



Modeling teachers' reactions to unexpectedness

Zoltán Kondé^{a,*}, Zoltán Kovács^{b,c}, Eszter Kónya^{c,d}

^a Institute of Psychology, Faculty of Art, University of Debrecen, Debrecen, Hungary

^b Institute of Mathematics, Faculty of Computer Science, Eszterházy Károly Catholic University, Eger, Hungary

^c MTA-Rényi-ELTE Research Group in Mathematics Education, 3300, Eger, Eszterházy tér 1, Hungary

^d Institute of Mathematics, Faculty of Science, University of Debrecen, Debrecen, Hungary

ARTICLE INFO

Keywords:

Mathematical unexpectedness
Cognitive load
Working memory
Schema

ABSTRACT

Background: Even experienced teachers make inconsistent classroom decisions in unexpected situations. From the cognitive load theory perspective, the effective handling of the novel, unexpected events by teachers depends on the cognitive load of the task, the teaching context in which the unexpectedness appears, and the available cognitive capacity.

Aims: Teachers' reactions to unexpected mathematical events, in particular the unexpectedness of the arithmetic calculation, was modeled, investigated experimentally, and explained within the theoretical framework of cognitive load theory.

Sample: 64 mathematics teacher trainees took part in the experiment.

Methods: In a dual-task arrangement, participants verified alternative answers to simple mathematical questions while memorizing task-irrelevant information. The answers represented low (schematic good responses), and high (unexpected good responses) processing load conditions, and control condition (incorrect responses). The memory load was low or high representing levels of extraneous load. The participants' cognitive capacity was estimated by a complex working memory span task.

Results: The verification of unexpected but correct answers was slow and more error-prone as compared with the processing speed and accuracy of schematic answers, presumably due to elevated processing (intrinsic) load. The increase in memory load resulted in slower and more inaccurate verifications. However, working memory capacity was found to mediate the extraneous load effect.

Conclusions: The results stress the importance of well-organized schemas for effective reactions to unexpected classroom events. Furthermore, it highlights the importance of accurately understanding and being aware of the impact of cognitive load on teachers to improve teaching practice.

Prologue

In a Hungarian seventh-grade classroom, the sum of the angles of convex polygons was concerned using the method of pattern finding. After determining the sum of the angles of quadrilaterals, pentagons, and hexagons, the teacher asked for the 23-sided polygon. One student, David, immediately responded verbally: "A triangle has 180, and then subtracting three from 23 is 20. So, we multiply 180 by 20 and add the result to 180." The teacher, unsure of the student's answer, ignored David's suggestion and continued the lesson with their line of thought.

David says the result is $(23 - 3) \cdot 180^\circ + 180^\circ$. David represented the pattern he discovered in a way that led to this calculation. According to the teacher's robust scheme, based on their studies and geometry

textbooks (i.e., the sum of interior angles in a convex n -sided polygon is $(n - 2) \cdot 180^\circ$), the answer is $(23 - 2) \cdot 180^\circ$. David's answer gave the same result, but it was unexpected for the teacher, who could not immediately decide whether it was right. Moreover, the teacher reported this uncertainty in her reflection after the lesson. David's answer was difficult to comprehend for two reasons: the student gave it orally, and the result was given in a different form than the one that fit the teacher's scheme.

1. Introduction

Teachers often use questions with high cognitive demands in a student-centered mathematics class, for example, asking students to

* Corresponding author. 4002, Debrecen, Pf. 400, Hungary.

E-mail addresses: konde.zoltan@arts.unideb.hu (Z. Kondé), kovacs.zoltan@uni-eszterhazy.com (Z. Kovács), eszter.konya@science.unideb.hu (E. Kónya).

<https://doi.org/10.1016/j.learninstruc.2023.101784>

Received 29 November 2022; Received in revised form 11 May 2023; Accepted 12 May 2023

Available online 23 May 2023

0959-4752/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

explain, justify, compare, generalize, and define (Lynch & Ames, 1971). After the teacher's question, the student formulates the answer. Then, the answer is evaluated and a decision must be made whether to incorporate the student's answer in the lesson, acknowledge it but put it aside, or ignore it (Mason, 2015). Making a consistent decision can be difficult if a mathematically unexpected answer is received from the student. The decision is made more difficult because the classroom environment affects teachers in complex ways: they want to follow the lesson plan, control what is happening in the classroom, and monitor students' level of comprehension and other unexpected situations that can disrupt the lesson (Feldon, 2007). The mathematics classroom can be a complex and dynamic environment where even the most prepared instructor might be surprised by unforeseen occurrences (Foster, 2015). For instance, in the case quoted, the teacher failed to take advantage of the unexpected situation, i.e., she did not demonstrate the equivalence of the two expressions; moreover, her in-the-moment reaction in the classroom was inadequate, as she recognized it after the lesson.

The current research originated from several such experiences; one is reported in the Prologue, where even mathematically well-prepared and experienced teachers make inconsistent classroom decisions reported by themselves after the class. A qualitative analysis of similar experiences (e.g., Kónya & Kovács, 2019; 2022) led to the assumption that decisions in unexpected situations are strongly influenced by the cognitive load on the teacher in the classroom, supposing that the teacher has an adequate mathematical and pedagogical background. Thus, this paper is guided by the research question of how a teaching situation's cognitive load influences the teacher's evaluative decisions.

To answer the research question, the authors designed a laboratory experiment to test their hypothesis; this quantitative approach is a novelty in their research. A simplified model situation was created that represents the classroom circumstances in which prompt but thoughtful reactions should be produced. The model situation offered a setting to observe reactions to a particular contingency that Mason (2015) calls mathematical unexpectedness under controlled experimental conditions. As in the Prologue, the experiment's unexpected situation arises from an unusual computational approach.

Different types of contingencies can affect the teacher's preconceived plan differently. However, the usual decisions (incorporate the answer, put it aside, or ignore it) require an accurate evaluation of the situation and deciding whether what the student is saying is correct. Our experiment focuses on this first stage of decision-making.

This article reviews research on contingency and unexpectedness in mathematics education and briefly introduces cognitive load theory. Next, the laboratory research design is presented, followed by a description and analysis of the results. Finally, we conclude with the pedagogical implications of the findings.

2. Theoretical background

2.1. Contingency and unexpectedness in mathematics instruction

Teachers' preparation for the lesson usually starts from textual sources (syllabus, textbook, workbook, online sources), based on which the lesson plan consists of a series of intended actions (Shulman, 1987). While the teacher's intended actions are planned, the students' reactions mostly are not. Rowland et al. (2005) describe contingency in mathematics classrooms as classroom events that are impossible to plan for, such as unexpected initiatives, responses, and questions from students. When learners express a mathematical idea, they take part in knowledge construction. The constructivist approach to learning encourages this perspective of the active participation of learners in the classroom (Thompson, 2020). To ignore such students' impulses or to dismiss them as "out of place" without examining them is to fail to engage the learner directly in active knowledge construction. The conclusion is not necessarily that the teacher is not interested in the issue the child has raised but that they may lack the behavioral resourcefulness or flexibility to

deviate from the lesson plan (Rowland et al., 2005). This segment of a math teacher's knowledge is related to what Mason et al. called *knowing-to-act in the moment* (Mason & Johnston-Wilder, 2004; Mason & Spence, 1999).

Schoenfeld (2010, 2013) outlines a general theory of in-the-moment decision-making. Schoenfeld's theory shows the factors that affect the success of a *goal-oriented decision-making* in any knowledge-intensive domain, such as teaching. Namely, a) the goals the individual is trying to achieve; b) the individual's knowledge; c) the individual's beliefs and orientations; and d) the individual's decision-making mechanism. One of these factors, d), is most closely related to the present research, as it is directly linked to the teacher's evaluative decision. The theory suggests that the decision-making mechanism is implemented in two ways depending on whether the conditions are familiar. If the conditions are familiar, people tend to react with well-practiced behaviors that can be explained by psychological constructs like schemas, scripts, etc. This situation also occurs when the learner gives the expected answer to the teacher's question. On the other hand, schematic reactions may be inappropriate if conditions suddenly deviate from the foreseeable; for example, a student gives an unexpected answer.

Contingent and unexpected situations are widely studied in the literature. For example, Rowland and Zazkis (2013) discuss three episodes from mathematics classes. In each example, the teacher was met with an unexpected circumstance that offered intriguing and fruitful learning opportunities. The authors emphasize the role of teachers' mathematical knowledge in coping with the contingency phenomenon. They argue that having relevant knowledge may result in an ability to evaluate students' initiatives immediately and open the gate for further exploration with students. They observe that, "knowledge of mathematics beyond the demands of the immediate curriculum offers some guidance to teachers in making an in-the-moment judgment of the mathematical potential of deviating from the intended instructional path" (p. 150).

In general, anything can be unexpected, which a teacher is not prepared for due to the lack of relevant cognitive schema or when a habitual way of reaction is not at hand. The sources of unexpectedness in the classroom can be diverse. It could be mathematical or pedagogical unexpectedness, but it could also be due to a student's behavior or a decline in student attendance (Mason, 2015). In the current paper, the focus is on what Mason calls mathematical unexpectedness. Mathematical unexpectedness appears whenever a new or contingent event is present in the context of mathematical problem solving or in a teaching situation, which could be unanticipated from the teachers' perspective. Professional expertise, which includes the content and the pedagogical content knowledge alongside the pedagogical knowledge, undoubtedly plays an essential role in enabling teachers to deal with mathematical unexpectedness (Venkat & Adler, 2014). Content knowledge means a deep understanding of the mathematics taught in school, while pedagogical content knowledge is the knowledge needed to make this subject matter accessible to learners. It includes, among other things, knowledge of students' mathematical conceptions, misconceptions, and subject-specific teaching strategies and representations. However, even though a teacher possesses all these elements of knowledge, as the example mentioned in the Prologue shows, other factors can also influence the teaching process.

Doyle (1986) argues that cognitive overload limits teachers' ability to adapt effectively to the complex classroom context. Feldon (2007) also stresses the impact of cognitive load on teacher performance, highlighting the role of automatic processes based on schemas in reducing cognitive load and pointing out the danger of teachers' cognitive overload. From the perspective of the dual process approach to cognition, he reviews observations of teachers' adaptive and maladaptive classroom performance that can be attributed to preexisting behavioral and cognitive schemas. He argues that cognitive load may mediate when behavior slips toward more automatic reactions. Cognitive overload, e.g., when facing the effects of task-irrelevant factors,

reduces the efficiency of controlling behavior consciously, which opens the place for automatic, effortless reactions.

Our study will present an approach to teachers' inconsistent classroom behavior from the cognitive load theory perspective.

2.2. Cognitive Load theory

Cognitive load theory is an instructional theory that aims to promote the effectiveness of educational methodology by considering the structural and functional characteristics of the cognitive system (Sweller, 1988; Sweller et al., 1998, 2019). The theory adopts the cognitive psychology approach to cognition, especially regarding limited information processing in working memory and the schematic organization of knowledge in long-term memory.

According to the prevailing theorizations, a multicomponent working memory system serves as a virtual platform for all purposeful, higher-order cognitive (e.g., learning, problem-solving, numerical cognition) activity (Baddeley, 2012; Cowan, 2014; Logie et al., 1994). The processing efficiency of new information in working memory is confined in terms of retention time and the amount of information that can be processed simultaneously. Due to limitations, control and regulation are required by the central executive or executive functions of working memory (Baddeley, 1996; Friedman & Miyake, 2017). The storage and processing capacity of working memory, especially the efficiency of executive functions, has primary importance for individual differences in higher-order cognition (Engle et al., 1999).

The controlled processes of working memory are slow and require effort. However, they can be enhanced when schematic contents of long-term memory are involved (Ericsson & Kintsch, 1995). Schemas are interacting elements constructed through learning that become automatic through repeated applications and can control processing in working memory (Sweller, 2003). Activating schemas in long-term memory and loading them into working memory eases the system from unnecessary processing and renders the cognitive activity involved in the task fast and effortless. Without schemas or using a less ordered or inappropriate schema, the processing becomes more demanding or resource-consuming for working memory (Kalyuga, 2010; Sweller, 2003; Würzberger et al., 2018). According to cognitive load theory, learning effectiveness is determined by the balance between the cognitive resources required (i.e., the load a cognitive activity imposes on working memory) and those available.

The overall cognitive load can be decomposed into different types of loads according to its sources (Sweller, 2010; Sweller et al., 1998). The processing requirement of a task is an inherent characteristic of the material involved. This arises from the complexity and structural organization of the information (i.e., the level of element interactivity) to be processed and represents the intrinsic aspect of cognitive load (Sweller, 1994, 2010b; Sweller & Chandler, 1994). Mathematical problems are often studied as prototypical examples of material with high-element interactivity (Ayres, 2001, 2006a; Sweller, 1988). However, even the processing of high-element interactivity material may impose only a reduced load on working memory when it is associated with well-organized schemas. In addition to intrinsic load, task-irrelevant, contextual information concerning how information is presented to and mediated by learners increases cognitive load.

When unrelated or interfering elements in parallel with task-related information are required to process, an unnecessary extraneous load is imposed on the working memory system. The sum of the intrinsic and extraneous loads determines the total load on the system. This supposed but disputed additive relationship is called the additivity hypothesis (Brünken, Plass, & Moreno, 2010; de Jong, 2010; Park et al., 2011; Sweller et al., 2011). When it exceeds the amount of resources available, cognitive overload occurs, which can prevent other resource-demanding processes necessary for learning, i.e., the so-called germane processing (de Jong, 2010; Sweller et al., 1998). Germane processing is related to mental exertion in processing novel information, which leads to schema

acquisition (i.e., learning) and automation. In later versions of the theory, it is no longer conceived as a type of capacity-demanding load but rather manages working memory resources between content and context-related (extraneous) mental activities (Sweller et al., 2019).

Cognitive load theory studies have highlighted various methodological factors, particularly in extraneous load, e.g., split-attention, redundancy, modality, and variability effects, which can impede the effectiveness of classroom learning (Sweller, 2010b; Sweller et al., 2019). Although under certain conditions, the increasing extraneous load can facilitate learning (Schnotz & Kürschner, 2007), the interfering effects of instructional design should be reduced or eliminated (Ayres, 2006a; Clark et al., 2006; Mayer & Moreno, 2010).

Cognitive load theory has traditionally focused on the learner-, content- and context-related factors that influence learning performance. In the present research, the scope of cognitive load theory has been extended to interpret teachers' reactions, specifically to unexpected classroom events. Returning to the Prologue, our question is, "Is it possible to explain the teacher's inappropriate reaction to the student's unexpected but correct answer from the cognitive load theory perspective?"

3. The present research

The present study investigates teachers' responsive behavior to unexpected mathematics classroom situations. Unexpected situations may arise in the classroom that causes the teacher to deviate from their teaching plan, which is triggered by students' non-schematic responses. These responses require the teacher to evaluate them before deciding on the continuation of the lesson. Our experiment is based on the general hypothesis that a teacher's evaluative reaction is influenced by cognitive load.

The cognitive load theory offered a general conceptual and methodological framework for the experimental investigation of reactions to unexpectedness. The efficiency of teachers' reactions to unexpectedness is supposed to be related to a) the processing demand generated by the unexpected, nonschematic event (intrinsic load); b) the additional load that the task-irrelevant contextual factors generate (extraneous load), see Fig. 1; and c) the amount of cognitive resources that can be devoted to the task (teachers' cognitive capacity).

The classroom situation in which mathematical unexpectedness occurs was modeled, and teachers' evaluative reactions were investigated under controlled experimental conditions. A verification task was given to the participants in which alternative (expected and unexpected) answers were required to be verified by pressing a key. The authors reasoned that even simple math problems might impose a relatively high processing demand when the problem elements are not associated with elaborated and automated schemas stored in long-term memory. Accordingly, slow and inaccurate reactions could be expected to nonschematic, unexpected answers when additional analytical steps are required for verification. Thus, *our first hypothesis: a fast and highly accurate verification is predicted when processing expected schematic information (low intrinsic load condition). In contrast, slower and more inaccurate verification is predicted for processing unexpected, nonschematic information (high intrinsic load condition).*

In classroom situations, many task-irrelevant but contextually related factors might impede proper reactions even to familiar events. The dual-task methodology (Brünken, Seufert, & Paas, 2010) was adapted for the experimentation to investigate the environmental (extraneous) effects on teachers' reactions to unexpectedness. To induce cognitive load, the primary verification task was implemented in a dual-task context in which a realistic, memory-demanding task served as a secondary task. Without any definite instruction, participants were encouraged to remember either simple or complex information potentially relevant for an ensuing recall. It was predicted that the secondary task would interfere with the primary task, generating extraneous load (Ayres, 2001). Thus, *our second hypothesis: under high extraneous load, the*

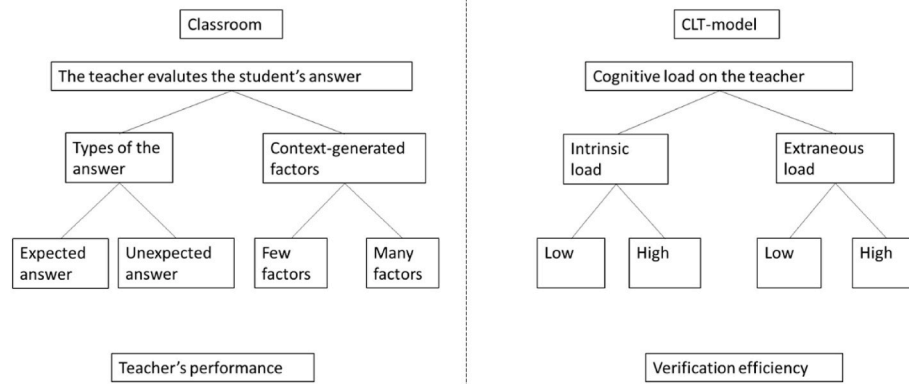


Fig. 1. The Cognitive Load Theory (CLT) model, representing the classroom situation.

accuracy and the reaction time of verifications are expected to decrease compared with that under low extraneous load.

While the verification performance is related to the schematic representations of the long-term memory, the secondary task affects the current load of the working memory. Therefore, in line with the additivity hypothesis of cognitive load theory, *our third hypothesis holds that primary task performance is also impaired under low and high extraneous conditions*. Specifically, increasing the working memory load imposed by the secondary task was expected to hinder the accuracy and the speed of verification of both schematic and nonschematic answers.

An interaction of load factors can also be expected, predicting a more significant decline in verification performance for nonschematic answers under high extraneous load conditions. This latter option is at odds with the additivity hypothesis, although it seems reasonable to assume that processing nonschematic answers requiring more working memory resources may be hampered, mainly when working memory capacity is engaged by task-irrelevant secondary information.

Furthermore, the decline in processing efficiency due to cognitive load is supposed to be a function of available resources. Since the primary math verification task and the secondary task are both thought to require working memory resources, the more capacity is available, the more effective dual-task performance can be expected. Thus, *our fourth hypothesis: better verification and more effective reactions to unexpectedness were expected for participants with higher working memory capacity*.

4. Methodology

4.1. Participants

Seventy mathematics teacher trainees took part in the experiment from two Hungarian universities. The subjects of our experiment are preparing to become secondary school mathematics teachers (in grades 5 to 12). In Hungary, these grades are taught by teachers who are qualified to teach two subjects, one of which is mathematics. All of the students included in the study had already completed at least one year of university studies in higher mathematics, however, they had not yet attended a prescribed teaching practice.

After data cleaning (for detailed explanation see in data preparation section), data from 64 people were included in the quantitative analyses. The participants were, on average, 26.64 ($SD = 8.85$) years old. Based on their training background, the participants were considered experts, assuming participants' mathematical skills would allow them to solve the math tasks used in the experiments with a high level of accuracy.

4.2. Data collection

The research was approved by the Hungarian United Ethical Review Committee for Research in Psychology (Ref. Number: 2022–30). The experimental script was programmed, and the test was customized with Inquisit Millisecond software (Inquisit Lab 5; Seattle, WA: Millisecond Software). The online data collection was administered by the Inquisit web application and conducted at a convenient time for the participants. Having a link, the participants logged in and obtained informed consent. Then, all participants completed an online experiment that lasted 40–50 min and took a 10-min online test, including a short survey with demographic questions, in a fixed order. The whole session lasted approximately an hour.

4.3. Procedure

A frame story was used to create an appropriate teacher perspective for the participants and to explain how the data collection process would work. The frame story was that a mathematics competition was being organized involving students from a class (competitors in the following). Competitors must solve mathematical equations, and their answers can be correct or false. The participant's task is to verify the competitors' answers. The contest includes two stages. In the group stage, the groups compete against each other, while in the individual stage, individual competitors compete against each other. At certain stations of the competition, participants had to answer questions about the performance of the groups or the individuals.

The instruction for participants included the frame story and the image¹ of the virtual class of 20 competitors, followed by an example of the mathematical task in the contest and a possible answer to it, and an explanation of how to use the response keys. Participants were asked to give quick but thoughtful responses to simulate a classroom-like situation that requires prompt reactions.

Following a practice trial, the group competition started. The first group of five competitors with portraits and nicknames was shown to the participant. Then, an equation was presented, and after 3 s, the first answer appeared, coupled with a randomly selected group member. This arrangement was shown until the participants' decisions. Finally, the participant indicated whether the answer was right or wrong by pressing the S or L key. Four answers (two were correct, two were false) followed each equation in a randomized order. The fifth answer was a repetition of a randomly selected previous answer.

There was no time limit for verification; no immediate feedback was given regarding whether the verification was correct. Although time

pressure may be a significant factor in the cognitive load experienced in educational settings, our aim was to investigate teachers' reactions to unexpectedness resulting from the atypical information content of the task rather than from suboptimal response conditions.

Following the verification of the answer given by the last competitor, a follow-up question appeared: How many correct answers were received from the group members? No feedback was given to the participants regarding their answers. Once the participants entered the number they thought was the correct answer, they moved on to the following equation by pressing the "space" bar. For each equation, the sequence of events was the same as described above (Fig. 2). One group received eight equations, and four groups competed. Each group consisted of five competitors, and a total of $8 \cdot 5 \cdot 4 = 160$ answers were given for verification in the first stage.

After a short break, the participants moved on to the individual stage, which differed only on the follow-up question. After verifying each group member's answer to a given equation, a portrait and the nickname of a competitor appeared on the screen, with the question: "Did this competitor give a correct or false answer?" Again, no feedback was given regarding whether the response given by the participant was correct or not. The participants also had to verify 160 answers in this stage.

4.4. Manipulation of the cognitive load

The cognitive load manipulation was based on the dual-task arrangement of the experiment in which the participant should handle mathematical information and simultaneously monitor the competition.

4.4.1. Manipulation of intrinsic load

The primary task was to solve linear equations, with one variable and operation only on one side and one number on the other. The representation of rational numbers in the equations and the answers were integer, ordinary, mixed, or decimal fractions. Sixty-four equations were created, each with two correct and two false answers (Table 1). The schematic correct answer gives the solution to the equation by performing one arithmetic operation. The alternative, nonschematic correct answer gives the solution that resulted from some additional operation compared to a simple one-step solution. These operations include:

- simplification of fractions (Equation 1),
- changing the representation of the fraction (Equations 2, 3, 4, 5),
- complicating the result by performing an unnecessary step (Equations 6, 7),
- using alternative notation (Equation 8).

The authors argue that verifying the straightforward one-step solution occurs more automatically, demanding fewer working memory resources. Accordingly, it represents the low intrinsic load condition.

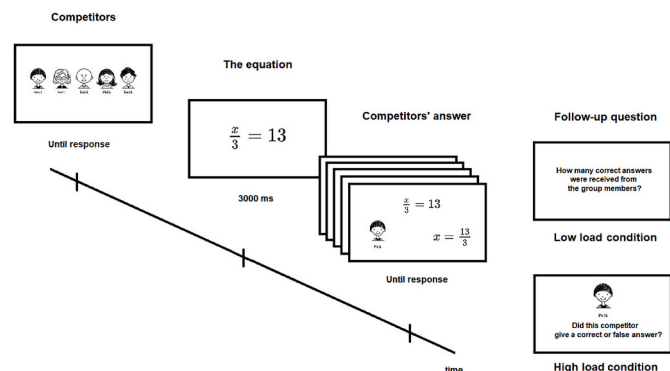


Fig. 2. Sequence of events in the experiment.¹

Table 1
Some examples of equations and answers.

Equation	Answer			
	Schematic correct (low intrinsic load)	Nonschematic correct (high intrinsic load)	False (control)	False (control)
1 $20x = 12$	$x = \frac{12}{20}$	$x = \frac{3}{5}$	$x = \frac{10}{6}$	$x = \frac{20}{12}$
2 $5x = 8$	$x = \frac{8}{5}$	$x = 1\frac{3}{5}$	$x = 40$	$x = 1.4$
3 $9x = 5$	$x = \frac{5}{9}$	$x = 0.5$	$x = \frac{9}{5}$	$x = 1.8$
4 $x + \frac{5}{4} = \frac{3}{4}$	$x = -\frac{2}{4}$	$x = -0.5$	$x = 2$	$x = \frac{15}{16}$
5 $x - 5.5 = \frac{15}{15}$	$x = 20.5$	$x = 20\frac{1}{2}$	$x = 10.5$	$x = 9.5$
6 $x - 4 = 5$	$x = 9$	$x = \frac{45}{5}$	$x = 1$	$x = -1$
7 $x - 10 = -4$	$x = 6$	$x = -6 $	$x = -6$	$x = \frac{4}{10}$
8 $x + 12 = 4$	$x = -8$	$x = \text{negative eight}$	$x = 8$	$x = \frac{1}{3}$

contrast, verifying a correct answer requiring additional operations not directly belonging to the equation-solving schema is more demanding and resource-consuming. Accordingly, the multistep solution represents the high intrinsic load condition.

The role of the two false answers was to give a probability of 0.5 that an answer was either correct or false. That is why the authors refer to this type of answer as a control. In terms of the load categories, 128 answers belonged to the control, 64 represented the low intrinsic load condition, and 64 represented the high intrinsic load condition. Note that the participants had to verify five answers for each equation, resulting in $64 \cdot 5 = 320$ answers. While four answers (two correct and two false) are assigned to each equation, the fifth answer is a random repetition of one of the answers presented earlier.

The equations in the experiment are designed to be straightforward exercises that university students with mathematics major have to solve reliably, so we assumed that the internal load is only affected by the number of steps needed to formulate the answers and ignored other possible effects, e.g., the appearance of typical errors and misconceptions. The schematic answers in this experiment match the Hungarian mathematics curriculum for grades 7 and 8. Students must solve degree-one equations at this level by modifying both sides equally. However, non-schematic solutions necessitate a flexible understanding of various rational number representations. In comparison, the Hungarian curriculum for grades 11 and 12 requires a thorough understanding of the decimal form of rational numbers. In summary, we can say that these students have the content knowledge necessary to solve the mathematics problems in the experiment.

4.4.2. Manipulation of extraneous load

The experimental manipulation of the extraneous load was implemented through a secondary task embedded into the frame story. In the first stage, the participants monitored the performance of the group in order to answer the follow-up question; they had to pay attention and keep in mind the number of correct answers given by the five members of the group. Since the correct answer was two or three, recalling this transparent (i.e., easily predictable) information was a less memory-demanding task. Consequently, the group competition represented the low extraneous load condition.

In contrast, in the second stage, the participants monitored the individual performance of the competitors. When answering the follow-up question, the working memory load was high, as each competitor's performance had to be remembered to recall a randomly selected competitor's performance. Accordingly, the individual competition

represented the high extraneous load condition.

Half of the equations coupled with related answers were presented in the low extraneous load condition, and the other half were presented in the high extraneous load condition.

4.5. Working memory capacity measurement

To assess working memory capacity, the Adaptive Operation Span was used as a subtest of the Adaptive Composite Complex Span (Gonthier et al., 2016). The test was developed to assess the working memory capacity of children and adolescents and, therefore, contains elementary addition tasks. A letter memory task is incorporated into a mathematical processing task. Participants are presented with additions to verify and letters to remember. After a series of addition-letter pairs, the participant must recall the letters in the correct order.

We favored this procedure in consideration of an apparent structural similarity between the test and the dual-task procedure applied in the present research. The authors made the original tasks slightly more complex by adding two-digit and one-digit numbers, resulting in crossing ten (e.g., 39 + 7 = 45 is true or false?), and the time constraint of verification was set to 1.5 s. This version consisted of six series with a variable number of consecutive addition-letter pairs, ranging from four to nine. The number of pairs presented in a series increased adaptively, meaning that the sequence length increased after the letters were recalled correctly and in the correct order but remained the same or decreased after an incorrect recall. The working memory capacity index is determined by the longest series on which the person was successfully tested. When indexing working memory capacity, neither the computational accuracy nor the reaction time of responses was considered.

5. Results

5.1. Data preparation

First, the overall verification accuracy was checked in the two stages of the experiment. At least a 75% verification accuracy rate was expected based on the pilot study results. For six participants, error rates higher than the performance threshold were observed in one of the experimental phases (accuracy ≤61%). Therefore, the data of these participants were excluded from further analyses². A sensitivity analysis using GPower 3.197 (Faul et al., 2007) showed that the given sample size is adequate to detect at least medium effect sizes ($\eta_p^2 > 0.079$ and $d > 0.29$) for all tests we were planning to conduct. $\alpha = 0.05$ and $\beta = 0.80$ were used and η_p^2 values of 0.01, 0.06 and 0.14 and d values of 0.20, 0.50 and 0.80 were regarded as small, medium and large, respectively (Cohen, 1988).

Next, the distribution of reaction time (RT) was inspected to identify outliers. Irrespective of the experimental conditions and whether a response was correct, RTs faster than 250 ms were treated as anticipatory reactions or indications of participants' inattention or inappropriate task commitment. We also used an upper limit of 60 s, approximately 3.5 SD above the global average. This trimming procedure resulted in a 0.4% data loss. Following the outlier eliminations, ca. 318.8 data were available per participant (instead of 320). In the 2 (extraneous load) × 3 (intrinsic load) × 5 (position) design, a total of 318.8/30 = 10.62 valid observations fell into each of the cells per participant. The data structure corresponds to a three-factor design, but in this study, only those analyses in which the position factor was not included were reported. Finally, median RTs and accuracy rates were calculated for each participant for each experimental factor level. A response was considered erroneous if it indicated a correct answer to a false or a false answer

¹ Source: <https://www.shutterstock.com/ru/image-vector/cute-smiling-faces-people-76786762>.

² Research data availability: <https://doi.org/10.48428/ADATTAR/PCABDH>

to correct. The overall accuracy rate was 93.1%. Thus, the calculation of the median RT for correct responses was based on 9.89 observations within cells.

5.2. Statistical analyses

A series of repeated-measures ANOVAs were conducted to examine the effects of experimental factors on response latencies and on accuracy rates. Beyond within-subject factors (types of loads), the continuous variable of the working memory capacity index was used as a covariate factor in some of the ANOVAs in order to test the moderating influence of working memory capacity on the effects of load factors. The results of within-subject tests were reported, and when the sphericity assumption was violated for a factor (based on Mauchly's test), ϵ -s were reported, and p values were calculated according to the Greenhouse-Geisser ($\epsilon < .75$) or Huynh-Feldt ($\epsilon > 0.75$) correction. In addition, correlational analyses were conducted to test the effect of working memory capacity on verification and secondary task performance.

5.3. The effects of cognitive load on the accuracy of verification

Table 2 summarizes the descriptive statistics for accuracy rates under different load conditions.

A 2 × 3 repeated-measures ANOVA was carried out with extraneous load (low vs. high) and intrinsic load (low vs. high vs. control) as within-subject factors. The main effect of the extraneous load factor was significant, with a large effect [$F(1, 63) = 8.76; p = .004, \eta_p^2 = 0.122$], indicating elevated error rates under high extraneous conditions. There was also a significant main effect of intrinsic load, and the effect size was large [$\epsilon = 0.566; F(2, 126) = 77.15; p < .001, \eta_p^2 = 0.55$]. The contrast analyses revealed that accuracy was approximately the same under the low intrinsic load and control conditions [$F(1, 63) = 2.77; p = .101, \eta_p^2 = 0.042$]. However, it was significantly lower under the high intrinsic load than under either the low load conditions [$F(1, 63) = 87.16; p < .001, \eta_p^2 = 0.58$] or the control condition [$F(1, 63) = 76.3; p < .001, \eta_p^2 = 0.548$].

The significant extraneous × intrinsic load interaction [$\epsilon = 0.621; F(2, 126) = 6.96; p = .006, \eta_p^2 = 0.099$] indicated that the effect of extraneous load on accuracy was different on the levels of intrinsic load. The accuracy difference between the high and low extraneous conditions differs not significantly for the high and low intrinsic load conditions [$F(1, 63) = 1.72; p = .194, \eta_p^2 = 0.027$]. However, the decrease in accuracy in the high and low intrinsic load conditions proved to be more pronounced than that in the control conditions [$F(1, 63) = 21.68; p < .001, \eta_p^2 = 0.256$ and $F(1, 63) = 9.66; p = .003, \eta_p^2 < 0.133$]. The bottom part of Fig. 3 illustrates the effects of load manipulations on accuracy.

5.4. The effects of cognitive load on verification latency

Table 3 summarizes the descriptive statistics for reaction times in milliseconds under different load conditions.

A similar 2 × 3 repeated-measures ANOVA was performed for reaction times with load types as within-subject factors. Again, the main effect of extraneous load [$F(1, 63) = 10.287; p = .002, \eta_p^2 = 0.14$] indicated longer response latencies under high extraneous load conditions. The main effect of intrinsic load [$\epsilon = 0.782; F(2, 126) = 164.155$;

Table 2
Mean accuracy rate (M) and standard deviations (SD) for the levels of intrinsic load in the low and high extraneous load conditions.

		N	Intrinsic load					
			Low		High		Control	
			M	SD	M	SD	M	SD
Extraneous load	Low	64	.964	.419	.868	.105	.959	.363
	High	64	.943	.485	.825	.148	.964	.359

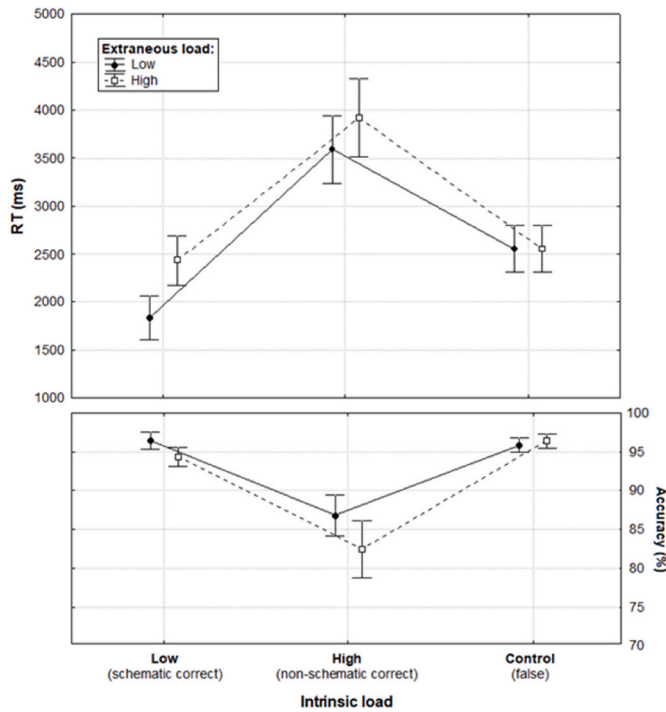


Fig. 3. Response time and response accuracy in the extraneous and intrinsic load conditions.

$p < .001$, $\eta_p^2 = 0.723$] was significant. The mean RT under the high load condition was slower than in the low [$F(1, 63) = 232.01$; $p < .001$, $\eta_p^2 = 0.786$] or in the control [$F(1, 63) = 135.93$; $p < .001$, $\eta_p^2 = 0.683$] condition. The RTs were slower in the control condition than in the low load condition [$F(1, 63) = 47.01$; $p < .001$, $\eta_p^2 = 0.427$]. In contrast with the accuracy, no significant intrinsic \times extraneous load interaction was found [$\epsilon = 0.676$; $F(2, 126) = 3.0$; $p = .074$, $\eta_p^2 = 0.045$], indicating that there were no differences in the effect of extraneous load across the levels of intrinsic load. The top part of Fig. 1 illustrates the effects of load manipulations on response latencies. The effect sizes were large for each significant effect.

5.5. The effects of working memory capacity on the efficiency of verification

The working memory capacity index (Table 4) was added as a covariate factor to the original 2 \times 3 model, which had the extraneous and intrinsic load as within-subject factors. It was done to see if the working memory capacity could mediate the load effects observed.

In RT data, the main effect of working memory capacity [$F(1, 62) = 0.330$, $p = .568$, $\eta_p^2 = 0.005$] was not significant, nor were its interactions with the load factors [F values > 0.43 , p values $> .514$, η_p^2 values < 0.037]. More importantly, introducing working memory capacity into the model, the main effect of extraneous load disappeared [$F(1, 62) = 0.0001$, $p = .988$, $\eta_p^2 = 0.001$], and the extraneous load \times intrinsic load interaction was not significant [$\epsilon = 0.678$; $F(2, 124) = 1.44$, $p = .24$, $\eta_p^2 = 0.023$]. However, the main effect of intrinsic load

Table 3 Mean RTs and standard deviations (SD) for the levels of intrinsic load conditions within the low and high extraneous load conditions.

		N	Intrinsic load					
			Low		High		Control	
			M	SD	M	SD	M	SD
Extraneous load	Low	64	1834.47	912.78	3589.94	1393.48	2551.23	987.05
	High	64	2431.89	1041.76	3922.01	1633.22	2557.23	997.13

remained significant, and the effect size was large [$\epsilon = 0.823$; $F(2, 124) = 17.338$, $p < .001$, $\eta_p^2 = 0.219$]. In the accuracy data, the main effect of working memory capacity [$F(1, 62) = 7.324$, $p = .009$, $\eta_p^2 = 0.106$] was significant, and its interactions with either of the load factors were not significant [F values > 3.243 , p values $> .071$, η_p^2 values < 0.05]. In line with the RT analyses, the main effect of extraneous load disappeared [$F(1, 62) = 3.093$, $p = .084$, $\eta_p^2 = 0.048$], and the main effect of intrinsic load remained significant [$\epsilon = 0.568$; $F(2, 124) = 12.947$, $p < .001$, $\eta_p^2 = 0.173$]. The extraneous \times intrinsic load was not significant [$\epsilon = 0.621$; $F(2, 124) = 0.106$, $p < .899$, $\eta_p^2 = 0.001$].

5.6. Performance in the secondary task

The secondary task performance was investigated separately during the two stages of the experiment. The accuracy index was calculated based on the rate of the correct responses to the follow-up questions representing low and high memory load conditions. The total number of questions was 32 in both stages. The accuracy of responses was compared with the participant's previously given response, regardless of whether it was correct. The descriptive data are in Table 5. The mean values of 0.890 and 0.875 indicate a not exclusively random strategy (i.e., the participant does not decide based on a guess) for responding to the follow-up questions. Note that modeling the exclusively random strategy with the binomial distribution ($n = 32$, $p = .5$) means the expected value for the accuracy index is 0.5, while $SD = 0.088$. Wilcoxon's signed rank test showed that the accuracy index in the secondary task did not differ significantly between the two stages [$W = 809.5$, $p = .564$]. The overall accuracy in both stages of the experiment indicated that the participants did not ignore the secondary tasks.

Correlation analyses were performed to investigate the role of working memory capacity in secondary tasks and the overall accuracy of verification (Table 6).

A significant, medium correlation appeared between the working memory score and the accuracy level of the verification tasks in both stages of the experiment. An elevated correlation coefficient was found between working memory capacity and secondary task accuracy which was more pronounced when the secondary task was more demanding.

Table 4 Descriptive statistics of the working memory capacity score.

	N	Min	Max	Mean	SD
working memory capacity score	64	4	9	6.67	1.415

Table 5 Descriptive statistics of the accuracy index in the secondary tasks.

	Accuracy index	
	First stage	Second stage
Valid	64	64
Mean	.890	.875
SD	.097	.126

Table 6

Correlation coefficients of working memory capacity scores with the accuracy of the verification and secondary tasks in the first and second stages of the experiment.

Load conditions		N	Correlation coefficients	
			Spearman's rho	sig (p)
Low extraneous load	verification task	64	.324	.009
	secondary task		.417	.001
High extraneous load	verification task	64	.377	.002
	secondary task		.562	.001

6. Discussion

In the present research, teachers' reactions to unexpected mathematical events originated from students' atypical answers were modeled and investigated in a dual-task experiment. Within the theoretical framework of cognitive load theory, the authors' general assumption was that the effective handling of unexpected events depends to a considerable extent on the cognitive requirement and the amount of cognitive resources that can be devoted to the cognitive activity. Accordingly, the efficiency of teachers' reactions to contingency was investigated as a function of two cognitive load types: the load imposed by task-relevant and task-irrelevant cognitive activities and the person's cognitive capacity.

In line with our expectations, the analysis of RT and accuracy data showed that the inherent characteristics of the mathematical task, i.e., the level of intrinsic load, which relates to the contingency in this research, significantly affected the efficiency of verification performance. Under a low intrinsic load, the evaluative responses were highly accurate and relatively fast. According to cognitive load theory, this result implies that processing typical schematic answers is primarily based on automatic recall of relevant schema (here, the commonly used one-step equation solution). The automatic, effortless application of schemas leads to highly effective (i.e., fast and accurate) verification. However, when an unexpected, correct but not schematic answer had to be verified, as a result of increased intrinsic load, the accuracy decreased, and the response time increased. This result indicates that processing unexpected, novel information does not rely solely on automatic recall and application of schemas, but additional analytical steps are required to obtain the correct answer. These nonautomatic, controlled operations lead to slower and, occasionally, incorrect evaluations, e.g., when analyses fail to provide the appropriate solution.

For the present purposes, the manipulation of the extraneous load was realized by incorporating an ecologically valid secondary task into the context of the verification task. The performance in the secondary task reached an acceptable level of accuracy, i.e., an accuracy rate far above the level of random responding. The accuracy of the responses for the more complex follow-up questions was comparable to that for the rather simple follow-up questions (.890 vs. .875, see Table 5). However, when it was required that a complex compound be kept in mind in parallel with the verification task, the increased working memory load resulted in slower and more error-prone verification. In contrast, verifications were fast and accurate when the secondary task imposed a low extraneous load. These findings confirmed our expectations that a resource-intensive secondary cognitive activity makes the primary task less effective.

As expected (Hypothesis 3), no interaction between intrinsic and extraneous load factors was observed in RT and accuracy data. The only significant interaction we found in accuracy was basically due to the null effect of extraneous load on processing false answers. Apart from this, the impact of extraneous load on RT and accuracy was approximately the same at low and high levels of the intrinsic load. The observed additivity of load factors supported the cognitive load theory hypothesis that cognitive load is derived from distinct sources or categories (de Jong, 2010; Sweller et al., 2011).

In formulating Hypothesis 3, we mentioned the possible interaction of the load factors. However, this was not observed, which we explain by the low complexity of the primary task. Further research may provide evidence that when using more complex mathematical problems as a primary task, processing accuracy and reaction time may suffer more degradation when the extraneous load is high.

Correlation and covariance analyses using the working memory capacity index confirmed that the amount of processing resources could be crucial for verification and secondary task performance, although to varying extents. The amount of working memory capacity correlated significantly and at a medium level with the accuracy of the secondary task in both low and high extraneous load conditions. In other words, the more information-processing resources available, the more effective information retention and retrieval in the secondary task. The amount of processing resources was also indicative of verification efficiency, albeit to a lesser extent than secondary task performance, as revealed by the weaker correlations.

The role of working memory capacity in secondary task performance received further support from the results of covariance analysis. Controlling for working memory capacity, the main effect and interactions of extraneous load factor disappeared in both RT and accuracy data. This finding suggests that the impact of extraneous load, which stems from the memory demand of the secondary task, is primarily mediated by the amount of processing resources available. The greater the working memory capacity is, the more efficient the secondary task performance and the less distracting the secondary task is for verification performance.

The intrinsic load's effect, however, remained significant when the effect of working memory capacity was partialled out. This fact demonstrates again that intrinsic and extraneous loads originate from different sources. While the amount of available resources mediated the effect of extraneous load on verification performance, the effect of intrinsic load was found to be at least partially independent of working memory capacity. The intrinsic load effect can be explained by the inherent characteristics of the material to be processed and its associations with the schemas stored in long-term memory. Accordingly, processing typical answers based on schemas stored in long-term memory imposes only a reduced load on working memory, as opposed to unexpected, nonschematic responses, which require additional analytical steps and working memory resources.

In sum, the present results demonstrated that delayed and more error-prone teachers' reactions could be expected in response to potentially valuable classroom events when they are novel, unexpected, and fall outside the prepared, schematic expectations. Furthermore, contextual factors irrelevant to the task may interfere with appropriate reactions even to expected events. From the current cognitive load theory perspective, the content- and context-related effects on behavioral data represent intrinsic and extraneous aspects of cognitive load. Our results suggest that extraneous load effects originating from contextual factors are mediated and can be explained by the amount of storage and processing capacity of working memory. On the other hand, the intrinsic load effect originated from the novel or nonschematic material, i.e., the unexpectedness of the content. This schema-related cognitive load cannot be explained solely by working memory capacity; instead, it is related to schema organization in long-term memory.

Pedagogical implications

As a pedagogical consequence, the authors consider accurate understanding and awareness of the impact of cognitive load on teachers to be important for improving teaching practice. The cognitive load on teachers during lessons can be intrinsic or extraneous. While the extraneous load is caused by context-generated factors that appear in the classroom and depends mainly on the person's working memory capacity, it can change little or none. However, based on the present research, the authors highlight the importance of well-organized

schemas stored in long-term memory in the effective reactions to contingent, unexpected events in the classroom. Teacher training and professional development programs for in-service teachers can play an essential role in building well-organized, rich schemas linked to particular curricular units. The reflective teacher recognizes the poor classroom decisions, as evident in the Prologue. The development of rich schemas over time, which can be helped by regular self-reflection in addition to learning as much as possible about students' typical answers, can reduce the intrinsic cognitive load. In the absence of proper schemas, there is a danger that a teacher will mishandle novel, unexpected, but otherwise potentially valuable student contributions.

Limitations and further research

The level of expertise of participants and the low complexity of mathematical content on which we investigated the load effects may be limiting factors for a broad generalization of the present results. The mathematical problems with low processing requirements could have been manageable tasks within the range of available processing capacity for experts, resulting in an underestimation of the role of working memory capacity in math-related processing. Further research is needed to investigate the load effects on the efficiency of skilled, expert-level performance guided by schemas in the mathematical domain and the possible mediating role of working memory. An additional empirical perspective for a cognitive load-related explanation of inconsistent reactions to unexpectedness is to involve subjective measurement techniques to estimate the extent of perceived load in parallel with behavioral measures (e.g., Ayres, 2006b).

CRedit author statement

Zoltán Kondé: Conceptualization, Methodology, Software, Formal analysis, Writing - Original Draft **Zoltán Kovács:** Conceptualization, Formal analysis, Writing - Original Draft, Investigation **Eszter Kónya:** Conceptualization, Investigation, Supervision, Writing - Review & Editing, Funding acquisition.

Acknowledgments

This study was funded by the Research Program for Public Education Development of the Hungarian Academy of Sciences (KOZOKT2021-16).

References

- Ayres, P. (2001). Systematic mathematical errors and cognitive load. *Contemporary Educational Psychology*, 26(2), 227–248. <https://doi.org/10.1006/ceps.2000.1051>
- Ayres, P. (2006a). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology*, 20(3), 287–298. <https://doi.org/10.1002/acp.1245>
- Ayres, P. (2006b). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, 16(5), 389–400. <https://doi.org/10.1016/j.learninstruc.2006.09.001>
- Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 49(1), 5–28. <https://doi.org/10.1080/713755608>
- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Brünken, R., Plass, J., & Moreno, R. (2010). Current issues and open questions in cognitive load research. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 253–272). Cambridge University Press. <https://doi.org/10.1017/CBO9780511844744.014>
- Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 181–202). Cambridge University Press. <https://doi.org/10.1017/CBO9780511844744.011>
- Clark, R. C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. Pfeiffer.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. In *Educational psychology review* (Vol. 26, pp. 197–223). <https://doi.org/10.1007/s10648-013-9246-y>. Issue 2.
- Doyle, W. (1986). Classroom organization and management. In M. C. E. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 392–431). Macmillan Publishing Company.
- Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309–331. <https://doi.org/10.1037/0096-3445.128.3.309>
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102(2), 211–245. <https://doi.org/10.1037/0033-295x.102.2.211>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist*, 42(3), 123–137. <https://doi.org/10.1080/00461520701416173>
- Foster, C. (2015). Exploiting unexpected situations in the mathematics classroom. *International Journal of Science and Mathematics Education*, 13(5), 1065–1088. <https://doi.org/10.1007/s10763-014-9515-3>
- Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, 86, 186–204. <https://doi.org/10.1016/j.cortex.2016.04.023>
- Gonthier, C., Thomassin, N., & Roulin, J. L. (2016). The composite complex span: French validation of a short working memory task. *Behavior Research Methods*, 48(1), 233–242. <https://doi.org/10.3758/s13428-015-0566-3>
- de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, 38(2), 105–134. <https://doi.org/10.1007/s11251-009-9110-0>
- Kalyuga, S. (2010). Schema acquisition and sources of cognitive load. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 48–64). Cambridge University Press. <https://doi.org/10.1017/CBO9780511844744.005>
- Kónya, E., & Kovács, Z. (2019). *Implementing problem solving in mathematics classes*. In I. Gebel, A. Kuzle, & B. Rott (Eds.) (pp. 121–128). WTM-Verlag <https://doi.org/10.37626/GA9783959871167.0>
- Kónya, E., & Kovács, Z. (2022). Management of problem solving in classroom context. *Center for Educational Policy Studies Journal*, 12(1), 81–101. <https://doi.org/10.26529/cepsj.89>
- Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition*, 22(4), 395–410. <https://doi.org/10.3758/BF03200866>
- Lynch, W. W., & Ames, C. (1971). *Individual cognitive demand schedule*. Indiana University. *Technical Report 4.2*.
- Mason, J. (2015). Responding in-the-moment: Learning to prepare for the unexpected. *Research in Mathematics Education*, 17(2), 110–127. <https://doi.org/10.1080/14794802.2015.1031272>
- Mason, J., & Johnston-Wilder, S. (2004). *Fundamental constructs in mathematics education*. Routledge. <https://doi.org/10.4324/9780203465387>
- Mason, J., & Spence, M. (1999). Beyond mere knowledge of mathematics: The importance of knowing-to act in the moment. *Educational Studies in Mathematics*, 38, 135–161. https://doi.org/10.1007/978-94-017-1584-3_7
- Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 131–152). Cambridge University Press. <https://doi.org/10.1017/CBO9780511844744.009>
- Park, B., Moreno, R., Seufert, T., & Brünken, R. (2011). Does cognitive load moderate the seductive details effect? A multimedia study. *Computers in Human Behavior*, 27(1), 5–10. <https://doi.org/10.1016/j.chb.2010.05.006>
- Rowland, T., Huckstep, P., & Thwaites, A. (2005). Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi. *Journal of Mathematics Teacher Education*, 8(3), 255–281. <https://doi.org/10.1007/s10857-005-0853-5>
- Rowland, T., & Zazkis, R. (2013). Contingency in the mathematics classroom: Opportunities taken and opportunities missed. *Canadian Journal of Science, Mathematics, and Technology Education*, 13(2), 137–153. <https://doi.org/10.1080/14926156.2013.784825>
- Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, 19(4), 469–508. <https://doi.org/10.1007/s10648-007-9053-4>
- Schoenfeld, A. H. (2010). *How we think: A theory of goal-oriented decision making and its educational applications*. Routledge. <https://doi.org/10.4324/9780203843000>
- Schoenfeld, A. H. (2013). Reflections on problem solving theory and practice. *The Mathematics Enthusiast*, 10(1 & 2), 9–34.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–21.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Sweller, J. (2003). Evolution of human cognitive architecture. *Psychology of Learning and Motivation*, 43, 215–266. [https://doi.org/10.1016/S0079-7421\(03\)01015-6](https://doi.org/10.1016/S0079-7421(03)01015-6)
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22, 123–138. <https://doi.org/10.1007/s10648-010-9128-5>

- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer Science & Business Media.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12(3), 185–233. https://doi.org/10.1207/s1532690xcil203_1
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31, 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Thompson, P. W. (2020). Constructivism in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 127–134). Springer International Publishing. https://doi.org/10.1007/978-3-030-15789-0_31.
- Venkat, H., & Adler, J. (2014). Pedagogical content knowledge in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 477–480). Springer Netherlands. https://doi.org/10.1007/978-94-007-4978-8_123.
- Wirzberger, M., Herms, R., Esmaili Bijarsari, S., Eibl, M., & Rey, G. D. (2018). Schema-related cognitive load influences performance, speech, and physiology in a dual-task setting: A continuous multi-measure approach. *Cognitive Research: Principles and Implications*, 3(1). <https://doi.org/10.1186/s41235-018-0138-z>
- Zoltán Kondé, University of Debrecen. He received his Ph.D. in psychology. His research interest includes creativity in mathematics and cognitive and attentional control.
- Zoltán Kovács, Eszterházy Károly Catholic University. He received his Ph.D. in pure mathematics. His research interest includes mathematical problem-solving and problem-posing.
- Eszter Kónya, University of Debrecen. She received her Ph.D. in mathematics education. Her research interest includes reasoning and proving and geometric reasoning.