



International Journal of Educational Methodology

Volume 6, Issue 4, 669 - 680.

ISSN: 2469-9632

<https://www.ijem.com/>

Tablet or Paper and Pen? Examining Mode Effects on German Elementary School Students' Computation Skills with Curriculum-Based Measurements

Stefan Blumenthal*

University of Rostock, GERMANY

Yvonne Blumenthal

University of Greifswald, GERMANY

Received: June 25, 2020 • Revised: September 18, 2020 • Accepted: October 13, 2020

Abstract: Progress monitoring of academic achievement is an essential element to prevent learning disorders. A prominent approach is curriculum-based measurement (CBM). Various studies have documented positive effects of CBM on students' achievement. Nevertheless, the use of CBM is associated with additional work for teachers. The use of tablets may be of help here. Yet, although many advantages of computer- or tablet-based assessments are being discussed in the literature (e. g. innovative item formats, adaptive testing, automated scoring and feedback), there are still concerns regarding the comparability of different assessment modes (paper-pencil vs. tablet). In the study presented, we analyze the CBM data of 98 fourth graders. They processed the exact same computation items once with paper and pen and once in a tablet application. The analyses point to comparable results in the test modes, although some significant deviations can be found at item level. In addition, the children report perceived benefits when working with the tablet.

Keywords: Progress monitoring, curriculum-based measurements, mode effects, tablet, elementary school.

To cite this article: Blumenthal, S., & Blumenthal, Y. (2020). Tablet or paper and pen? Examining mode effects on German elementary school students' computation skills with curriculum-based measurements. *International Journal of Educational Methodology*, 6(4), 669-680. <https://doi.org/10.12973/ijem.6.4.669>

Introduction

Progress Monitoring of academic skills with Curriculum-Based Measurements

The repeated collection of student performance data has established itself in Germany as an important and independent field of research (e.g. Voß & Gebhardt, 2017). The term progress monitoring covers all those instruments and methods that aim at a formative evaluation of instruction and intervention in school contexts and have been sufficiently tested regarding their psychometric suitability (Klauer, 2014). The aim of this approach is to map and document learning processes to detect learning deficits as early as possible, to draw conclusions about effective pedagogical interventions and to implement data-based decisions. Progress monitoring is thus an essential element of school-based prevention of learning disorders in Germany (Hartke, 2005; Huber & Grosche, 2012; Kretschmann, 2009) and also forms a basis for evidence-based pedagogical action in individual cases (Blumenthal et al., 2019).

A very prominent approach to the diagnosis of learning trajectories can be seen in the so-called curriculum-based measurements (CBM; Deno, 1985, 2003; Fuchs & Fuchs, 1986; Hosp et al., 2007). CBM are short tests which are intended to map learning progress in individual specific competencies in different learning areas (Stecker et al., 2008). Originally, CBM were developed in the USA in the 1970s as an evaluation tool to assess the effectiveness of special education measures. The special feature of the procedures is the equivalent combination of their psychometric quality - to ensure accurate and valid data - and their simple implementation - in the sense of simple application and evaluation as well as economic implementation - to ensure high-frequency use in educational practice (Deno, 2003).

Various studies point to positive effects on student performance when CBM are used as a progress monitoring tool (Foegen et al., 2007; Fuchs & Fuchs, 1986; Shapiro et al., 2006; Souvignier & Förster, 2011; Souvignier et al., 2014; Stecker et al., 2005). CBM are particularly common in the academic fields of reading, writing and mathematics at elementary school level and, to some extent, in the secondary sector (Foegen et al., 2007; McMaster & Espin, 2007; Stecker et al., 2008; VanDerHeyden & Burns, 2008; Wayman et al., 2007).

* Corresponding author:

Stefan Blumenthal, University of Rostock, Department for Special Education, August-Bebel-Straße 28, 18055 Rostock, Germany.

✉ stefan.blumenthal@uni-rostock.de



Against the background of the previously reported findings on the effectiveness of CBM, especially in connection with digitally supported processing of student data, and the advantages and availability (at least in the private sphere) of modern digital media, the question arises as to which extent do digitally collected results differ from those collected as paper-pencil-based versions (so-called mode effects; Hensley, 2015; Huff & Sireci, 2001).

Literature Review

While the international state of research shows great potential for the use of CBM in the school field, there is currently only a small and slowly growing range of procedures in the German-speaking world (Hesse & Latzko, 2017; Voß & Gebhardt, 2017). This contradiction is probably primarily to be seen in the multitude of challenges in the conception and evaluation of progress monitoring instruments. In addition to the required evidence for common status diagnostic instruments (objectivity, reliability, and validity), further requirements apply to progress monitoring instruments. This is to ensure that the results can be interpreted correctly over time. According to Fuchs (2004), in addition to psychometric suitability, aspects relevant to practice are particularly important. The applicability, usefulness, and effectiveness of CBM in school practice are of importance here.

While various ways of constructing and evaluating instruments have already been described in the German literature (Gebhardt et al., 2015; Voß et al., 2017; Wilbert & Linnemann, 2011), the usability and applicability of CBM is currently receiving particular attention. This is linked to questions of the successful implementation of progress monitoring tools in school practice (Hebbecke & Souvignier, 2018; Voß, 2014). In this context, the significance of modern digital media in the school setting is discussed (Fuchs, 2004).

A computer-/tablet-based progress monitoring tool offers many advantages for teachers (see table 1). On the one hand, the frequent use of printed CBM is very cost-intensive, which is problematic in the context of limited printing quotas in schools. In addition, the repeated use of CBM requires time to collect, evaluate and interpret the data. In a survey of German teachers conducted by Voß (2014), the participants rated the regular use of paper & pencil CBM as a significant (although justifiable against the background of the gain in knowledge) additional expenditure. A further problem area is the possible restriction of consistency in evaluation within the school staff which could lead to a limited objectivity of evaluation and interpretation (Fuchs et al., 1994).

Digital assessments (whether with the computer or the tablet) lead to a higher time-efficiency. Especially data evaluation, presentation, and documentation, which would be the task of teachers, can thus be automated and run in a much more resource-efficient manner (Fuchs, 2004). In addition, it is much easier to provide support for data interpretation.

In Germany, some websites are available to support the progress monitoring of students. The website www.quop.de (Souvignier et al., 2014), has been extensively tested and evaluated. The present research shows that computer-based tests are very well accepted by practitioners and lead to higher learning gains (Förster & Souvignier, 2014; Souvignier et al., 2014). Previous experience shows similar tendencies for the internet platforms www.levumi.de (Mühling et al., 2017) and www.lernlinie.de (Voß et al., 2020).

Table 1. Arguments for and against paper-pen-based and tablet-based assessment

Arguments for a paper-pen-based approach	Arguments for a tablet-based approach
<ul style="list-style-type: none"> • Familiar working tool in schools • Low availability of tablets in schools • Continuous maintenance for tablets is necessary 	<ul style="list-style-type: none"> • Cost-effective acquisition of current devices is possible • Intuitive work • Positive impact on children with learning disabilities (Genz & Bresges, 2017) • Increased motivation of students (Bremer & Tillmann, 2014) • Use of less resources compared to paper-based work • Innovative task format, adaptive assessment, and automated feedback possible (Bridgeman, 2009; Redecker & Johannessen, 2013) • Efficacy of computer-based data analysis (Fuchs et al., 1988, 1989)

The question of mode effects (here the influence of different administrative conditions on the results of digital as opposed to analogue tests) is important in order to assess whether the results obtained can be interpreted as equivalent (American Educational Research Association, 2014). To test for these differences, the two modes of implementation must be subjected to an empirical comparison. The findings indicate comparability when the results and distributions of students are similar across both formats (Pomplun et al., 2002).

There is already a comprehensive body of research on mode effects in paper-pencil-based and computer-based tests in general (Bennett et al., 2008; Cayton-Hodges et al., 2015; Hensley, 2015; Huff & Sireci, 2001; Poggio et al., 2005; Pommerich, 2004; Wang, et al., 2007). However, the research findings are not uniform (Hensley, 2015). Some studies show that the mode of administration has little or no impact on student performance (Hargreaves et al., 2004; Peak,

2005; Wang et al., 2007), while others point to advantages for paper-pencil-based testing (Bennett, 2003). There are also studies that show better performance in computer-based procedures (Clariana & Wallace, 2002; Pomplun et al., 2002). There are also limitations in the findings in that many studies refer to higher grades or adulthood, that different test areas (mathematics skills, reading or writing skills) and test formats were investigated, that PCs were generally used but not mobile devices, that most comparative studies did not investigate mode effects at item level but only at the overall test level, and that the use of CBM was only occasionally specifically investigated in these studies. Against this background, the questions relevant to the present paper arise.

Methodology

Research Goal

The presented study aims to expand the given research corpus. For this purpose, results of primary school children will be analyzed comparatively in a thoroughly evaluated CBM for the assessment of addition and subtraction competencies of primary school children (Sikora & Voß 2017) in a paper-pencil-based and a tablet-based test. The following questions guide the research:

1. Can paper-pencil-based CBM also be implemented as a tablet version (applicability)?
2. Are there differences in the results of paper-pencil-based and tablet-based CBM procedures (mode effects)?
3. Does the tablet-based implementation have advantages over the paper-pencil-based version (advantages of digital media)?

Sample and Data Collection

A total of $N = 98$ children from five fourth grades took part in the study. The gender ratio was approximately balanced with 54.1 % girls ($N = 53$). Disease-related data losses of individual children were recorded at individual test times. Due to data protection objections of the participating schools and parents, no personal analysis of the origin and socioeconomic status could be carried out.

As dependent variable the results of the children in the addition and subtraction CBM of Sikora and Voß (2017) were used. To construct the CBM, clustered item pools were generated for each item position of a test according to mathematical-didactic as well as empirical aspects ($N = 24$). From these, tasks were randomly selected and transferred to the corresponding test versions. In this way, ten different but structurally similar CBM versions could be created for the fourth school year. A comprehensive longitudinal study ($N = 463$ primary school students) showed the psychometric and diagnostic suitability of the method (Sikora & Voß, 2017). According to this, all items meet the requirements of a one-dimensional Rasch model ($0.7 \leq \text{Infit} \leq 1.3$; Bond & Fox, 2015) and show satisfactory item difficulties and selectivity. Reliability (Cronbach's $\alpha = .86$; range from .84 - .88) as well as validity coefficients (significant correlation of $r = .64$ with standardized mathematics test; significant correlation $r = .31$ with standardized reading and spelling test) are also adequate. The sensitivity to change of the developed CBM could be shown by means of multi-level modelling (Sikora & Voß, 2017).

In addition to the CBM, the children completed a questionnaire. The questionnaire contained a total of 21 items relating to the scales "use of the tablet", "motivation to use the tablet", "(perceived) advantage of using the tablet", "motivation to use the tablet in general", "use of the tablet at home", "attitude towards mathematics" and "use of the tablet at school" (see Table 2). The items were presented as ratings on a 4-level Likert scale with the extreme values "I fully agree" (1) and "I do not agree" (4). In the context of the study presented here, only selected findings from the student survey are discussed in detail.

Table 2. Summary of student questionnaires

Scale	Number of items	Cronbach's α	Example
Working during the tablet-based test	3	.56	It was easy to work with the tablet.
Motivation during the tablet-based test	4	.87	It is more interesting to solve the tasks with the tablet.
(perceived) advantages of using tablets	4	.71	I felt comfortable whilst working on difficult tasks with the tablet.
Motivation of tablet usage in general	2	.78	I would like to work more often with the tablet in Math.
Tablet usage at home	5	.61	I also use a tablet in my free time.
Attitude towards mathematics	2	.39	I feel bad when I work on a difficult math problem.
Usage of tablets in school	1	-	I often work with tablets in school.
Total	21	.77	

In addition to the data listed above, teachers were also asked to provide the half-yearly grades for each child.

The tests were administered in the middle of the fourth school year within one school week as part of a daily exercise. On two consecutive days, the versions for addition were used first in paper-pencil-based or tablet-based form. One part of the class received the paper version, the other part the tablet version. On the next test day this was exchanged accordingly. On the following two school days, the versions for subtraction were used in the same way.

Analyzing of Data

The Shapiro-Wilk-test reveals that none of the data is normally distributed (addition paper & pencil: $W(74) = .502$, $p < .001$; addition tablet: $W(74) = .486$, $p < .001$; subtraction paper & pencil: $W(74) = .425$, $p < .001$; subtraction tablet: $W(74) = .367$, $p < .001$). For data analysis, correlations between the results of the paper-pencil-based and tablet-based tests were calculated - separately for the areas of addition and subtraction. Using a graphical model test as well as the inferential statistical analysis of the different test modes, differences in the results were checked for. A comparison of the reliability of the paper-pencil-based and tablet-based tests and their correlations with the semi-annual score was determined as a further value for assessing possible mode effects (Hensley, 2015).

Findings / Results

Results at the Overall Test Level

The comparison of the distributions of the paper-pencil-based vs. tablet-based test values does not show any significant differences (see Table 3). This applies to both the addition and subtraction CBM. This assumption is further supported by a graphical estimation of measurement invariance. The items were plotted on the X- and Y-axes separately according to their implementation mode (see Figure 1). Since the item parameters run along the bisectors, it can be assumed that the test mode is measurement invariant. Inference-statistically, no significant differences can be shown either in the areas of addition or subtraction (see Table 3).

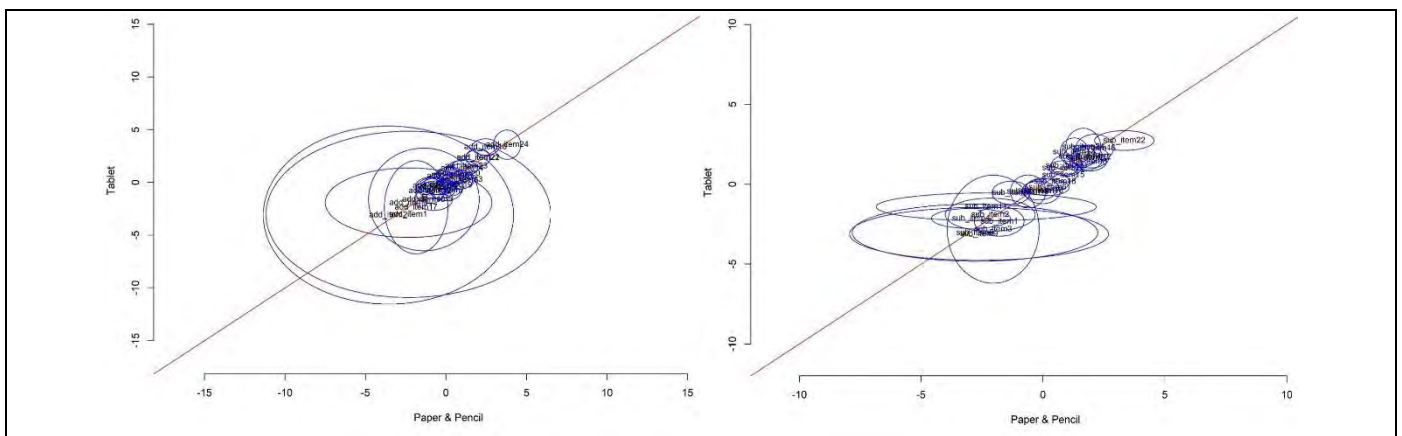


Figure 1. Graphic model test on measurement invariance for the CBM addition (left) and subtraction (right)

According to Cohen (1988), the correlations between the test modes are high (see Table 3), but significantly lower for the addition CBM ($r = .63$, $N = 87$) than for the subtraction CBM ($r = .84$, $N = 85$) (z-test over the correlation coefficients: $z = -3.09$, $p = .001$).

Table 3. Descriptive statistics, correlations, and differences between both test modes for the CBM addition and subtraction

	CBM addition		CBM subtraction	
	Paper & Pencil	Tablet	Paper & Pencil	Tablet
Mean	17.59	16.80	14.88	13.40
Standard deviation	3.93	4.09	4.82	5.22
Minimum	4	4	4	2
Maximum	24	24	24	24
Wilcoxon-test over test mode	$z = 1.78$, $p = .076$		$z = 0.96$, $p = .340$	
Correlation between both test modes	$r = .63$, $N = 87$		$r = .84$, $N = 85$	

The correlations with the half-year grade in mathematics as an external criterion indicate negligible differences (addition: tablet $r = -0.60$ vs. paper-pencil $r = -0.59$; subtraction: tablet $r = -0.53$ vs. paper-pencil $r = -0.58$; z-test over the correlation coefficients: $z = 0.46$, $p = .032$). The correlations are negative since the evaluation of the raw scores in the CBM and the grades are opposite.

Results on Item Level

The proportion of matching results between the two test modes for each item is on average $m = 77.1\%$ ($sd = 12.4\%$) for the addition range. For item 22, for example, the agreement is only 52.6% (see Table 4). The task is to solve $29600 + 53700 = \underline{\quad}$. The distribution of correct and incorrect solutions for the item is shown in Table 4. According to this, the paper-pencil form is more likely to provide a correct solution.

Table 4. Distribution of correct and false answers for item 22 ($29600 + 53700 = \underline{\quad}$)

	Ratio of false answers in %	Ratio of correct answers in %
Paper-based test mode	38	35
Tablet-based test mode	48	26

The highest percentage of identical results for the addition is found for item 2 (97.7%) (see Table 5). Similar values are found for the subtraction area ($m = 80.1\%$, $sd = 8.6\%$, $min = 65.0\%$, $max = 95.0\%$).

Table 5. Percentage of identical answers of the paper-based vs. tablet-based test runs

	Addition	Subtraction
Item 1	96.6%	85.0%
Item 2	97.7%	87.5%
Item 3	92.0%	95.0%
Item 4	82.8%	80.0%
Item 5	89.7%	90.0%
Item 6	77.0%	65.0%
Item 7	80.5%	75.0%
Item 8	75.9%	77.5%
Item 9	79.3%	95.0%
Item 10	90.8%	75.0%
Item 11	83.9%	87.5%
Item 12	86.2%	85.0%
Item 13	57.5%	70.0%
Item 14	58.6%	67.5%
Item 15	67.8%	77.5%
Item 16	75.0%	92.5%
Item 17	83.9%	82.5%
Item 18	70.1%	75.0%
Item 19	78.8%	67.5%
Item 20	70.2%	77.5%
Item 21	59.3%	71.8%
Item 22	52.6%	82.5%
Item 23	67.9%	82.5%
Item 24	76.3%	77.5%

Note: The darker the background the stronger the conformity.

Results of the Student Survey

The student survey indicates that the use of tablets at school is rather rare. The statement "At school I often work on the tablet." was rated an average of $m = 3.51$ ($sd = 0.85$; on a 4-level Likert scale) by the children. Nevertheless, the handling of the tablet-based test in class seems realistic. The children stated that they found it easy to handle the tablet in the test situation ($m = 1.21$, $sd = 0.47$) and that they could see everything well ($m = 1.31$, $sd = 0.65$), problems were only reported sporadically ($m = 3.47$, $sd = 0.94$).

The answers of the students indicate a tendency that the tablet-based test is perceived as more advantageous. While the distribution of the answers to the questions according to the difficulty of the tasks on the tablet is very balanced (see Figure 2, left), the majority of the respondents assume that they can solve the tasks more quickly with the tablet (see Figure 2, right).

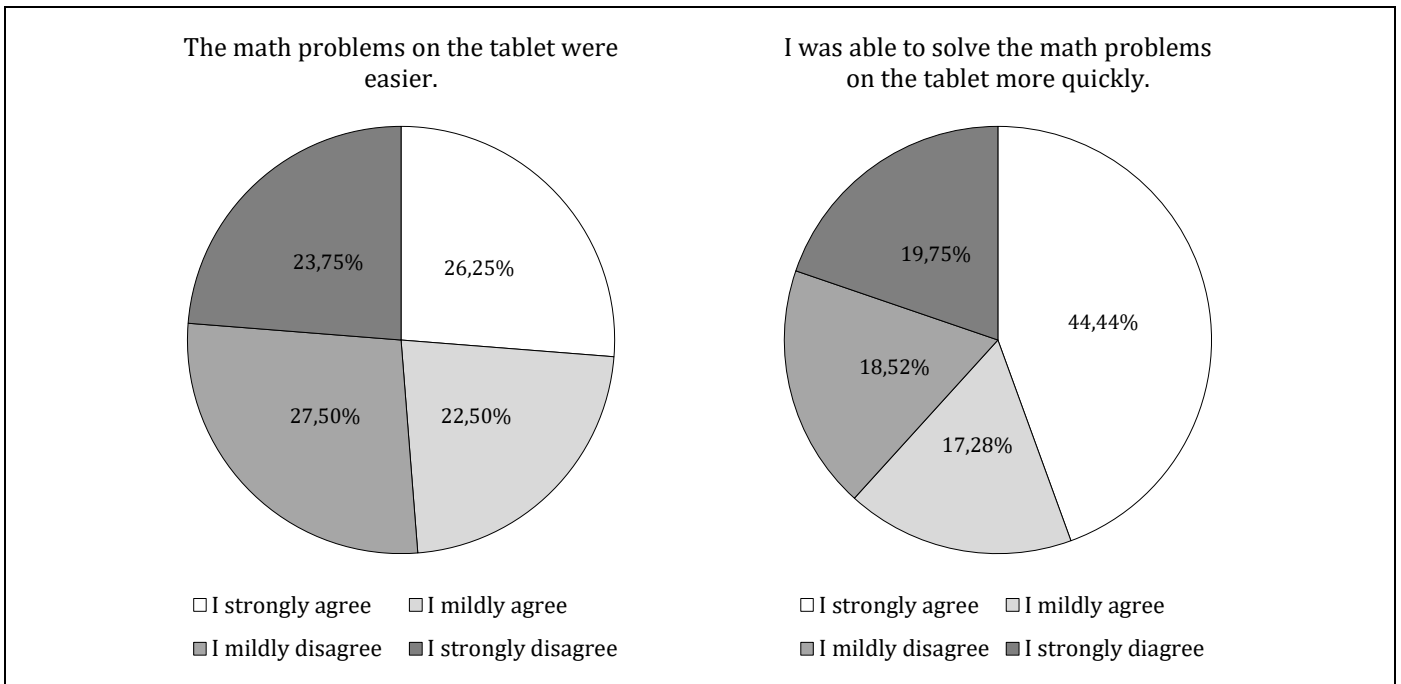


Figure 2. Distribution of answers regarding the aspect "(perceived) advantages of using tablets"

A differentiated consideration of the distribution of responses to the aspect "(perceived) advantage of using a tablet" against the background of a rather negative attitude towards mathematics points to further advantages of tablet-based testing (see Figure 3). Slightly more than half of the children surveyed said that difficult math tasks made them feel uncomfortable ("I fully agree" and "I tend to agree"), and nearly half of them indicated that this feeling did not occur when testing with the tablet.

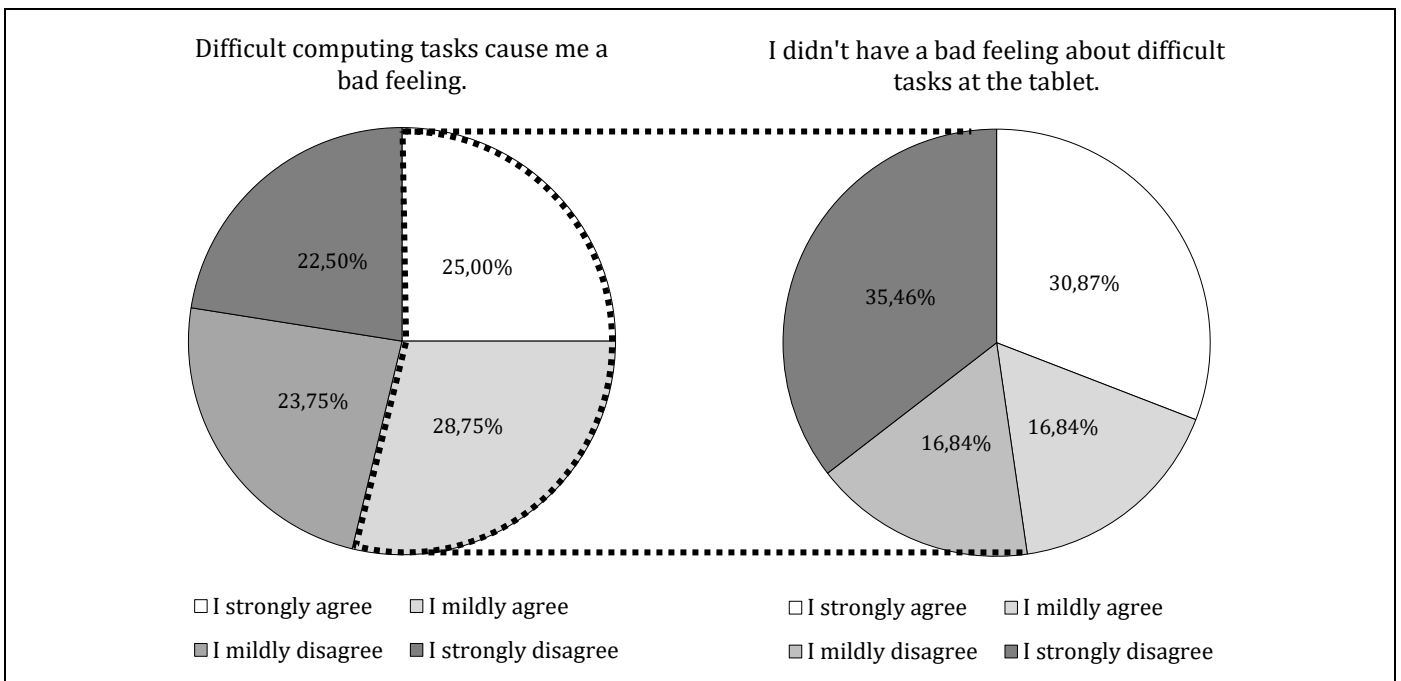


Figure 3. Differentiated analysis of the distribution of answers to the aspect "(perceived) advantage of tablet use" against the background of a rather negative attitude towards mathematics

Discussion

Against the background of the importance of CBM-based progress monitoring of students' performance, but also the challenges of its implementation in practice (additional time and resource-related expenditure), this article elaborated on the importance of digitally supported progress monitoring using tablets. Especially we focus on the question of

possible discrepancies between the different test modes (here: paper-pencil-based vs. tablet-based). The aim of the present study was therefore to examine whether paper-pencil-based CBM can also be implemented as a tablet version (usability), whether differences between the test conditions become apparent (mode effects) and whether a tablet-based implementation has advantages over the paper-pencil-based version (advantages of digital media).

Basically, the applicability of tablet-based CBM in the classroom seems possible. This is supported by the findings, according to which many German children have access to a tablet at home (Educational Media Research Association Southwest, 2016). The student survey conducted here shows an even clearer picture. Around 64% of the children questioned in this study have a tablet at home and use it at least one day a week. The tablet therefore is a familiar medium for children. Accordingly, the students examined found it easy to operate the devices in the test situation.

To clarify the question of the extent to which paper-pencil-based CBM can be carried out as a tablet version, various analyses were conducted in this study. Following Hensley (2015), it must be shown that the mean values and standard deviations are comparable for both forms of test performance. The same applies to the reliabilities and correlations with external criteria. Only when these criteria are met the use of a new evaluation method can be considered equivalent (American Educational Research Association, 2014). The descriptive results determined here at the overall test level show high similarities for both test modes. No significant differences could be detected by variance analysis or by using a graphical model test to determine the measurement invariance. This is consistent with existing research findings (Poggio et al., 2005; Pommerich, 2004; Wang et al., 2007). We also found high correlations between the test modes. The correlations with the half-year mark in mathematics are also comparable. These findings taken together indicate that there are no mode effects between the test methods used here and that the results can be considered equivalent in each case. In a differentiated analysis at item level, however, in some cases quite large deviations in the students' results could be registered (up to just under 50 %). These differences can be seen to varying degrees between the areas of addition and subtraction. As an example, it was shown that items with large numbers are more likely to be incorrect when they are processed on the tablet. The same significant effects of the test mode are also reported earlier (Bennett et al., 2008; Hensley et al., 2016; Jerrim, 2016). There are two main lines of argumentation to explain these differences (Hensley, 2015): a) differences due to the technology used or b) differences due to subject factors. In terms of technology, results could be affected by the fact that there was no way to edit and change tasks retroactively within the CBM app (Vispoel, 2000). Differences due to personal factors are also conceivable here (Lottridge et al., 2011; Poggio et al., 2005). Since the processing time was limited to 15 minutes regardless of the test mode, weaker students in particular were no longer able to complete the last and most difficult tasks. This reduces the internal validity of the presented findings. Overall, however, the deviations in individual items do not have a serious influence on the overall test results.

Finally, the advantages of a tablet-based test compared to a paper and pencil test need to be clarified (Cozad & Riccomini, 2016). In general, it can be said that the motivation of the children was significantly higher when they worked on the tablet. Interestingly, several students stated that they were able to complete the tasks on the tablet faster (almost 50 %) or that the tasks were easier (almost 60 %). However, the data collected indicate that, on average, the children completed about the same number of items in both test conditions (about 22-23 items). In contrast to the results of the survey, it is shown that the children solve more items correctly on average in the paper-pencil-based test (addition: $m_{\text{Paper}} = 17.59$ vs. $m_{\text{Tablet}} = 16.80$; subtraction: $m_{\text{Paper}} = 14.88$ vs. $m_{\text{Tablet}} = 13.40$). However, these deviations are not statistically significant. Even if the advantages of the tablets described by the children cannot be reproduced, these perceived advantages – together with the more economical test implementation and evaluation – speak more in favor of the use of tablet-based CBM in schools. This could be because children nowadays are confronted and use a wide range of digital media (Arslan-Cansever, 2019). Particularly relevant in this context seems to be the finding that about half of those students who stated that difficult mathematics tasks made them feel uncomfortable (just over 50 %) did not had this uncomfortable feeling when they completed the tasks on the tablet. Thus, working with the tablets seems to reduce a high emotional pressure of the children when working on the items. This is particularly important in view of the limited time available for processing when using CBM (Hosp et al., 2007). This is a key factor as math anxiety have a big impact on math performance (e. g. Kesici & Bindak, 2019). Another advantage of a tablet application is the possibility to determine the time needed to solve each task automatically, so that it is not necessary to limit the processing time. However, there are special circumstances for schools in Germany. Although there has been a marked increase in the frequency of regular private computer use in recent years (in 2003, 82% of the German 15-year-olds surveyed stated that they used the computer regularly in a domestic context, in 2006 the proportion was 90% and in 2012 almost 100%; Organisation for Economic Cooperation and Development, 2015; Senkbeil & Wittwer, 2007), the use of computers in class is significantly lower. For example, only 23 % (2003) and 31 % (2006) of students reported using computers regularly in class (Senkbeil & Wittwer, 2007). Although this percentage rose to just under 70 % by 2012, Germany is still below the OECD average (Organisation for Economic Cooperation and Development, 2015). The present study also shows that German children rarely use a tablet in class ($m = 3.51$, $sd = 0.85$ on a 4-level Likert scale).

One approach to explain this is that there are not enough digital terminals available in schools to enable all students to work with them. Even though the number of computers, notebooks or tablets has increased in recent years (Hensley, 2015), in 2006 an average of 12 students still shared a PC or 174 children a notebook in the school setting (German

Federal Ministry of Education and Research, 2006). A better supply of German schools with digital end devices is to be achieved through the recently published funding program of the German Federal Ministry of Education and Research (2019).

Conclusion

The present study shows that a tablet-based application can significantly reduce the perceived disadvantages of using CBM in everyday school life (no additional financial and time expenditure for teachers and less pressure solving computation tasks for students), with comparable results for students.

Moreover, the findings indicate that the processing of the CBM on the tablet has the impression of an advantage among the students. They find the items simpler and a considerable proportion of students who feel emotional stress when solving hard computational tasks, do not have this bad feeling when working on the tablet.

The corona pandemic has also highlighted the need for efficient digital infrastructures in schools and for digital educational content (Casale et al., 2020; Nobel et al., 2020).

The funding directive for the improvement of technical equipment at German schools described above is promising here (German Federal Ministry of Education and Research, 2019). This provides the technical framework for digitally supported progress monitoring. However, it is of far greater importance to make teachers at German schools aware of the advantages of using CBM in class and for accompanying interventions, thus ensuring increased demand.

Suggestions

CBM are a tool to make educational decisions based on reliable and objective data (Jung et al., 2018; Voß & Blumenthal, 2019). In this way the instruction or intervention is evaluated, and a modification can be initiated if necessary (e.g. Stecker et al., 2005). The positive effects of progress monitoring have been reported several times in the international literature. In Germany, the concept of data-based decision-making based on progress monitoring of student performance has only partially gained acceptance in school practice. There are several explanations for this, from which suggestions can be derived.

First, it should be noted that German teachers have different views on data relevant to school and related decisions than teachers from the USA (Blumenthal et al., submitted). Accordingly, the demand for CBM in Germany is certainly lower. In general, questions regarding the successful implementation of CBM in German schools do not yet seem to have been sufficiently clarified and should be in the focus of further research (Hebbecker, & Souvignier, 2018).

Another suggestion is that teachers in Germany should be made aware about the idea of data-based decision-making. This topic should already be given a stronger focus in university education (Wagner et al., 2017), but also to teachers already working in school practice (Espin et al., 2017). In the literature there is a lack of evaluated concepts for further training and coaching of teachers for a safe handling of data in the sense of a data-based decision-making approach. Initial findings from Espin et al. (2017) should be further expanded by systematic research.

Furthermore, the design and evaluation of instruments for reliable and valid progress monitoring in Germany should be a priority in research (Voß & Gebhardt, 2017).

Digital test environments offer a great advantage for supporting the application and implementation of data-based decision-making in everyday school life. They reduce the additional work involved in test execution and evaluation and allow automated interpretation of test results in terms of comprehensive student performance profiles and tailored support information (Hosp et al., 2007; Kingston & Nash, 2011). Researchers should work on possibilities to support the effective data use of teachers; digital media can play a significant role here.

Limitations

Finally, we would like to point out limitations of the results presented here. The CBM were worked on here by the students in a time of 15 minutes each. Usually, however, a time limitation to a few minutes is provided. Whether the findings are transferable even for such short processing periods remains open. It must also be indicated that this study was designed as a cross-sectional study. Whether comparable results can be achieved in a longitudinal design cannot be conclusively clarified either.

The findings reported here relate exclusively to computation tasks in addition and subtraction. A generalization of the findings to other types of assessments is not justified. Important hints for the selection and development of apps as an assessment tool give, for example, Cayton-Hodges et al. (2015).

References

- American Educational Research Association. (2014). *Standards for educational and psychological testing*. AERA.
- Arslan-Cansever, B. (2019). Investigation of third grade students' views on media literacy. *International Journal of Educational Methodology*, 5(1), 265-273. <https://doi.org/10.12973/ijem.5.1.265>
- Bennett, R. E. (2003). *Online assessment and the comparability of score meaning*. ETS.
- Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B. & Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *The Journal of Technology, Learning, and Assessment*, 6(9), 1-38.
- Blumenthal, Y, Hartke, B. & Voß, S. (2019). The role of evidence-based practice in German special education — State of research and discussion. *Education Sciences*, 9(2), 1-11. <https://doi.org/10.3390/educsci9020106>
- Casale, G., Börnert-Ringleb, M., & Hillenbrand, C. (2020). Fordern auf Distanz? Sonder-pädagogische Unterstützung im Lernen und in der emotional-sozialen Entwicklung während der Schulschließungen 2020 gemäß den Regelungen der Bundesländer [Promoting at a distance? Special educational support in learning and emotional and social development during the school closures in 2020 in accordance with the regulations of the federal states]. *Zeitschrift für Heilpädagogik*, 71(5), 254-267.
- Cozad, L. E., & Riccomini, P. J. (2016). Effects of digital-based math fluency interventions on learners with math difficulties: a review of the literature. *The Journal of Special Education Apprenticeship*, 5(2), 1-19.
- German Federal Ministry of Education and Research (2006). *IT-Ausstattung der allgemeinbildenden und berufsbildenden Schulen in Deutschland. Bestandsaufnahme 2006 und Entwicklung 2001 bis 2006* [IT-equipment in German schools. Stocktaking from 2006 and development from 2001 to 2006]. https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/it-ausstattung_der_schulen.pdf
- German Federal Ministry of Education and Research (2019). *Wissenswertes zum digitalpakt schule* [Useful information on the digital reform in schools]. <https://www.bmbf.de/de/wissenswertes-zum-digitalpakt-schule-6496.php>
- Cayton-Hodges, G. A., Feng, G. & Pan, X. (2015). Tablet-based math assessment: What can we learn from math apps? *Educational Technology & Society*, 18(2), 3-20.
- Clariana, R. & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology*, 33(5), 593-602. <https://doi.org/10.1111/1467-8535.00294>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219-232. <https://doi.org/10.1177/001440298505200303>
- Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education*, 37(3), 184-192. <https://doi.org/10.1177/00224669030370030801>
- Espin, C. A., Wayman, M. M., Deno, S. L., McMaster, K. L., & Rooij, M. de (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/lrpd.12123>
- Foegen, A., Jiban, C. & Deno, S. L. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education*, 41(2), 121-139. <https://doi.org/10.1177/00224669070410020101>
- Foerster, N. & Souvignier, E. (2014). 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>
- Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, 33(2), 188-192.
- Fuchs, L. S., Fuchs, D. & Hamlett, C. L. (1994). Strengthening the connection between assessment and instructional planning with expert systems. *Exceptional Children*, 61(2), 138-146.
- Fuchs, L. S. & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53(3), 199-208. <https://doi.org/10.1177/001440298605300301>
- Gebhardt, M., Heine, J.-H., Zeuch, N. & Förster, N. (2015). Lernverlaufsdiagnostik im Mathematikunterricht der zweiten Klasse: Raschanalysen und Empfehlungen zur Adaption eines Testverfahrens für den Einsatz in inklusiven Klassen [Progress monitoring for mathematics in second grade: Analysis and recommendations for the application of an diagnostic tool in inclusive classrooms]. *Empirical special education/ Empirische Sonderpädagogik*, 3, 206-222.

- Grünke, M. (2006). Zur Effektivität von Fördermethoden bei Kindern und Jugendlichen mit Lernstörungen. Eine Synopse vorliegender Metaanalysen [Efficacy of interventions for students with learning disabilities. A synopsis of meta-analyses]. *Childhood and development/ Kindheit und Entwicklung*, 15(4), 239-254. <https://doi.org/10.1026/0942-5403.15.4.239>
- Hargreaves, M., Shorrocks-Taylor, D., Swinnerton, B., Tait, K. & Threlfall, J. (2004). Computer or paper? That is the question: does the medium in which assessment questions are presented affect children's performance in mathematics? *Educational Research*, 46(1), 29-42. <https://doi.org/10.1080/0013188042000178809>
- Hartke, B. (2005). Schulische prävention – Welche maßnahmen haben sich bewährt [Prevention of learning difficulties – Which interventions stand the test]? In S. Ellinger & M. Wittrock (Hrsg.), *Sonderpädagogik in der Regelschule. Konzepte, Forschung, Praxis* [Special Education in Mainstream Schools. Concepts, Research, Practice] (pp. 11-37). Kohlhammer.
- Hattie, J. (2013). *Lernen sichtbar machen* [Visible learning]. Schneider-Verlag.
- Hebbecke, K. & Souvignier, E. (2018). Formatives Assessment im Leseunterricht der Grundschule – Implementation und Wirksamkeit eines modularen, materialgestützten Konzepts [Formative assessment of reading skills in primary schools – Implementation and efficacy of a modular concept]. *Journal for Educational Science/ Zeitschrift für Erziehungswissenschaft*, 21, 735-765. <https://doi.org/10.1007/s11618-018-0834-y>
- Hensley, K. K. (2015). *Examining the effects of paper-pencil-based and computer-based modes of assessment on mathematics curriculum-based measurement*. [Doctoral dissertation, University of Iowa] Iowa Research Online. <https://doi.org/10.17077/etd.iresh1q>
- Hensley, K., Rankin, A., & Hosp, J. (2017). Comparing student performance on paper- and computer-based math curriculum-based measures. *Assistive Technology*, 29(3), 140-145. <https://doi.org/10.1080/10400435.2016.1212129>
- Hesse, I. & Latzko, B. (2017). *Diagnostik für Lehrkräfte* [Assessment for teachers] (3rd ed.). Budrich.
- Hosp, M. K., Hosp, J. L. & Howell, K. W. (2007). *The ABC's of CBM: A practical guide to curriculum-based measurement*. Guilford.
- Huber, C. & Grosche, M. (2012). Das response-to-intervention-Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik [The response-to-intervention model as basis of an inclusive paradigm shift in special needs education]. *Journal for Therapeutic Pedagogy/ Zeitschrift für Heilpädagogik*, 8, 312-322.
- Huff, K. L. & Sireci, S. G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice*, 20(3), 16-25. <https://doi.org/10.1111/j.1745-3992.2001.tb00066.x>
- Jerrim, J. (2016). PISA 2012: how do results for the paper and computer tests compare? *Assessment in Education: Principles, Policy & Practice*, 23(4), 495-518. <https://doi.org/10.1080/0969594X.2016.1147420>
- Jung, P.-G., McMaster, K. L., Kunkel, A. K., Shin, J. & Stecker, P. M. (2018). Effects of data-based individualization for students with intensive learning needs: A meta-analysis. *Learning Disabilities Research & Practice*, 33(3), 144-155. <https://doi.org/10.1111/ldrp.12172>
- Kesici, A., & Bindak, R. (2019). Does mathematics anxiety have any impact on secondary school pupils' friend choices? *International Journal of Educational Methodology*, 5(1), 109-116. <https://doi.org/10.12973/ijem.5.1.123>
- Kingston, N. & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28-37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Klauer, K. J. (2014). Formative Leistungsdiagnostik: Historischer hintergrund und weiterentwicklung zur lernverlaufsdiagnostik [Formative achievement assessment: Historical background and evolution into progress monitoring]. In M. Hasselhorn, W. Schneider & U. Trautwein, U. (Eds.), *Lernverlaufsdiagnostik* [Progress monitoring] (Vol. 12., pp. 1-17). Hogrefe.
- Kretschmann, R. (2009). Prävention [Prevention]. In B. Ahrbeck & M. Willmann (Eds.), *Pädagogik bei Verhaltensstörungen* [Pedagogy for behavioral disorders] (pp. 237-246). Kohlhammer.
- Lottridge, S. M., Nicewander, W. A. & Mitzel, H. C. (2011) A comparison of paper and online tests using a within-subjects design and propensity score matching study. *Multivariate Behavioral Research*, 46(3), 544-566. <https://doi.org/10.1080/00273171.2011.569408>
- McMaster, K. & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *Journal of Special Education*, 41 (2), 68-84. <https://doi.org/10.1177/00224669070410020301>

- Nobel, K., Barwasser, A., Melzer, C., & Grünke, M. (2020). Webbasiertes distanzlernen - Erfahrungen im rahmen einer pilotstudie während der Covid-19-schulschließungen [Web-based distance learning - Experience in a pilot study during the Covid-19 school closures]. *Journal for therapeutic pedagogy/ Zeitschrift für Heilpädagogik*, 71(9), 465-479.
- Educational Media Research Association Southwest (2016). *KIM-Studie 2016. Kindheit, Internet, Medien. Basisstudie zum Medienumgang 6- bis 13-Jähriger in Deutschland* [KIM- Survey 2016. Childhood, internet, media. Study on media behavior and habits of 6- to 13- year-olds in Germany]. https://www.mpfs.de/fileadmin/files/Studien/KIM/2016/KIM_2016_Web-PDF.pdf
- Mühling, A., Gebhardt, M. & Diehl, K. (2017). Formative diagnostik durch die onlineplattform Levumi [Formative assessment with the website Levumi]. *Computer Science Spectrum/ Informatik Spectrum*, 40(6), 556-561. <https://doi.org/10.1007/s00287-017-1069-7>
- Organisation for Economic Cooperation and Development. (2015). *Students, computers and learning: Making the connection*. https://www.oecd-ilibrary.org/education/students-computers-and-learning_9789264239555-en <https://doi.org/10.1787/9789264239555-en>
- Peak, P. (2005). *Recent trends in comparability studies*. Pearson educational measurement. http://images.pearsonassessments.com/images/tmrs/tmrs_rg/TrendsCompStudies.pdf
- Poggio, J., Glasnapp, D. R., Yang, X. & Poggio, A. J. (2005). A comparative evaluation of score results from computerized and paper & pencil mathematics testing in a large-scale state assessment program. *Journal of Technology, Learning, and Assessment*, 3(6), 1-30.
- Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *Journal of Technology, Learning, and Assessment*, 2(6), 1-44.
- Pomplun, M., Frey, S. & Becker, D. F. (2002). The score equivalence of paper-and pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement*, 62(2), 337-354. <https://doi.org/10.1177/0013164402062002009>
- Redecker, C. & Johannessen, O. (2013). Changing assessment. Toward a new assessment paradigm using ICT. *European Journal of Education*, 47(1), 79-95. <https://doi.org/10.1111/ejed.12018>
- Senkbeil, M. & Wittwer, J. (2007). Die Computervertrautheit von jugendlichen und wirkungen der computernutzung auf den fachlichen kompetenzerwerb [Media habits of adolescents and impacts of computer usage on skills development]. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Eds.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* [PISA 2006, the results of the third international comparative study] (pp. 277-308). Waxmann.
- Shapiro, E. S., Keller, M. A., Lutz, J. G., Edwards Santoro, L. & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment*, 24(1), 19-35. <https://doi.org/10.1177/0734282905285237>
- Sikora, S. & Voß, S. (2017). Konzeption und Güte curriculumbasierter messverfahren zur erfassung der arithmetischen leistungsentwicklung in den klassenstufen 3 und 4 [Conception and psychometric properties of curriculum-based measurements for arithmetical skills in grade 3 and 4]. *Empirical special education/ Empirische Sonderpädagogik*, 9(3), 236-257.
- Souvignier, E. & Förster, N. (2011). Effekte prozessorientierter diagnostik auf die entwicklung der lesekompetenz leseschwacher viertklässler [Impacts of process-oriented assessment on reading skills of poor readers in 4th grade]. *Empirische Sonderpädagogik*, 3(3), 243-255.
- Souvignier, E., Förster, N. & Schulte, E. (2014). Wirksamkeit formativen assessments - Evaluation des ansatzes der lernverlaufsdiagnostik [Efficacy of formative assessment- Evaluation of progress monitoring]. In M. Hasselhorn, W. Schneider & U. Trautwein (Eds.), *Lernverlaufsdiagnostik* [Progress Monitoring] (Vol 12, pp. 221-237). Hogrefe.
- Stecker, P. M., Lembke, E. S. & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision making. *Preventing School Failure: Alternative Education for Children and Youth*, 52(2), 48-58. <https://doi.org/10.3200/psfl.52.2.48-58>
- 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>
- VanDerHeyden, A. M. & Burns, M. K. (2008). Examination of the utility of various measures of mathematics proficiency. *Assessment for Effective Intervention*, 33(4), 215-224. <https://doi.org/10.1177/00131640021970600>
- Vispoel, W. P. (2000). Reviewing and changing answers on computerized fixed-item vocabulary tests. *Educational and Psychological Measurement*, 60(3), 371-384. <https://doi.org/10.1177/00131640021970600>

- Voß, S. & Blumenthal, Y. (2019). *Data-based decision-making. Zum konstrukt und verständnis datenbasierter förderentscheidungsprozesse im unterricht* [Concept and insights on data-based decision-making for instructional purposes]. [Paper presentation]. Conference of Empirical Special Education Research, University of Siegen, Germany. <https://doi.org/10.13140/RG.2.2.14823.80802>
- Voß, S. & Gebhardt, M. F. (2017). Themenheft Verlaufsdiagnostik [Special issue on progress monitoring]. *Empirische Sonderpädagogik*, 9(4), 95-97.
- Voß, S. (2014). *Curriculumbasierte messverfahren im mathematischen erstunterricht – Zur güte und anwendbarkeit einer adaption US-Amerikanischer Verfahren im deutschen Schulsystem* [Curriculum-based measurement in first grade – Psychometric properties and adaptability of an US-American assessment tool in the German school system]. SVH.
- Voß, S., Blumenthal, Y., Ehrich, K. & Mahlau, K. (2020). Multimodale diagnostik als ausgangspunkt für spezifische förderung – eine darlegung am beispiel der rechtschreibung [Multimodal assessment as starting point for special support – Explanation using the example of orthography]. *Journal for Curative Education/ Zeitschrift für Heilpädagogik*, 71, 88-99.
- Voß, S., Sikora, S. & Mahlau, K. (2017). Vorschlag zur Konzeption eines curriculumbasierten messverfahrens zur erfassung der rechtschreibleistungen im grundschulbereich [Suggestion of a curriculum-based measurement for spelling skills in elementary schools]. *Journal for therapeutic pedagogy/ Empirische Sonderpädagogik*, 9(2), 184-194.
- Wagner, D. L., Hammerschmidt-Snidarich, S. M., Espin, C. A., Seifert, K., & McMaster, K. L. (2017). Pre-service teachers' interpretation of CBM progress monitoring data. *Learning Disabilities Research & Practice*, 32(1), 22-31. <https://doi.org/10.1111/ldrp.12125>
- Wang, S., Jiao, H., Young, M. J., Brooks, T. & Olson, J. (2007). A meta-analysis of testing mode effects in grade K-12 mathematics tests. *Educational and Psychological Measurement*, 67(2), 219-238. <https://doi.org/10.1177/0013164406288166>
- Wayman, M. M., Wallace, T., Wiley, H. I., Tichdt, R. & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, 41(2), 85-120. <https://doi.org/10.1177/00224669070410020401>
- Wilbert, J. & Linnemann, M. (2011). Kriterien zur analyse eines tests zur lernverlaufsdiagnostik [Criteria for the analysis of a progress monitoring tool]. *Empirical special education/ Empirische Sonderpädagogik*, 3(3), 225-242.