Effects of Data-Based Decision-Making on Low-Performing Readers in General Education Classrooms: Cumulative Evidence From Six Intervention Studies

Martin T. Peters, MSc[1] [iD], Natalie Förster, PhD[1] [iD],
Karin Hebbecker, PhD[1], Boris Forthmann, PhD[1],
and Elmar Souvignier, PhD[1]

Abstract

In most general education classrooms in Germany, students with and without special educational needs are taught together. To support teachers in adapting instruction to these heterogeneous classrooms, we have developed learning progress assessment (LPA) and reading instructional materials, the *Reading Sportsman* (RS), in line with the theoretical framework of data-based decision-making, which has led to beneficial effects in several studies. However, data from these studies have not been analyzed to examine effects for low-performing readers. Data within and across six quasi-experimental studies conducted by our team in Grades 2 to 4 were reanalyzed to examine the effects of LPA on students whose reading fluency scores were at or below the 25th percentile ($n = 1,346$ students from 264 classes). In each study, students had been assigned to a control group (with business-as-usual instruction), an LPA group only, or an LPA-RS group (i.e., with teachers provided with LPA and the RS). Separate multilevel regression results were integrated by means of meta-analytical methods to investigate the consistency of results. Overall, findings from the single studies indicated no positive effects of LPA with or without the RS compared with the control group. The integrated analyses showed small positive effect trends on reading fluency and intrinsic reading motivation.

Keywords

data-based decision-making, learning progress assessment, differentiated reading instruction, reading comprehension, reading fluency

For teachers, adapting instruction to students' needs is a daily challenge, and this challenge increases when students in the same class show large interindividual differences, necessitating differentiated instruction (Mullis et al., 2017). Such situations are common in Germany because ratification of the Convention on the Rights of Persons with Disabilities by the UN has resulted in many students who were previously taught in special education classes now attending general education classes. Results from the latest Progress in International Reading Literacy Study (PIRLS) indicate that the large achievement gap between high- and low-performing students in Germany has widened further (Mullis et al., 2017). Moreover, the number of fourth graders in Germany with only basic reading skills has nearly doubled since 2001 and is exceeded by only four European countries. In light of the undeniable importance of reading skills for participating in the information society of the 21st century, and the increasing number of students for whom learning to read is challenging, teachers in general education

must pay special attention to low-performing students who may suffer from a reading or learning disability (LD). This is especially true in countries such as Germany, where students with or at-risk for LDs are not routinely identified or provided specialized instruction within schools. The responsibility for meeting the needs of these students lies primarily with the classroom teacher, with support from specialists occurring only rarely and on an unsystematic basis (Tarelli et al., 2012).

Over the past several years, our research team has examined the extent to which data-based decision-making (DBDM; Ikemoto & Marsh, 2007; Mandinach, 2012) can help classroom teachers manage the large within-class

[1]University of Münster, Germany

Corresponding Author:
Martin T. Peters, Institute of Psychology in Education, University of Münster, Fliednerstrasse 21, Münster 48149, Germany.
Email: martin.peters@uni-muenster.de

heterogeneity of students in general education classrooms. As a reliable basis for DBDM, we equipped teachers of Grades 2 to 4 with a tool for learning progress assessment (LPA) in reading (Souvignier et al., 2021), either alone or with *Reading Sportsman* (RS) reading instructional materials, that were peer-mediated and designed to enhance differentiation of instruction. Results revealed that implementation of LPA both with and without teachers' additional use of the RS resulted in improvements in the reading performance of students as a whole (Förster et al., 2018; Förster & Souvignier, 2014a, 2015). Unfortunately, these studies did not isolate effects for the most vulnerable students, that is, students who were struggling to acquire adequate reading skills. Examination of effects for this group of students is essential because research has revealed mixed results on the effects of differentiated instruction on low-performing students. For instance, Gersten et al. (2009) found little evidence for the effectiveness of providing differentiated instruction for struggling readers. Furthermore, D. Fuchs et al. (2010) observed that systematic differentiation was only rarely fully realized in general education classrooms, and that low-performing students often were not adequately supported (c.f., Pressley et al., 2001). However, Coyne et al. (2018) found that students who experienced reading difficulties benefited more from an intervention that enabled differentiating instruction compared with general classroom instruction. In their comprehensive literature review, Gersten et al. (2017) found that small group-differentiated reading interventions have a positive impact on reading fluency and reading comprehension. Given these mixed findings, we deemed it worthwhile to reexamine data from our previous studies to determine effects for low-performing readers. Before describing these studies in more detail, we discuss the theoretical framework underlying the use of DBDM as an approach for differentiating instruction in general education classrooms.

DBDM

Data-based decision making provides a general theoretical framework for providing teachers with an approach to successfully manage the large heterogeneity of student reading ability and to adapt instruction to students' needs. This process is defined as repeated systematic collection and analysis of data (also to monitor and evaluate instruction effects) that are used to guide a range of (instructional) decisions to help improve the success of students and schools (L. S. Fuchs, 2004; Ikemoto & Marsh, 2007).

Highlighting the potential of this framework, Keuning et al. (2019) found that primary school teachers who were trained in DBDM effected higher mathematics and spelling skills for their students than teachers who were not trained in DBDM. Similarly, Carlson et al. (2011) found that extensive teacher training on the interpretation and use of data

led to higher student achievement in mathematics and showed a positive effect trend on reading. At the same time, research also has highlighted that DBDM can be challenging for teachers, as it requires them to repeatedly assess and document student data (Mandinach, 2012), to correctly interpret (graphed) data (Espin et al., 2017), and to make instructional adjustments based on data (Förster et al., 2018; D. Fuchs et al., 2010; Stecker et al., 2005). Therefore, providing support for teachers at the different steps of DBDM is necessary (Espin et al., 2017; Kippers et al., 2018). For example, L. S. Fuchs and Vaughn (2012) proposed equipping teachers with both computer-based assessments and support for differentiating instruction. Providing teachers with such supports has been shown across several studies to yield positive effects on student learning (e.g., Förster et al., 2018; Förster & Souvignier, 2015; L. S. Fuchs et al., 1990; Keuning et al., 2019; Staman et al., 2017).

Data Collection for DBDM With LPA

To make timely instructional decisions within DBDM, teachers need assessments of students' learning progress that are highly feasible and easy to implement with limited resources. Thus, for general education, computer-based approaches have been recommended (Mandinach, 2012). In Germany, the digital platform *quop* (Souvignier et al., 2021) was developed to provide LPA for general education. In LPA, students complete a series of equivalent tests during the school year at intervals of 3 weeks, and teachers receive information about their students' progress. While LPA shares important aspects with the curriculum-based measurement (CBM) approach (e.g., repeated measurement of student progress using equivalent tests), single tests within LPA are designed to provide differentiated information about component processes of a respective skill (e.g., reading accuracy, reading fluency, and reading comprehension) to inform instructional decisions (Förster & Souvignier, 2011).

In previous research, teachers have judged the computer-based LPA assessments to be feasible and to provide useful information (Förster & Souvignier, 2015). Moreover, studies have shown that teachers who use LPA effect greater reading growth in their students than teachers who assess reading performance only once at the beginning of the school year. For instance, Förster and Souvignier (2014a) found that the use of LPA resulted in significantly higher reading comprehension compared with a control group (CG; $d = .24$). A similar effect of LPA occurred in another study using a different reading comprehension measure ($d = .18$; Förster & Souvignier, 2015). Furthermore, Förster et al. (2018) found that students whose teachers had access to a combination of LPA and RS had a significantly higher achievement in reading fluency compared with a CG ($d = .30$).

The Challenge of Differentiating Reading Instruction

The positive effects that emerge from providing teachers with progress data on student achievement are mainly ascribed to teachers' use of data to adapt instruction to students' needs (Förster et al., 2018; Schildkamp et al., 2019; van Geel et al., 2016). This is in line with findings showing that students' progress in reading depends on how well instruction matches students' abilities (Connor et al., 2004), and that reading fluency and reading comprehension instruction have specific effects on the targeted competences (Seuring & Spörer, 2010). It is also in line with research that suggests that there are different subtypes of developing readers, and that each group should receive instruction that is guided by assessment data and tailored to their needs. For instance, Buly and Valencia (2002) identified 10 profiles of students who failed state reading assessments, and Pierce et al. (2007) pointed to four clusters of students who did not read fluently. Both claimed that students in the different profiles or groups should receive assessment-based individualized instruction. In short, students with different reading skills need different reading instruction, and the teacher's task is to decide who needs what and to realize the differentiation.

Developmental models of reading comprehension (e.g., Perfetti et al., 2005) suggest that the ability to read and understand written text unfolds in sequential steps. Readers need to reach a certain level of word-recognition accuracy before becoming more fluent readers (Karageorgos et al., 2020), and they need to read at a sufficient pace to have enough cognitive capacity to understand what they read (Wolf & Katzir-Cohen, 2001). Reading accuracy, reading fluency, and reading comprehension can efficiently be fostered by different instructional approaches.

Reading accuracy can be effectively promoted by syllable-based reading, which can be employed as a bridge between single-letter recoding and automatic word recognition (Müller et al., 2017; Müller & Richter, 2017). Low-performing readers should be supported in decoding words accurately and quickly, freeing resources to focus on meaning. This can be fostered by (repeatedly) reading aloud, such as via instructional approaches like repeated reading (Samuels, 1979; Therrien, 2004). Once students have become fluent readers, they should be instructed to use reading strategies to increase their reading comprehension (National Institute of Child Health and Human Development, 2000). One example of an evidence-based method that includes teaching and learning reading strategies is reciprocal teaching (Palincsar & Brown, 1984).

Certainly, applying the optimal instructional method for each child in a classroom is challenging for teachers, even if assessment data are available. As early as the 1990s, the research team of Lynn and Douglas Fuchs evaluated different approaches for supporting teachers, including skills analysis (L. S. Fuchs et al., 1990), expert system instructional consultation (L. S. Fuchs et al., 1992), and instructional recommendations (L. S. Fuchs et al., 1994). Recent research in general education has shown that differentiation of instruction can be supported by providing teachers with assessment information and with instructional materials that is evidence based, feasible, and ready to use in entire classes (Förster et al., 2018; Hondrich et al., 2015). In addition, the use of instructional approaches that are (mainly) student centered and do not permanently require instruction or supervision by a teacher seems effective.

DBDM and Student Motivation

Although most research on DBDM has focused on achievement outcomes, DBDM has the potential to also influence students' motivation. From a theoretical perspective, DBDM might be expected to lead to improvement in students' reading motivation, as it satisfies basic psychological needs (Ryan & Deci, 2000). For example, seeing one's own improvement when looking at a progress graph might promote the experience of competence; working together with peers might increase feelings of relatedness; and being involved in decisions about next steps of instruction might satisfy the need for autonomy. However, two studies looking at motivational outcomes associated with student progress monitoring have shown mixed results. Förster and Souvignier (2014a) found that CG students and students whose progress was monitored by their teachers did not differ in their reading motivation, and that students' reading motivation even decreased when they reflected on their progress. In contrast, Calhoon and Fuchs (2003) found that working with graphs that visualized students' learning progress in mathematics increased student motivation to work hard.

Purpose and Research Questions

In sum, research has supported the use of DBDM as a method for differentiating instruction in general education classrooms, but effects of differentiation on low-performing students have been mixed. In our own research, we have found that LPA can assist general education teachers in differentiating classroom reading instruction. Results of this research have shown that teachers' use of LPA leads to improved reading performance for students as a whole; however, effects for low-performing readers have not been isolated. Given the doubts raised about the effectiveness of differentiation for low-performing students, examination of the specific effects of LPA for low-performing readers is in order.

This study represents a meta-analytic reanalysis of data across previous studies conducted to examine the effects of LPA on student reading performance. The specific aim of the reanalysis is to isolate effects of LPA with and without

Table 1. Grades, Measures, Conditions, and Sample Sizes of the Single Studies Ordered Per Grade.

Reference	Grade	Study	Sample	Measures		Conditions			n	
				RF/RC	IM/EM	CG	LPA	LPA-RS	Total	PR $\leq$ 25
Peters et al. (in press)	2	1	1	✓	✓	✓	✓	✓	619	150
Förster et al. (2018)[a]	3	2	2	✓	✓	✓	✓	✓	1,205	287
Hebbecker & Souvignier (2018)	3	3	4	✓	✓	—	✓	✓	936	221
Förster & Souvignier (2015)	3	4	5	✓	—	✓	✓	—	958	222
Förster et al. (2018)[a]	4	2	3	✓	—	✓	✓	✓	828	199
Förster & Souvignier (2014a)	4	5	6	✓	✓	✓	✓	—	599	145
Förster & Souvignier (2011)	4	6	7	✓	—	✓	✓	—	534	122
									5,679	1,346

Note. RF = reading fluency; RC = reading comprehension; IM = intrinsic reading motivation; EM = extrinsic reading motivation; CG = control group; LPA = learning progress assessment; LPA-RS = combination of learning progress assessment and *Reading Sportsman* (RS); PR = Percentile rank.
[a]Some of these data stem from a larger study by Kawohl (2015).

material-based support on the performance of students who struggle in reading. To achieve this aim, we analyzed data both within and across six similarly designed LPA intervention studies that our team had conducted to increase our sample size, and to obtain cumulative evidence as well as a reliable estimation of the overall effects for low-performing readers.

Our specific purpose was to examine whether equipping teachers with a computerized LPA tool positively affected the reading fluency, reading comprehension, and reading motivation of low-performing readers (Research Question 1). Further, we examined whether providing the RS to differentiate reading instruction increased the effects of LPA for low-performing readers (Research Question 2).

Our hypothesis was that the findings from our classwide studies would generalize to low-performing students and that the reading fluency, reading comprehension, and intrinsic (but not extrinsic) reading motivation development of students whose teachers had LPA information would exceed that of students whose teachers had less assessment information. This hypothesis was based on previous studies that have shown positive effects of progress monitoring on reading achievement (e.g., Stecker et al., 2005) and reading motivation (Calhoon & Fuchs, 2003). We further hypothesized that a combination of LPA and the RS would have additional positive effects on the outcomes of low-performing students.

Method

Participants and Design

We used data from six intervention studies conducted by our research team between 2011 and 2018 in general primary education, each study lasting the duration of one school year. All studies included a condition in which teachers were equipped with the LPA tool to monitor their

students' reading progress (LPA group). The LPA group was compared with a business-as-usual CG and/or a group in which teachers received a combination of the LPA tool and the RS to differentiate reading instruction based on students' LPA results (LPA-RS group). One study included students from Grades 3 and 4, which we treated as two separate samples in the analysis, resulting in seven independent samples from six studies with students from second to fourth grade (see Table 1).

In all studies, a quasi-experimental pretest-posttest design was realized. Assignment to the different conditions was made after informing teachers about the general study design and considering the schools' technical equipment and consent to use the LPA tool and the RS. LPA and LPA-RS conditions differed slightly across studies, as in some studies, teachers received additional teacher training on how to use the LPA or on how to differentiate instruction with the RS. To increase power, we considered all groups that used the LPA tool but had no access to the RS as the LPA condition, independent of additional trainings, and all groups that used the LPA tool *and* the RS as the LPA-RS condition. This decision was based on the DBDM process, which differentiates the phases of data assessment and interpretation (supported by the LPA tool) from the instructional decision-making phase (supported by the RS).

Reading fluency and reading comprehension were assessed in all samples, and reading motivation was assessed in four samples at the beginning and end of the school year. In three samples, the LPA condition was compared with a CG but not with the LPA-RS condition (see Table 1). All teachers received the results of the pretests shortly after data collection; thus, all teachers had access to student achievement data at the beginning of the school year.

Participation in the primary studies was voluntary, and informed consent was obtained from parents. Overall, *N* = 5,679 students from 264 classes participated in the studies (see Table 1). As no formal process of identifying students

Table 2. Demographic Information of Teachers of the Single Studies.

Reference	Grade	Study	Sample	Age: M (SD)	% Female	Years of teaching experience: M (SD)
Peters et al. (in press)	2	1	1	46.97 (10.07)	90.9	18.90 (9.15)
Förster et al. (2018)	3/4	2	2/3	50.41 (10.05)	85.7	24.11 (11.61)
Hebbecker & Souvignier (2018)	3	3	4	49.20 (9.50)	86.4	21.00 (10.30)
Förster & Souvignier (2015)[a]	3	4	5	—	86.6	—
Förster & Souvignier (2014a)	4	5	6	45.33 (9.85)	90.24	16.96 (9.63)
Förster & Souvignier (2011)[b]	4	6	7	—	100	—

Note. Age and years of teaching experience were collected in all studies except Förster and Souvignier (2015) and Förster and Souvignier (2011).
[a]Gender information was only available for 69.67% of teachers. [b]Gender information was only available for 75% of teachers.

with LDs exists within the German school system, we could not select formally identified students with or at-risk for LDs for our reanalysis. Thus, given the age of the participants in the studies, we chose to identify low-performing readers based on code-related reading difficulties, more specifically, based on reading fluency scores.

According to a common cutoff, we identified low-performing readers as students below the 25th percentile on a reading fluency test (Suggate, 2016). The resulting sample consisted of $N = 1,346$ students (M age = 8.82, $SD = 0.80$; 44.2% female) and $N = 264$ teachers. Most students spoke German (71.7%) or German and another language (15.9%) at home. Compared with standardized norm scores, the average percentile ranks of the different samples included in our analyses varied approximately between 10 and 16 or below 1.0 to 1.3 standard deviations, respectively (exact percentile ranks could not be determined due to deviating test times of the norming sample). The great majority of teachers was female (89.9%). On average, teachers were 47.97 years old and had 20.24 years of teaching experience (see Table 2).

LPA. Students in the LPA and LPA-RS conditions completed eight equivalent computer-based tests (10–15 min) at intervals of 3 weeks. Before the first test, students received instructions on how to use the tool and completed a subsequent tutorial. In the following weeks, teachers assisted students in logging into the system and then students completed the tests on their own. Tests for second graders assessed the efficiency of reading comprehension processes at the word, sentence, and text levels using a word/pseudoword discrimination task, a sentence verification task, and a text completion task, respectively (Förster & Kuhn, 2020). The tests in Grades 3 and 4 consisted of a maze task, in which every seventh word was deleted. After the maze task, students could see the complete text and answered a number of multiple-choice comprehension questions. Half of the questions could be answered using information that was explicitly presented in the text. The other half required students to draw inferences (Förster & Souvignier, 2011).

For all tests, teachers received differentiated assessment information in a teacher menu, including reading accuracy, reading fluency, and reading comprehension information. Individual results could be compared with a norm curve including all students who had completed the respective LPA tests. All words, sentences, texts, questions, and distractors were systematically constructed following the same strict rules for all eight tests in each grade, and the tests have been found to be equivalent (Förster & Kuhn, 2020; Förster & Souvignier, 2014b). For Grade 2, correlations of the tests with standardized achievement tests (word level: $r = .64$; sentence level: $r = .75$; text level: $r = .62$) were higher than correlations with a standardized test on mathematics ($r = .43$). Furthermore, retest reliability (word level: $.61 < \alpha < .72$; sentence level: $.65 < \alpha < .69$; text level $.65 < \alpha < .69$) and split-half reliability (word level: $.82 < \alpha < .89$; sentence level: $.86 < \alpha < .92$; text level: $.65 < \alpha < .83$) were acceptable to high. For Grades 3 and 4, correlations of the tests with standardized reading achievement tests ($r = .71$) were higher than correlations with standardized tests on mathematics ($r = .37$), and internal consistencies were acceptable to high ($.77 < \alpha < .90$).

Participation in the LPA testing was examined for the different studies. Due to illness or organizational reasons (e.g., class trip), the full LPA information across all eight tests was not available for every student for all teachers. The following percentages indicate how many of the students in the single studies completed all eight LPA tests during the school year: Study 1: 92.6%; Study 2: 86.8%; Study 3: 95.0%; Study 4: 64.7%; Study 5: 63.3%; Study 6: 77.5%. Taken together, fidelity was high across studies, and information about reading progress was available to teachers for most students at most measurement points.

RS reading instructional materials. Teachers in the LPA-RS condition received reading instructional materials called the RS, which has been developed by our team. The RS contained evidence-based reading methods to differentiate reading instruction in terms of reading accuracy, reading fluency, and reading comprehension. Furthermore, it consisted of

three methods, each with three levels of difficulty. In all methods, students with similar achievement levels worked in pairs. Teachers were encouraged to use the LPA reading accuracy, reading fluency, and reading comprehension results to differentiate reading instruction (i.e., to match students to the appropriate method and difficulty level and to create pairs) and to check the LPA results every 3 weeks and form new pairs if needed.

Within the RS, to train reading accuracy, a method named "reading slalom" was provided. In this method, students alternately read a series of words aloud after identifying and underlining syllables. While one student underlined the syllables and read the words out loud, the other student read along silently and gave feedback to his or her partner. Syllable-based reading was found to be effective to foster students' reading accuracy (Müller et al., 2017; Müller & Richter, 2017).

Then, to foster reading fluency, the RS provided the "reading sprinter" method. It was based on the method of repeated reading (Samuels, 1979). In this method, students read aloud a number of words or a text for a predefined time, taking turns as the silent reader (i.e., the trainer) or the oral reader (i.e., the sportsman). The trainer's task was to monitor the sportsman's reading and to note reading mistakes. After reading, both students counted the number of words the sportsman read correctly. After taking turns, the initial sportsman read the same text once again, and they again counted the reading mistakes and number of words read correctly. Overall, students took the sportsman and the trainer role twice for the same text. Reading the same text repeatedly can improve word recognition (LaBerge & Samuels, 1974), and help students to realize that they can improve through practice.

Then, RS provided students who could read fluently with the "reading tandem" method rather than the reading sprinter activities. The reading tandem was a paired version of the reciprocal teaching method (Palincsar & Brown, 1984). Before working with the reading tandem, teachers taught two to four reading strategies (e.g., recognizing and clarifying difficult words, summarizing, and making predictions). Then, two students worked together and took turns either reading a text aloud and applying reading strategies to the text (i.e., the student in the back of the tandem) or reading along silently and instructing the other student to use the reading strategies (student in the front of the tandem).

The fidelity with which the RS was implemented could be observed only in Study 2 (i.e., Samples 2 and 3). Trained research assistants rated how many aspects of the RS were implemented correctly in one lesson. The observations showed that all aspects of the reading sprinter were implemented correctly in all observed classrooms except one aspect in one classroom. Thus, especially the reading sprinter was implemented with high fidelity (Förster et al., 2018).

Measures

Across all primary studies and all conditions, students' reading fluency and reading comprehension were assessed at the beginning and end of the school year by trained student assistants using standardized achievement tests. In four samples, reading motivation was also assessed at both measurement points with a questionnaire (see Table 1).

Reading fluency. In Samples 2, 3, 5, 6, and 7, the Salzburger Lese-Screening to assess basal reading skills for Grades 1 to 4 (SLS 1–4; Mayringer & Wimmer, 2003) was used, and in Samples 1 and 4, a more recent version of the test that can be administered in Grades 2 to 9 was used (SLS 2–9; Wimmer & Mayringer, 2014). In both versions, a list of 70 sentences is presented, and students are asked to indicate within 3 minutes whether the sentence is correct (e.g., "Bananas are blue"). The sentences become increasingly longer and more complex. Both versions are highly reliable ($r_{xx}' \geq .90$) and valid ($r \geq .81$ with a reading-aloud measure). To compare contrasts across measures and samples, we used z-standardized sum scores in our analyses.

Reading comprehension. In Sample 1, reading comprehension was assessed using the ELFE II test (Ein Leseverständnistest für Erst- bis Siebtklässler; Lenhard et al., 2017). This test assesses reading comprehension at the word, sentence, and text levels. At the word level, students see 75 pictures and must select the corresponding and correct word for the picture out of four options within 3 minutes. At the sentence level, students read 36 sentences with a missing word in each sentence, and they have to select the missing word out of five choices. At the text level, students read 26 short texts and are asked to choose a sentence (out of four sentences) that best fits the story. Retest reliability was found to be satisfactory ($r = .78$), and the correlation with reading fluency was found to be high ($r = .91$). In Samples 5 and 7, a previous version of the same test was used (ELFE; Lenhard & Schneider, 2006). Internal consistencies for word, sentence, and text comprehension subtests of this test ranged from .86 to .96. With respect to decoding speed, criterion validity was high ($r = .79$).

In Samples 2, 3, 4, and 6, reading comprehension was assessed using a shortened version of the Hamburger reading comprehension test (HAMLET; Lehmann et al., 2006), as administration of the original test would have taken two lessons. This HAMLET-S (Förster & Souvignier, 2014a) consisted of two continuous nonfictional texts and one noncontinuous text (i.e., a table). In addition to the comprehension questions from the initial version of the test, six further questions were developed, considering the construction rules for items of the original version of the test, resulting in a total of 22 items. As for the original version, internal consistencies were high ($\alpha = .87$), and correlations with teacher

judgments of reading comprehension ranged from .58 to .65. Again, reading comprehension was scored and used in all samples as *z*-standardized sum scores.

Reading motivation. In Samples 2 and 6, the Habitual Reading Motivation Questionnaire (Möller & Bonerad, 2007) was used; it has five items to assess intrinsic reading motivation (e.g., "I enjoy reading books"; α = .71−.76) and four items to measure extrinsic reading motivation (e.g., "I love being the best at reading"; α = .75−.81). In Samples 1 and 4, the Reading Motivation Questionnaire (Stutz, 2014) was used, with four items measuring intrinsic reading motivation (e.g., "I read because I can imagine so much"; α = .70−.81) and three items assessing extrinsic reading motivation (e.g., "I read because it is important to me to always be the best in reading"; α = .78−.80). Items for both questionnaires and all subscales were rated on a 4-point Likert-type scale from 1 (*does not apply to me*) through 4 (*applies to me*). Intrinsic and extrinsic motivation scores were calculated as *z*-standardized averages across the items pertaining to each of the respective scales.

Data Analysis

Multivariate normality for all samples could be rejected according to Mardia's skewness coefficient (all *p*s < .002). Mardia's multivariate kurtosis coefficient was further significant in four out of the seven samples (all *p*s < .009) but not in Sample 3 ($b_{2,4}$ = 25.70; B = 1.03; p = .300), Sample 4 ($b_{2,8}$ = 81.91; B = 0.97; p = .330), and Sample 6 ($b_{2,8}$ = 82.73; B = 1.26; p = .210). The analysis of the missing data pattern revealed that the overall amount of missing data was below 10% for six samples, with a range from 2.33% (Sample 6) to 9.39% (Sample 1). Only Sample 2 had a considerably larger average percentage of missing data with 27.66%. This amount in Sample 2 was driven by a large amount of missing values on the motivational measures (all percentages > 44.25%). The percentages of missing values for each variable ranged between 0% and 17.33% across the other data sets. Given that multivariate normality was violated for all samples, it was further indicated by Jamshidian and Jalal's two-step procedure (Jamshidian et al., 2014) that there was not sufficient evidence to reject the assumption of missing completely at random in five samples because of a nonsignificant Hawkin's test (all *p*s > .067). For Sample 4, the same conclusion was drawn based on the combination of Hawkin's test and nonparametric test of homoscedasticity (Sample 4: Hawkin's test *p* = .012; nonparametric test of homoscedasticity *p* = .053). For Sample 6, the amount of missingness was too small (i.e., too few patterns of missing data) to apply Jamshidian and Jalal's procedure. Based on these missing data patterns, applying multiple imputation as technique for handling missing values was well justified.

Multiple imputation was implemented using the package mice (Van Buuren & Groothuis-Oudshoorn, 2011) for statistical software R (R Core Team, 2020). According to literature recommendations, we used 40 imputed data sets (e.g., Azur et al., 2011). To consider the effects of a multilevel clustering on standard errors (i.e., students were nested in classrooms), we used Arellano's method (Arellano, 1987) to estimate a robust covariance matrix (i.e., standard errors are on the diagonal) as it is implemented in the plm package (Millo, 2017). In addition, normality violations were addressed by relying on robust HC4 standard errors (Cribari-Neto, 2004). All estimated models used the pretest as a control variable when the posttest was predicted by group contrasts in a linear model framework. Of note, imputation was performed for all relevant variables in a sample, but linear model coefficients of the regression analyses were separately estimated for each dependent variable. The model coefficient *b* represents the effect size and was based on the same regression model in each of the primary studies. That is, in each sample, the dependent variable (T2) was predicted by its initial level (T1) and a dummy-coded (i.e., binary) independent variable to contrast the respective conditions (e.g., CG vs. LPA). Given that the dependent variable was standardized in each regression model, the regression coefficient for the dummy variable (i.e., the effect size) represents the mean difference in standardized units between the conditions when statistically controlling for the initial level of the dependent variable. Hence, the interpretation of the standardized model coefficient *b* is analogous to the interpretation of standardized mean differences (e.g., Cohen's *d*). According to Acock (2014), *b* < .20 is considered a weak, .20 < *b* < .50 a moderate, and *b* > .50 a strong effect. All regression analyses were also performed without outliers (see Table 4). Finally, estimates of contrasts between CG, LPA, and LPA-RS conditions and the respective standard errors were pooled by meta-analytical methods as implemented in metafor (Viechtbauer, 2010).

Results

In the following, we will first report the results of the comparison of the CG and LPA conditions, before we present the findings for the comparison of the CG and the LPA-RS conditions. In each of the paragraphs, we first report the findings for the single studies and then refer to the meta-analytical summary of the primary studies.

The regression analyses for the single studies revealed no significant differences between the CG and LPA conditions in reading fluency, reading comprehension, or reading motivation (Research Question 1, see Tables 3 and 4). Similarly, the meta-analytical summary of the primary studies revealed no significant differences between the CG and LPA conditions and relatively low effect sizes were found.

Table 3. Parameters and Standard Errors of the Regression Analyses of the Single Studies Including Only Low-Performing Students.

Grade	Study	Sample	Contrast	RF b	RF SE	RC b	RC SE	IM b	IM SE	EM b	EM SE
2	1	1	CG vs. LPA	-.02	.20	.02	.25	.21	.20	-.11	.19
			CG vs. LPA-RS	-.08	.21	-.35	.23	.16	.21	.05	.23
			LPA vs. LPA-RS	-.06	.24	-.36	.22	-.04	.22	.16	.22
3	2	2	CG vs. LPA	.17	.13	.24	.15	.29	.31	-.03	.32
			CG vs. LPA-RS	.48**	.14	-.05	.17	.46	.40	-.11	.42
			LPA vs. LPA-RS	.32*	.14	-.29*	.14	.17	.19	-.09	.21
3	3	4	LPA vs. LPA-RS	.02	.13	-.22	.13	-.30	.15	-.08	.13
3	4	5	CG vs. LPA	.06	.12	.07	.10	—	—	—	—
4	2	3	CG vs. LPA	.17	.18	.21	.19	—	—	—	—
			CG vs. LPA-RS	.04	.19	.10	.20	—	—	—	—
			LPA vs. LPA-RS	-.13	.16	-.11	.18	—	—	—	—
4	5	6	CG vs. LPA	.10	.11	.16	.15	-.03	.15	-.24	.16
4	6	7	CG vs. LPA	.08	.15	.11	.15	—	—	—	—

Note. RF = reading fluency; RC = reading comprehension; IM = intrinsic reading motivation; EM = extrinsic reading motivation; CG = control group; LPA = learning progress assessment; LPA-RS = combination of learning progress assessment and *Reading Sportsman* (RS).
*p < .05. **p < .001. Results are ordered by grade.

Table 4. Parameters and Standard Errors of the Regression Analyses of the Single Studies Without Outliers, Only Including Low-Performing Students.

Grade	Study	Sample	Contrast	RF b	RF SE	RC b	RC SE	IM b	IM SE	EM b	EM SE
2	1	1	CG vs. LPA	-.09	.19	-.01	.21	.15	.20	-.14	.19
			CG vs. LPA-RS	-.09	.22	-.43*	.22	.22	.22	.07	.22
			LPA vs. LPA-RS	.00	.22	-.43*	.18	.06	.23	.22	.22
3	2	2	CG vs. LPA	.19	.13	.20	.16	.27	.26	-.02	.31
			CG vs. LPA-RS	.47***	.14	-.15	.17	.49	.38	-.10	.45
			LPA vs. LPA-RS	.28*	.13	-.35*	.14	.22	.22	-.08	.22
3	3	4	LPA vs. LPA-RS	-.05	.14	-.22	.13	-.30	.15	-.08	.13
3	4	5	CG vs. LPA	.05	.12	.06	.10	-	-	-	-
4	2	3	CG vs. LPA	.13	.18	.17	.20	-	-	-	-
			CG vs. LPA-RS	.02	.19	.08	.20	-	-	-	-
			LPA vs. LPA-RS	-.10	.16	-.09	.19	-	-	-	-
4	5	6	CG vs. LPA	.10	.13	.12	.15	-.07	.16	-.26	.16
4	6	7	CG vs. LPA	-.01	.17	.12	.16	-	-	-	-

Note. Results are ordered by grade. RF = reading fluency; RC = reading comprehension; IM = intrinsic reading motivation; EM = extrinsic reading motivation; CG = control group; LPA = learning progress assessment; LPA-RS = combination of learning progress assessment and *Reading Sportsman* (RS).
*p < .05. **p < .001.

However, the results showed effect trends in the expected directions. For example, low-performing students of the LPA treatment group showed trends toward a greater learning growth in reading fluency ($b = .10$, 95% confidence interval [CI] = [−.20, .40]) and reading comprehension ($b = .13$, 95% CI = [−.18, .45]) than low-performing students in the CG. Likewise, students of the LPA group tended to have a greater growth in intrinsic reading motivation ($b = .12$, 95% CI = [−.39, .63]) and less growth in extrinsic

reading motivation ($b = −.15$, 95% CI = [−.66, .36]) than students who received regular reading instruction (see Figure 1).

When investigating whether the combination of LPA and RS had additional effects on the reading achievement and reading motivation of low-performing students (Research Question 2), only findings from Sample 2 revealed significant results (see Tables 3 and 4): In this sample, students in the LPA-RS group showed significantly greater learning
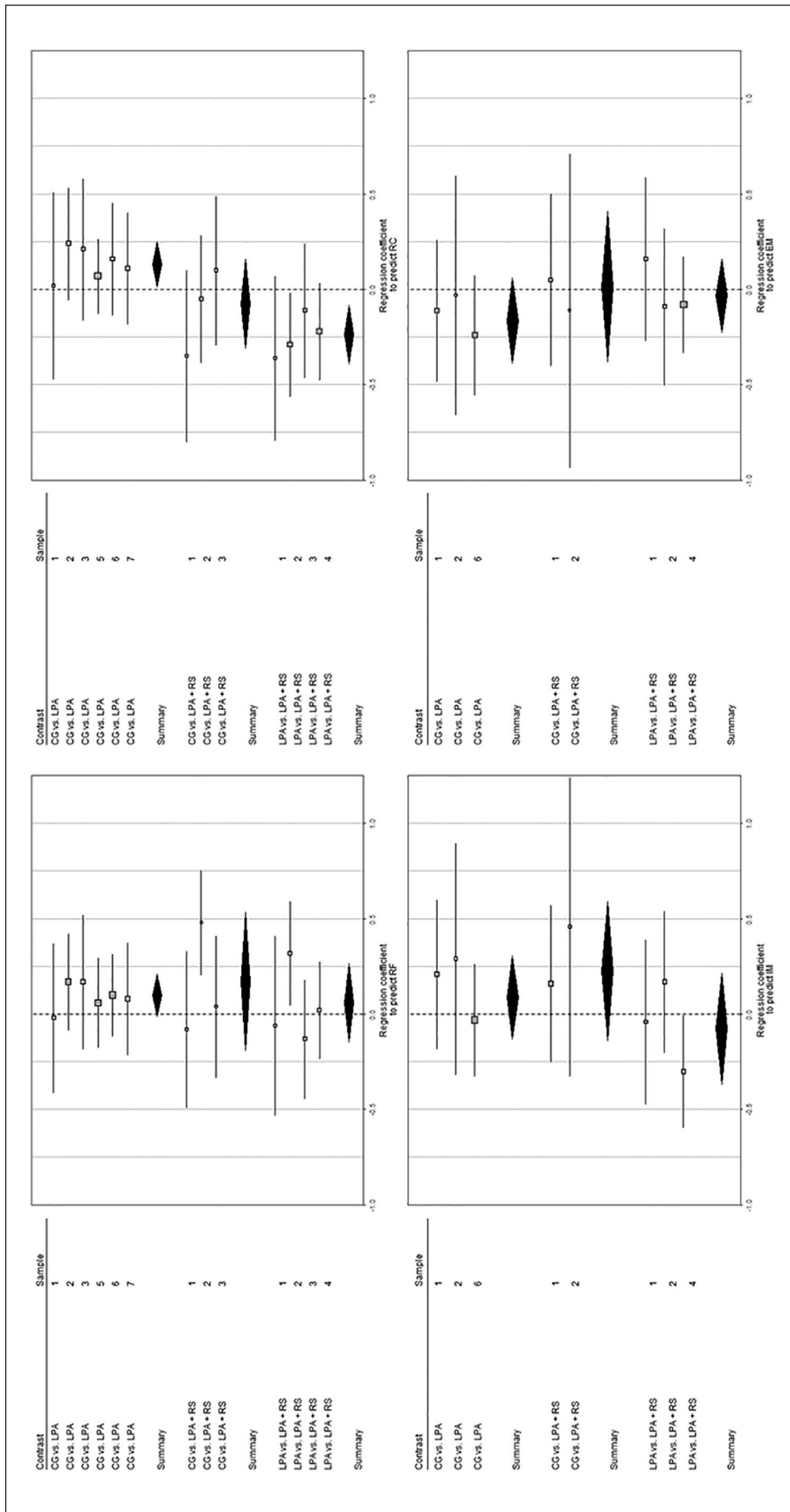
Figure 1. Forest plots displaying results of the single studies and meta-analyses only including low-performing students.
Note. The size of the squares represents the weight received by a study based on its inverse of the squared standard error. RF = reading fluency; RC = reading comprehension; IM = intrinsic reading motivation; EM = extrinsic reading motivation; CG = control group; LPA = learning progress assessment; LPA-RS = combination of LPA and *Reading Sportsman* (RS).

growth in reading fluency than the CG ($b = .48$, $SE = .14$, $p < .001$) and the LPA group ($b = .32$, $SE = .14$, $p = .02$). For reading comprehension, however, students in the LPA group showed significantly more growth than those in the LPA-RS group ($b = -.29$, $SE = .14$, $p = .04$).

The meta-analytical approach revealed no significant differences between the LPA-RS and the CG or LPA conditions, respectively. However, some effect trends were found again (see Figure 1). Compared with the CG, low-performing students in the LPA-RS group tended to show more growth in reading fluency ($b = .19$, 95% CI = [−.28, .66]) and intrinsic reading motivation ($b = .26$, 95% CI = [−.46, .99]). Surprisingly, low-performing students of the LPA group seemed to benefit more in reading comprehension than low-performing students whose teachers had access to both LPA and the RS ($b = -.24$, 95% CI = [−.63, .15]).

Discussion

In this study, we reanalyzed data from previous studies to determine whether the positive effects found for the implementation of a DBDM approach in general education can be generalized to low-performing readers. More specifically, we merged and reanalyzed data from seven samples drawn from six previous studies and used a meta-analytic approach to examine effects of low-performing readers. We differentiated effects of LPA with or without the RS, and we examined both achievement and motivation outcomes. Overall, the results of our analyses revealed that the use of LPA tended to be associated with higher reading growth, but effects were nonsignificant and smaller than those found in the larger samples of all students. For instance, the effect sizes for reading fluency and reading comprehension varied between $d = .12$ and $.24$ for the whole class, but were found to be $b = .10$ and $.13$ for reading fluency and reading comprehension, respectively, for low-performing students. The low-performing students, however, tended to benefit from LPA with respect to their intrinsic reading motivation ($b = .12$), whereas no such effect was found for the whole class.

When combined with the RS, effects were still nonsignificant for low-performing readers, but effect sizes for reading fluency increased slightly to $b = .19$, indicating that providing teachers with reading instructional materials might be a promising first step to improve low-performing students' reading fluency and reading motivation. Surprisingly, the combination of LPA with the RS may not be sufficiently intensive to accelerate the reading comprehension of students initially low-performing. This pattern of findings suggests that the successful implementation of DBDM via the combination of LPA and the RS may lead to differentiation of instruction in general education classrooms and an overall improvement in achievement. However, these results may not apply to low-performing students, including those with or at-risk for LDs in reading.

The positive trends in the data suggest that providing teachers with LPA and the RS to differentiate instruction might be a small step in the desired direction, yet the lack of significant results suggests that effecting performance gains in students with reading difficulties requires much more than classroom-level differentiated instruction. In accordance with the observation by D. Fuchs et al. (2010), our results would suggest that these students are likely in need of more intensive, specific, and structured instruction. For instance, Wanzek and Vaughn (2007) identified at least 100 sessions of reading intervention (i.e., 20 weeks of daily intervention) as a proxy for interventions that have a positive impact on student achievement. In our studies, low-performing students might not have worked enough with the RS to reveal large positive effects on their reading achievement.

Possible Explanations of the Findings

One possible explanation for the nonsignificant differences between conditions and for the relatively low effect sizes may be that the methods used to promote reading accuracy, reading fluency, and reading comprehension were—apart from being evidence based—also selected based on the extent to which they were student centered and could be implemented in pairs in the classroom. This decision was primarily based on the consideration that in almost all classrooms, only one teacher is responsible for all students. Thus, if instruction is not directed to the whole classroom at once and is to be differentiated for students with different abilities, students need to complete the exercises independently. Studies, however, show that student-centered approaches are less effective for subsamples of low-performing students (Connor et al., 2004).

Accordingly, pairs of students that were formed by teachers and who worked together in our studies were supposed to have similar reading levels, which implies that the students with the poorest reading skills worked together. This may have resulted in the low-performing students being overwhelmed with the task of doing the reading exercises *and* monitoring their partners' reading progress. Overall, the DBDM approach provides the chance to adapt instruction to students' needs and is, on average, successful for promoting students' reading abilities. Moreover, particularly when average- and high-achieving students work together in pairs, the teacher gets the opportunity to focus on the low-performing students and provide teacher-guided instruction. Yet, our findings indicate that it would be useful to make teachers more aware of the special needs of their low-performing students and to include intensive, highly structured, and teacher-guided reading instruction into the

RS. Such an approach would help students apply the methods correctly and would thus be beneficial for fidelity.

One method that was likely implemented with high fidelity is the reading sprinter. In one of our studies investigating the effects of LPA and a combination of LPA and the RS, classroom observations showed that this method was implemented as intended (Förster et al., 2018). The reading sprinter employs the method of repeated reading, which has been shown to positively affect students' reading fluency and self-concept (Lee & Yoon, 2017; Samuels, 1979; Therrien, 2004). Given that the method is intended to foster basal reading skills, it might be more suitable for low-performing readers in primary school than the reading tandem method, which involves monitoring or using reading strategies. Nevertheless, the effects of different instructional approaches should be further studied, as some studies indicate that repeated reading might not be the appropriate method for all low-performing students (e.g., Chard et al., 2009). For instance, it should be investigated whether repeated reading is more effective when a low-performing reader works together with a higher performing fluent model.

Information on students' reading progress only leads to positive effects if the data are actually used to inform practice (Espin et al., 2017); however, research also indicates that using data for instructional decision-making is the most ambitious step in the DBDM process (Kippers et al., 2018; Stecker et al., 2005). Furthermore, it is especially difficult to realize a systematic differentiation of instruction in heterogeneous classes when general education teachers are not supported by special educators (D. Fuchs et al., 2010; L. S. Fuchs & Vaughn, 2012). In this context, our results resemble past findings showing that DBDM is a challenge for teachers, and they need further support in incorporating it into their daily routines (Spear-Swerling & Cheesman, 2012; Staman et al., 2017). To be more concrete, support for teachers needs to address (a) working with assessment data, which is usually not a central component in teacher education (Darling-Hammond, 2006) and (b) the use of data to adapt instruction, which is not as widespread as it could be (Brunner et al., 2005). Moreover, teacher training should also (c) focus on making teachers more familiar with evidence-based methods for reading instruction, which are only rarely used in everyday instructional practice in Germany (Bremerich-Vos et al., 2017).

Limitations

When interpreting the findings of this study, it is important to consider the distinction between implementing programs *as intended* versus *as conducted* (Century & Cassata, 2016). The objective of many studies is to investigate the effects of programs, given the precondition that the core components of the innovative concepts were implemented *as intended*, that is, when they are implemented with high treatment fidelity (c.f. Stecker et al., 2005). The second perspective, in contrast, is more concerned with evaluating the effects of an implementation *as conducted*, that is, to evaluate what actually happened in schools and to what extent desired effects of the transferred concept are attained. Our study focuses on the second approach, as teachers were asked to use the provided concepts in realistic practical settings, and no further support was provided (e.g., personnel) apart from the LPA tool and the RS. Clearly, evaluating implementation as conducted increases the ecological validity but is usually associated with a lack of experimental control and a decrease of internal validity.

Given that we could not realize a consistent control of the general instructional processes across all schools for all six studies, we do not know to what extent LPA and the RS were used as intended. Classroom observations were only conducted in Study 2 (i.e., Samples 2 and 3), which should be considered as a limitation of our study. Yet, the available data show that LPA information was available for most of the students at all measurement points, and the classroom observations indicated that teachers were using the RS as intended. To better understand why the effects were found to be quite low, systematic classroom observations would be helpful in future studies. Another limitation is that the LPA tests between Grade 2 and Grades 3 and 4 differed, which could be a reason for different effects between different grades.

In interpreting the results, readers must also keep in mind that all six studies were conducted in Germany. Assumptions about reading instruction in the CG are based on representative data from an international study (Tarelli et al., 2012) and a recent observational study that indicates that teachers of second-grade classrooms in general primary school in Germany only rarely use evidence-based methods in everyday instructional practice (Peters et al., 2021). As previously described, general education in Germany is typically characterized by one teacher being responsible for the entire class with little and unsystematic support from other professionals (e.g., special education teachers, school psychologists). While the average class size in Germany (i.e., 22 students to one teacher) is similar to the average class size in the EU, the percentage of teachers who lack support from additional specialized professionals is much higher in Germany (Tarelli et al., 2012). Therefore, our findings might not be generalizable to education systems in which more professional resources are available.

A limitation of our study concerns the fact that information about how many students of our sample were diagnosed with a reading or learning disability is missing for the primary studies. Yet, the estimated average percentile ranks of the different samples (percentile ranks 10–16) indicate that our samples most likely included a substantial number of students who would be considered reading disabled according to Suggate (2016).

It should be noted that all analyses did not correct for unreliability of the measures (e.g., Schmidt & Hunter, 1996); consequently, the estimated group contrasts may represent slight underestimates of the true effects. While such corrections are possible in common applications of meta-analytical methods, the analyses for this study were already quite complex because we accounted for statistical control of the pretest measures, violations of normality, multilevel clustering, and missing data. Furthermore, we used *z*-standardized sum scores of the reading comprehension measure. A latent variable approach using several indicators for reading comprehension could have been useful.

Implications and Future Prospects

As mentioned in the introduction, heterogeneity in German general education classrooms is increasing, and this is accompanied by the concern that low-performing students might be left behind because systematically differentiated instruction is only rarely found in general education settings (D. Fuchs et al., 2010; Tarelli et al., 2012). The results of our study are especially important for countries such as Germany, where students with learning difficulties are placed back into general education classrooms, but often with no accompanying systematic instruction from LD specialists. It seems that implementing DBDM with support for teachers (via LPA and LPA plus the RS) in general classroom settings is possible and can help to prevent low-performing students from being left behind. Nevertheless, it becomes apparent that merely implementing classroom-level differentiation is not enough to meet the needs of low-performing students and teachers need support that goes beyond the concepts we provided. Most importantly, future studies should consider factors that affect the success of DBDM implementation. For instance, a closer contact between researchers and teachers (e.g., by continuous professional development or coaching) could facilitate teachers' knowledge and skills about DBDM. At the same time, researchers could profit from teachers' practical experience. Moreover, Hoogland et al. (2016) identified three important factors that affect whether the implementation of DBDM is successful: (a) the presence of a DBDM culture in schools (e.g., collective responsibility and respect for teachers' autonomy and professionalism), (b) the facilitation of resources (e.g., sufficient time), and (c) the development of teachers' knowledge and skills on DBDM. Thus, the use of LPA and the RS to differentiate instruction might need to be adapted to the needs of each school. Overall, teachers seem to need more resources to implement LPA and to use the RS, which should include evidence-based and student-centered approaches (e.g., time or personnel).

The six primary studies investigated effects of DBDM for one school year. Taking into consideration that transferring innovative concepts into schools requires a substantial and sustainable change (Gottfredson et al., 2015; Humphrey et al., 2016), it would be interesting to study long-term effects over more school years in the future. For instance, teachers could receive continuous training on how to read and interpret graphed data that display information on students' learning progress (Espin et al., 2017). Another way to further improve DBDM could be to establish data teams, in which teachers collaboratively analyze data and discuss instructional adjustments (Schildkamp et al., 2019). As several studies from Fuchs and Fuchs and colleagues show, providing teachers with clear suggestions on how to utilize data for instructional decision-making is effective (Stecker et al., 2005). Taken together, a combination of reading instructional materials, teacher training, and cooperation among school teams may be a promising approach for implementing DBDM routines; however, more research is needed on how teachers can be best supported in this domain.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iDs

Martin T. Peters [ID] https://orcid.org/0000-0001-5267-6609
Natalie Förster [ID] https://orcid.org/0000-0003-0634-5993

Open Practices

All data and syntax can be found in the Open Science Framework and are available via the following link: https://osf.io/nj6tu/?view_only=96cc0e4fb15b4f478f4c9cd4571d4387. Materials have not been made publicly available. The study's design and analysis plan were not preregistered.

References

Acock, A. C. (2014). *A gentle introduction to Stata* (4th ed.). Stata Press.

Arellano, M. (1987). Computing robust standard errors for within group estimators. *Oxford Bulletin of Economics and Statistics*, *49*(4), 431–434. https://doi.org/10.1111/j.1468-0084.1987.mp49004006.x

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, *20*(1), 40–49. https://doi.org/10.1002/mpr.329

Bremerich-Vos, A., Stahns, R., Hußmann, A., & Schurig, M. (2017). *Förderung von Leseflüssigkeit und Leseverstehen*

im Leseunterricht [Fostering reading fluency and reading comprehension in reading instruction]. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe, & R. Valtin (Eds.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* [IGLU 2016. Reading competences of primary school students in Germany in international comparison] (pp. 279–296). Waxmann.

Brunner, C., Fasca, C., Heinze, J., Honey, M., Light, D., Mandinach, E. B., & Wexler, D. H. (2005). Linking data and learning: The grow network study. *Journal of Education for Students Placed at Risk*, *10*(3), 241–267. https://doi.org/10.1207/s15327671espr1003_2

Buly, M. R., & Valencia, S. W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis*, *24*(3), 219–239. https://doi.org/10.3102/01623737024003219

Calhoon, M. B., & Fuchs, L. S. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial and Special Education*, *24*(4), 235–245. https://doi.org/10.1177/07419325030240040601

Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, *33*(3), 378–398. https://doi.org/10.3102/0162373711412765

Century, J., & Cassata, A. (2016). Implementation research: Finding common ground on what, how, why, where, and who. *Review of Research in Education*, *40*(1), 169–215. https://doi.org/10.3102/0091732X16665332

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

Connor, C. M., Morrison, F. J., & Katch, L. E. (2004). Beyond the reading wars: Exploring the effect of child-instruction interactions on growth in early reading. *Scientific Studies of Reading*, *8*(4), 305–336. https://doi.org/10.1207/s1532799xssr0804_1

Coyne, M. D., Oldham, A., Dougherty, S. M., Leonard, K., Koriakin, T., Gage, N. A., Burns, D., & Gillis, M. (2018). Evaluating the effects of supplemental reading intervention within an MTSS or RTI reading reform initiative using a regression discontinuity design. *Exceptional Children*, *84*(4), 350–367. https://doi.org/10.1177/0014402918772791

Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis*, *45*(2), 215–233. https://doi.org/10.1016/S0167-9473(02)00366-3

Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education*, *57*(3), 300–314. https://doi.org/10.1177/0022487105285962

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/19.1111&ldrp.12123

Förster, N., Kawohl, E., & Souvignier, E. (2018). Short- and long-term effects of assessment-based differentiated reading instruction in general education on reading fluency and reading comprehension. *Learning and Instruction*, *56*, 98–109. https://doi.org/10.1016/j.learninstruc.2018.04.009

Förster, N., & Kuhn, J.-T. (2020). *Ice is hot and water is dry—Developing equivalent reading tests using rule-based item design*. Unpublished manuscript. Institute for Psychology in Education, University of Münster.

Förster, N., & Souvignier, E. (2011). Curriculum-based measurement: Developing a computer-based assessment instrument for monitoring student reading progress on multiple indicators. *Learning Disabilities: A Contemporary Journal*, *9*, 21–44.

Förster, N., & Souvignier, E. (2014a). Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. *Learning and Instruction*, *32*, 91–100. https://doi.org/10.1016/j.learninstruc.2014.02.002

Förster, N., & Souvignier, E. (2014b, August). *Learning progress assessment in reading: How can we construct and evaluate tests to monitor student reading progress?* [Paper presentation]. Special Interest Group Assessment and Evaluation of the European Association of Research on Learning and Instruction 6th Biennial Meeting, Madrid, Spain.

Förster, N., & Souvignier, E. (2015). Effects of providing teachers with information about their students reading progress. *School Psychology Review*, *44*, 60–75. https://doi.org/10.17105/SPR44-1.60-75

Fuchs, D., Fuchs, L. S., & Stecker, P. (2010). The "blurring" of special education in a new continuum of general education placements and services. *Exceptional Children*, *76*(3), 301–323. https://doi.org/10.1177/001440291007600304

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, *33*, 188–192. https://doi.org/10.1080/02796015.2004.12086241

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement, using a reading maze task. *Exceptional Children*, *58*, 436–450. https://doi.org/10.1177/001440299205800507

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children*, *60*, 518–537. https://doi.org/10.1177/001440299406000605

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1990). The role of skills analysis in curriculum-based measurement in math. *School Psychology Review*, *19*, 6–22. https://doi.org/10.1080/02796015.1990.12087335

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities*, *45*(3), 195–203. https://doi.org./10.1177/0022219412442150

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, D. W. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention in the primary grades*. Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance.

Gersten, R., Newman-Gonchar, R. A., Haymond, K. S., & Dimino, J. (2017). *What is the evidence base to support reading interventions for improving student outcomes in grades 1–3?* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.

Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science: The Official Journal of the Society for Prevention Research*, *16*(7), 893–926.

Hebbecker, K., & Souvignier, E. (2018). Formatives Assessment im Leseunterricht der Grundschule—Implementation und Wirksamkeit eines modularen, materialgestützten Konzepts [Formative assessment in reading instruction in primary school—Implementation and effectiveness of a modular, material-based concept]. *Zeitschrift für Erziehungswissenschaft*, *21*, 735–765. https://doi.org/10.1007/s11618-018-0834-y

Hondrich, A. L., Hertel, S., Adl-Amini, K., & Klieme, E. (2015). Implementing curriculum-embedded formative assessment in primary school science classrooms. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 353–376. https://doi.org/10.1080/0969594X.2015.1049113

Hoogland, I., Schildkamp, K., van der Kleij, F., Heitink, M. C., Kippers, W. B., Veldkamp, B. P., & Dijkstra, A. M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and Teacher Education*, *60*, 377–386. https://doi.org/10.1016/j.tate.2016.07.012

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). *Implementation and process evaluation (IPE) for interventions in education settings: A synthesis of the literature*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Review_Final.pdf

Ikemoto, G. S., & Marsh, J. A. (2007). *Cutting through the "data-driven" mantra: Different conceptions of data-driven decision making* In P. A. Moss (Ed.), *Evidence and decision making: Yearbook of the national society of education* (pp. 105–131). Wiley-Blackwell.

Jamshidian, M., Jalal, S., & Jansen, C. (2014). MissMech: An R package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR). *Journal of Statistical Software*, *56*(6), 1–31. https://doi.org/10.18637/jss.v056.i06

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*. https://doi.org/10.1016/j.cogdev.2020.100949

Kawohl, E. (2015). Diagnosebasierte individuelle Leseförderung in der Grundschule [Assessment-based individualized reading instruction in primary school] [Doctoral dissertation]. University of Münster.

Keuning, T., van Geel, M., Visscher, A., & Fox, J.-P. (2019). Assessing and validating effects of a data-based decision-making intervention on student growth for mathematics and spelling. *Journal of Educational Measurement*, *56*(4), 757–792. https://doi.org/10.1111/jedm.12236

Kippers, W. B., Poortman, C. L., Schildkamp, K., & Visscher, A. J. (2018). Data literacy: What do educators learn and struggle with during a data use intervention? *Studies in Educational Evaluation*, *56*, 21–31. https://doi.org/10.1016/j.stueduc.2017.11.001

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

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. https://doi.org/10.1177/0022219415605194

Lehmann, R. H., Peek, R., & Poerschke, J. (2006). *Hamburger Lesetest für 3. und 4. Klassen (HAMLET 3-4)* [Hamburger Reading test for grade 3 and 4]. Hogrefe.

Lenhard, W., Lenhard, A., & Schneider, W. (2017). *ELFE II. Ein Leseverständnistest für Erst- bis Siebtklässler* [ELFE II. A reading comprehension test for 1st grade to 7th grade students]. Hogrefe.

Lenhard, W., & Schneider, W. (2006). *Ein Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6)* [A reading comprehension test for 1st grade to 6th grade students (ELFE1-6)]: Hogrefe.

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, *47*(2), 71–85. https://doi.org/10.1080/00461520.2012.667064

Mayringer, H., & Wimmer, H. (2003). *SLS—Salzburger Lese—Screening für die Klassenstufen 1—4* [SLS—Salzburger reading-screening for grades 1-4.]. Huber.

Millo, G. (2017). Robust standard error estimators for panel models: A unifying approach. *Journal of Statistical Software*, *82*(3), 1–27. https://doi.org/10.18637/jss.v082.i03

Möller, J., & Bonerad, E.-M. (2007). Fragebogen zur habituellen Lesemotivation [Habitual reading motivation questionnaire]. *Psychologie in Erziehung Und Unterricht*, *54*, 259–267.

Müller, B., & Richter, T. (2017). *Förderung hierarchieniedriger Leseprozesse* [Fostering basal reading skills]. In M. Philipp (Ed.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (pp. 236–251). Juventa.

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*, 1–9. https://doi.org/10.3389/fpsyg.2017.01635

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/

National Institute of Child Health and Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. U.S. Government Printing Office.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117–175. https://doi.org/10.1207/s1532690xci0102_1

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme

(Eds.), *The science of reading: A handbook* (pp. 227–247). Blackwell.

Peters, M. T., Forthmann, B., Förster, N., & Souvignier, E. (2021). *Business-as-usual reading instruction in 2nd grade in Germany: Teacher-guided and one for all*. Unpublished manuscript. Institute of Psychology in Education, University of Münster.

Peters, M. T., Hebbecker, K., & Souvignier, E. (in press). Effects of providing teachers with tools for implementing assessment-based differentiated reading instruction in second grade. *Assessment for Effective Intervention*.

Pierce, M. E., Katzir, T., Wolf, M., & Noam, G. G. (2007). Clusters of second and third grade dysfluent urban readers. *Reading and Writing: An Interdisciplinary Journal*, *20*(9), 885–907. https://doi.org/10.1007/s11145-007-9058-x

Pressley, M., Allington, R. L., Wharton-McDonald, R., Block, C. C., & Morrow, L. M. (2001). *Learning to read: Lessons from exemplary first-grade classrooms*. Guilford Press.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ryan, R. M., & Deci, E. L. (2000). Self-Determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher*, *32*, 403–408.

Schildkamp, K., Poortman, C. L., Ebbeler, J., & Pieters, J. M. (2019). How school leaders can build effective data teams: Five building blocks for a new wave of data-informed decision making. *Journal of Educational Change*, *20*(3), 283–325. https://doi.org/10.1007/s10833-019-09345-3

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, *1*(2), 199–223. https://doi.org/10.1037/1082-989X.1.2.199

Seuring, V. A., & Spörer, N. (2010). Reziprokes Lehren in der Schule: Förderung von Leseverständnis, Leseflüssikeit und Strategieanwendung [Reciprocal teaching in schools: Fostering reading comprehension, reading fluency and strategy use]. *Zeitschrift für Pädagogische Psychologie*, *24*, 191–205. https://doi.org/10.1024/1010-0652/A000016

Souvignier, E., Förster, N., Hebbecker, K., & Schütze, B. (2021). *quop: An effective web-based approach to monitor student learning progress in reading and mathematics in entire classrooms*. In S. Jornitz & A. Wilmers (Eds.), *International perspectives on school settings, education policy and digital strategies: A transatlantic discourse in education research* (pp. 283–298). Budrich.

Spear-Swerling, L., & Cheesman, E. (2012). Teachers' knowledge base for implementing response-to-intervention models in reading. *Reading and Writing*, *25*, 1691–1723. https://doi.org/10.1007/s11145-011-9338-3

Staman, L., Timmermans, A., & Visscher, A. J. (2017). Effects of a data-based decision making intervention on student achievement. *Studies in Educational Evaluation*, *55*, 58–67. https://doi.org/10.1177/0022487117704170

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

Stutz, F. (2014). *Reading motivation, reading amount, and reading comprehension in the early elementary years* [Doctoral dissertation]. University of Potsdam. https://publishup.uni-potsdam.de/frontdoor/index/index/docId/7805

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, *49*(1), 77–96. https://doi.org/10.1177/0022219414528540

Tarelli, I., Lankes, E.-M., Drossel, K., & Gegenfurtner, A. (2012). *Lehr- und Lernbedingungen an Grundschulen im internationalen Vergleich* [Conditions of teaching and learning in primary schools in international comparison.]. In W. Bos, I. Tarelli, A. Bremerich-Vos, & K. Schwippert (Eds.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 137–173). Waxmann.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education*, *25*(4), 252–261. https://doi.org/10.1177/07419325040250040801

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

van Geel, M., Keuning, T., Visscher, A. J., & Fox, J. P. (2016). Assessing the effects of a school-wide data-based decision-making intervention on student achievement growth in primary schools. *American Educational Research Journal*, *53*(2), 360–394. https://doi.org/10.3102/0002831216637346

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48. https://doi.org/10.18637/jss.v036.i03

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, *36*(4) 541–561. https://doi.org/10.1080/02796015.2007.12087917

Wimmer, H., & Mayringer, H. (2014). *Salzburger Lese-Screening für die Schulstufen 2-9 (SLS 2 9)* [Salzburger Reading-Screening for grade 2-9.]. Huber.

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading*, *5*(3), 211–239. https://doi.org/10.1207/S1532799XSSR0503_2