Shifts in student attention on algorithmic and creative practice tasks

Mathias Norqvist[1,3] · Bert Jonsson[2,3] · Johan Lithner[1,3]

Abstract

In mathematics classrooms, it is common practice to work through a series of comparable tasks provided in a textbook. A central question in mathematics education is if tasks should be accompanied with solution methods, or if students should construct the solutions themselves. To explore the impact of these two task designs on student behavior during repetitive practice, an eye-tracking study was conducted with 50 upper secondary and university students. Their eye movements were analyzed to study how the two groups shifted their gaze both within and across 10 task sets. The results show that when a solution method was present, the students reread this every time they solved the task, while only giving minute attention to the illustration that carried information supporting mathematical understanding. Students who practiced with tasks without a solution method seemed to construct a solution method by observing the illustration, which later could be retrieved from memory, making this method more efficient in the long run. We discuss the implications for teaching and how tasks without solution methods can increase student focus on important mathematical properties.

Keywords Mathematical reasoning · Eye tracking · Solution strategies · Consecutive tasks

1 Introduction

As eye-tracking equipment has become more accessible for researchers in mathematics education during the last decades, more and more studies that utilize this technique are being published (Strohmaier et al., 2020). By using eye tracking, the present study aims at getting a deeper understanding of how students' mathematical reasoning is affected over a series of similar tasks. Previous studies have shown that constructing solution methods during the learning phase can be more efficient at later testing than utilizing provided solution methods (e.g., Jonsson et al., 2014; Karlsson Wirebring et al., 2015; Norqvist et al., 2019b).

✉ Mathias Norqvist
mathias.norqvist@umu.se

[1] Department of Science and Mathematics Education, Umeå University, 901 87 Umea, Sweden

[2] Department of Applied Educational Science, Umeå University, Umea, Sweden

[3] Umeå Mathematics Education Research Centre, Umeå University, Umea, Sweden

However, these studies have not considered how students' attention shifts over consecutive practice tasks, which can provide indications as to why constructing solutions seems to be superior for later test scores compared to using predefined solutions. The intervention in the present study is based on a research framework on imitative and creative reasoning (Lithner, 2008) and uses eye tracking methodology to compare how repeated practice, over a series of similar consecutive mathematics tasks, by constructing solutions or utilizing provided solutions may affect students' solution strategies.

2 Background

Often much of the classroom time in mathematics, especially in the higher grades, is spent solving tasks provided by a textbook or by the teacher (Mullis et al., 2012). To a large extent, these tasks can be solved by applying an already known solution method or by applying or adjusting a given solution template (Boesen et al., 2014; Newton & Newton, 2007; Shield & Dole, 2013; Thompson et al., 2012; Van Zanten & van den Heuvel-Panhuizen, 2018; Stacey & Vincent, 2009). For example, in a study of common secondary textbooks from Australia, Canada, Finland, India, Ireland, Nepal, Scotland, Singapore, South Africa, Sweden, Tanzania, and the USA, Jäder et al. (2020) found that 79% of the tasks could be solved by imitating procedures provided in the book, 13% could be completed by mainly applying given procedures but making some minor modifications, and only 9% of the tasks required the construction of solution methods. The focus on given solution templates (routine task solving) can give students the impression that the core of mathematics is to solve the task by using a predefined formula or by selecting the appropriate (and only) solution from a plethora of formulas that they have tried to memorize (Hiebert & Grouws, 2007). Using predefined learning templates can fast-track the solving process of specific tasks through "a fixed set of step-by-step procedures for solving a (mathematics) problem" (Fan & Bokhove, 2014, p. 486) without (almost) any understanding of the underlying mathematical properties. For example, it is possible to teach 9-year-old children to find the derivative for simple polynomial functions, even though they do not grasp what a function or derivative is. Although the use of routinized templates, commonly denoted as algorithms, is associated with high reliability, speed, and a reduction in working memory load, it is critical that mathematics in school also provides tools for non-routine problem solving and thus for students to become proficient problem solvers (e.g., NCTM, 2011; Niss & Højgaard, 2019; Niss & Jensen, 2002; Skolverket, 2011). In this regard, mathematics contains competencies related to mastering general problem-solving strategies, self regulation, and constructive beliefs, in addition to resources such as routine procedures (Schoenfeld, 1985). Challenging the students to reason and argue mathematically could be one way to emphasize mathematical skills that could be beneficial for problem solving and mathematical understanding (Schoenfeld, 1985; Sidenvall, 2019). Brousseau (1997) argued that the ideal mathematical learning environment involves students taking responsibility for the solution process. However, this a-didactical approach requires tasks or activities where the students initially lack methods to solve the task, tasks that, by extension, give rise to a *productive struggle* where students have to explore and reason (Hiebert & Grouws, 2007). The beneficial effects of productive struggle have also been confirmed in memory research, where the struggle that takes place during encoding, and subsequently facilitates memory retention, is denoted as desirable difficulty (e.g., Bjork & Bjork, 2011).

2.1 Eye tracking in mathematics education

Studies in mathematics education utilizing eye tracking have been increasing rapidly over the last decade (Strohmaier et al., 2020). Eye tracking has been used to study a multitude of topics, such as arithmetic, geometry, reading and word problems, reasoning, and representations (for an overview, see Strohmaier et al. (2020)), and is a quite objective measure compared to interviews, in the sense that it is hard to control eye movements to comply with what a subject thinks that a researcher wants (e.g., Susac et al., 2014). In a study on how students read different mathematical representations, Andrá et al. (2015) have argued that eye tracking can be a useful tool to characterize differences in how students navigate through different mathematical stimuli. Eye tracking has also provided evidence that individual students read, and reread, tasks differently, depending on achievement. For example, De Corte et al. (1990) showed that, independent of ability, students generally read through a complex task just as quickly the first time, while rereading (parts of the task) takes more time for low achievers than for the high achievers. Some eye-tracking studies have also focused on reasoning and strategy choice when solving mathematical tasks. Crisp et al. (2012) used eye tracking to investigate students' strategy choice when deriving a function from a table of values and concluded that neither the choice of solution strategy nor the success rate depended on the student's mathematical background. Obersteiner and Tumpek (2016) and Obersteiner and Staudinger (2018) both investigated strategy choices among university students when comparing or adding fractions and concluded that their participants adapted their solution strategy depending on the given task. In the present study, we investigate if students adapt their strategies when meeting consecutive imitative or creative tasks with the same solution method (i.e., if we can see changes in foci as they become familiar with the solution method), as well as if this adaption will impact which information they utilize in similar tasks with other solution methods.

2.2 Framework of mathematical reasoning

Mathematical reasoning is one of the competencies that is regarded as important if one is to develop proficiency in mathematics and thereby be able to justify and make well-founded mathematical decisions (e.g., Kilpatrick et al., 2001). However, many decisions are not mathematically justified and well founded. These poor decisions are most likely influenced by the many tasks, with predefined solution methods, that students encounter and work with during their schooling (Hiebert, 2003). Although those tasks, to some extent, require reasoning, they lack the requirement of decisions or justifications for which method to use to solve the tasks (Jäder et al., 2020; Stacey & Vincent, 2009).

As presented above, many authors have characterized differences between non-routine problem solving and routine task solving, where the latter is likely to lead to rote learning. A framework by Lithner (2008, 2017) emphasizes the distinction between, on one hand, reasoning in constructing, and on the other hand, imitating, solutions. This framework has, for example, been used in research to (a) analyze student reasoning (e.g., Aaten et al., 2017; Mumu & Tanujaya, 2019; Rofiki et al., 2017; Sukirwan et al., 2018), (b) analyze effects of student reasoning (e.g., Jonsson et al., 2014; Norqvist, 2018), or (c) characterize textbook tasks or teaching (e.g., Bergqvist & Lithner, 2012; Brehmer et al., 2016; Mac an Bhaird et al., 2017).
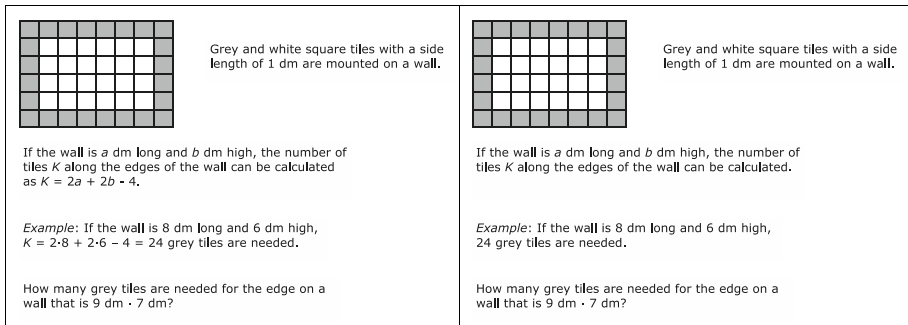
Fig. 1 Example of AR task (left) and CMR task (right)

Within the framework, imitative reasoning occurs when a person uses a known or given solution method to (try to) solve a task. Lithner (2008) identifies two sub-types of imitative reasoning, *memorized* and *algorithmic*. Memorized reasoning concerns recalling a memorized fact (e.g., recalling the answer to 4×8 or the number of sides in polygons). Three versions of algorithmic reasoning are identified: *familiar* (recalling a memorized procedure), *delimiting* (searching for procedures by superficial clues), and *guided* (using a procedure provided by another person or a text). The latter version is focused on in the present study, in the form of written tasks accompanied with solution methods. This type of algorithmic information (e.g., an algorithm for long division, the formula for calculating the volume of a cylinder or worked examples providing templates for solving linear equations) seems to be common in mathematics textbooks (e.g., Jäder et al., 2020; Newton & Newton, 2007; Shield & Dole, 2013). Tasks that are possible to solve by guided algorithmic reasoning will be denoted AR tasks (see Fig. 1).

The second type of reasoning that Lithner (2008) describes is *creative mathematical reasoning* (CMR). This reasoning includes the (re)construction of a new or forgotten solution method that is mathematically founded and justified. CMR does not require extraordinary creativity or genius. Instead, it can be part of an ordinary mathematical solution process as long as the reasoner constructs a, for her or him, new reasoning sequence. Constructing (parts of) a new solution method also provides opportunities for a productive struggle that can be necessary for deeper processing of the mathematics (e.g., Hiebert & Grouws, 2007; Jonsson et al., 2016). A task without included information on how to solve it is by definition not an AR task. If, in addition, the task is not of a common type where a specific solution algorithm is likely to have been memorized earlier by the student solving it, then it is likely to require CMR (Boesen et al., 2010). In this study, a task that does not include a solution method and is not likely (by the researchers' experiences and earlier analyses of Swedish teaching) to be of a common type where a complete solution usually is memorized by students will be denoted CMR task (see Fig. 1). We could not determine what reasoning they actually use without disturbing the data collection of this study, but earlier studies using similar tasks have shown that students are more inclined to use CMR in similar tasks than in AR tasks.

2.2.1 Past results utilizing the framework

Previous studies using Lithner's (2008) framework have shown that students who have to construct their own reasoning sequences and solution methods perform better during

follow-up tests when compared to students who follow given solution methods (Jonsson et al., 2014, 2020; Norqvist, 2018; Norqvist et al., 2019b). There are also indications that brain activity differs between those who engage in CMR compared to AR (Karlsson Wirebring et al., 2015; Wirebring et al., 2021). Results from Hershkowitz et al. (2017) also indicate that students who engage in CMR become knowledge agents in the classroom, and as such, gather followers in the mathematical learning situation.

To solve a CMR task, it is necessary to consider the relevant mathematical properties represented by for example an illustration, which in turn enhances understanding of the task. In contrast, to solve an AR version of the same task, it is not necessary to consider the illustration, since sufficient information for solving the task can be found in a provided formula and/or example. Hence, a formula/example in an AR task can be applied in a mechanical way, potentially leading to rote learning, although it is also possible to consider an illustration to obtain an enhanced understanding of the solution method. However, Norqvist et al. (2019b) showed that students practicing with AR tasks disregarded illustrations and focused mainly on the provided formula/example. In contrast, participants practicing with CMR tasks focused, to a relatively large extent, on the illustrations (Norqvist et al., 2019b).

The purpose of Norqvist et al. (2019b) was to identify main similarities and differences in task solving reasoning. The assumption was that AR and CMR tasks (i.e., that tasks with or without solution templates) will invite different types of reasoning, as indicated by the time spent gazing in areas of different task information (e.g., an illustration will be important for constructing a solution when no template is given). The eye fixations of the three first similar tasks from each of the ten presented task sets (comprised of ten subtasks each) were analyzed using a cluster analysis method where the participants were automatically grouped into subgroups depending on the dwell time in each of the areas of information. The Norqvist et al. (2019b) study extracted eye fixation of a few similar tasks and thus identified group-specific (AR/CMR) sub-clusters. However, this approach was purely data-driven and reduced the original (n) observation into (g) groups, respectively, and thus provided a time-independent average snapshot of how students' eye fixations were clustered. This focus on *momentary* snapshots of student eye movements in Norqvist et al. (2019b) raises the question of short-term *dynamic* changes in eye fixations over consecutive tasks, which Norqvist et al. (2019b) did not address. The present study will therefore investigate the eventual changes in eye fixations that could occur when meeting consecutive tasks, similar to what can happen when working with a mathematics textbook.
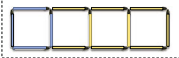
3 Aim and hypotheses

The purpose of the study is to analyze short-term dynamic changes, i.e., how eye fixations change over time, as the students (perhaps) adopt their reasoning and learning strategies to the specific format of CMR and AR tasks. Firstly, it is of interest in itself to understand more about students' learning processes. Secondly, understanding these processes may shed further light on why and under what conditions learning by CMR may be more effective than learning by AR. Thirdly, we know that over long periods (in the magnitude of months and years), a main focus on practicing by non-routine problem solving relative to rote learning affects students' strategies positively (e.g., Boaler, 2002; Hiebert, 2003; Schoenfeld, 1985), but it is not known if there are dynamic effects

over short periods of time. The setting of the study can be seen as a controlled micro version of an ordinary classroom, providing specific information about effects on learning strategies from repeatedly solving similar tasks.

This is studied by observing students' gaze both over consecutive mathematics tasks where only the numerical value differs (within task sets), and between different mathematics tasks where the solution method differs (across task sets). Based on the previous studies and the above reasoning, sets of hypotheses for short-term dynamic changes concerning similar tasks (within task sets) and different tasks (across task sets) are posed in conjunction with "time on task" hypotheses. With support from Norqvist et al. (2019b), we focused the hypotheses on four areas of interest, formula, example, illustration, and question (see Fig. 2), in the present study denoted with subscripts, such as $formula_{AR}$ and $formula_{CMR}$, to differentiate between the two conditions. For within task sets (same subtasks, different numerical values), it is hypothesized that:

1. Within AR task sets, students will gradually focus less on the illustration, $formula_{AR}$, and $example_{AR}$.

    *   Argument: Students will focus on the parts that are most useful for solving the task (i.e., the solution method given by $formula_{AR}$ and $example_{AR}$) and, through the repeated exposure, they will learn the method and subsequently retrieve the method from memory rather than reread it. The gradual decline in the focus of the illustration is caused by the discovery that since an algorithmic solution method is provided, it is not necessary to consider properties of the illustration.

2. Within CMR task sets, students will gradually focus less on the illustration, $formula_{CMR}$, and $example_{CMR}$.

    *   Argument: Students will learn that the illustration is most useful; however, across subtasks, they will be able to retrieve it from memory without attending to the image. The gradual decline in the focus of the $formula_{CMR}$ and $example_{CMR}$ is caused by the discovery that they do not carry any essential information.

    For across-tasks sets (different tasks), it is hypothesized that:

Fig. 2 Example of areas of interest for an AR-task (left) and a CMR-task (right) where the numbers indicate the different areas of interest; illustration (1), description (2), formula (3), example (4), and question (5). Dashed lines and numbers (1–5) were not visible to the participants. Reprinted from "Investigating algorithmic and creative reasoning strategies by eye tracking," by M. Norqvist, B. Jonsson, J. Lithner, T. Qwillbard and L. Holm, 2019, *Journal of Mathematical Behavior, 55*, 100,701, p. 5. Copyright 2019 by Elsevier

3. Across AR task sets, students will gradually focus less on the illustration but have a continued focus on the formula, example, and question.

    - Argument: Since all AR-tasks have similar format, the students will learn which information will be most useful to solve the task type (even though the new task set includes new information and illustrations), and for each new task set, they will use the corresponding new solution method in formula$_{AR}$ and example$_{AR}$.

4. Across CMR task sets, students will gradually focus less on the formula$_{CMR}$ and example$_{CMR}$ but have a continued focus on the illustration and question.

    - Argument: Since all CMR tasks have similar format, the students will learn that the formula$_{CMR}$ and example$_{CMR}$ contain less useful information in this task type. They will, however, need to consider mathematical properties of the illustration to construct a new solution method for each new task set.

    For the time-on-task analyses, it is hypothesized that:

5. Although students in both groups will consider all information in the first tasks of the first task sets, CMR tasks will take a longer time than AR tasks at the beginning of each task set.

    - Argument: The task format is new to the students, and they need to start by identifying which parts to use and how. It is generally quicker to use a provided method than to construct one.

6. Students who practice with CMR tasks decrease their time on task within a task- set, while students who practice with AR tasks do not display the same decrease.

    - Argument: Although the AR solution method may be learnt through repeated exposure, the solutions constructed by the struggle associated with CMR task solving will be more effectively consolidated in memory and, after a few subtasks within a task set, quickly recalled instead of constructed again.

4 Method

To study the hypotheses, an experiment was designed in which students practiced with either AR or CMR tasks. Students' focus on different types of task information, related to their task-solving strategies, were recorded using eye tracking.

4.1 Participants

Fifty participants, comprising students from an upper secondary school and university students from a clinical psychology program in Sweden, with a mean age of 23.0 years ($SD = 3.2$), were recruited to participate in the study. The participants can be considered to be mid to high achievers. The study was approved by the Regional Ethical Review Board and written informed consent was obtained from all participants in accordance with the Declaration of Helsinki. Students who did not participate in all sessions were removed from the data, which left 48 participants (23 AR and 25 CMR) for the analysis.

4.2 Experimental design

In a between-group design, the students participated in three sessions: (1) background information gathering and cognitive testing, (2) mathematical practice, and (3) mathematical test. Eye-tracking equipment was used during session 2.

4.2.1 First session

During the first session, the students provided some background information (e.g., gender and age) and completed two cognitive tests: working memory and fluid intelligence. Cognitive ability and its relation to learning mathematics is well established. In a large prospective study of 70,000 students, it was found that general intelligence could explain up to 58.6% of the variation in performance on national tests at the age of 16 years (Deary et al., 2007). Fluid intelligence is part of general intelligence and is recognized as a causal factor when experiencing non-familiar situations in general (Valentin Kvist & Gustafsson, 2008; Watkins et al., 2007) and solving mathematical tasks in particular (Floyd et al., 2003; Taub et al., 2008). Ravens Advanced Progressive Matrices (APM) (Raven & Raven, 1991), which is the most common test for fluid intelligence, consists of 48 regular test items and 12 practice items. That task was self-paced but with a maximum of 25 min available (see Norqvist et al. (2019b) for more information).

Moreover, it is well known that working memory influences math performance (De Smedt et al., 2009; Passolunghi et al., 2008; Raghubar et al., 2010). For the measure of complex working memory, the operation span task was used (Unsworth et al., 2005). In the operation span task, the students performed mathematical operations; after each operation, a letter was displayed, and they were instructed to remember each presented letter. After a sequence, which varied between 2 and 7 operations and letters, they were asked to recall the letters in the order they were presented. Measures of internal consistency (Cronbach's alpha) were extracted from a larger pool of data and found to be 0.84 for Ravens APM and 0.83 for operation span. Table 1 displays the mean values and standard deviation for the CMR and AR groups respectively.

Although the AR group was numerically stronger regarding both measures, an independent t-test revealed no significant differences between groups for the operation span tasks and Ravens matrices, $t(48)=1.34$, $p=0.18$ and $t(48)=1.07$, $p=0.29$, respectively. Note that the high standard deviation for operation span task in the CMR group is mainly driven by the low score for one participant in the CMR group. The correlation between operation span and Ravens matrices was found to be 0.42, $p=0.003$, and was therefore collapsed into one measure of cognitive proficiency index (CPI). In order to control for individual differences in task-specific abilities and general cognitive abilities as well as gender, the participants were matched according to their CPI index and gender and assigned to two independent groups, AR and CMR.

Table 1 Mean values and standard deviations of cognitive test scores for the two conditions

	Ravens matrices		Operation span	
	AR	CMR	AR	CMR
Mean	11.88	11.00	58.68	54,92
Standard deviation	2.90	2.90	6.53	12.33

AR algorithmic reasoning, *CMR* creative mathematical reasoning

4.2.2 Second session

In the second session, 1 week later, the two groups practiced, using either tasks with solution templates (AR tasks) or tasks without solution templates (CMR tasks) (see Fig. 2).[1] All tasks were presented on a computer screen and students answered by typing in their answer via a keyboard. The tasks were divided into areas of interest (see Fig. 2). The illustration, description, and question were identical in the AR and CMR tasks. However, the formula and example differed between the two conditions, containing solution methods for AR but only non-essential information for CMR. The reason for including the latter was to make the two tasks similar in layout, and to provide comparable areas of interest for eye-tracking analyses. The participants were provided with 11 different task sets. Each task set comprised 10 subtasks that were identical, apart from different numerical values in the question, and the allotted time for each set was 5 min. For example, the subtasks with matchstick squares (Fig. 2) asked for 6 squares, 100 squares, 40, 9, 12, 8, 30, 13, 20, and 11 squares. If a student solved all 10 subtasks within the allotted time, more subtasks were resampled from the previous subtasks to ensure a practice time of 5 min per task set for both conditions. Task set 11 proved to be too difficult in both practice conditions and was therefore removed from the analysis, which left the first 10 task sets available for analysis.

During the practice session, participants' eye movements were recorded by a desk-mounted Eyelink 1000, sampling at 500 Hz. To impose constant viewing distance, a chinrest was used and placed about 75 cm from the screen. Eye movements were recorded monocularly, typically using the right eye, and in-plane gaze spatial resolution of about one degree, which is sufficient as the areas of interest were separated by three to four degrees. Before each session the equipment was calibrated to ensure correct measurements. Fixations were defined as non-blink inter-saccadic intervals based on the manufacturers default settings. The saccade definitions consisted of an amplitude change of $0.15°$, velocity $> 30°/ms$ and an acceleration threshold of $8000°/ms^2$. Only fixations with durations longer than 50 ms were considered for further analysis (for more details, see Norqvist et al., 2019b). Fixations in the five separate areas of interest (i.e., illustration, description, formula, example, and question) were separated. Time spent gazing in the specified areas of interest (dwell time) and time spent trying to solve a task (time on task) were calculated automatically by the eye-tracking software.

4.2.3 Third session

In the third session, 1 week after the practice session, the students solved tasks that could be solved with the same solution methods as during practice. The performance differences following the learning session showed that practice scores were in favor of the AR-group, $t(32) = 5.56$, $p < 0.001$, $d = 1.58$, (86% vs. 57% correct answers), while test scores were in favor of the CMR-group, $t(46) = -2.44$, $p = 0.019$, $d = 0.71$, (28% vs. 45% correct answers) as previously reported in Norqvist et al. (2019b). However, in this study, the test scores were not analyzed and will therefore not be discussed further. For more information on test scores and their implications, see Norqvist et al. (2019b).

---

[1] All task sets can be found in Norqvist et al. (2019a).

4.3 Analyses

Since the CMR participants, on average, solved fewer tasks due to differences in solution speed and time restraints, we initially investigated whether the group equivalence in cognitive ability was maintained across subtasks. Non-parametric Mann–Whitney U tests of those participants who completed subtask 10 for each of the 10 task sets revealed no significant effects ($U = 65.0$–197.0; $p = 0.887$–0.978). Hence, the two groups still were comparable at the end of each task set with respect to cognitive ability. The results of hypotheses 1–4 are presented as the average proportional time spent gazing at a specific area of interest with a 95% confidence interval. The advantage of using confidence intervals is that the displayed effects are likely to exist in the population (Ranstam, 2012), and are to be interpreted as covering the true value in 95 out of 100 studies. In the present context, the confidence intervals provide a conservative estimate of group differences and reduce the risk for a type I error, hence deciding that there is a difference between AR and CMR groups when it is not (du Prel et al., 2009). To address hypotheses 5–6, the mean difference in *time on task* between the two conditions was calculated and presented for all 100 subtasks. The calculated time differences for each task set are based on the participants who solved all subtasks in the specific task set. The results are presented as descriptive graphs (hypotheses 1–4) and a table (hypotheses 5–6) and reflect the underlying distributions regarding eye fixations within and across task sets and group differences in speed of task solving.

5 Results

The results from the present study focusing on the learning conditions will be presented with three different foci connected to the six hypotheses. The first set of analyses for the two learning conditions encompass hypotheses 1 and 2 and include students' eye-tracking behavior *within task sets* with subtasks that only differ numerically. The second set of analyses for the two learning conditions encompass hypotheses 3 and 4 and include students' eye-tracking behavior *across task sets* with main focus on the first subtask in each task set. The third set of analyses encompass hypotheses 5 and 6 and include how the two learning conditions differ in *time on task* within and across each task set.

5.1 Behavior within task sets (same subtasks, different numerical values)

5.1.1 The AR group

The dwell-time pattern is similar for all task sets: a proportionately short time on the illustration during the first subtask, which in the latter subtasks decreased to almost zero (Fig. 3a). Dwell time on the formula$_{AR}$ and example$_{AR}$ was comparatively high (about 30–40%) and fairly constant throughout the subtasks (Fig. 3 b and c), while dwell time on the question rose slightly throughout the task set (Fig. 3d). The gradual decline regarding the illustration confirms the first part of hypothesis 1, but there was no apparent decrease in focus on the formula and example within the sets, which disconfirms the second part of hypothesis 1.
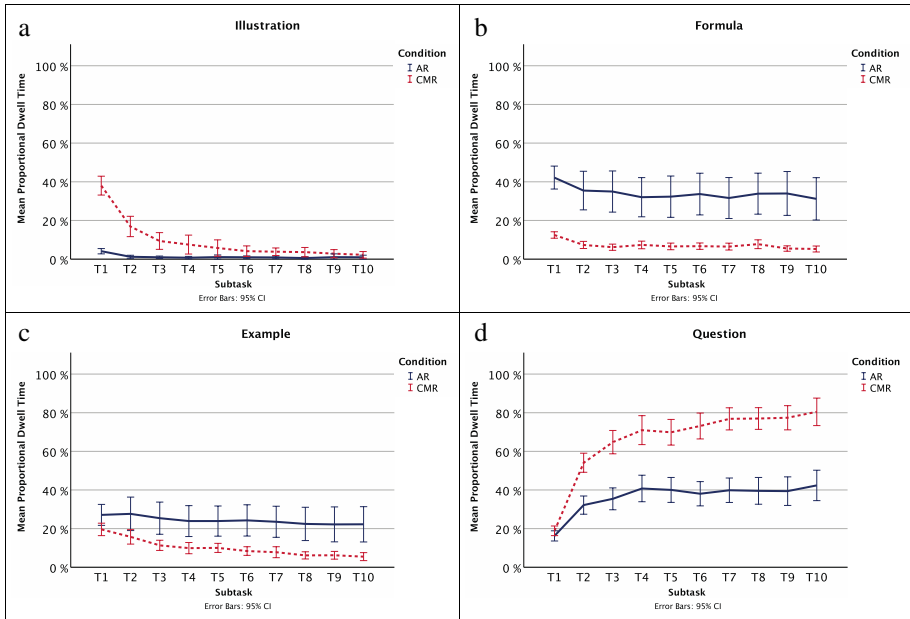
Fig. 3 **a–d** Mean proportional dwell time within task sets 1–10, on illustration, formula, example, and question. Solid line — AR; dashed line — CMR
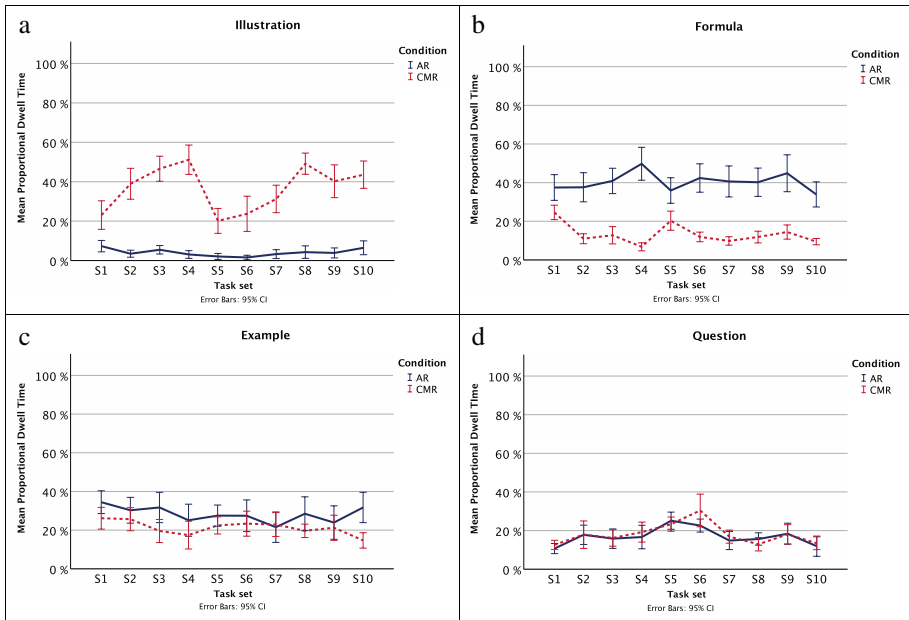
5.1.2 The CMR group

The CMR group spent proportionally more dwell time on the illustration in the first four subtasks within each task set (Fig. 3a). Simultaneously, dwell time on the formula$_{CMR}$ decreased from subtask 1 and was consistently below the 10% level (Fig. 3b). The dwell time on the example$_{CMR}$ gradually decreased (Fig. 3c). Hypothesis 2 is confirmed.

Both groups showed an increase in proportional dwell time on the question, as the dwell time in other areas decreased, but this increase was more pronounced in the CMR group (Fig. 3d), especially from subtask 1 to subtask 2.

5.2 Behavior across task sets (different tasks)

5.2.1 The AR group

The analysis did not reveal decreased dwell time on the illustration, but rather a consistently low level, disconfirming the first part of hypothesis 3. This lack of change across task might be due to floor effects. Hence, the proportion of dwell time focusing on the illustration was low already at task set 1. In fact, the AR group devoted less than 10% of the dwell time across all task sets to the illustration (Fig. 4a). While the AR group spent a low proportion of their time on the illustration, they spent a comparably high proportion of their solution time on the formulas$_{AR}$ and examples$_{AR}$ in all task sets (Fig. 4 b and c), confirming the second part of hypothesis 3.

Fig. 4 **a–d** Proportional dwell time across subtask 1 of all task sets on illustration, formula, example, and question. Solid line — AR, dashed line — CMR

5.2.2 The CMR group

The participants maintained a fairly low but consistent focus on the formulas$_{CMR}$ and examples$_{CMR}$. Hence, regarding change across task sets, the first part of hypothesis 4 was disconfirmed. A consistently high proportion of dwell time on the illustration and (on a lower level) on the question across all task sets (Fig. 4 a and d) confirms the second part of hypothesis 4. Notably, there was a large variation in dwell time on the illustration in the CMR group, depending on task set, probably due to the varying task difficulty and complexity of the illustrations.

5.3 Speed of task solving

An analysis of time on task within each task set indicates that practicing by CMR can be more efficient in the long run, not only with respect to post-test performance, as found earlier (Jonsson et al., 2014), but also concerning practice time. Since practicing by AR includes using a given solution method, it is as expected initially faster than practicing by CMR where solutions need to be constructed (see Table 2, subtasks T1–T3), confirming hypothesis 5. Although solving CMR tasks on average was slower at the beginning of each task set (except task set 6), CMR was slightly faster than AR at the end of the task sets (Table 2, subtasks T5–T10), which confirms hypothesis 6.

Table 2 Mean absolute time difference in ms (CMR–AR) includes only participants who solved subtask 10 (hence, there are different individuals in each set)

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Set 1	4116	16010	2611	2702	2574	2513	2378	1874	1169	504
Set 2	15609	11092	614	512	3217	842	1362	494	1270	70
Set 3	13230	2131	2552	142	1604	1588	684	150	439	634
Set 4	25054	5279	544	774	1389	996	523	499	3351	2674
Set 5	8985	5084	2616	732	99	2435	2111	955	1603	3428
Set 6	6244	1328	3700	4204	4513	5559	2865	5164	4275	3867
Set 7	21898	1642	1008	1485	1309	45	1842	7012	1792	146
Set 8	27889	1188	1854	264	486	2861	536	580	373	195
Set 9	14166	7006	1570	1293	1087	2929	2137	76	782	388
Set 10	25199	3850	6988	3123	2915	744	2884	1288	916	2997

Blue/white text — AR faster than CMR; red/black text — CMR faster than AR

Task set 6 appears as an outlier in that the CMR-group is faster from the start of the task set. This is potentially a contextual effect since there were similarities in the illustrations in task sets 5 and 6.

In summary, for all six hypotheses:

1. Within AR task sets, students gradually focused less on the illustration but maintained a high focus on formula$_{AR}$ and example$_{AR}$.
2. Within CMR task sets, students gradually focused less on the illustration, formula$_{CMR}$, and example$_{CMR}$.
3. Across AR task sets, students maintained a low focus on the illustration and a high focus on formula$_{AR}$ and example$_{AR}$.
4. Across CMR task sets, students maintained a low focus on the formula$_{CMR}$ and example$_{CMR}$ and a high focus on the illustration and question.
5. Practicing by AR was initially faster than practicing by CMR.
6. CMR was slightly faster than AR at the end of the task sets.

6 Discussion

During the years in school, in mathematics classes and as homework, students solve a lot of tasks, many of which are designed to practice a given solution method or algorithm. In a sense, this study can be considered a micro version of ordinary classroom teaching, with a focus on how students use available information in consecutive mathematics tasks. The present study indicates that students who solve tasks with given solution methods do not, at least not to a high degree, internalize useful information. Moreover, the experimental setting in the present study provides a link between educational design and learning processes (i.e., the relation between task design and how students' focus is affected by repeated practice) that mathematics education largely lacks (Niss, 2007).

6.1 Within task sets

In contrast to the second part of the first hypothesis, the AR group maintained a high focus on the formula$_{AR}$ throughout each task set, which would indicate that the given solution method was not internalized and subsequently not recalled from memory, even though the formula$_{AR}$ was the same as in the previous subtask. If textbooks mainly provide tasks that promote AR (Jäder et al., 2020; Shield & Dole, 2013), this could be unproductive for students who use textbooks as the main source of practice tasks.

At the same time, it is evident from our results that, as hypothesized, students who practice with CMR tasks will shift focus from the illustration to the updated information (i.e., the question) after a solution method has been constructed. That the problem-solving process transforms to a more routinized solution method is perhaps exactly what we would hope for as learning occurs (Hiebert, 2003). Additionally, Brousseau (1997) predicted that activities that involve struggling with the mathematical ideas should lead to a deeper knowledge than just learning by rote, something that other researchers also claim (Bjork & Bjork, 2011; Hiebert & Grouws, 2007; Kapur, 2010). The fact that the CMR students also outperform the AR students in post-tests (e.g., Jonsson et al., 2014; Norqvist et al., 2019b) supports Brousseau's theory. The CMR group also decreased their attention to the formula$_{CMR}$ and example$_{CMR}$, as hypothesized, as they realized that these areas contain no essential information.

6.2 Across task sets

Our results also provide indications that students who are given solution methods disregard other meaning-building information even though the tasks change, partly confirming the third hypothesis. Looking at the first subtask across all task sets, we can see that the AR group (i) maintained a comparably high proportion of their dwell time focused on the formula$_{AR}$ and example$_{AR}$ and (ii) consistently allocated only a small proportion of their dwell time to the illustration. The finding that students disregarded the illustration almost instantly was not expected, especially since no information about the first tasks was known in advance. Hence, with no initial focus on the illustration, no gradual shift in attention from the illustration could occur, partly disconfirming hypothesis 3. As a teacher, one would probably hope for the students to reflect upon how given solution methods work, but the lack of focus on the illustration is an indication that spontaneous reflection is unlikely. Rather, students who meet AR tasks seem to fall into a habit of utilizing given solution methods without asking themselves why they work. This could be one explanation for why working with procedural tasks solely is not beneficial for conceptual understanding (Brousseau, 1997).

When students worked with CMR tasks, they initially displayed a diverse gaze pattern when new task types were presented, partially confirming the fourth hypothesis. However, every time they encountered a new task set, they continued to focus on the areas of interest containing non-essential information. Hence, when engaging in CMR practice, they initially in every task set, seemed to evaluate where the important information was located. Brousseau (1997) argued that one important aspect of learning from solving tasks is that the student takes full responsibility for the solution process and having to analyze the task is part of this responsibility. Since an analysis of the task, which takes the mathematical properties into consideration, is not necessary for a correct solution when a solution method is given, students who practice by AR can disregard this analytical process.

6.3 Time on task

Although the AR group who are presented with solution methods solve tasks faster initially compared to the CMR group, our results indicate that the initial struggle for students without given solution methods has a later payoff as shown by faster task solving for participants in the CMR group, confirming hypotheses 5 and 6. It seems that the mathematical struggle associated with CMR tasks (Jonsson et al, 2016) leads to short-term dynamic changes as reflected by more effective problem solving within and across task sets and as shown in Norqvist et al. (2019b), better performance 1 week later. These findings are in line with both the retrieval effort hypothesis showing that more effortful encoding leading to more effective retrieval (Bjork & Bjork, 2011; Pyc & Rawson, 2009) and Brousseau's (1997) theory, and Hiebert and Grouws's (2007) notion of productive struggle. There are, of course, ways of designing AR tasks that would make them initially slower than the corresponding CMR task (for example, by making the solution method more complex than needed), but this is not usually the way solution methods are presented in school.

6.4 Method discussion

Although fewer participants in the CMR group solved all 10 subtasks in each task set in the allotted time, these students did not differ with respect to cognitive ability. However, we are aware that the reduced number of participants could potentially affect the external validity if only high achieving students in the CMR group solved all tasks. However, the CMR students remaining at the end of each task set were not the same for each task set and cognitive abilities in this reduced group did not differ from the large CMR group. With that said, the sample is most likely relatively strong in mathematics. Unfortunately, it was not possible in the present study to extract school marks, but in a previous study, the correlation between upper secondary students' cognitive ability and overall school marks was 0.49, $p < 0.001$ (Jonsson et al., 2021). In another study using AR and CMR tasks, we included participants from both basic and advanced math tracks at upper-secondary school in a series of three separate experiments (altogether 273 participants) covering both within-subject and between-subject designs. However, when assessing performance one week later, no main effects of math tracks were obtained in neither of the experiments (Jonsson et al., 2020).

Moreover, solving tasks with the head placed in a chinrest can, of course, be discussed in terms of ecological validity, since a restrained head could potentially impact the ability to concentrate. On the other hand, the strict experimental design provides a high degree of internal validity, hence establishing causality between task type (AR, CMR) and eye-tracking behavior.

It should be noted that eye-tracking studies often are based on the eye–mind hypothesis (i.e., what the eye has in focus is also what is being processed by the mind; see Just & Carpenter, 1980). This hypothesis might not, however, be completely true in mathematics education. The eye-tracking equipment measures foveal visual tracking, while it cannot capture participants' ability to detect and process information in the periphery. For example, Schindler and Lilienthal (2019) complemented eye tracking with a stimulated recall interview and showed that the observed student did refresh their memory of geometric information while just glancing in the correct direction of the given illustration. It seems plausible that a similar phenomenon could be present here. For instance, glancing and maintaining the image "in mind" simultaneously as they process information in other areas of interest by

using the peripheral vision seems likely to achieve. Arguably, tasks containing no formula and example (CMR tasks) would benefit more from glancing at the illustration, potentially underestimating the obtained difference between AR and CMR participants regarding the illustration (see Figs. 3a and 4a).

Although there are some limitations to the eye-tracking technique, it can provide insight into student behavior and mathematical reasoning, difficult to obtain by other methods. The present study provides detailed information on what students focused on, regarding AR and CMR tasks, which in turn gives insights into students' reasoning and how it develops.

6.5 Implications and conclusions

Applying the results from this study to classroom teaching is, of course, not straightforward. In a classroom situation, there are indefinitely more variables to consider, but there is one thing that probably is true for both our microversion of teaching, and the macroversion that takes place in the classroom. When students solve tasks with given solution methods, many of them may make a habit of using the formulas without consideration to why they work, while students who are not given solution methods must regard all given information from the start of a new task. The results indicate that students who have solution methods available tend to read these methods repeatedly instead of recalling them from memory, something that students practicing with CMR tasks seem to do. Repeated retrieval from memory is in itself a powerful tool to use when learning (e.g., Karpicke & Roediger, 2007; Wiklund-Hornqvist et al., 2014), while reading available algorithms seems to elicit no such retrieval. Norqvist (2018) showed that provided explanations together with given solution methods do not yield better test scores than the types of AR tasks in this experiment, and the habit of quickly finding and applying solution methods, indicated by the AR group's eye movements in this study, might be why. If students do not reflect on why formulas, solved examples, or algorithms work, it is harder to remember and transfer them when needed in a new situation or task, as shown by Jonsson et al. (2020). The short-term effects for AR in this study may have a long-term influence if students' start thinking that doing mathematics only involves imitating a predefined formula without having to know why it works.

Practice tasks without complete solution methods, which encourage a suitable struggle with mathematical properties during task solving, could be just what is needed for students to not fall into the habit of un-reflected use of solution methods. Since, for many students, the aim of mathematics seems to be to get correct answers to tasks rather than understanding the solution idea, the reflection does not come naturally. Thus, as seen by the analysis of eye movements in this study, tasks that promote creative reasoning could help direct attention to the important mathematical properties and structure of the task that we want students to learn and remember. The objective measure of students' attention that eye tracking can provide has shown to be valuable in this study. Without eye tracking we would not have gotten this detailed view of which information students utilize or insight into the lack of internalization for the AR condition. Useful data on students' task solving strategies and how they change over time can be obtained by other methods, for example written tests or think-aloud interviews. However, without eye tracking, it would hardly have been possible to quantify students' reasoning foci and carry out the statistical analyses above to test this study's hypotheses. More generally, one of the many strengths of eye tracking is the possibility to quantify human attention. Hence, eye tracking, can play an important part in future studies, by complementing more traditional research methods with additional insight into students' attention and thinking processes.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Aaten, A. B., Deprez, J., Roorda, G., & Goedhart, M. (2017). *Undergraduates' reasoning while solving integration tasks: Discussion of a research framework*. Paper presented at the CERME 10.

Andrá, C., Lindström, P., Arzarello, F., Holmqvist, K., Robutti, O., & Sabena, C. (2015). Reading mathematics representations: An eye-tracking study. *International Journal of Science and Mathematics Education, 13*(2), 237–259. https://doi.org/10.1007/s10763-013-9484-y

Bergqvist, T., & Lithner, J. (2012). Mathematical reasoning in teachers' presentations. *Journal of Mathematical Behavior, 31*(2), 252–269. https://doi.org/10.1016/j.jmathb.2011.12.002

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Boaler, J. (2002). *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning (Rev. and expanded ed.)*. L. Erlbaum.

Boesen, J., Lithner, J., & Palm, T. (2010). The relation between types of assessment tasks and the mathematical reasoning students use. *Educational Studies in Mathematics, 75*, 89–105. https://doi.org/10.1007/s10649-010-9242-9

Boesen, J., Helenius, O., Lithner, J., Bergqvist, E., Bergqvist, T., Palm, T., & Palmberg, B. (2014). Developing mathematical competence: From the intended to the enacted curriculum. *Journal of Mathematical Behavior, 33*, 72–87. https://doi.org/10.1016/j.jmathb.2013.10.001

Brehmer, D., Ryve, A., & Van Steenbrugge, H. (2016). Problem solving in Swedish mathematics textbooks for upper secondary school. *Scandinavian Journal of Educational Research, 60*(6), 577–593. https://doi.org/10.1080/00313831.2015.1066427

Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Kluwer Academic Publishers.

Crisp, R., Inglis, M., Mason, J., & Watson, A. (2012). Individual differences in generalisation strategies. *Research in Mathematics Education, 14*(3), 291–292. https://doi.org/10.1080/14794802.2012.734981

De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. *Journal of Educational Psychology, 82*(2), 359. https://doi.org/10.1037/0022-0663.82.2.359

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology, 103*(2), 186–201. https://doi.org/10.1016/j.jecp.2009.01.004

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13–21. https://doi.org/10.1016/j.intell.2006.02.001

du Prel, J.-B., Hommel, G., Röhrig, B., & Blettner, M. (2009). Confidence interval or p-value? Part 4 of a series on evaluation of scientific publications. *Deutsches Arzteblatt International, 106*(19), 335–339. https://doi.org/10.3238/arztebl.2009.0335

Fan, L., & Bokhove, C. (2014). Rethinking the role of algorithms in school mathematics: A conceptual model with focus on cognitive development. *ZDM-Mathematics Education, 46*(3), 481–492. https://doi.org/10.1007/s11858-014-0590-2

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools, 40*(2), 155–171. https://doi.org/10.1002/pits.10083

Hershkowitz, R., Tabach, M., & Dreyfus, T. (2017). Creative reasoning and shifts of knowledge in the mathematics classroom. *ZDM-Mathematics Education, 49*(1), 25–36. https://doi.org/10.1007/s11858-016-0816-6

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 371–404). Information Age.

Hiebert, J. (2003). What research says about the NCTM standards. In J. Kilpatrick, W. G. Martin, D. Schifter, & National Council of Teachers of Mathematics. (Eds.), *A research companion to principles and standards for school mathematics* (pp. 5–23). National Council of Teachers of Mathematics.

Jäder, J., Lithner, J., & Sidenvall, J. (2020). Mathematical problem solving in textbooks from twelve countries. *International Journal of Mathematical Education in Science and Technology, 51*(7), 1120–1136. https://doi.org/10.1080/0020739x.2019.1656826

Jonsson, B., Norqvist, M., Liljekvist, Y., & Lithner, J. (2014). Learning mathematics through algorithmic and creative reasoning. *Journal of Mathematical Behavior, 36*, 20–32. https://doi.org/10.1016/j.jmathb.2014.08.003

Jonsson, B., Kulaksiz, Y. C., & Lithner, J. (2016). Creative and algorithmic mathematical reasoning: Effects of transfer-appropriate processing and effortful struggle. *International Journal of Mathematical Education in Science and Technology, 47*(8), 1206–1225. https://doi.org/10.1080/0020739x.2016.1192232

Jonsson, B., Wiklund-Hörnqvist, C., Stenlund, T., Andersson, M., & Nyberg, L. (2021). A learning method for all: The testing effect is independent of cognitive ability. *Journal of Educational Psychology, 113*(5), 972–985. https://doi.org/10.1037/edu0000627

Jonsson, B., Granberg, C., & Lithner, J. (2020). Gaining mathematical understanding: The effects of creative mathematical reasoning and cognitive proficiency. *Frontiers in Psychology, 11*(3666). https://doi.org/10.3389/fpsyg.2020.574366

Just, M. A., & Carpenter, P. A. (1980). A theory of reading — From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354. https://doi.org/10.1037/0033-295x.87.4.329

Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science: An International Journal of the Learning Sciences, 38*(6), 523–550. https://doi.org/10.1007/s11251-009-9093-x

Karlsson Wirebring, L., Lithner, J., Jonsson, B., Liljekvist, Y., Norqvist, M., & Nyberg, L. (2015). Learning mathematics without a suggested solution method: Durable effects on performance and brain activity. *Trends in Neuroscience and Education, 4*(1–2), 6–14. https://doi.org/10.1010/j.tine.2015.03.002

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57*(2), 151–162. https://doi.org/10.1016/j.jml.2006.09.004

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. National Academy Press.

Lithner, J. (2017). Principles for designing mathematical tasks that enhance imitative and creative reasoning. *ZDM-Mathematics Education*, 1–13. https://doi.org/10.1007/s11858-017-0867-3

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics, 67*(3), 255–276. https://doi.org/10.1007/s10649-007-9104-2

MacanBhaird, C., Nolan, B. C., O'Shea, A., & Pfeiffer, K. (2017). A study of creative reasoning opportunities in assessments in undergraduate calculus courses. *Research in Mathematics Education, 19*(2), 147–162. https://doi.org/10.1080/14794802.2017.1318084

Mackintosh, N., & Mackintosh, N. J. (2011). *IQ and human intelligence*: Oxford University Press.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics* (ED544554). TIMSS & PIRLS International Study Center.

Mumu, J., & Tanujaya, B. (2019). Measure reasoning skill of mathematics students. *International Journal of Higher Education, 8*(6), 85–91. https://doi.org/10.5430/ijhe.v8n6p85

NCTM. (2011). *Common core state standards for mathematics*. National Council of Teachers of Mathematics. Commission on Standards for School Mathematics.

Newton, D., & Newton, L. (2007). Could elementary mathematics textbooks help give attention to reasons in the classroom? *Educational Studies in Mathematics, 64*(1), 69–84. https://doi.org/10.1007/s10649-005-9015-z

Niss, M., & Jensen, T. H. (2002). *Kompetencer og matematiklæring: Idéer og inspiration til udvikling af matematikundervisning i Danmark* (Vol. 18): Undervisningsministeriet.

Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics, 102*(1), 9–28. https://doi.org/10.1007/s10649-019-09903-9

Niss, M. (2007). Reactions on the state and trends in research on mathematics teaching and learning. From here to Utopia. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (Vol. 2, pp. 1293–1312). Information Age.

Norqvist, M. (2018). The effect of explanations on mathematical reasoning tasks. *International Journal of Mathematical Education in Science and Technology, 49*(1), 15–30. https://doi.org/10.1080/0020739X.2017.1340679

Norqvist, M., Jonsson, B., & Lithner, J. (2019a). Eye-tracking data and mathematical tasks with focus on mathematical reasoning. *Data in Brief, 25*, 104216. https://doi.org/10.1016/j.dib.2019.104216

Norqvist, M., Jonsson, B., Lithner, J., Qwillbard, T., & Holm, L. (2019b). Investigating algorithmic and creative reasoning strategies by eye tracking. *Journal of Mathematical Behavior, 55*, 100701. https://doi.org/10.1016/j.jmathb.2019.03.008

Obersteiner, A., & Staudinger, I. (2018). How the eyes add fractions: Adult eye movement patterns during fraction addition problems. *Journal of Numerical Cognition, 4*(2), 317–336. https://doi.org/10.5964/jnc.v4i2.130

Obersteiner, A., & Tumpek, C. (2016). Measuring fraction comparison strategies with eye-tracking. *ZDM-Mathematics Education, 48*(3), 255–266. https://doi.org/10.1007/s11858-015-0742-z

Passolunghi, M. C., Mammarella, I. C., & Altoè, G. (2008). Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades. *Developmental Neuropsychology, 33*(3), 229–250. https://doi.org/10.1080/87565640801982320

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*(4), 437–447. https://doi.org/10.1016/j.jml.2009.01.004

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*(2), 110–122. https://doi.org/10.1016/j.lindif.2009.10.005

Ranstam, J. (2012). Why the P-value culture is bad and confidence intervals a better alternative. *Osteoarthritis and Cartilage, 20*(8), 805–808. https://doi.org/10.1016/j.joca.2012.04.001

Raven, J., & Raven, J. (1991). [Court, J. H. (1991)]. *Manual for Raven's progressive matrices and vocabulary scales: Section, 1*. Oxford Psychologists Press.

Rofiki, I., Nusantara, T., & Chandra, T. (2017). Exploring local plausible reasoning: The case of inequality tasks. *Paper Presented at the Journal of Physics: Conference Series, 943*, 012002.

Schindler, M., & Lilienthal, A. J. (2019). Domain-specific interpretation of eye tracking data: Towards a refined use of the eye-mind hypothesis for the field of geometry. *Educational Studies in Mathematics, 101*(1), 123–139. https://doi.org/10.1007/s10649-019-9878-z

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.

Shield, M., & Dole, S. (2013). Assessing the potential of mathematics textbooks to promote deep learning. *Educational Studies in Mathematics, 82*(2), 183–199. https://doi.org/10.1007/s10649-012-9415-9

Sidenvall, J. (2019). Literature review of mathematics teaching design for problem solving and reasoning. *Nordisk matematikkdidaktikk, NOMAD:[Nordic Studies in Mathematics Education], 24*(1), 51–74.

Skolverket. (2011). *Läroplan, examensmål och gymnasiegemensamma ämnen för gymnasieskola 2011*. Skolverket: Fritze distributor.

Stacey, K., & Vincent, J. (2009). Modes of reasoning in explanations in Australian eighth-grade mathematics textbooks. *Educational Studies in Mathematics, 72*(3), 271. https://doi.org/10.1007/s10649-009-9193-1

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics, 104*(2), 147–200. https://doi.org/10.1007/s10649-020-09948-1

Sukirwan, D., & Herman, T. (2018). Analysis of students' mathematical reasoning. *Paper presented at the Journal of Physics: Conference Series, 948*, 012036.

Susac, A., Bubic, A., Kaponja, J., Planinic, M., & Palmovic, M. (2014). Eye movements reveal students' strategies in simple equation solving. *International Journal of Science and Mathematics Education, 12*(3), 555–577. https://doi.org/10.1007/s10763-014-9514-4

Taub, G. E., Floyd, R. G., Keith, T. Z., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly, 23*(2), 187–198. https://doi.org/10.1037/1045-3830.23.2.187

Thompson, D., Senk, S., & Johnson, G. (2012). Opportunities to learn reasoning and proof in high school mathematics textbooks. *Journal for Research in Mathematics Education, 43*, 253–295. https://doi.org/10.5951/jresematheduc.43.3.0253

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505. https://doi.org/10.3758/bf03192720

Valentin Kvist, A., & Gustafsson, J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence, 36*(5), 422–436. https://doi.org/10.1016/j.intell.2007.08.004

Van Zanten, M., & van den Heuvel-Panhuizen, M. (2018). Opportunity to learn problem solving in Dutch primary school mathematics textbooks. *ZDM-Mathematics Education, 50*(5), 827–838. https://doi.org/10.1007/s11858-018-0973-x

Watkins, M. W., Lei, P. W., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence, 35*(1), 59–68. https://doi.org/10.1016/j.intell.2006.04.005

Wiklund-Hornqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology, 55*(1), 10–16. https://doi.org/10.1111/sjop.12093

Wirebring, L. K., Wiklund-Hornqvist, C., Stillesjo, S., Granberg, C., Lithner, J., Andersson, M., Jonsson, B. (2021). Effects on brain activity after creative mathematical reasoning when considering individual differences in cognitive ability. *biorxiv,* 2021.03.30.437492. https://doi.org/10.1101/2021.03.30.437492