levumi).



**Fig. 2** Distribution of individual weekly learning growth in the Levumi reading tests SiL-Levumi, WoL-Levumi, Pseudo-WoL-Levumi and SinnL-Levumi for the intervention group of poor readers and the comparison group of poor readers over the second half of the school year 2017/2018 ( $n=64$ )

In the SiL-Levumi, there was a significant difference in weekly learning growth between intervention group ( $M=0.64$ ) and comparison group ( $M=0.35$ ),  $F=7.650$ ,  $df\ 1$ ;  $p<0.001$ ,  $\eta_p^2 = 0.110$ . In the WoL-Levumi, no significant differences in the learning growth between the intervention group ( $M=0.30$ ) and the comparison group ( $M=0.31$ ) could be proven ( $F=0.041$ ,  $df\ 1$ ,  $p=0.84$ ). In Pseudo-WoL-Levumi, the students in the comparison group ( $M=0.49$ ) had a significantly higher learning growth than the students in the intervention group ( $M=0.25$ ),  $F=8.806$ ;  $df\ 1$ ,  $p<0.001$ ,  $\eta_p^2 = 0.124$ . In the SinnL-Levumi, no significant differences in learning growth between the intervention group ( $M=0.38$ ) and the comparison group ( $M=0.18$ ) were observed ( $F=2.457$ ,  $df\ 1$ ,  $p=0.12$ ).

A second indication of the benefit of additive reading intervention is a comparison between the weekly growth of the students in the intervention group in the regular teaching phase and the phase with regular teaching and additive reading instruction. Fig. 3 shows the trend of the mean scores of the SiL-Levumi of the lowest readers. In the SiL-Levumi, the average growth of correctly read syllables per measurement



**Fig. 3** Trend of mean sum scores of SiL-Levumi in the intervention group ( $n = 32$ )

point in the regular teaching phase is 0.49 syllables per measurement point ( $b = 0.49$ ,  $SE_b = 0.08$ ,  $R^2 = 0.86$ , 95% CI [0.31, 0.67]). On average, the students in the intervention group can increase by about one correctly read syllable over all two survey times by participating in regular teaching. In the phase with regular teaching and additive reading instruction, the learning growth in syllable reading increases to 0.69 per week ( $b = 0.69$ ,  $SE_b = 0.15$ ,  $R^2 = 0.88$ , 95% CI [0.23, 1.12]).

In the regular teaching phase, the students also improved on average in the Levumi tests WoL-Levumi ( $b = 0.18$ ,  $SE_b = 0.042$ ,  $R^2 = 0.73$ , 95% CI [0.08, 0.28]), Pseudo-WoL-Levumi ( $b = 0.25$ ,  $SE_b = 0.066$ ,  $R^2 = 0.67$ , 95% CI [0.09, 0.40]) and SinnL-Levumi ( $b = 0.30$ ,  $SE_b = 0.103$ ,  $R^2 = 0.54$ , 95% CI [0.052, 0.54]). In addition to SiL-Levumi, the students in the intervention group also had a higher learning growth in WoL-Levumi ( $b = 0.44$ ;  $SE_b = 0.106$ ;  $R^2 = 0.85$ ; 95% CI [0.10, 0.78]) and SinnL-Levumi ( $b = 0.45$ ;  $SE_b = 0.084$ ;  $R^2 = 0.91$ ; 95% CI [0.18, 0.72]) in the phase with additive reading intervention compared to the phase without additional intervention. Only for the Pseudo-WoL-Levumi, no growth in learning could be proven for the intervention phase ( $b = 0.31$ ;  $SE_b = 0.14$ ;  $R^2 = 0.62$ ; 95% CI [-0.14, 0.75]).

## 7 Discussion

In this study, the CBM online platform Levumi was able to present the learning processes of second-class students well over a period of 20 weeks. The students in the intervention group of poor readers, whose learning processes were monitored at high frequency with the Levumi online platform, had a higher learning growth in the Levumi reading fluency tests (R-CBM) SiL-Levumi (syllables), WoL-Levumi (words) and the reading comprehension test SinnL-Levumi (CBM maze) in the phase with additive reading instruction. It can also be positively emphasised that between 25 (SinnL-Levumi) and 31 (SiL-Levumi) of the students in the intervention group recorded a positive learning growth in reading. Nevertheless, even in this group, a few students did not have any growth in learning. The poor readers, who were not tested at high frequency with the Levumi tests and did not receive additive reading instruction, had lower learning growth in the same period than the students

had in the intervention group. These results illustrate the heterogeneous learning development within learning groups in primary school classes. On the positive side, a large number of poor readers were able to improve their reading skills with an additive reading instruction, especially as a lower growth in learning was to be expected during the instruction period due to the summer months (seasonal effect; Christ et al. 2010).

When interpreting the results, it must be taken into account that the study is based on a small sample. Moreover, the effectiveness of the additive reading instruction cannot be assessed due to the chosen study design. The focus of the current study is on the documentation of learning development over the second school semester.

The results show a strong dispersion of the individual learning growth within the phase of regular teaching and also within the phase of regular teaching and additive reading instruction. The lowest readers seem to benefit to varying degrees from regular teaching and additive reading instruction. For school practice, this means that a long-term use of CBM is an important basis for identifying struggling readers and developing individualized reading instruction in heterogeneous primary school classes (Bremerich-Vos et al. 2017). The CBM data not only provide information on the content of instruction, but also on whether instruction is effective. This data-based feedback can help teachers monitor the learning process of students with learning difficulties in smaller steps and plan appropriate instructions for adaptive teaching (Espin et al. 2017).

In addition, Levumi reading tests were shown to be valid instruments to measure reading fluency and reading comprehension. The covariations of the raw values of the Levumi reading tests and the corresponding values of ELFE II showed strong to medium significant relationships for the sample of  $n = 114$  students in second classes. As expected, the SinnL-Levumi correlates higher with the ELFE II subtest at sentence level than with the subtest at word level. Furthermore, the pretest results in Levumi show medium correlations with the posttest results of ELFE II. The Levumi reading tests therefore seem to have adequate predictive power in the documentation of reading development. The posttest results of the Levumi reading tests and ELFE II correlate less strongly than the pretest results. This is intentional, as the aim of the Levumi tests is to provide a broad picture of the lower performance range in particular.

Considering the chosen study design, another conclusion of the results of this study is that especially for students with low reading ability, the combination of regular teaching and additive instruction seems to have a positive influence on the development of reading fluency and reading comprehension at sentence level. In contrast to Hebbeck and Souvignier (2018), the results indicate that the combination of CBM with additive reading intervention can make a positive contribution to the learning success of primary school students with low reading skills.

The aim of the study was to demonstrate a first verification of the use of the CBM platform Levumi in longitudinal section with previously instructed surveyors and instructors. A first limitation is that a proof in school practice is still pending. A second limitation is that the students could not be observed any further, because the children went on their summer holidays after the posttest. Furthermore, this study deliberately resorted from a random sample and selected schools that agreed

to have learning development tested weekly by students over a half-year period. In this study, three schools with similar catchment areas were considered. These are characterized by a high degree of children with an immigrant background who had a great need for reading instruction.

The results show that the web-based CBM platform Levumi can map learning outcomes in reading for second-class students in a longitudinal section and achieves moderate till strong positive correlation with a standardized reading test. The inclusion and schooling of students with an immigrant background will further increase the heterogeneity of primary schools in reading (Bremerich-Vos et al. 2017). The results of this study are positive examples of the benefits of CBM for planning adaptive reading instruction in heterogeneous primary school classes. Thus, the study has created important prerequisites for researching the effectiveness of CBM and data-based decision making in combination with reading instruction in further studies.

**Funding** Open Access funding provided by Projekt DEAL.

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

## References

- Ardoin, S.P., Christ, T.J., Morena, L.S., Cormier, D.C., & Klingbeil, D.A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*(1), 1–18. <https://doi.org/10.1016/j.jsp.2012.09.004>.
- Bos, W., Valtin, R., Hußmann, A., Wendt, H., & Goy, M. (2017). IGLU 2016: Wichtige Ergebnisse im Überblick. 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* (pp. 13–28). Münster: Waxmann.
- Bremerich-Vos, A., Wendt, H., & Hußmann, A. (2017). Bausteine adaptiven Leseunterrichts angesichts gewachsener Heterogenität. 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* (pp. 298–314). Münster: Waxmann.
- Christ, T.J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading: an evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review, 39*(3), 447–462.
- Dummer-Smoch, L., & Hackethal, R. (2016). *Kieler Leseaufbau* (9th edn.). Kiel: Veris-Verlag.
- 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>.
- Espin, C.A., Wayman, M.M., Deno, S.L., McMaster, K.L., & de Rooij, M. (2017). Data-based decision-making: developing a method for capturing teachers' understanding of CBM graphs. *Learning Disabilities Research & Practice, 32*(1), 8–21. <https://doi.org/10.1111/ldrp.12123>.

- Förster, N., & Souvignier, E. (2011). Curriculum-based measurement: developing a computer-based assessment instrument for monitoring student reading progress on multiple indicators. *Learning Disabilities*, 9(2), 21–44.
- Fuchs, L. S. (2004). The past, present and future of curriculum-based measurement research. *School Psychology Review*, 33(2), 188–192.
- Fuchs, L. S. (2017). Curriculum-based measurement as the emerging alternative: three decades later. *Learning Disabilities Research & Practice*, 32(1), 5–7. <https://doi.org/10.1111/ldrp.12127>.
- 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.
- Gebhardt, M., Diehl, K., & Mühlhling, A. (2016). Online Lernverlaufsmessung für alle SchülerInnen in inklusiven Klassen. [www.LEVUMI.de](http://www.LEVUMI.de). *Zeitschrift für Heilpädagogik*, 67(10), 444–454.
- Gold, A., Behrendt, S., Lauer-Schmaltz, M., & Rosebrock, C. (2013). Förderung der Leseflüssigkeit in dritten Grundschulklassen. In C. Rosebrock & A. Bertschi-Kaufmann (Eds.), *Literalität erfassen: Bildungspolitisch, kulturell, individuell* (pp. 203–218). Weinheim: Beltz.
- Graney, S. B., Martínez, R. S., Missall, K. N., & Aricak, O. T. (2010). Universal screening of reading in late elementary school. *Remedial and Special Education*, 31(5), 368–377. <https://doi.org/10.1177/0741932509338371>.
- Hasenäcker, J., & Schroeder, S. (2017). Syllables and morphemes in german reading development: evidence from second graders, fourth graders, and adults. *Applied Psycholinguistics*, 38(3), 733–753. <https://doi.org/10.1017/S0142716416000412>.
- Hebbecke, K., & Souvignier, E. (2018). Formatives Assessment im Leseunterricht der Grundschule – Implementation und Wirksamkeit eines modularen, materialgestützten Konzepts. *Zeitschrift für Erziehungswissenschaft*, 21(4), 735–765. <https://doi.org/10.1007/s11618-018-0834-y>.
- Jungjohann, J., & Gebhardt, M. (2019). *SinnL-Levumi – Test der Leseflüssigkeit zur Lernverlaufsdagnostik – „Silben lesen“ der Onlineplattform www.levumi.de*. Trier: ZPID. <https://doi.org/10.23668/psycharchives.2462>. Verfahrensdokumentation aus PSYNDEX Tests-Nr. 9007767 und Silbenlisten.
- Jungjohann, J., DeVries, J. M., Gebhardt, M., & Mühlhling, A. (2018a). Levumi: a web-based curriculum-based measurement to monitor learning progress in inclusive classrooms. In K. Miesenberger & G. Kouroupetroglou (Eds.), *Computers helping people with special needs*. Lecture notes in computer science, (Vol. 10896, pp. 369–378). Cham: Springer. [https://doi.org/10.1007/978-3-319-94277-3\\_58](https://doi.org/10.1007/978-3-319-94277-3_58).
- Jungjohann, J., DeVries, J. M., Mühlhling, A., & Gebhardt, M. (2018b). Using theory-based test construction to develop a new curriculum-based measurement for sentence reading comprehension. *Frontiers in Education*. <https://doi.org/10.3389/educ.2018.00115>.
- Jungjohann, J., Diehl, K., & Gebhardt, M. (2019). *SiL-Levumi – Test der Leseflüssigkeit zur Lernverlaufsdagnostik – „Silben lesen“ der Onlineplattform www.levumi.de*. Trier: ZPID. <https://doi.org/10.23668/psycharchives.2462>. Verfahrensdokumentation aus PSYNDEX Tests-Nr. 9007767 und Silbenlisten.
- Jungjohann, J., Gegenfurtner, A., & Gebhardt, M. (2018c). Systematisches Review von Lernverlaufsmessung im Bereich der frühen Leseflüssigkeit. *Empirische Sonderpädagogik*, 10(1), 100–118.
- Jungjohann, J., Gebhardt, M., Diehl, K., & Mühlhling, A. (2017). *Förderansätze im Lesen mit LEVUMI*. Technische Universität Dortmund: Dortmund. <https://doi.org/10.17877/DE290R-18042>. Accessed May 14, 2019.
- Keuning, T., van Geel, M., & Visscher, A. (2017). Why a data-based decision-making intervention works in some schools and not in others. *Learning Disabilities Research & Practice*, 32(1), 32–45. <https://doi.org/10.1111/ldrp.12124>.
- Klauer, K. J. (2014). Formative Leistungsdiagnostik: Historischer Hintergrund und Weiterentwicklung zur Lernverlaufsdagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Eds.), *Lernverlaufsdagnostik* 1st edn. Tests and Trends, (Vol. 12, pp. 1–17). Göttingen: Hogrefe.
- KMK (2005). *Bildungsstandards im Fach Deutsch für den Primarbereich. (Jahrgangsstufe 4); [Beschluss vom 15.10.2004]*. Neuwied: Luchterhand. Beschlüsse der Kultusministerkonferenz
- Knoepke, J., Richter, T., Isberner, M.-B., Naumann, J., & Neeb, Y. (2014). Phonological recoding, orthographic decoding, and comprehension skills during reading acquisition. *Zeitschrift für Erziehungswissenschaft*, 17(3), 447–471. <https://doi.org/10.1007/s11618-014-0560-z>.
- 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>.
- Lenhard, W., Lenhard, A., & Schneider, W. (2017). *ELFE II – ein Leseverständnistest für Erst- bis Siebtklässler: Version II* (1st edn.). Hogrefe Schultests. Göttingen: Hogrefe.
- Mühlhling, A., Jungjohann, J., & Gebhardt, M. (2019). Progress monitoring in primary education using Levumi: a case study. In H. Lane, S. Zvacek & J. Uhoimobhi (Eds.), *CSEDU 2019. Proceedings*

- of the 11th international conference on computer supported education, 2–4 May, 2019, Heraklion, Greece (pp. 137–144). <https://doi.org/10.5220/0007658301370144>.
- Müller, B., Križan, A., Hecht, T., Richter, T., & Ennemoser, M. (2013). Leseflüssigkeit im Grundschulalter: Entwicklungsverlauf und Effekte systematischer Leseförderung. *Lernen und Lernstörungen*, 2(3), 131–146. <https://doi.org/10.1024/2235-0977/a000039>.
- Müller, B., Richter, T., Karageorgos, P., Krawietz, S., & Ennemoser, M. (2017). Effects of a syllable-based reading intervention in poor-reading fourth graders. *Frontiers in Psychology*, 8, 1635. <https://doi.org/10.3389/fpsyg.2017.01635>.
- Munser-Kiefer, M. (2012). Leseförderung mit (meta-)kognitiven Strategien in der Grundschule? Befunde und Schlussfolgerungen aus einer Interventionsstudie. *Zeitschrift für Grundschulforschung*, 5(2), 122–135.
- Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.
- Ritter, C., & Scheerer-Neumann, G. (2009). *PotsBlitz – Potsdamer Lesetraining. Förderung der basalen Lesefähigkeiten* (1st edn.). Köln: ProLog Therapie- und Lernmittel.
- Rosebrock, C., & Nix, D. (2006). Forschungsüberblick: Leseflüssigkeit (Fluency) in der amerikanischen Leseforschung und -Didaktik. *Didaktik Deutsch*, 20, 90–112.
- Rosebrock, C., Rieckmann, C., Nix, D., & Gold, A. (2010). Förderung der Leseflüssigkeit bei leseschwachen Zwölfjährigen. *Didaktik Deutsch*, 28, 33–58.
- Samuels, S. J. (2012). Reading fluency. Its past, present and future. In C. L. Z. Blachowicz, K. Lems & T. V. Rasinski (Eds.), *Fluency instruction. Research-based best practices* (2nd edn., pp. 3–16). New York: Guilford.
- Schmidt, C., & Liebers, K. (2016). Formatives Assessment an Grundschulen – Praxis und Bedingungsfaktoren. In K. Liebers, B. Landwehr, A. Marquardt & K. Schlotter (Eds.), *Lernprozessbegleitung und adaptives Lernen in der Grundschule. Forschungsbezogene Beiträge*. Jahrbuch Grundschulforschung, (Vol. 19, pp. 133–138). Wiesbaden: Springer VS.
- Shin, J., Deno, S. L., & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *The Journal of Special Education*, 34(3), 164–172.
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: review of research. *Psychology in the Schools*, 42(8), 795–819. <https://doi.org/10.1002/pits.20113>.
- Vogt, F., & Brühwiler, C. (2014). Adaptive Lehrkompetenz für den Umgang mit Heterogenität. *Lehren & lernen*, 40(8/9), 47–51.
- Voß, S., & Gebhardt, M. (2017). Editorial Schwerpunktthema: Verlaufsdiagnostik in der Schule. *Empirische Sonderpädagogik*, 9(2), 95–97.
- Voß, S., Blumenthal, Y., Mahlau, K., Marten, K., Diehl, K., Sikora, S., et al. (2016). *Der Response-to-Intervention-Ansatz in der Praxis. Evaluationsergebnisse zum Rügener Inklusionsmodell* (1st edn.). Münster: Waxmann.
- Walter, J. (2008). Adaptiver Unterricht erneut betrachtet: Über die Notwendigkeit systematischer formativer Evaluation von Lehr- und Lernprozessen und die daraus resultierende Diagnostik und Neudefinition von Lernstörungen nach dem RTI-Paradigma. *Zeitschrift für Heilpädagogik*, 59(6), 202–215.
- Walter, J. (2010). Lernfortschrittsdiagnostik am Beispiel der Lesekompetenz (LDL): Messtechnische Grundlagen sowie Befunde über zu erwartende Zuwachsraten während der Grundschulzeit. *Heilpädagogische Forschung*, 36(4), 162–176.
- Walter, J. (2011). Die Entwicklung eines auch computerbasiert einsetzbaren Instruments zur formativen Messung der Lesekompetenz. *Heilpädagogische Forschung*, 37(3), 106–126.
- Wember, F. B., & Greisbach, M. (2018). Zwischen individueller Leseförderung und inklusivem Unterricht – Didaktische Aspekte der Inklusion im Förderschwerpunkt Lernen. In R. Benkmann & U. Heimlich (Eds.), *Inklusion im Förderschwerpunkt Lernen* 1st edn. Inklusion in Schule und Gesellschaft, (Vol. 9, pp. 133–213). Stuttgart: Kohlhammer.
- Wilbert, J. (2014). Instrumente zur Lernverlaufsdiagnostik: Gütekriterien und Auswertungsherausforderungen. In M. Hasselhorn, W. Schneider & U. Trautwein (Eds.), *Lernverlaufsdiagnostik* 1st edn. Tests und Trends, (Vol. 12, pp. 281–308). Göttingen: Hogrefe.