Learning Disabilities: A Contemporary Journal 17(1), 5-28, 2019

Quantity Recognition in Structured Whole Number Representations of Students With Mathematical Difficulties: An Eye-Tracking Study

Maike Schindler
Eveline Bader
University of Cologne

Achim J. Lilienthal
Örebro University

Florian Schindler
TU Dortmund University

Alfred Schabmann
University of Cologne

Quantity recognition in whole number representations is a fundamental skill children need to acquire in their mathematical development. Despite the observed correlation to mathematics achievement, however, the ability to recognize quantities in structured whole number representations has not been studied extensively. In this article, we investigate how students with mathematical difficulties (MD) differ from typically developing (TD) students in quantity recognition in structured whole number representations. We use eye tracking (ET), which can help to identify students' quantity recognition strategies. In contrast to methods that include collecting verbal answers and reports, ET avoids an additional verbalization step, which may be affected by poor language skills and by low meta-cognitive abilities or memory issues when monitoring, recalling, and explaining one's thoughts. We present an ET study with 20 students of which ten were found to have MD in initial tests (using qualitative and quantitative diagnostics). We used ET glasses, which allow seeing the students' view of the scene with an augmented visualization of the gaze point projected onto the scene. The obtained gaze-overlaid videos also include audio data and records of students' answers and utterances. In our study, we did not find significant differences between the error rates of MD and TD students. Response times, however, were longer for students with MD. The analysis of the ET data brought students' quantity recognition strategies to light, some of which were not found in previous research. Our analyses revealed differences in the use of these quantity recognition strategies between MD and TD students. Our study illustrates the power of ET for investigating students' quantity recognition strategies and the potential of ET to support MD students.

Keywords: **Mathematical Difficulties, Structured Whole Number Representations, Quantity Recognition, Abacus, Dot-Field, Eye Tracking.**

*Please send correspondence to: Maike Schindler, Jun.-Prof. Dr., University of Cologne, Department of Special Education and Rehabilitation, Classen-Kappelmann-Str. 24, 50931 Köln, Germany, Phone: +49 (0) 221 93317-420, Email: maike.schindler@uni-koeln.de.

Introduction

Quantity recognition in whole number representations, i.e., to grasp sets of elements in whole number representations and to say how many they are, constitutes an essential mathematics skill needed to develop a number sense and thus appears to be critical for students' mathematical development (e.g., Clements, 1999; Scherer et al, 2016). For quantity recognition, subitizing, the ability to fast and accurately grasp smaller quantities (Schleifer & Landerl, 2011), and one-by-one counting are important (Lindmeier & Heinze, 2016). However, students furthermore need to be able to perceive adequate partitions of sets and number structures, which relate to the base-10 system (Lindmeier & Heinze, 2016; Obersteiner et al., 2014). Therefore, external representations such as the 100-abacus or the 100-dot-field (see Fig. 1) are largely used in mathematics classrooms (Gaidoschik, 2015; Schipper, 2011). To perceive numbers in such external number representations, students may recognize larger numbers by perceiving partitions as groups—and make use of structures, such as 50, 10s, or 5s (Obersteiner et al., 2014)—which, in turn, may support students' understanding of the decimal system and its underlying structures (Benz, 2014; Clements, 1999).

Several studies have shown that abilities such as using number-groups or understanding the base-10 number system may pose severe problems to students with mathematical difficulties (MD; Moser Opitz et al., 2016). Yet, even though there is a larger number of studies addressing subitizing among students with MD (e.g., Moeller et al., 2009; Fischer et al., 2008; Schleifer & Landerl, 2011), there is only a limited number of studies investigating how students with MD recognize quantities in structured whole number represenations such as the 100-dot-field or the 100-bead-abacus, and how their strategies differ from students without MD (see Rottmann & Schipper, 2002 for an example). Some researchers report about MD students' issues using such representations (e.g., Gaidoschik, 2015; Rottmann & Schipper, 2002; Schipper, 2011)—and studies such as Obersteiner et al.'s (2013; 2014) suggest that the ability to recognize quantities in structured whole number representations correlates with mathematics achievement—suggesting that students with MD may have lower performances or even problems with quantity recognition in structured whole number representations. Together, these findings indicate that MD students may have problems using adequate strategies when determining quantities in whole number representations.

However, previous research has not yet investigated systematically how students with MD recognize quantities in structured whole number representations. This is striking since—on a practical level—such representations are frequently used as learning aids for students with MD (e.g., Wartha & Schulz, 2012). It is not clear to what extent students with MD actually profit from such representations, and how they perceive quantities on them. On a theoretical level, it is questionable if and how students with MD differ from TD students in quantity recognition on whole number representations: If problems in subitizing "cascade" into larger problems with quantity recognition (see Geary, 2013), and if MD students profit from the structures given in supportive representations by using them to recognize quantities. Previous research indicates that MD students tend to have pattern recognition difficulties (Ashkenazi et al., 2013) and a lack of conceptual number knowledge needed to compose and decompose sets of quantities (Starkey & McCandliss, 2014). These findings suggest that MD students may have problems "seeing" and using structures in structured whole number representations. However, their quantity recognition in whole number representations itself has rarely been under investigation. Therefore, we see a research gap to investigate how students with MD use whole number representations, how they perceive the quantities, and what quantity recognition strategies they use.

The aim of the current study is to evaluate whether students with MD differ from TD students in quantity recognition in structured whole number representations, in particular on the 100-bead abacus and the 100-dot-field, which are frequently used to facilitate students' learning (see, e.g., Gaidoschik, 2015). Our study uses ET, which has been used in previous studies to investigate students with MD (e.g., Moeller et al., 2009) and proven to provide opportunities for understanding strategies and thoughts of students with MD in particular (Schindler & Lilienthal, 2018). We investigate error rates, response times, and especially students' strategy use identified through ET, and analyze group differences between a MD and TD group.

Mathematical Difficulties (MD)

Learning difficulties in mathematics are an important topic in practice and research. However, to date, there is no consensus on a definition or term characterizing the group of students having difficulties in mathematics (Scherer et al., 2016). Some researchers speak of *mathematical learning disabilities*, others of (severe) *mathematical difficulties*, or *developmental dyscalculia* depending on different national educational contexts and research traditions (e.g., mathematics education, special education, pedagogy, psychology, or neuro-biology) (see Scherer et al., 2016.; Moser Opitz et al., 2016). Medical models—as reflected in the International Classification of Diseases (ICD-11; WHO, 2018) and the Diagnostic and Statistical Manual of Mental Disorders (DSM-V; APA, 2018)—label a *disorder*; namely a *developmental learning disorder with impairment in mathematics* (ICD-11) or rather a *specific learning disorder* in mathematics (DSM-V). Even though the ICD and earlier versions of DSM support an IQ-discrepancy-model (where disorder in mathematics is diagnosed based on a discrepancy between the person's IQ and math performance), recent research (e.g., Kuhn et al., 2013) suggests not to distinguish children who fulfil the discrepancy and those who do not, since the two groups "do not show qualitatively different cognitive patterns in counting, subitizing, or magnitude comparison" (p. 244). In this article, we follow this idea and do not distinguish according to students' intelligence level. We regard *students with mathematics difficulties* (following Moser Opitz et al., 2016; short: MD students) as those students who encounter difficulties with a certain set of mathematical problems both on a conceptual and procedural level (Moser Opitz et al., 2016; Scherer et al., 2016): These difficulties concern basic arithmetic such as verbal counting (e.g., counting by groups and counting principles), grouping, degrouping, the base-10 number system and understanding the place value, understanding the meaning of operations, solving word problems; as well as factual knowledge, fact retrieval, and (deficits in) working memory.

Quantity Recognition

Subitizing and conceptual subitizing as prerequisite

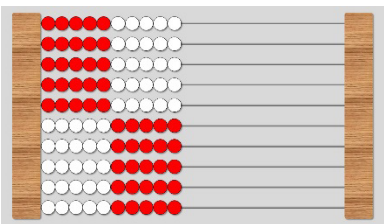
Determining quantities—to grasp sets of items and say how many they are—is a fundamental skill for children to learn. Young children already have the ability to *subitize*—"[t]he ability of recognizing a number of briefly presented items without actually counting" (Fischer et al., 2008). Besides counting, which is far from trivial for young children and requires different abilities which need to be learned (Gelman & Gallistel, 1986), also *conceptual subitizing* is a crucial ability needed to determine quantities. Conceptual subitizing means that students make use of patterning abilities, that they perceive sets through structuring them in subsets (Clements, 1999). This involves abilities such as composing and decomposing as well as understanding the concepts of number and part-whole-relationship (Starkey & McCandliss, 2014). Conceptual subitizing, unlike the above-mentioned (perceptual) subitizing, does not develop spontaneously. It involves the use of mathematical concepts, especially of number, and, thus, must be *learned* and also fostered in school (Clements, 1999)—especially for those students who may encounter difficulties in mathematics learning. Conceptual subitizing helps children to develop arithmetic strategies and develop a sophisticated number concept (Clements, 1999; Steffe & Cobb, 1988) and is involved when students work with external representations displaying quantities.

The differences in counting vs. conceptual subitizing resemble processes in visual search known as pop-out vs. serial search (e.g., Chiu et al., 2014; Williams et al., 1997): Whereas in serial search, "observers allocate complete attentional resources discretely and wholly to individual objects one at a time" (Chiu et al., 2014, p. 331), in pop-out (or parallel search), "the unique target object that differs from distractor objects by the only feature (e.g., color, orientation, etc.) in the array appears to pop-out" (Chiu et al., 2014, p. 331). Likewise, in counting, students focus on the given items one after the other; in (conceptual) subitizing, they perceive a quantity at one glance, through perceiving (some of) their characteristics directly. A similar distinction can be made for the development of word reading in childhood: Whereas young children tend to read words character by character, older students or adults typically recognize words "in an instant" and thus have fewer fixations while reading, fixating primarily on long rather than short words (see Miller & O'Donnell, 2013; Rayner, 2009). In other words, fewer fixations go along with a higher efficiency both in determining quantities and reading words with increasing age.

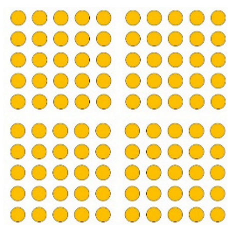
Structured whole number representations

For supporting mathematics learning, external representations play a fundamental role (Duval, 2006; Verschaffel et al., 2010). External representations visualize mathematical objects or relationships and, thus, may support students' conceptual development (Duval, 2006), especially the development of the concept of number (Obersteiner et al., 2014). The properties of numbers can be visualized through external representations in a way that "relationships between numbers can be highlighted by grouping the items rather than by using arbitrary arrangements" (Obersteiner et al., 2014, p. 354).

Two of the most popular external representations used as visualizations of the number range up to 100, which are frequently used in German-speaking countries, are the 100-bead abacus and the 100-dot-field (Gaidoschik, 2015; Landerl et al., 2017). The *100-bead abacus* (also called 100-frame, Fig. 1) has ten horizontal poles with ten beads each. In particular, each pole contains five red and five white (or blue) beads, beginning with five red beads, followed by five white ones. Poles six to ten have reversed colors to represent the 50 structure. The *100-dot-field* (Fig. 1) uses the same structures as the abacus but represents them slightly differently. Here, the dots are all in the same color (e.g., yellow or red), but there is a bigger gap between the fifth and the sixth dot of each row and between the fifth and sixth row. Both representations are extensions of smaller representations: the 20-bead abacus (also: 20-frame) and 20-dot-field (e.g., Gerlach et al., 2015; Kaufmann & Wessolowski, 2015), which only involve two poles/rows of beads/dots.



Computerized version of the
100-bead abacus
(as presented in this study)



Computerized version of the
100-dot field
(as presented in this study)

Figure 1. **Structured whole number representations used in this study.**

Schooling and intervention programs use structured whole number representations such as the 100-bead abacus and the 100-dot-field with the aim to support students' internalization of the structures of the base-10 system and to finally help students to "see" these structures even when the external representations are not physically available (Gerlach et al., 2015; Wartha & Schulz, 2012). Empirical studies have shown that the fruitful use of structured whole number representations is, however, not self-explanatory for students—especially for younger students and students with difficulties. Lindmeier and Heinze (2016) found significant differences between first-graders and adults, in particular, different strategy use, for instance, in the 10-dot-field and the 20-bead abacus (e.g., differences in counting, subitizing, or using structures). As to be expected, first-graders counted more often and subitized less often than adults. Adults used structure-based strategies more often. Obersteiner et al. (2014) studied, among others, first-graders' quantity recognition in a 20-dot-field and found that many students were able to subitize conceptually and to use the structures of 10 and 5. Finally, Rottmann and Schipper (2002) compared high and low achievers' use of the 100-dot-field in addition and subtraction tasks and found "that most high achievers do not use (or no longer use) the material. Low achievers, on the contrary, use it in a way which turns out not to be helpful at all for them"

(p. 51). Taken together, these studies and their results suggest that external representations may be fruitful support for students, but are not necessarily self-explanatory, especially for students with MD.

Related work: Quantity recognition of students with MD

Most research addressing quantity recognition of students with MD focuses on *subitizing* and *counting* of small sets up to eight items. Research on subitizing has, for instance, investigated students' *mistakes or error rates*: Whereas some researchers found that students with MD tend to make more mistakes than TD students in subitizing and counting tasks (e.g., Ashkenazi et al., 2013; Fischer et al., 2008), other studies indicate that error rates in subitizing and counting are very low for both students with and without MD (e.g., Schleifer & Landerl, 2011). Furthermore, students with MD tend to have longer *response times* in subitizing and counting tasks (Ashkenazi et al., 2013; Moeller et al., 2009; Fischer et al., 2008), and especially the slopes appear to be steeper for students with MD (Ashkenazi et al., 2013; Moeller et al., 2009). For counting, Schleifer and Landerl (2011) found that response times were comparable between the groups. Taken together, these findings suggest that students with MD may need to count even for very small quantities up to three, whereas students without MD appear to be more able to recognize them without counting. Further, it appears that students with MD may be slower in counting than TD students.

Eye tracking has offered further insights: Moeller et al. (2009) found in a case study addressing both subitizing and counting that one student with MD had substantially more fixations and another one substantially longer fixations than a TD control group, indicating what they call "a back-up counting strategy even for small object sets" (p. 371), which may explain the longer response times found in Ashkenazi et al.'s (2013) study in counting tasks.

In addition to subitizing and counting, previous research on students with MD addressed *conceptual subitizing*. Most preschool children, who have a less developed number sense and not yet a well-developed "conceptual understanding of a cardinal value as composed of subsets" (Starkey & McCandliss, 2014, p. 135), cannot yet subitize conceptually, because they lack the number concepts (Clements, 1999). This suggests that students with MD, who are delayed in the development of conceptual arithmetic understanding (Moser Opitz, 2013), may have deficits in conceptual subitizing as well. Indeed, Starkey and McCandliss (2014) found that the conceptual subitizing ability correlates with arithmetic competence. Thus, it can be expected that students with low arithmetic competence make less use of patterns and structures when perceiving quantities. Ashkenazi et al.'s (2013) as well as Olkun and Akkurt-Denizli's (2015) studies—investigating response times and error rates—confirmed that students with MD may not benefit from a structured presentation of dots in the same way as TD students.

The present study

Our study is interested in the question if and how MD students differ from TD students in quantity recognition in structured whole number representations. For pursuing this question, we use—besides error rates and response times— ET as a tool. In previous studies, ET has proven to be useful to analyze students' strategies working with such structured representations (Lindmeier & Heinze, 2016; Rottmann & Schipper, 2002; Schindler & Lilienthal, 2018). In particular, Lindmeier and Heinze (2016) conclude that their "study shows that eye-tracking data can be used to access different strategies when solving number tasks in structured representations" (p. 7). Obersteiner et al. (2014) furthermore point out that assessing children's strategies on structured whole number representations is very challenging and that a "promising approach that could be used in future studies is to combine computerized tasks and the recording of eye movements to assess children's use of external representations" (p. 369). We follow this suggestion and use ET in this study to study MD students' quantity recognition in whole number representations.

We ask (1) whether students with MD make *more mistakes* in quantity recognition tasks on the 100-bead abacus and the 100-dot-field than TD students, and (2) whether *response times* of students with MD in quantity recognition tasks on the 100-bead abacus and the 100-dot-field are longer than those of TD students.

Further, we investigate differences between students with and without MD based on ET. We ask (3) whether students with MD use different *strategies* for quantity recognition (in the 100-bead abacus and the 100-dot-field) than TD students.

Method

Participants and diagnostics of MD

We use data from a research project with 20 fifth-grade students, aged between 10;0 and 11;11 (see Tab. 1), in a German comprehensive school with medium school socioeconomic status.[1] The study took place in the first weeks of fifth grade after the students had finished the German primary school after grade four. Our diagnostic procedure comprised both qualitative diagnostic interviews and a standardized test.

Students with MD were found through *qualitative diagnostic interviews* addressing MD (following Wartha & Schulz, 2012). Such interviews were conducted with all students in the participating school whose prior mathematics grades were poor ("not passed") or who had already shown signs of mathematical difficulties in prior schooling. An experienced active teacher and researcher, who had exams and expertise both in special education and dyscalculia/MD, and who had more than 10 years of teaching experience was involved in this project and conducted the qualitative individual interviews (approx. 40 mins per student). The mathematical problems addressed topics that typically cause difficulties among students with MD: stabilized counting strategies due to a lack of understanding of operations and numbers, a lack of operative strategies even for simple addition and subtraction problems, a lack of understanding of the number range up to 100, and difficulties with the base-10 system (see, e.g., Moser Opitz, 2013; Wartha & Schulz, 2012). The interviews followed a diagnostic guideline but were, still, adaptive in a sense that problems were left out or

1 In this part of Germany, the schools are assigned "location types" based on the schools' students' achievements in mathematics, German, and English, the ratio of students with immigration background, and the percentage of unemployment/welfare-collection among parents/legal guardians. The "location types" are ranked 1 (favorable preconditions) to 5 (unfavorable preconditions). The participating school was ranked 3 (medium).

simplified if they led to overextension of the students. For each problem, the students were first asked to note it and, then, to *think aloud* (Ericsson & Simon, 1980), i.e., to verbalize their thoughts. Verbalization was facilitated by the teacher asking questions such as “How did you proceed?”. The analysis of the interviews did not intend to measure MD through a certain score but asked what difficulties and individual strategies the students showed. MD were assigned when students had severe difficulties in the addressed mathematical topics (following Wartha & Schulz, 2012). Nine students who showed MD in the diagnostic interviews participated in the study; as well as eleven TD students (picked by the teachers).

We conducted a *standardized arithmetic paper-pencil speed test* (HRT; Haffner et al., 2005) with all 20 students in a classroom setting. For the purpose of this study, only the first part of HRT, which can be used solely for diagnosing MD (at PR≤10; Haffner et al., 2005), including simple mental arithmetic, was administered (similar to Moeller et al., 2009; Schleifer & Landerl, 2011). Four subtests address arithmetic operations (addition, subtraction, multiplication, and division; e.g., 5 + 3 = _ or 3 x 5 = _). The two further subtests contain completion tasks (e.g., _ - 2 = 6) and symbolic quantity comparison (e.g., 18 _ 7 (correct response: >)).

In the standardized test, we figured that the performances of one student of the MD group and two students of the TD group were assigned a percentile rank of 14 and 16 resp., which is within the “at risk zone” (Haffner et al., 2005, p. 20). Thus, we conducted qualitative interviews also with the two low performing students of the TD group. The analysis revealed that one of the students showed MD (following Wartha & Schulz, 2012) and thus was counted to the MD group, whereas the other student did not show such difficulties. Accordingly, both groups of students finally comprised ten students. Among the ten students with MD, there were four with special educational needs (in learning, social and emotional development, and physical development).

Table 1. ***Participants of the study.***

Participants	**Number of students**	**Age** Mean (SD)	**Gender** Numbers Female/Male	**Special Needs** Number of students	**Mother tongue**	**HRT (T-Value)** Mean (SD)
MD Group	10	10;10 (0;6)	6F 4M	4	7 German 3 Other	26.2 (10.7)
TD Group	10	10;7 (0;5)	4F 6M	0	8 German 2 Other	49.3 (6.4)
All	20	10;8 (0;6)	10F 10M	4	15 German 5 Other	37.8 (14.5)

Apparatus

The experiments were carried out with the wearable eye tracker Tobii Pro Glasses 2. For this explorative study, we decided not to use a remote eye tracker. Remote eye trackers, connected to the monitor, have certain advantages, especially with regard to analyzing measures such as mean fixation durations, etc. in large quantitative studies. However, in our explorative study, we wanted to investigate students' gazes in more detail. We wanted to observe, for instance, when students' gazes went out of the monitor, and where exactly (to the interviewer, or similar), which was of interest in this explorative study. A remote eye tracker would not have tracked these gazes. The use of glasses further offered us the opportunity to record the video of the scene view, as well as ET and audio data in a time-synchronized manner. This prevented an additional triangulation with external audio data, which is required when remote eye-trackers are used. For our study, ET glasses did not have a critical disadvantage as compared to remote eye trackers. Using glasses in combination with tasks presented on a monitor is, indeed, not unusual in mathematics education research (see, e.g., Shvarts, 2018).

The ET device Tobii Pro Glasses 2 is fitted with a high-resolution scene camera, a microphone, and ET sensors. The scene camera records an HD video stream (1920 x 1080 pixels) of the participants' view with a field of view of 82 deg. (horizontal) x 52 deg. (vertical). Infrared illuminators and ET sensors (cameras pointing towards the participant) afford to track the eye gaze with a sampling rate of 50 Hz up to 100Hz by identifying the displacement of corneal reflections from the detected center of the pupil (corneal reflection ET). Due to the optical ET technique and its low weight (45 grams), the glasses are very unobtrusive. Stimuli were presented on a 24" monitor driven at a refresh rate of 60 Hz with a resolution set to 1920 x 1080 pixels. During the experiment the students sat still on a firm chair with a viewing distance of approximately 0.5 m and stable viewing angle.

Tasks and procedure

Representations and tasks. We used a computerized version of the 100-bead abacus (Fig. 1). The numbers were systematically chosen so that all ones and tens were included once; plus 100. This led to eleven tasks (arranged according to size: 7, 15, 20, 31, 43, 54, 68, 76, 89, 92, 100) which were presented in randomized order. The computerized 100-dot field used yellow dots (Fig. 1). The same numbers as in the abacus were used but in different randomization.

Procedure. The students were tested individually in a quiet room. They were seated in front of an external computer monitor, wearing the ET glasses. For adjusting the eye tracker, a one-point calibration was conducted. Then, before the students started working on the tasks, they first saw a picture of the respective representation (100-bead abacus, 100-dot field) and were asked to describe it. This was followed by two practice trials, whose exact numbers were not used in further test trials. The students were instructed to name the number of dots in every task as fast and correct as possible. In between the tasks, the students were instructed to fixate a star in the middle of the screen before the next task appeared. Response times were recorded from the appearance of the number stimulus until the students gave their answers. The students received no response whether their answers were correct or incorrect. Verbal answers were recorded through the audio-recorder of the ET glasses. The responses were then rated either correct or incorrect in the analysis, thus a maximum of one error was assigned for each task.

ET Data and Analysis

Data. Our ET data analysis uses gaze-overlaid videos (videos recorded by the ET glasses overlaid with eye gazes presented as a semi-transparent dot in the video) since our aim was to analyze student strategies in as much detail as possible. We refrained from comparing ET measures (e.g., fixation durations) since our main aim was not to investigate statistical differences between the groups' eye movement measures, but to reveal what strategies they used and if those differed between groups. In the gaze-overlaid videos, all gaze points of the students were included. For instance, we did not differentiate between "real fixations" (which need to be detected with a certain threshold, typically 200-300ms, Holmqvist et al., 2011) and shorter gazes, because qualifying eye movements in this way has the disadvantage that the analysis becomes dependent on the threshold value. In our study, we did not intend to exclude certain eye movements but wanted to take all of them into account to infer student strategies.

Data Analysis. We analyzed the data in four stages. *First stage:* We assigned strategies to the clips of gaze-overlaid videos inductively, following Mayring's (2014) qualitative content analysis in an inductive manner. For instance, when a student was to determine the number 89 on the abacus and focused, one after the other, on every last bead in each row, from top to bottom, we assigned that (s)he was counting rows (see Schindler, 2019, for an example). The first three analysis steps (see Tab. 2 for an example) were (1) *description* of the student's *eye movements* in the video, (2) *paraphrasing* the content-bearing semantic elements in the description relevant for identifying student strategies, and transposing them to a uniform stylistic level, and (3) *category development*, i.e., inductively assigning categories to the data with according descriptions.

Table 2. ***Data analysis steps: Example.***

Description of eye movement	Paraphrase/Transpose	Category
Gazes one after the other on the following beads: 50, 10, 20, 29/30, 39/40, 50, 60, 70, 80, 89, 88, 87, 86, 85, 84, 83, 82, 81 *Quick saccade over the abacus* *Gazes one after the other on the following beads: 10, 20, 30, 40, 50, 60, 70, 80, 89, 88, 87, 86, 85, 84, 83, 82, 81* *Short outlier, then "89"*	Short orientation (50) Counting rows Counting the ones Re-counting rows Re-counting the ones Answer "89"	Main strategy: Counting rows Counting ones for determining the ones

Second stage: After having categorized all data, we refined the set of categories in a *category revision* step. This revision step resulted in six categories of strategies, describing the main strategies that the students used to determine quantities. These categories apply to all tasks used in this study. These categories were:

1) **Counting ones**: Students counted every dot/bead given in the task. That means the students focused on every single dot/bead, not only counted the incomplete rows.
2) **Counting fives**: Students counted "fives", that means groups of five, to determine the quantity. Here, the students focused on one dot/bead in every group of five. For instance, when determining 92, they counted 18 fives and added 2.
3) **Counting rows**: Students counted the given rows. Here, they looked at the rows one after the other—from top to bottom or vice versa.
4) **Subitizing the biggest unit**: Students grasped the biggest unit in one glance. For instance, when determining 68, they perceived 50 in one glance. The students either gazed over the respective biggest unit (here: 50) briefly or not at all, and started to look at the rest (here: 18) directly.
5) **Last row**: This strategy occurred in particular when counting 92. The students focused only on the two given dots in the last row and stated directly that it was 92.
6) **Subtraction**: This strategy occurred when counting 89 and 92 in particular. Here, the students focused on the missing dots/beads, and subtracted.

Third stage: For the subsequent statistical analysis, we estimated a-priori which strategies can be theoretically expected for each task (see Tab. 3). The categories *counting ones* and *counting fives* are applicable in all tasks. *Counting rows* is applicable in all tasks but one (7), where there is no full row given. *Subitizing the biggest unit* applies to 100 (where the biggest unit is 100), 92, 89, 76, 68, and 54 (where the biggest unit is 50), and to 43, 31, 20, where the units of 30 or 20 can be subitized. *Subtraction* and *last row* are applicable in tasks 89 and 92; in other tasks, they would not be beneficial. Theoretically, one could, for example, even subtract in task 7, through counting the remaining 93 and then subtracting them from 100, but since this would be very inefficient, *subtraction* is an unlikely strategy here.

Table 3. ***Theoretically expected strategies: A-priori estimation.***

	Quantity										
	7	**15**	**20**	**31**	**43**	**54**	**68**	**76**	**89**	**92**	**100**
Subtraction									x	x	
Last row									x	x	
Subitizing biggest unit			x	x	x	x	x	x	x	x	x
Counting rows		x	x	x	x	x	x	x	x	x	x
Counting fives	x	x	x	x	x	x	x	x	x	x	x
Counting ones	x	x	x	x	x	x	x	x	x	x	x

Note: Crosses "x" indicate expected strategies for the respective tasks (quantities).

Fourth stage: Based on the developed set of categories, two raters coded all gaze-overlaid data independently from each other. Both raters were researchers acquainted with this kind of research: They had been involved in the data generation in this (or a similar) ET study and had analyzed similar data (not these data in particular) in different ways before. We finally calculated the interrater reliability using Cohen's kappa (Cohen, 1960). The interrater agreement was 0.96, which can be considered excellent or almost perfect (Landis & Koch, 1977).

Statistical Analysis. For the statistical analysis, the program SPSS 25® was used.

Error rates: As the error rates were low – both for abacus and field – and non-normally distributed (see Tab. 4), we followed the recommendations provided by Kubinger, Rasch and Yanagida (2011) and used the *principal component test* (see Läuter, Glimm, & Kropf, 1998). This test is suited for small sample size and non-normal data with doubtful homogeneity of variances. The null hypothesis for this test is that both means are the same for the two groups.

Reaction times: For identifying possible differences in reaction times between the MD and TD group, we conducted a repeated measurement (condition x group) ANOVA with the response times (condition means abacus vs. field; group means MD vs. TD group).

Student strategies: For comparing students' strategy use, we carried out a Mann-Whitney U test for every strategy separately for the two conditions (abacus, field) to analyze differences between MD and TD in total use of strategies (sum of particular strategies over all quantities). Results were alpha adjusted (Bonferroni-Holm). On a single item (task) level, given the large number of comparisons and the small sample size, we used effect sizes for the comparison rather than significance testing, drawing on *Cramérs V*.

Results

Error rates

Error rates were generally low (Tab. 4). On average, the students made about one (10%) error. Although the TD group made even slightly more errors than the MD group, no significant differences between the groups were found both for abacus and field errors ($t\,(16.54) = 0.67$, $p = .21$).

Table 4. ***Errors. Comparison of MD and TD. Mean (SD), skewness, kurtosis, Shapiro-Wilks Test for normality (W)***

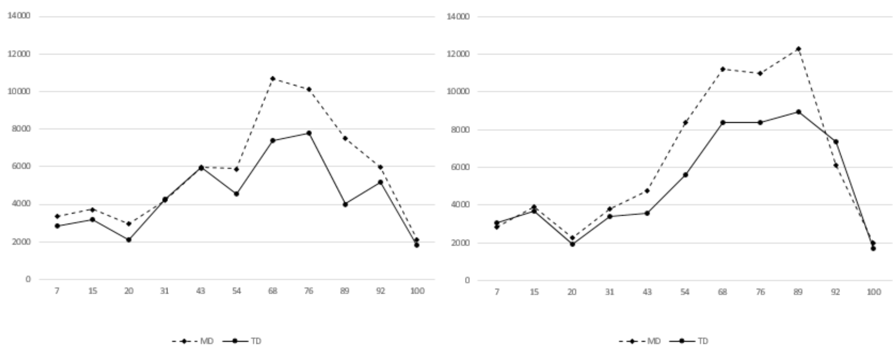
	MD				TD			
	M (SD)	*skewness*	*kurtosis*	*W*	*M (SD)*	*skewness*	*kurtosis*	*W*
Abacus	0.8 (1.40)	1.66	2.05	.65**	1.0 (1.25)	1.72	3.42	.77**
Field	1.0 (0.94)	0.99	1.18	.84*	1.5 (1.43)	1.70	3.96	.81*

Note: * $p < .05$, ** $p < .01$

Response times

In this section, we compare the reaction times between MD and TD students. A repeated measurement (condition x group) ANOVA with the response times (collapsed separately for both conditions as sum scores over all items) revealed an insignificant main effect for the condition (abacus/field). This means that there was no significant difference between response times on the abacus vs. field. However, the main effect for group was significant ($F\,(3,18) = 5.89$, $p = .026$, $\eta_p^2 = .25$), indicating significantly higher total response times for the MD group. The interaction term (condition x group) was not significant.

Separate repeated measurement ANOVAS for the two conditions (abacus, field) revealed that the response times were different for different stimuli (quantities) in every condition. (abacus: $F\,(10, 180) = 23.72$, $p < .001$, $\eta_p^2 = .57$; field: $F\,(10, 180 = 10.20$, $p < .001$, $\eta_p^2 = .36$). However, there was no number x group interaction, indicating that the profiles of response times over the items (quantities) were essentially parallel for MD and TD (Fig. 2), and only one univariate statistic for the comparison of groups was significant (abacus quantity 54).



Note: left figure: field; right figure: abacus.

Figure 2. **Response times (median reaction times) for 11 different stimuli (quantities).**

Results based on ET

Students' strategy use over all tasks. Table 5 informs about what theoretically expected strategies were used/not used in the respective tasks (quantities) (blank parts indicate that the respective strategies were not used).

Due to missing assumptions for parametric testing (i.e., extremely skewed data), we carried out a Mann-Whitney U test for every strategy separately for the two conditions (abacus, field) to analyze differences between MD and TD students in total use of strategies (sum of particular strategies over all quantities). Alpha adjusted (Bonferroni-Holm) results revealed no differences in the use of strategies in the abacus condition. In the field condition, TD students applied "subitizing" in significantly more tasks (quantities) than MD students ($U = 7.50$, $p = .03$, $d = 1.78$; Tab. 5).

Table 5. ***Sum of students' use of strategies over all quantities.***

		Subtraction		Last row		Subitizing Biggest Unit		Counting Rows		Counting Fives		Counting Ones	
		M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Abacus	MD	0.13	(0.35)	0.75	(0.46)	2.75	(1.49)	5.25	(2.43)	2.00	(2.07)	0.13	(0.35)
	TD	0.86	(1.07)	0.71	(0.76)	4.00	(1.91)	4.29	(1.70)	1.00	(0.58)	0.14	(0.38)
	d	0.63		0.12		1.55		0.53		0.59		0.03	
Field	MD	0.67	(0.50)	0.33	(0.50)	2.78	(2.05)	4.78	(2.59)	2.33	(3.04)	0.11	(0.33)
	TD	0.75	(0.46)	0.88	(0.35)	5.38	(1.41)	2.63	(1.06)	1.25	(0.46)	0.13	(0.35)
	d	0.14		1.02		1.78	*	0.96		0.05		0.02	

Note: * $p < .05$

Table 6. ***Use of strategies. Percentage of children having applied particular strategies and Cramérs V for the comparison of TD and MD.***

		Abacus						Field					
		Subtr	Last row	Subitiz.	C Rows	C 5s	C 1s	Subtr	Last row	Subitiz.	C Rows	C 5s	C 1s
7	MD					90.00	10.00					90.00	10.00
	TD					80.00	20.00					90.00	10.00
	Cramérs V					*.14*	*.14*					*.00*	*.00*
15	MD				70.00	30.00					70.00	30.00	
	TD				88.89	11.11					70.00	30.00	
	Cramérs V				.23	.23					.00	.00	
20	MD			70.00	30.00					**60.00**	30.00	10.00	
	TD			90.00	10.00					**100.00**	0.00	0.00	
	Cramérs V			.25	.25					**.50**	.42	.23	
31	MD			33.33	66.67					10.00	70.00	20.00	
	TD			40.00	60.00					50.00	50.00	0.00	
	Cramérs V			.07	.07					.44	.20	.33	
43	MD			**0.00**	90.00	10.00				20.00	70.00	10.00	
	TD			**50.00**	50.00	0.00				33.33	66.67	0.00	
	Cramérs V			**.58**	.44	.23				.15	.04	.22	
54	MD			20.00	70.00	10.00				11.11	66.67	22.22	
	TD			60.00	40.00	0.00				40.00	60.00	0.00	
	Cramérs V			.41	.30	.23				.33	.07	.36	

68	MD			20.00	70.00	10.00			**30.00**	**60.00**	10.00
	TD			20.00	80.00	0.00			**88.89**	**11.11**	0.00
	Cramérs V			.00	.12	.23			**.60**	**.51**	.22
76	MD			22.22	66.67	11.11			**33.33**	44.44	22.22
	TD			22.22	77.78	0.00			**88.89**	11.11	0.00
	Cramérs V			.00	.12	.24			**.57**	.37	.35
89	MD	**10.00**	0.00	10.00	**70.00**	10.00	60.00		10.00	20.00	10.00
	TD	**60.00**	10.00	30.00	**0.00**	0.00	70.00		20.00	10.00	0.00
	Cramérs V	**.52**	.23	.25	**.73**	.23	.11		.14	.14	.23
92	MD	**0.00**	60.00	0.00	30.00	10.00		**40.00**	10.00	40.00	10.00
	TD	**37.50**	62.50		0.00	0.00		**88.89**	11.11	0.00	0.00
	Cramérs V	**.50**	.03		.40	.22		**.51**	.02	.49	.22
100	MD			100.00					100.00		
	TD			100.00					100.00		
	Cramérs V			.00					.00		

Note: White background of cells indicates that the particular strategies were theoretically expected for the respective tasks (quantities). Gray shading, if the strategies were not theoretically expected for the respective tasks (quantities). Zeros omitted if occurred in both groups. Bold text indicates big effects (*Cramérs V* > .50).

Students' strategy use per task: Comparing the strategies on a single item (quantity) level revealed differences in the frequency of strategy use between MD and TD students (Tab. 6). Given the large number of comparisons and the small sample size, we rely on the effect sizes for the comparison rather than significance testing. Differences with Cramérs V > .50 (big effect) revealed that certain strategies were applied substantially more often by TD students than by MD students. This is the case for subitizing the biggest unit (43 abacus; 20, 68, 76 field), subtraction (89, 92 abacus), and last row (89 abacus). That means that TD students used the strategies subitizing, subtraction, and last row more often in the mentioned tasks than MD students. On the other side, MD students counted rows more often than TD students, which effect sizes of 0.51 (76 field) and 0.73 (89, abacus) hint at.

Discussion

The aim of this study was to evaluate whether students with MD differ from TD students in quantity recognition in structured whole number representations. To pursue this aim, we conducted an explorative study including a group of ten students with MD and ten TD students. We investigated error rates, response times, and students' strategy use based on qualitative ET video data.

Quantity recognition in structured whole number representations—general results

Quantity recognition in representations such as the 100-bead abacus and the 100-dot field is a complex process and involves subitizing and counting as well as conceptual knowledge about numbers, partitions, and number structures. In our experiments with grade 5 students, we found that both, students with MD and TD students, had generally low error rates. This suggests that students at this age are generally able to grasp numbers in such representations in one or another way—which may be related to the fact that such mathematical tasks and their representations are generally introduced in grade 2 in German classrooms (Gaidoschik, 2015; Krauthausen, 2018). Thus, students in grade 5 may be experienced using them. Our results indicate that it takes the students less time to determine smaller quantities as well as bigger quantities close to 100 (shortest for 100). So, unlike for smaller numbers of unstructured dots, where response times tend to increase with number (e.g., Schleifer & Landerl, 2011), the response times for quantity recognition at the 100-bead abacus and 100-dot field do not appear to increase in a monotonous way with number. This connects to Obersteiner et al.'s (2014) finding for the 20-bead abacus that "although response times depended on the number of dots to some extent, this dependency was substantially reduced by the structure of the twenty-frame" (p. 365). In our study, we found that numbers such as 68, 76, and 89 required the longest time for processing. This is explainable through the complex processes involved to grasp the particular numbers of tens and ones.

Our study brought many student strategies to light. Especially strategies such as subitizing the biggest unit of 20 or 30, or counting fives even for quantities such as 92, which leads to long response times were, to the best of our knowledge, not found in previous studies. In this regard, our work might be the starting point for a string of similar future investigations that collect further data with the aim to understand students' quantity recognition better.

Quantity recognition in structured whole number representations in students with MD

Our study results indicate that students with MD differ in their quantity recognition in whole number representations from TD students. These differences were not perceivable in the *error rates*, which were very low for both groups of students and actually slightly *higher* for the TD students. This relates to Schleifer and Landerl's (2011) finding that error rates for subitizing and counting up to eight were low for both groups of students. Our findings are not in line with Ashkenazi et al.'s (2013) as well as Olkun and Akkurt-Denizli's (2015) findings that MD students had higher error rates than their controls in unstructured and structured dot counting. The low error rates in our study may be a result of students' prior school experiences with the given representations, whereas Ashkenazi et al. and Olkun and Akkurt-Denizli's used canonic (dice-like) representations which are not necessarily familiar to the students.

In our study, the *response times* of students with MD were significantly longer in total, meaning that students with MD needed in sum significantly more time for all tasks than the TD students. This relates to research on subitizing, counting, and conceptual subitizing, where researchers found similarly longer response times for MD students (e.g., Moeller et al., 2009; Fischer et al., 2008; Ashkenazi et al., 2013; Olkun & Akkurt-Denizli, 2015). The differences on a task level were generally not significant in our study. Whether this is due to the small sample size will have to be investigated in future studies.

The *ET data analysis* revealed differences between the MD students' strategy use as compared to the TD students. On a general level, with respect to the total use of strategies (i.e., the sum of strategies over all quantities), TD students applied the strategy *subitizing the biggest unit* significantly more often in the 100-dot-field than MD students. This means that TD students more often grasped a certain structure (e.g., 50 or 20) in a glance than MD students to determine quantities. This relates to the finding that the conceptual subitizing ability correlates with arithmetic competence (Starkey & McCandliss, 2014), which suggests that students with low arithmetic competence make less use of patterns and structures when perceiving quantities. Our results are in line with these findings.

On a task level, the TD students used the strategies *subitizing the biggest unit* (43 abacus; 20, 68, 76 field), *subtraction* (89, 92 abacus), and *last row* (89 abacus) more often in the mentioned tasks than MD students. On the other side, MD students counted rows more often than TD students (89 abacus; 76 field). The TD students tended to use elaborated strategies more often, in which they made use of adequate and task-relevant structures. For instance, in 76 on the 100-dot-field, the TD students tended to grasp 50 at a glance and counted rows from 50 onwards, whereas MD students tended to count all rows from the beginning. MD students did less often make use of the structured representation of 50 in this task. As an overall trend, TD students tended to use more elaborated strategies, where they made use of facilitating structures, and which were efficient in the respective tasks. MD students predominantly counted rows. This indicates that they made use of the structure of ten. As compared to the TD students, however, they did not use different strategies flexibly, but rather repeatedly tended to use the counting rows strategy for many tasks. MD students partially also used the *counting fives* strategy, for instance, when determining 92, they counted 18 fives and added 2. Of course, such a strategy is less efficient, slower, and more error-prone than other strategies, such as *counting rows*.

With respect to the tasks and their potential to differentiate between MD and TD students, we found that for 7 and 15, the use of strategies of MD and TD students was (almost) identical. In 20 and 31 on the 100-dot-field, the TD students tended to grasp the tens (20/30) more often as a whole (*subitizing biggest unit*) than the MD students. Similarly, in 43 and 54 on the 100-abacus, the TD students tended to grasp the tens more often as a whole (*subitizing biggest unit*), whereas the MD students tended to *count rows*. In 68 and 76 on the 100-dot-field, the TD students tended to grasp 50 as a whole (*subitizing biggest unit*), whereas the MD students tended to *count rows*. This effect was not perceivable in the 100-abacus (see below for an explanation). For 89 and 92, we found that the TD students more often used *subtraction* or *last row*, with some big effects in the statistical analysis, whereas the MD students counted rows more often. Finally, for 100 all students used the same strategy: they perceived 100 in one glance, through *subitizing*. In sum, for the smaller quantities (7 and 15), the strategies used by students with MD did not differ much from those of TD students. However, they differed noticeably for all the other quantities (except for 100) with some big effects for the comparison of MD and TD. Only for one of these tasks, also response times had shown significant differences between the MD and TD group (abacus 54). Overall, it can be said that MD students tended to count rows (and fives) more often and that they less often used *subitizing the biggest unit*, *subtraction*, and *last row*. This finding meets previous findings that MD students tend to be less flexible in their strategy use and tend to have problems both on a conceptual and procedural level (Moser Opitz et al., 2016; Scherer et al., 2016), which may hinder them to use strategies such as *subitizing biggest unit*, *subtraction*, and *last row*. In this respect, the current findings are in line with previous findings.

When comparing students' strategies on the 100-dot-field and the 100-abacus, we found that in both representations, MD students tended to count rows. However, TD students did not always make use of more efficient strategies in both representations. For instance, for 68 and 76 on the abacus, TD students did seldom make use of the 50 (*subitizing biggest unit*) but counted all rows instead—similar to the MD students. On the contrary, for 68 and 76 on the 100-dot-field, the majority of TD students did make use of the 50 and counted onwards from 60. It might be that the structure of the 100-dot-field facilitated them to grasp 50, which was more difficult on the 100-abacus. Possibly, the gap between the fifth and the sixth dot on the 100-dot-field helped the students more than the reversed colors in the 100-abacus to make use of 50. Future research should investigate such differences between these representations further.

On the use of the 100-bead abacus and 100-dot field—for facilitating students' learning

Our findings indicate that students with MD partially use the properties of the structured representations, which may help them to compensate for possible deficits in subitizing and counting (e.g., Schleifer & Landerl, 2011). However, we found that still in grade 5, some students with MD do not use facilitating structures as intended, and partially use inefficient, error-prone, and time-consuming strategies such as counting 5s.

Thus, our findings are in line with previous research that showed that even young students can be able to use structures in structured representations (Lindmeier & Heinze, 2016; Obersteiner et al., 2014), but that especially the structures in such representations are not easy to use for students with MD (Rottmann & Schipper, 2002). We see a research gap addressing the question of how MD students' quantity recognition in structured whole number representations can be fostered so that they become more able to use the facilitating structures in these representations. This appears to be crucially important to study since these representations are frequently used in order to foster students with MD and facilitate their learning (e.g., Wartha & Schulz, 2012). We hope that future research will address this gap.

Eye Tracking for investigating MD

On a methodological level, the results of this study indicate that *error rates* in quantity recognition tasks in whole number representations may not be an appropriate indicator for MD: We did not find significant differences between MD and TD students with the given number of 10 MD and 10 TD students. Rather, the TD students made slightly *more* mistakes. *Response times* did not show significant differences at the task level either. However, in a holistic view, considering the sum of response times over all tasks, the differences were significant between the groups. This indicates that the sum of response times over a series of tasks may give researchers hints on possible MD. Still, we used only one particular series of tasks in this study, and the question arises of how many tasks would be sufficient to reliably identify MD among students. Answering this question is outside the scope of this article and should be investigated in future studies. Finally, our study demonstrated the potential of ET as a research method. Due to the complexity of the mathematical activity studied in our project, we did not limit ourselves to analyzing statistical measures, such as saccade lengths or fixation durations, but analyzed the raw video data using content analysis and worked out categories inductively. By doing so, we found a multitude of student strategies of which several were not found in previous research. ET appears to be especially beneficial for students with MD, who may not be confident to explain their strategies in think-aloud protocols (Ericsson & Simon, 1980), may suffer from anxieties, or may even not consciously think about subconscious mental processes (Schindler & Lilienthal, 2018).

Imagining an extension of the key ideas of our study brings a compelling application to mind: The ET-based research method used in this paper could be further developed as a tool for teachers that allows identifying the vector of strategies available to a student, enabling more individually targeted teaching. Developing such an ET-based tool requires to automatize (at least in parts) the analysis that we carried out in this study manually. If a full automatization can be reached, even applications for a computer could be constructed which present a sequence of tasks that adapt to the individual strategies identified for the particular student in each task.

Limitations

One limitation of our statistical analysis is the small sample size of ten TD students and ten students with MD. However, this was an explorative study, using ET to identify students' strategies in the tasks presented. Low numbers of participants are not unusual in ET studies – as for instance in Moeller et al. (2009), who investigated eight (TD) vs. 2 (MD) students in subitizing and counting tasks. For this kind of methodology, which requires substantial data processing, our sample size was reasonable. Further, our study had partially an explorative character, aiming to find student strategies and differences between TD and MD children. Our study revealed that there are (at least) six categories of strategies that students use to determine quantities on the abacus and field. Through our inductive approach, we found these strategies, some of which were, to the best of our knowledge, not found in previous studies. This is a novel finding and was the base for our statistical analysis of differences between MD and TD students.

Even though the data in our study did—due to the small sample size—not consistently allow for significance testing, its findings are a springboard for further research. Further studies based on bigger samples will overturn the limitations of this study.

References

American Psychiatric Association (APA) (Ed.) (2018). Diagnostic and statistical manual of mental disorders (DSM–5). http://www.dsm5.org/ Accessed 2018-01-14.

Ashkenazi, S., Mark-Zigdon, N., & Henik, A. (2013). Do subitizing deficits in developmental dyscalculia involve pattern recognition weakness? *Developmental Science, 16*, 35–46.

Baroody, A. J. (1986). Counting ability of moderately and mildly handicapped children. *Education & Training of the Mentally Retarded, 21*, 289–300.

Benz, C. (2014). Identifying quantities – Children's constructions to compose collections from parts or decompose collections into parts. In U. Kortenkamp, B. Brandt, C. Benz, G. Krummheuer, S. Ladel, & R. Vogel (Eds.), *Early Mathematics Learning* (pp. 189–203). New York: Springer.

Chiu, E., Rigoli, L., & Spivey, M. (2014). Eye-tracking investigation of visual search strategies when mediated by language. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36). https://escholarship.org/uc/item/6311g331

Clements, D. H. (1999). Subitizing: what is it? Why teach it? *Teaching Children Mathematics, 5*, 400–405.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational studies in mathematics, 61*, 103–131.

Ericsson, A. & Simon, H. A. (1980). Protocol analysis: Verbal aspects as data. *Psychological Review, 87*, 215–251.

Fischer B, Gebhardt C, Hartnegg, K. (2008). Subitizing and visual counting in children with problems in acquiring basic arithmetic skills. *Optometry & Vision Development, 39*, 24–29.

Gaidoschik, M. (2015). Einige Fragen zur Didaktik der Erarbeitung des „Hunderterraums". *Journal für Mathematik-Didaktik, 36*, 163–190.

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22*, 23–27.

Gelman, R., & Gallistel, C. R. (1986). *The child's understanding of number.* Cambridge, MA: Harvard University Press.

Gerlach, M., Fritz, A., Ricken, G., Schmidt, S. (2015). *Kalkulie – Diagnose- und Trainingsprogramm für rechenschwache Kinder.* Berlin, Germany: Cornelsen.

Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *Heidelberger Rechentest (HRT 1-4).* Göttingen, Germany: Hogrefe.

Kaufmann, S. & Wessolowski, S. (2015). *Rechenstörungen – Diagnose und Förderbausteine.* Seelze, Germany: Kallmeyer.

Krauthausen, G. (2018). Einführung in die Mathematikdidaktik – Grundschule (4.Auflage). Berlin, Germany: Springer Spektrum.

Kubinger, K. D., Rasch, D., & Yanagida, T. (2011). *Statistik in der Psychologie: vom Einführungskurs bis zur Dissertation.* Hogrefe Verlag.

Kuhn, J. T., Raddatz, J., Holling, H., & Dobel, C. (2013). Dyskalkulie vs. Rechenschwäche: Basisnumerische Verarbeitung in der Grundschule. *Lernen und Lernstörungen, 2*, 229–247.

Landerl, K., Vogel, S., & Kaufmann, L. (2017). *Dyskalkulie.* München, Germany: UTB.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Läuter, J., Glimm, E., & Kropf, S. (1998). Multivariate tests based on left-spherically distributed linear scores. *The Annals of Statistics*, *26*, 1972–1988.

Lindmeier, A., & Heinze, A. (2016). *Strategies for recognizing quantities in structured whole number representations – A comparative eye-tracking study.* Paper presented at 13th International Congress on Mathematical Education (ICME-13), 2016.

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution.* Klagenfurt, Austria: Beltz.

Miller, B., & O'Donnell, C. (2013). Opening a window into reading development: Eye movements' role within a broader literacy research framework. *School Psychology Review*, *42*, 123–139.

Moeller, K., Neuburger, S., Kaufmann, L., Landerl, K., & Nuerk, H. C. (2009). Basic number processing deficits in developmental dyscalculia: Evidence from eye tracking. *Cognitive development*, *24*, 371–386.

Moser Opitz, E. (2013). *Rechenschwäche, Dyskalkulie: Theoretische Klärungen und empirische Studien an betroffenen Schülerinnen.* Bern, Switzerland: Haupt.

Moser Opitz, E., Freesemann, O., Prediger, S., Grob, U., Matull, I., & Hußmann, S. (2016). Remediation for students with mathematics difficulties: An intervention study in middle schools. *Journal of Learning Disabilities*, *50*, 724–736.

Obersteiner, A., Reiss, K., Ufer, S., Luwel, K., & Verschaffel, L. (2014). Do first graders make efficient use of external number representations? The case of the twenty-frame. *Cognition and Instruction*, *32*, 353–373.

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction*, *23*, 125–135.

Olkun, S., & Akkurt-Denizli, Z. (2015). Using basic number processing tasks in determining students with mathematics disorder risk. *Dusunen Adam: Journal of Psychiatry & Neurological Sciences*, *28*, 47–57.

Rottmann, T., & Schipper, W. (2002). Das Hunderter-Feld—Hilfe oder Hindernis beim Rechnen im Zahlenraum bis 100?. *Journal für Mathematik-Didaktik*, *23*, 51–74.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.

Scherer, P., Beswick, K., DeBlois, L., Healy, L., & Opitz, E. M. (2016). Assistance of students with mathematical learning difficulties: how can research support practice?. *ZDM*, *48*, 633–649.

Schindler, M. (2019). *Student counting rows and ones on abacus 89*. Retrieved from: https://youtu.be/C2mlnkxHBeg

Schindler, M., & Lilienthal, A. J. (2018). Eye-tracking for studying mathematical difficulties—also in inclusive settings. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 115–122). Umeå, Sweden: PME.

Schindler, M., Bader, E., Lilienthal, A.J., Schindler, F., & Schabmann, A. (2019). Quantity recognition in structured whole number representations of students with mathematical difficulties: An eye-tracking study. *Learning Disabilities: A Contemporary Journal, 17*, 5– 28.

Schipper, W. (2011). Vom Calculieren zum Kalkulieren – Materialien als Lösungs- und als Lernhilfen. In A. Steinweg (Eds.), *Medien und Materialien. Tagungsband des AK Grundschule in der GDM* (pp. 71–85). Bamberg, Germany: University of Bamberg Press.

Schleifer, P., & Landerl, K. (2011). Subitizing and counting in typical and atypical development. *Developmental Science*, *14*, 280–291.

Shvarts, A. (2018). A dual eye-tracking study of objectification as student-tutor joint activity appropriation. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 171–178). Umeå, Sweden: PME.

Starkey, P., & Cooper, R. G. (1995). The development of subitizing in young children. *British Journal of Developmental Psychology, 13*, 399–420.

Starkey, G. S., & McCandliss, B. D. (2014). The emergence of "groupitizing" in children's numerical cognition. *Journal of Experimental Child Psychology, 126*, 120–137.

Steffe, L. P. & Cobb, P. (1988). *Construction of Arithmetical Meanings* and *Strategies.* New York: Springer.

Verschaffel, L., De Corte, E., de Jong, T., & Elen, J. (Eds.) (2010). *Use of representations in reasoning and problem solving. Analysis and improvement.* London, UK: Routledge.

Wartha, S., & Schulz, A. (2012). *Rechenproblemen vorbeugen. Grundvorstellungen aufbauen. Zahlen und Rechnen bis 100.* Berlin, Germany: Cornelsen.

Williams, D. E., Reingold, E. M., Moscovitch, M., & Behrmann, M. (1997). Patterns of eye movements during parallel and serial visual search tasks. *Canadian Journal of Experimental Psychology*, *51*, 151–164.

World Health Organisation (WHO) (Ed.) (2018). ICD-11: International classification for diseases for mortality and morbidity statistics (ICD-11 MMS) 2018 version. http://www.who.int/classifications/icd/. Accessed 2018-09-10.