



# The Influence of Interactive Learning Materials on Solving Tasks That Require Different Types of Mathematical Reasoning

Marija Kaplar<sup>1</sup> · Slaviša Radović<sup>2</sup> · Kristina Veljković<sup>3</sup> · Ksenija Simić-Müller<sup>4</sup> · Miroslav Marić<sup>5</sup>

Received: 20 September 2019 / Accepted: 3 January 2021 / Published online: 26 January 2021  
© Ministry of Science and Technology, Taiwan 2021

## Abstract

The purpose of this study is to analyse the effects of the Interactive Learning Materials Triangle (iLMT) on the learning and knowledge retention of 12-year-old students. The iLMT is a digital version of the standard school learning materials in Serbia, and is characterized by a high degree of interactivity and immediate feedback during the learning process. We conducted an experiment to explore whether iLMT influences student success in solving mathematical tasks that require different types of mathematical reasoning. Based on previous extensive research by Lithner, 4 types of tasks are discussed: high relatedness answer, high relatedness algorithm, local low relatedness, and global low relatedness. The study involved 633 students and 13 teachers of mathematics, equally distributed in control and test groups. The main findings indicate that student success on a knowledge test for high relatedness answer and local low relatedness tasks for the test group was significantly higher than for the control group. On the knowledge retention test, students in the test group outperformed students in the control group at high relatedness algorithm and local low relatedness tasks. Our results also suggest that, even when learning materials are carefully digitalized with the use of available technological advantages, student success in global low relatedness tasks may still be lacking.

**Keywords** Elementary education · GeoGebra · Interactive learning environments · Mathematical reasoning

---

✉ Marija Kaplar  
marija.radojicic@uns.ac.rs

## Introduction

During the past decade, there has been evident progress in applications of digital materials in learning mathematics, as educators have become more aware of the importance of such instructional tools in educational practice. In line with this progress, teaching equipment (both hardware and software) has been modernized at all levels, from primary schools to higher education (De Witte & Rogge, 2014; Takači, Stankov, & Milanovic, 2015). Furthermore, extreme situations such as the COVID-19 pandemic highlight the importance of having digital versions of learning materials available. Digital learning materials span a wide range in terms of form and content, but a question common to all is how effective they are.

Although learning environments have changed over the years, the long-standing problem of mathematics education is that students often apply routine thinking, while educators expect them to develop greater understanding, make complex conceptual connections, and be problem solvers (Bergqvist & Lithner, 2012; Hiebert, 2003; Lampert, 1988; Lithner, 2008). The development of different types of student mathematical reasoning is an important aim of contemporary mathematics education. Special attention is given to creative mathematical reasoning that will prepare students for solving non-routine mathematical problems and life-long learning (Freudenthal, 1987; Granberg & Olsson, 2015; Norqvist, Jonsson, Lithner, Qwillbaird, & Holm, 2019). In one of the early studies, Lithner (1998) stresses that students often focus on what is familiar to them and on what they can easily recall at a superficial level, avoiding the use of creative mathematical reasoning (perceived as a complex thinking activity). It is imperative to foster and improve students' creative reasoning, but in the same time, it is the most difficult task for educators (Freudenthal, 1987; Norqvist, 2018; Norqvist et al., 2019).

Researchers have shown that students have more opportunities to learn facts and practise simple procedures to solve problems than to engage in more extensive learning processes and deeper cognitive strategies (Boesen et al., 2014; Hiebert, 2003; Jonassen, 2000). As a consequence, students encounter difficulties when trying to solve complex mathematics tasks and tasks that require creative mathematical reasoning (Boesen, Lithner, & Palm, 2006; Lithner, 2004; Palm, Boesen, & Lithner, 2005). However, experimental studies have shown that students who were involved with solving complex tasks and developing creative mathematical reasoning outperform students who were exposed to routine problems only (Jonassen, 2000; Kapur, 2011). These results indicate that high-level mathematical reasoning can be learned and complex mathematical skills can be developed, but this is a multifaceted and complex educational challenge that goes beyond one educational system, and more comprehensive research of effectiveness of different teaching strategies is needed (Freudenthal, 1987; Lampert, 1988; Lithner, 2015). Moreover, insights about the effects of interactive learning materials on student success in solving mathematics tasks that require different types of reasoning (as described below) are still lacking.

Bearing in mind the above-mentioned issues, the purpose of this study is to investigate the effects of an interactive learning environment that is a digital version of standard school learning materials in Serbia. Our interactive learning environment is based on a particular instructional design model and use of available tools (mostly GeoGebra), with the purpose of helping students to better understand mathematical

concepts and properties. We will investigate to what extent the design of such a learning environment supports students in solving tasks that require different types of reasoning.

## Theoretical Framework

In an extensive body of work, Lithner (1998, 2000, 2004, 2015) explored the relationship between mathematical tasks that students encounter and type of reasoning they need to solve these tasks. He defined mathematical reasoning as “the line of thought, ways of thinking adopted in order to produce assertions and reach conclusions” (Lithner, 2000, p. 166). Lithner (2008) gives two basic types of reasoning: imitative reasoning (IR) and creative mathematics reasoning (CMR) and then divides these into subtypes. IR is defined as reasoning connected to routine tasks, where the respondent imitates procedure or facts memorized from the learning materials. IR includes two subtypes: memorized reasoning (MR) and algorithmic reasoning (AR).

MR takes place when recalling a complete answer or when the strategy implementation consists only of writing it down. Usually in this kind of reasoning, previously learned material has been just recalled. For example, for the question: “How many  $\text{dm}^2$  are in  $\text{m}^2$ ?”, a student can give the correct answer merely by recalling information from the book without deeper insight into the concept of area. Similarly, this type of reasoning is usually present when students memorize mathematical proofs without being able to explain them.

AR is reasoning based on repetition of already learned familiar procedures, which lead to the conclusion. Actually, this kind of reasoning is based on recalling the solution algorithm, and usually is applied in calculation tasks or tasks which do not include the need for creating a new solution. In this kind of tasks, the major challenge for student is to choose the appropriate algorithm. For example, in the task: “Find the area of a triangle if  $a = 5$  and  $h_a = 3$ ”, a student can easily find the area only by applying an algorithm without really understanding what  $h_a$ ,  $a$ , or area are. Hence, AR may be applied even with weak understanding of a topic (Lithner, 2008).

While MR and AR are characterized with a lack of conceptual thinking and deep insight into intrinsic mathematical properties of elements, these are indispensable in the case of CMR. Students usually apply CMR when they are faced with a new task and need to create a new reasoning algorithm (novelty). Moreover, it involves both plausibility (arguments which support strategy choice) and mathematically founded arguments (based on understanding of mathematical properties of the objects involved in the reasoning).

In an attempt to design an outline for task classification according to appropriate mathematical reasoning, Boesen, Lithner, and Palm (2010) present a framework for classifying tasks, which relies on the study of Palm et al. (2005). The classification is made according to the similarity between tasks that students have to solve, and tasks and information that can be found in textbooks and other learning materials. It includes the following steps: (a) defining the task solution; (b) describing tasks through variables such as information about mathematical components and various hints; (c) identification of the rules, theorems, described facts, exercises, and tasks in learning materials which have the same (and similar) solutions; and (d) conclusion about relatedness between tasks and learning material. If there are at least three exercises, examples,

rules, etc. with similar characteristics as the task in the learning materials, it will be classified as high relatedness, while in all other cases, it will be classified as a low relatedness task (Boesen et al., 2010).

Within high relatedness tasks, two subtypes are defined: high relatedness answer (HRA1) and high relatedness algorithm (HRA2). HRA1 are tasks whose solutions require textual answers or facts that can be found in learning materials. HRA2 are tasks whose solutions are reached through implementation of algorithms available in learning materials. HRA1 and HRA2 correspond to MR and AR, respectively (Boesen et al., 2010).

Low relatedness tasks are divided in two subtypes, local low relatedness (LLR) and global low relatedness (GLR). LLR are tasks whose solutions can be reached through modification or combination of different algorithms available in learning materials. GLR tasks require creating a new (for the student) solution, which is to a lesser extent based on previously seen algorithms and facts. Usually, these tasks are new for reasoner, and are often presented in real-life context. The framework from Boesen et al. (2010), together with connections to the Lithner's (2008) types of reasoning, is presented in Table 1. We note that the above categorization is closely aligned to levels of cognitive demand (National Council of Teachers of Mathematics [NCTM], 2014), where high relatedness corresponds to low-level tasks, and low relatedness corresponds to higher-level tasks. Although there are various task categorizations according to students' reasoning, we used the classification given by Boesen et al. (2010) in our research. Their work presents the task taxonomy based on the similarity between tasks and information that can be found in textbooks, which best fits the aims of the present study.

## Conclusions of Previous Studies

The type of a task can be an important trigger, which initiates the kind of reasoning students apply (Jäder, Sidenvall, & Sumpter, 2017). When students are faced with a task that has similarities to a task they have already seen in the textbook or in class, they will try to recall the ways of solving it or relate to a suitable algorithm, i.e. students will apply MR or AR (Boesen et al., 2010). However, when students are faced with a completely unfamiliar problem, they usually reason in a way that can be characterized as CMR. Furthermore, in such a situation, students sometimes apply IR but without success (Jäder et al., 2017). In the conclusion of their paper, Boesen et al. (2010) write that students in their study achieved higher scores on problems that required the use of IR.

Bergqvist (2007) points out that school tests in mathematics in Sweden are created so that over 70% of the tasks can be solved by applying only MR and AR, while tasks that require CMR are often neglected. In addition to this, Bergqvist (2007) notes that many teachers create tests that rely heavily on IR. This phenomenon emphasizes the important role of learning environments and learning content, i.e. ways in which the material is presented and the type of outcome that is expected of students. Lithner and Bergqvist (Bergqvist, 2007; Lithner, 2000, 2004, 2008) stress the importance of the approach in presenting mathematical content. They acknowledge that difficulties in learning CMR are with respect to lack of (a) reflection on learning, (b) mathematical arguments, and (c) mathematical foundation during the learning process (Bergqvist & Lithner, 2012). In a study by Jonassen (2000), one group of students was exposed to

**Table 1** The connection between types of tasks and corresponding reasoning

| Types of reasoning | Task categories | Brief description                 | Task example  |
|--------------------|-----------------|-----------------------------------|---|
| IR                 | MR              | High relatedness answer (HRA1)    | Type of task where a concrete response is expected, which pupils have already had the opportunity to hear/see in class.   |
|                    | AR              | High relatedness algorithm (HRA2) | Type of task where direct application of some of the algorithms presented during instruction is expected.   |
| CMR                |                 | Local low relatedness(LLR)        | Type of task that can be solved by adapting some of the already seen algorithms to the given problem or by combining several familiar algorithms.   |
|                    |                 | Global low relatedness(GLR)       | Type of task where the expected response is based to a very small extent on previously presented algorithms or solutions.   |
|                    |                 |                                   | <p>What is the sum of the measures of interior angles in a triangle? (Students were previously introduced to the sum of the interior angles in a triangle during the learning process.)</p> <p>Find the measure of the interior angle of a triangle whose other two angles are <math>58^\circ</math> and <math>45^\circ</math>. (Students were previously introduced to the appropriate algorithm for solving this kind of task and they exercised similar tasks.)</p> <p>Calculate measures of angles in a triangle ABC, where angle <math>\alpha = 44^\circ</math> and the angle in the intersection of bisectors of angles <math>\alpha</math> and <math>\beta</math> is <math>114^\circ</math>. (In this case, students were previously acquainted with the properties of the bisector and the sum of the interior angles of a triangle, and they used both in various tasks but they had not previously encountered a task that combines these algorithms, as this one does.)</p> <p>On a school break, Mark has to go to the bakery or the book store. He knows that the angle between the school, bakery and book store is <math>56^\circ</math>, and the angle between the bakery, book store and school is <math>58^\circ</math>. Which is closer to Mark, the bakery or book store? (In this case, students were previously familiar with the relationship of angles and sides in a triangle but had not seen a similar task in teaching or learning materials. To solve this problem it is necessary to translate the problem from real to mathematical, recognize that they can apply the rule for the ratio of angles to sides in a triangle, to successfully apply it, and to evaluate it.)</p> |

IR imitative reasoning, MR memorized reasoning, AR algorithmic reasoning, CMR creative mathematics reasoning

content that requires AR, while another group of students was engaged with materials that require more CMR. Students who were oriented toward more CMR content performed better on the final exam.

Granberg and Olsson (2015) present the advantages of using the software GeoGebra for developing CMR in students. As one of the main advantages of GeoGebra, the authors stress the importance of immediate feedback (the so-called creative feedback). Creative feedback does not give simple clues for the next step or the final answer, but helps students visualize their ideas, enabling them to connect mathematical properties with reasoning. In such a case, a direct answer is not given, but every move in the applet is visualized to allow students to test and verify/falsify their own ideas in solving tasks. Creative feedback (1) fosters students to experiment and evaluate their strategies in order to construct understanding (Olsson, 2017), and (2) allows students to evaluate and interpret their conclusions (Olsson, 2017). Earlier findings indicate that GeoGebra software, due to these advantages, is a suitable environment for developing CMR (Granberg & Olsson, 2015). Dragging, another advantage of the GeoGebra, allows discovering invariant properties and investigating geometry problems using various strategies, such as exploring, conjecturing, validating, and justifying (Arzarello, Olivero, Paola, & Robutti, 2002). Dragging enables students to feel “motion dependency”, which can be related with logical dependency within the mathematical context (Baccaglioni-Frank & Mariotti, 2010), and contribute to the formation of abstract knowledge (Leung, 2008). Finally, GeoGebra supports students in linking together geometric interpretation, algebraic properties, and arithmetic interaction through the interactive representation of mathematical objects and concepts. To the extent of our knowledge, there is a lack of investigations of the effects of learning materials with the GeoGebra applets on imitative reasoning.

## Present Study

This study is the continuation of a project that analysed the influence of the Interactive Learning Materials Triangle (iLMT) on students' attitudes toward mathematics, mathematics classes, and textbooks (Radović, Radojičić, Veljković & Marić, 2020). In a previous study (Radović et al., 2020), we emphasized that students recognize interactive learning environments such as iLMT as a positive and welcome material in mathematics learning. Although the overall impact of iLMT was highly positive, authors stated that a more detailed study was needed, which would give better insight into students' understanding of the material and students' reasoning.

The iLMT content is similar to the standard learning environment, i.e. official mathematics textbooks for sixth grade (same definitions, theorems, and tasks), but the material is presented in digital and interactive form. In particular, iLMT is based on certain instructional design models and utilizes information-communication technologies such as GeoGebra, which can contribute to the development of different forms of mathematical reasoning in students (Arzarello et al., 2002; Baccaglioni-Frank & Mariotti, 2010; Bergqvist & Lithner, 2012; Granberg & Olsson, 2015).

Mathematics classrooms in Serbia are still, to a large extent, knowledge-oriented, with the practices being largely teacher-centred (Radišić & Baucal, 2018). The students' role is, for the most part, receptive, as lecturing is the dominant form of teaching (Mincu, 2009; Radišić & Baucal, 2018). The official curriculum for the sixth grade in

Serbia states that mathematics should be taught so that it enables students to develop abstract and critical thinking and apply the acquired knowledge and skills in further education. However, the prescribed learning outcomes related to triangles can only be met by activating IR, and it is noticeable that the largest number of tasks in textbooks is related to specific calculations and implementation of known algorithms for solving tasks. Only about 30% of tasks in regular textbooks can initiate the development of students' CMR (in regard to the content used in this study).

Having in mind the advantages of GeoGebra applets and digital learning materials, the aim of this study is to investigate the effect of the developed interactive learning environment iLMT on student success in solving mathematical tasks that require different types of mathematical reasoning. Considering that iLMT presents a digital, student-oriented, and interactive version of textbook materials used in the standard environment (i.e. control group), but grounded in a particular instructional design model and the use of available tools (mostly GeoGebra), this study provides a new and different contribution from similar previous studies (Granberg & Olsson, 2015; Jonassen, 2000). Unlike other studies that tested materials specifically created to foster a certain type of reasoning, we investigated the effect of digital versions of standard learning materials presented with available advantages of information-communication technologies. In this paper, we will explore the influence of iLMT on four groups of tasks introduced earlier in this paper: HRA1, HRA2, LLR, and GLR. Thus, the main research questions underlying this article are as follows:

1. Are there any significant effects of group (control or test), gender, or their interaction on student success on HRA1, HRA2, LLR, and GLR tasks, on knowledge and knowledge retention tests?
2. Are there any significant effects of group, gender, or their interaction on student success in solving tasks that belong to two different types of reasoning (IR and CMR) on knowledge and knowledge retention tests?

## Design of the Interactive Learning Materials Triangle

The iLMT environment is developed as a supplementary teaching material in digital and interactive form. The main principles for design of interactive learning environments, on which the iLMT is grounded, are indicated by Chiu and Churchill (2016) and Radović et al., (2020). The instructional design model for interactive learning of mathematics consists of providing a learning environment encouraging and valuing the following: (1) experimentation with ideas and receiving feedback, (2) making students' mathematical thinking and understanding of problem situations public, (3) giving students autonomy for making and testing conjectures and playful exploration of mathematics ideas, (4) allowing direct manipulation of objects and linking theory with applications, and (5) linking the symbolic to the visual. iLMT is based on the regular teaching material (text book) used by students, coupled with main advantages of GeoGebra applets and interactivity, which should empower students to explore mathematical concepts and foster use of different types of reasoning (Lampert, 1988). The applets are used as a supporting tool in the process of developing mathematical ideas and reasoning. Moreover, with creative feedback, GeoGebra helps



students to visualize their ideas and supports them in developing understanding of mathematical properties and in algebraic interpretation. This kind of visual creative feedback allows students to interpret their conclusions with respect to cognitive models and reasoning they develop.

The content of iLMT is in line with the national curriculum (Radović et al., 2020). Special attention in creation of iLMT was also paid to recommendations given in Lithner's extensive research about mathematical reasoning (Lithner, 2000, 2004, 2008). Creation of materials is supervised and checked by experts in the field of methodology of teaching and education technology from the Faculty of Mathematics at the University of Belgrade and elementary school teachers. After approval by experts, final versions of the instructional materials were presented in a digital format.

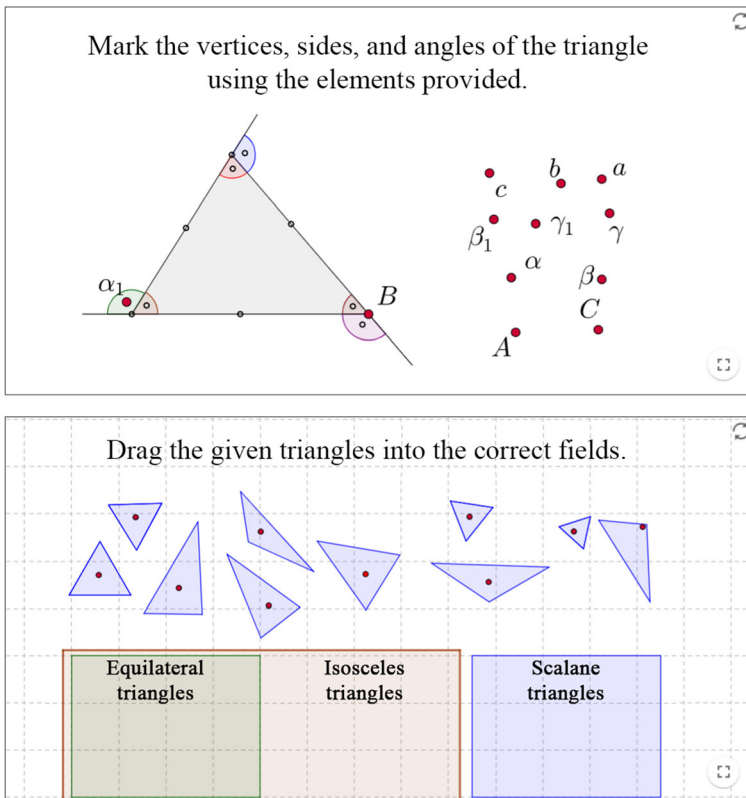
### The Content of iLMT

The units in iLMT are designed for students in elementary school, age twelve. Teaching materials cover triangle geometry, which involves the following teaching units: (1) the concept of a triangle, (2) elements and types of triangles based on side lengths, (3) the sum of the interior angles of triangles and triangle types based on angle size, (4) the sum of exterior angles and the ratio of sides and angles in the triangle, and (5) the triangle inequality. There were 5 lessons that introduced new mathematical concepts and 4 lessons that contained examples of and exercises for those concepts. Content of the learning materials in iLMT presents a digital and interactive version of the materials that was used by students in the control group. We have calculated that 17% of the iLMT tasks can be classified as HRA1, 30% as HRA2, 30% as LLR, and 23% of them as GLR.

Each iLMT learning unit consists of three sections: introductory, main, and final part. The introductory part of a lesson serves to remind students of essential concepts and relationships they learned in previous lessons. It is combined with GeoGebra applets to review previous material and to give motivation for the new content (Hohenwarter, Hohenwarter, & Lavicza, 2008). Various types of materials are used to address IR: students repeat definitions, annotate triangles, fill in the gaps in sentences, or make appropriate moves in applets. For example, in the material helping the acquisition of the definition of triangle, in the test group, students have the opportunity to fill in the gaps in the definition, choose the correct definition from a few options, answer questions about triangles, and mark correct statements about triangles. In all cases, students can repeat tasks as many times as they want, and have immediate feedback on whether a given answer is correct or not. There is also a "check" button in this type of task, which allows displaying the correct answer.

The introductory part of the second lesson also contains interactive tasks about triangle annotation where MR is required, like the ones in Fig. 1. Furthermore, in the first task in Fig. 1 (upper picture), students annotate a triangle by moving letters to their appropriate places (HRA1, i.e. MR required). If a student moves the letter to the appropriate place in the triangle, the letter fits perfectly and then blinks green. But if the student moves the letter to the wrong place, the letter is not matched. As this is practising essential concepts they learned in previous lessons, immediate feedback helps them visualize their activity and increase their knowledge after each mistake. In that manner, through assisting students in remembering how to annotate a triangle, MR





**Fig. 1** Interactive tasks in the introductory part of the lesson: *The sum of the interior angles and types of triangles depending on the size of sides*

should be triggered. Differently from the iLMT, in the regular school context, students do similar tasks on paper. Students try to solve the task from the textbook, and then, the teacher gives feedback on the blackboard.

In the second task (lower picture), students in the test group should classify triangles by their side lengths (MR and AR required). In this type of task, students have to know the classification of triangles by side length (HRA1, i.e. MR required) and to apply that definition (HRA2, i.e. AR required). They drag triangles to the appropriate group of equilateral, isosceles, and scalene triangles. If a student makes a mistake (triangle cannot fit into the group), it will reset and an additional theoretical hint will appear. Such hints remind students of the previously learned definition of different types of triangles, which contributes to MR development. Such feedback does not give simple clues for the next step or the final answer, but helps students to repeat the experiment and make a correct conclusion. During the entire time, the student has information about whether something is correct or incorrect. Students may try until they find the correct answer (which should foster AR through HRA2 task). As in the first task, in the regular school context, students solve the problem on paper and the teacher presents correct solutions at the blackboard.

In the main part of the lesson, the goal of the unit is achieved with the assistance of appropriate illustrations, texts, and interactive applets. These three types of learning aids help students to acquire new concepts, to reveal relationships that are presented in

the unit, and to develop different types of reasoning. For example, in one of the interactive applets, students can drag angles to one vertex and form the straight angle (Fig. 2). Students might reshape the triangle and repeat the experiment with different angle sizes, to reveal that the sum of interior angles in a triangle is constant. During a learning period, students can repeat the experiment as many times as they want. Here, students have to form a conclusion with the support of GeoGebra, and get creative feedback on every move from the applet. This approach aims to enhance the development of CMR (through LLR and GLR task), as students are faced with an unfamiliar situation where they have to discover a rule and make an abstract conclusion. While students using iLMT form conclusions about the sum of interior angles of a triangle by experimenting, in the regular school context, the teacher presents the theorem of the sum of interior angles of a triangle on the blackboard from the regular textbook.

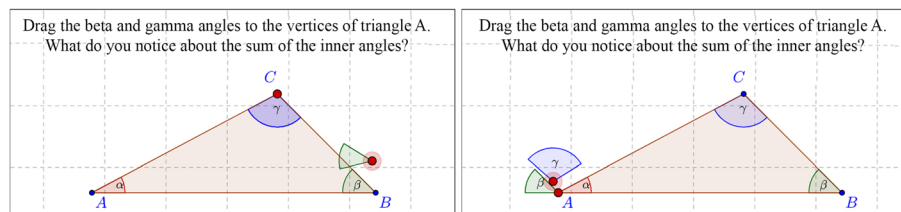
The final part of each iLMT unit includes exercises, quizzes, questionnaires, and games. The main objective of the final part of the lesson is to review what was just learned. Each teaching unit contains examples placed in an authentic context and adjusted to the age and background of the students.

Overall, the learning design (within introductory, main, and final parts of a lesson) in iLMT is student-oriented, which encourages students to explore materials, interact with GeoGebra applets, and enter answers in different forms. Although students are faced with multifaceted and complex mathematics properties, they are supported with corresponding mathematical visualization and creative feedback suitable for the development of CMR (Granberg & Olsson, 2015; Lithner, 2008). For example, the applet shown in Fig. 3 prompts students to move vertices and angles' sides to conjecture when it is possible to create a triangle depending on angle size and distances between points. Students can move vertices on the base and change angle sizes by using sliders or just by moving sides with a mouse. They can repeat the experiment as many times as they want but conceptual knowledge should be the outcome of their activities.

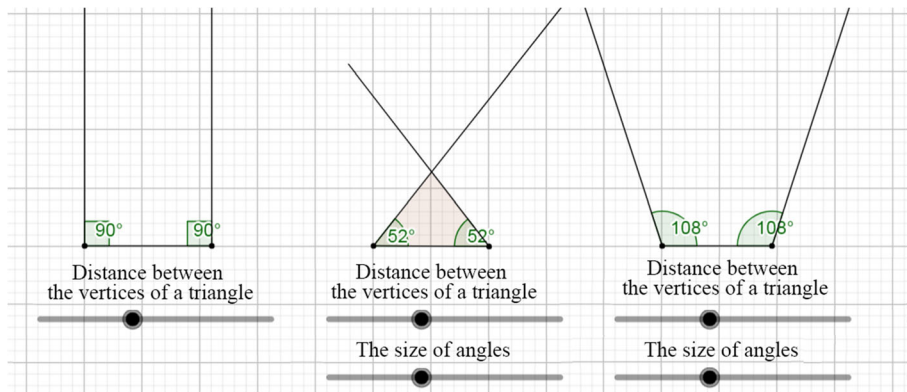
## Research Method

### Participants

Participants in this study were sixth-grade students. The study included 633 students (320 boys and 313 girls) and 13 teachers from 11 primary schools in Serbia. Nine schools participated with one teacher, and two schools participated with two teachers.



**Fig. 2** GeoGebra interactive applet in the main part of the lesson: *The sum of the interior angles and types of triangles depending on the size of angles*



**Fig. 3** GGIA that helps discover when it is possible to form a triangle depending on angle size and side length

Students were divided into two groups at random—test (320 students: 165 boys and 155 girls) and control (313 students: 155 boys and 158 girls) groups.

## Procedure

The experimental period lasted twelve school classes (period of three weeks), taking into account that the students had 4 classes of mathematics per week. Every school class lasted 45 minutes. In the test group, students during regular class used iLMT in a digital classroom where every student has access to a computer. iLMT materials were used during the entire time, where the teacher guided students through materials from the introduction to the final part of the lesson. Students were instructed to the iLMT materials, were supervised in their work by the teacher, and were given instructions when needed. Students in the control group learned about the topic using regular textbooks during the regular classroom setting, where teacher manages the class and directs students to work. Students in the control group were learning the same content (theorems, definitions, similar tasks, similar exercises, examples) as the test group, but without the assistance of interactive learning materials. In the control group, the teacher guided students through teaching materials given in regular textbooks. For example, when learning new content, students in the test group first do experiments using interactive applets, and then make conclusions together with the teacher. In the control group, the teacher presents the same content through discussion with students with appropriate explanations and proofs on the blackboard. Also, while exercising, students in the test group got creative feedback from the iLMT and could try solving tasks again using different strategies. However, students in the control group, after trying to solve a task, would get feedback from the teacher only in the form of a correct answer on the blackboard.

Before the experimental period, teachers had undergone training to ensure the same treatment toward students in all groups. Several meetings were held for teachers to present their suggestions for the modification of iLMT. Minor changes were made to the material to make it fit into their daily teaching practice and material preference. Teachers were given detailed instructions regarding the entire procedure of the study.

## Data Collection and Measuring Instruments

The data for this study were collected over the period of three months. Prior to the experiment, in order to measure students' prior knowledge, grades (learning marks) from the previous learning period were collected. Students received grades on a 1–5 interval scale (1 denotes insufficient, 2 sufficient, 3 good, 4 very good, and 5 excellent knowledge). During the first class after the experimental period, students from both groups solved problems on a KT (knowledge test). Two months after the experimental period, a RT (retention test) was conducted.

The RT and KT were created with the aim of covering all four types of tasks (HRA1, HRA2, LLR, GLR) and measuring students' knowledge and retention of learned material. All of the tasks given at KT and RT were categorized according to the presented theoretical framework. Formulation of tasks and relevance of task within RT and KT were constructed and checked by professors from the Faculty of Mathematics at University of Belgrade and experienced teachers of mathematics, as a result of joint discussions (Radović et al., 2020). Cronbach's alpha for both tests is equal to 0.71. Afterward, a pilot study was conducted, which involved 60 students who had used iLMT for two weeks prior to KT, and 15 out of the 60 were also interviewed. Final adjustments and minimal changes were made to the tests after the pilot study.

There were 9 tasks on the KT: one (11%) HRA1 task, 3 (33%) HRA2 tasks, 3 (33%) LLR tasks, and 2 (22%) GLR tasks. On the RT, there were 7 problems: one (14%) HRA1 task, 3 (43%) HRA2, 2 (29%) LLR, and 1 (14%) GLR task (as shown in Table 2). On the KT, students should have shown that they mastered the complete studied area, including both basic and more demanding learned concepts. On the other hand, on the RT, we expected from students to show basic knowledge and understanding of the concepts. Consequently, the RT contained fewer examples. These two tests do not measure the same level of students' knowledge, so their scores are not comparable. For each problem, after an analysis of students' open answers, we recorded students' solutions as a dichotomous variable denoted with 0—not solved and 1—solved. Student success was measured by their success rate on each task group, as well as for the tasks that require IR and CMR. Success rate was calculated as  $100 \times \text{Number Of Successes} / \text{Total Number Of Tasks}$ . Examples and number of tasks, their description, and possible values of success rate, on each of four types of tasks, on both the KT and RT are given in Table 2.

## Data Analysis

Pretest grades were summarized by mean and standard deviation (sd) for female and male students in the control and test groups. Effects of gender, group, and their interaction on grades were tested with Brunner nonparametric two-way ANOVA (Brunner, Konietzschke, Pauly, & Puri, 2016). Success rates on four types of tasks and on tasks that require IR and CMR, on both KT and RT, were graphically represented with a pyramid bar graph, for students in the control and test groups. Brunner nonparametric two-way ANOVA was used for testing the effects of group, gender, and their interaction on success rate while solving HRA1, HRA2, LLR, and GLR tasks, on KT and RT. Testing the effects of group, gender, reasoning type, and their interaction on success rates on KT and RT was done using Brunner-Langer

**Table 2** The description of two tests (KT and RT) with the number of tasks, task examples, task description, and possible values of success rates, for four task types

| Task group | KT            |  | RT  |
|------------|---------------|--|---|
| HRA1       | N (%)         | 1 task (11%)   | 1 task (14%)  |
|            | Task example: | <i>What is a triangle? Write a definition or your own explanation.</i>   | <i>How can we categorize triangles by their sides?</i>  |
|            | Description   | Students have the opportunity to learn the definition of a triangle and the triangle types in class. They could give an answer based on their knowledge or from memory without understanding of the definition.  |   |
|            | Success rate  | 0 or 100   | 0 or 100  |
| HRA2       | N (%)         | 3 tasks (33%)  | 3 tasks (43%)   |
|            | Task example: | Task example: <i>Two interior angles of a triangle are <math>\alpha = 32^\circ</math> and <math>\beta = 72^\circ</math>. Calculate the third angle and determine to which groups of triangles this triangle belongs based on the angles. Write down the procedure.</i>   | Task example: <i>Can you compose a triangle of three matchsticks with lengths of 1 cm, 3 cm, and 7 cm? Explain.</i> |
|            | Description   | Here, it was expected that students would give answers based on an algorithm they have seen in class. In both groups (test and control), an algorithm for solving this kind of task was presented.   |   |
|            | Success rate  | 0, 33, 67 or 100   | 0, 33, 67 or 100  |
| LLR        | N (%)         | 3 tasks (33%)  | 2 tasks (29%)   |
|            | Task example: | <i>Calculate measures of angles in a triangle ABC, where angle <math>\alpha = 44^\circ</math> and the angle in the intersection of bisectors of angles <math>\alpha</math> and <math>\beta</math> is <math>114^\circ</math>.</i>   | <i>The angle at the base of an isosceles triangle is <math>35^\circ</math>. What is its vertex angle?</i>           |
|            | Description   | The solution of this task requires a combination of algorithms that were presented in class. Understanding of algorithms and relations among them is necessary. We expected from students to recall appropriate algorithms and make purposeful connections between them. |   |
|            | Success rate  | 0, 33, 67, or 100  | 0, 50, or 100   |

Table 2 (continued)

| Task group    | KT   |               | RT           |
|---------------|--|---------------|--------------|
| GLR           | N (%)  | 2 tasks (22%) | 1 task (14%) |
| Task example: | <p><i>On a school break, Mark has to go to the bakery or the book store. He knows that the angle between the school, bakery and book store is 56°, and the angle between the bakery, book store and school is 58°. Which is closer to Mark, bakery or book store, and why? Draw a picture and write down an explanation.</i></p> <p><i>In the match with Nadal, Novak scored his ace points twice with his service. Both times, Novak served from the right corner. In the first service, he hit Nadal's right-hand corner and in the second from the same position, Nadal's left-hand corner. Both times the serve was at the same speed.</i></p> <p><i>a) When and why did the ball travel a longer distance?</i></p> <p><i>b) When did the ball travel longer? Why? (Explain using knowledge about triangles)</i></p> |               |              |
| Description   | For this type of task, students have not seen similar problems in their materials. They had to understand the requirements of the task and relate them with mathematical concepts and algorithms they knew. They had to create their own solutions without relying on similar examples.  |               |              |
| Success rate  | 0, 50, or 100  |               |              |

IR and CMR scores were calculated as a sum of scores on HRA1 and HRA2, and on LLR and GLR task types, respectively. Then, success rates on tasks that require IR and CMR were calculated as described above. Possible values of success rates on IR tasks were 0, 25, 50, 75, and 100 on both tests. Possible values of success rates on CMR tasks were 0, 20, 40, 60, 80, and 100 on the KT and 0, 33, 67, and 100 on the RT

nonparametric mixed ANOVA (Brunner, Domhof, & Langer, 2002). ANOVA-type test statistic was used for testing the effects of within-subject factors (reasoning type) and ANOVA-type statistic with modified box approximation for testing the significant main effects and interactions involving only the between-subject factors (group and gender). Post hoc analysis was performed using the Brunner-Munzel test for independent samples (Brunner & Munzel, 2000) and the Munzel-Brunner exact rank test for paired samples (Munzel & Brunner, 2002).

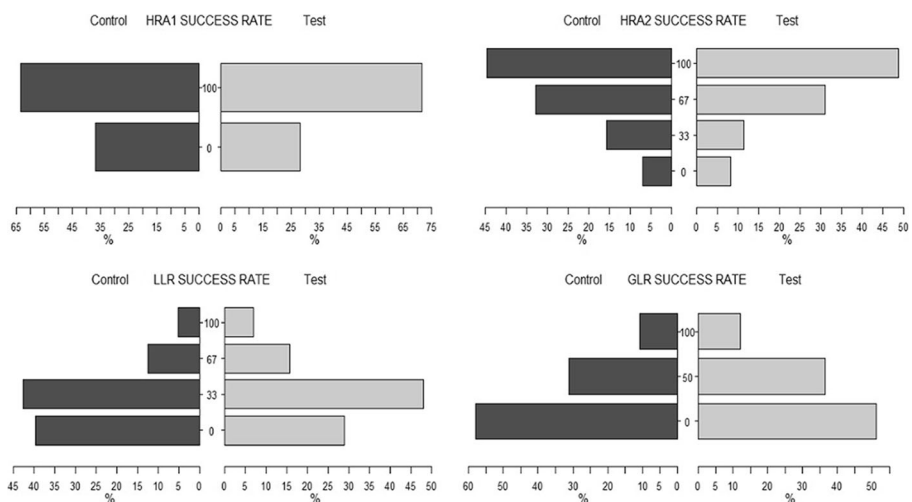
$P$  values  $< 0.05$  were considered statistically significant. Data analysis was performed in statistical software R, version 3.5.1 (using R packages *stats*, *nortest* (Gross & Ligges, 2015), *nparLD* (Noguchi, Gel, Brunner, & Konietzschke, 2012), *rankFD* (Konietzschke, Friedrich, Brunner, & Pauly, 2016), and *lawstat* (Gastwirth et al., 2017).

## Results

Students' pretest knowledge was measured with grades they had received before the experimental period started. Mean (sd) grade in the test group was 3.5 (1.2) and 3.4 (1.2) in the control group. Mean (sd) grade for girls was 3.6 (1.2) and 3.3 (1.2) for boys. The effect of gender on students' grades was statistically significant ( $F = 18.224$ ,  $df_1 = 1$ ,  $df_2 = 625.18$ ,  $p < 0.001$ ). Mean rank of grades of female students was 346.3, and 288.3 for male students, indicating higher grades for female than male students. Success rates of students in both control and test groups, while solving HRA1, HRA2, LLR, and GLR tasks on KT, are graphically represented in Fig. 4.

Statistically significant effects on success rates on these task types for both tests and mean ranks of groups are given in Table 3.

On the KT, by looking at the significant differences between success rates and mean ranks of the groups, we can conclude that students in the test group were more



**Fig. 4** Success rates of students in the control and test groups while solving HRA1, HRA2, LLR, and GLR tasks on KT



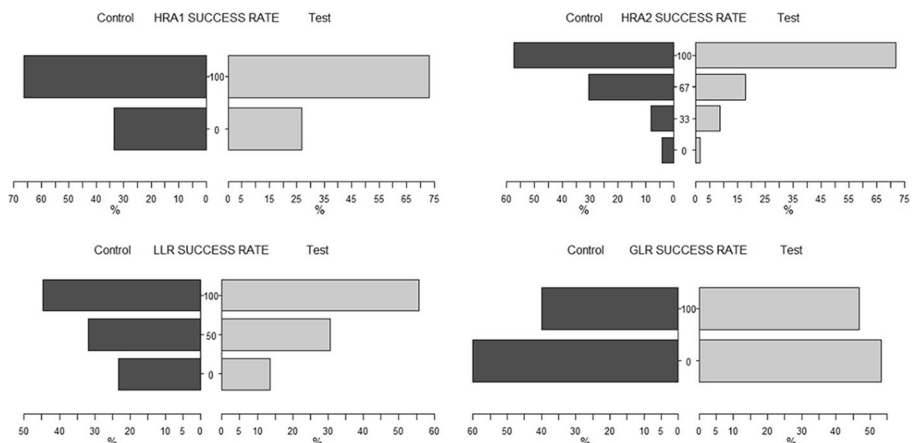
**Table 3** Statistically significant effects on students' success rates on HRA1, HRA2, LLR, and GLR tasks

| Task group | Test | Sign. effect | ANOVA test                               | Mean rank            |
|------------|------|--------------|--|----------------------|
| HRA1       | KT   | Group        | $F=4.961, df_1=1, df_2=607.08, p=0.026$  | C: 295.4<br>T: 321.3 |
| HRA2       | KT   | Gender       | $F=6.656, df_1=1, df_2=608.22, p=0.010$  | M: 292.0<br>F: 325.8 |
| LLR        | KT   | Group        | $F=7.540, df_1=1, df_2=600.91, p=0.006$  | C: 289.5<br>T: 326.2 |
| HRA1       | RT   | Gender       | $F=7.840, df_1=1, df_2=593.46, p=0.005$  | M: 289.8<br>F: 320.9 |
| HRA2       | RT   | Group        | $F=13.416, df_1=1, df_2=588.98, p<0.001$ | C: 283.1<br>T: 325.6 |
|            |      | Gender       | $F=22.765, df_1=1, df_2=588.98, p<0.001$ | M: 278.0<br>F: 333.1 |
| LLR        | RT   | Group        | $F=10.764, df_1=1, df_2=593.66, p=0.001$ | C: 283.4<br>T: 325.3 |
|            |      | Gender       | $F=8.977, df_1=1, df_2=593.66, p=0.003$  | M: 286.4<br>F: 324.4 |

C control group, T test group, M male, F female

successful in solving HRA1 and LLR tasks than students in the control group, as well as that girls were more successful in solving HRA2 tasks than boys.

Success rate of students in the control and test groups, while solving HRA1, HRA2, LLR, and GLR tasks on RT, is graphically represented in Fig. 5. On the RT, students in the test group were more successful in solving HRA2 and LLR tasks than students in the control group, and girls were more successful in solving HRA1, HRA2, and LLR tasks than boys. On the KT or RT, there was no significant interaction effect



**Fig. 5** Success rates of students in the control and test groups while solving HRA1, HRA2, LLR, and GLR tasks on RT

of group and gender on students' success rates. There were no statistically significant effects on students' success on GLR tasks on either test.

### Reasoning Type

Statistically significant results and mean ranks on tasks that require IR and CMR on the KT and RT are given in Table 4. On both tests, students were more successful in solving tasks that require IR than CMR, and students in the test group were more successful than students in the control group. On RT, girls were more successful than boys in solving both IR and CMR tasks.

### Analysis of Gender Effect

As it was stated, female students had higher pretest grades. To distinguish between the effect of mathematics knowledge (measured by grades) and real gender differences, significant gender effect was further analysed. We redid Brunner-Langer nonparametric mixed ANOVA, by including the grade as one of the factors. Procedure and results are described in more detail in the Supplement.

**Task Types.** On the KT, effect of gender on students' success rate on HRA2 tasks is not statistically significant (see Table 5 in the Supplement). On RT, gender has a statistically significant effect on success rate on HRA2, but not on success rates on HRA1 and LLR task types. Girls solved HRA2 tasks on the RT more successfully than boys.

**Success Rates on Tasks That Require IR and CMR on RT.** On RT, both boys and girls were more successful in solving tasks that require IR than CMR, and girls were more successful than boys when solving both IR and CMR tasks.

**Table 4** Statistically significant effects on students' success rates on tasks that require IR and CMR

| Test | Sign. effect   | ANOVA test                                 | Mean rank               |
|------|----------------|--|-------------------------|
| KT   | Reasoning type | $F=1363.875, df_1=1, df_2=\infty, p<0.001$ | IR: 828.0<br>CMR: 403.7 |
|      | Group          | $F=5.056, df_1=1, df_2=625.07, p=0.025$    | C: 593.8<br>T: 637.9    |
| RT   | Reasoning type | $F=294.471, df_1=1, df_2=\infty, p<0.001$  | IR: 723.9<br>CMR: 495.1 |
|      | Group          | $F=12.692, df_1=1, df_2=622.17, p<0.001$   | C: 570.3<br>T: 646.3    |
|      | Gender         | $F=16.154, df=1, df_2=622.17, p<0.001$     | M: 567.8<br>F: 653.0    |

C control group, T test group, M male, F female

## Discussion

### Groups of Tasks

**Knowledge Test.** The results of this study show that students in the test group outperformed students in the control group in solving HRA1 and LLR. The use of iLMT with interactive text fields, the option to repeat a task multiple times, availability of correct answers, and various types of applets, influenced the development of cognition needed for solving HRA1. This indicates that the usage of GeoGebra as part of iLMT has potential for development of IR.

Higher success of students in the test group in solving LLR is another positive effect of iLMT. LLR presents part of CMR, which is an important factor for deeper understanding of concepts in mathematics and their application in real-life situations (Bergqvist & Lithner, 2012; Jonassen, 2000). Our result is in line with earlier studies that discuss positive effects of GeoGebra on CMR (Granberg & Olsson, 2015; Olsson, 2017, 2018). We consider the main properties of iLMT that contribute to the development of LLR to visualize students' thinking, students' autonomy, the possibility of repetition of experiments, and immediate creative feedback that does not give direct answer but encourages students to explore and make conclusions (Granberg & Olsson, 2015).

Despite the existence of high level of interactivity, immediate feedback, and advantage of GeoGebra applets, a positive effect of iLMT on solving GLR tasks is still lacking. The issue may be in the small number of tasks that support GLR in learning materials (both textbook and iLMT). Regardless of interactivity and fast feedback, more time should be dedicated to GLR tasks. Our results can be relevant to educators and more broadly to educational policymakers because results indicate that material in school textbooks does not sufficiently foster students' ability to solve GLR tasks. Our findings are in line with the results from the PISA 2018 assessment, where only 5% of the students in Serbia (OECD average 11%) could model complex situations mathematically and apply higher-level reasoning (Organization for Economic Cooperation and Development [OECD], 2019).

Many countries support introducing digital and interactive materials in learning mathematics (Pepin, Choppin, Ruthven, & Sinclair, 2017). Furthermore, educational regulation in Serbia requires that each textbook must contain an electronic supplement as additional learning material. However, to the extent to our knowledge, there is a lack of evaluation of their effectiveness. Our results indicate that even when learning materials are digitalized with many technological advantages, GLR reasoning may still be lacking. Having in mind the importance of GLR in interactive learning material development and future research, special attention should be given to the improvement of iLMT regarding GLR tasks.

**Retention Test.** The results showed that the students in the test group were more successful in solving HRA2 and LLR tasks than students in the control group. There was no statistical difference in solving HRA2 on the KT, but there was on the RT. This indicates that iLMT has a positive influence on students' knowledge durability of learned algorithms. Better success of the students in

the test group on both the KT and RT in solving LLR also indicates that iLMT is a suitable environment for training students for solving tasks that require LLR. Such results point out that students who used iLMT had deeper insight into algorithms, which allowed for combining them in LLR tasks. This is in line with other studies that support usage of GeoGebra in tasks related to CMR (Granberg & Olsson, 2015; Olsson, 2017). Another advantage of iLMT is student-oriented teaching. While students' active role in the process of learning mathematical content in the control group was incidental, it was a widespread practice in the test group. Despite advantages provided by technology, as in the KT, there were no differences in students' success on the RT between the test and control groups in GLR tasks. Such a result indicates that even carefully created digitalized materials may not make a significant effect on students' success and knowledge durability in solving GLR tasks.

### Findings Related to Reasoning Types

On both the KT and RT, reasoning scores of students in the test group were significantly higher than reasoning scores of students in the control group, and IR scores were higher than CMR, which is in line with other similar research (Bergqvist, 2007; Boesen et al., 2010; Lithner, 2000). Higher success of students in the test group testifies to the positive effect of iLMT. On the other hand, better success on tasks that require IR than CMR indicates the need for improving iLMT. Considering that the iLMT content is similar to that in the official mathematics textbooks in Serbia, these results indicate the need for fostering CMR content in both textbooks and iLMT. Our findings are in line with PISA results, which show that about 40% of the students in Serbia can only use their mathematical knowledge and skills in a familiar context with information given explicitly when applying routine procedures (OECD, 2019). Development of CMR requires that students deal with unfamiliar problems (Hiebert & Grouws, 2007), with the help of supportive creative feedback (Granberg & Olsson, 2015). The main change in iLMT, according to literature (Boesen et al., 2014; Hiebert & Grouws, 2007; Jonassen, 2000), can be in the ratio of IR and CMR content, where advantage will be given to CMR materials.

### Gender Differences

The study found that female students solved HRA2 tasks on the RT more successfully than male students. This result is in line with Sumpter (2015) who stated that girls tend to use familiar strategies. Also, girls are considered diligent and hardworking and more prone to use strategies their teacher had showed them (Brandell & Staberg, 2008).

Furthermore, on the RT, the success rate of female students on IR and CMR was higher than that of male students. On the other hand, there is vast research that testifies the advantage of males in solving mathematical tasks (Brandell, Leder, & Nyström, 2007; Brandell & Staberg, 2008). In line with literature, there are many factors that can have an impact on gender differences such as learners' attitudes, motivation, stereotype threat in mathematics tests, differences in socialization, and the impact of socio-economic variables. Having in mind the delicate area of gender differences with all those

conditions, it is expected that the change of learning environment can have an important role, which could be the case in applying iLMT. More comprehensive investigation is required to make a reliable assertion about gender differences.

## Conclusion and Recommendation for Further Studies

In this study, we explored the influence of the iLMT on students' success in solving four groups of tasks: HRA1, HRA2, LLR, and GLR. Considering that iLMT presents a digital, student-oriented, and interactive version of textbook materials used in the standard environment (i.e. control group), this study provides a new and different contribution from similar previous studies that tested digital materials specifically created to foster a certain type of reasoning.

Taken altogether, the findings of this study suggest one major contribution. Based on the discussed results, iLMT had a positive impact on students' success on both the KT and RT to certain extent. Moreover, it represents suitable instructional material for teaching students for solving tasks that require different types of reasoning, with potential for improvement. Other scholars could use our approach and design principles, as well as our results to design even more effective, efficient, and enjoyable learning environments to teach mathematical concepts.

Use of digital learning materials in teaching mathematics is very delicate, and it must be done carefully and with special attention to the kind of mathematical reasoning it is designed for. However, we emphasize the great potential of GeoGebra applets, especially for improving students' success in solving tasks that require CMR. We emphasize that the mere digitization of content will not bring significant effects on student success in solving all types of tasks. Our study showed that even carefully created digitized materials may not have significant effects on students' success in the solving GLR tasks, if they are based on content that is insufficiently related to GLR. In addition to the many benefits of GeoGebra, we believe that more content in the textbook and interactive materials must be dedicated to the GLR tasks. Keeping in mind the wide usage of GeoGebra applets and rapid digitization of materials, we believe that these results can be useful to many educators. These results can be important primarily for scholars and teachers interested in improving critical thinking, abstract understanding, and facilitating connections between authentic learning and mathematical theory. However, we call for further studies to research connections between these aspects of mathematical understanding and interactive learning environments.

In terms of limitations, our research has been focused on teaching units about triangles, while other topics were not considered. However, it is possible that the significant effect of interactivity on the learning was due to the geometry topic. It would be interesting to witness the effect of iLMT (or effects of implemented characteristics of the instructional design) on different mathematics topics (for example algebra) or different student age. We also note that a short-term intervention could cause the halo effect of novelty (Radović et al., 2020). The question remains of what would happen if students used iLMT or similar materials over a longer time period.

Our research also indicates the greater success of girls in some cases. Further research may explore the fact that girls outperformed boys in solving HRA2 and to bring more light on the result that girls outperformed boys on the RT in solving tasks that require IR or CMR. We have not been able to provide more clarity to these interesting results.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10763-021-10151-8>.

**Funding** This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, by Project No. III-47003.

## Compliance with Ethical Standards

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

- Arzarello, F., Olivero, F., Paola, D., & Robutti, O. (2002). A cognitive analysis of dragging practises in Cabri environments. *Zentralblatt für Didaktik der Mathematik*, 34(3), 66–72.
- Baccaglini-Frank, A., & Mariotti, M. A. (2010). Generating conjectures in dynamic geometry: The maintaining dragging model. *International Journal of Computers for Mathematical Learning*, 15(3), 225–253.
- Bergqvist, E. (2007). Types of reasoning required in university exams in mathematics. *The Journal of Mathematical Behavior*, 26(4), 348–370.
- Bergqvist, T., & Lithner, J. (2012). Mathematical reasoning in teachers' presentations. *The Journal of Mathematical Behavior*, 31(2), 252–269.
- Boesen, J., Helenius, O., Lithner, J., Bergqvist, E., Bergqvist, T., & Palm, T. (2014). Developing mathematical competence: From the intended to the enacted curriculum. *Journal of Mathematical Behavior*, 33, 72–87.
- Boesen, J., Lithner, J., & Palm, T. (2006). *The relation between test task requirements and the reasoning used by students*. Department of Mathematics and Mathematical Statistics, Umeå universitet.
- 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.
- Brandell, G., Leder, G., & Nyström, P. (2007). Gender and mathematics: Recent development from a Swedish perspective. *ZDM*, 39(3), 235–250.
- Brandell, G., & Staberg, E.-M. (2008). Mathematics. A female, male or gender neutral domain? A study of attitudes among students at secondary level. *Gender and Education*, 20(5), 495–509.
- Brunner, E., Domhof, S., & Langer, F. (2002). *Nonparametric analysis of longitudinal data in factorial experiments*. New York: John Wiley & Sons Google Scholar.
- Brunner, E., Konietzschke, F., Pauly, M., & Puri, M. L. (2016). Rank-based procedures in factorial designs: Hypotheses about nonparametric treatment effects. *Statistical Methodology (Series B)*, 79(5), 1463–1485.
- Brunner, E., & Munzel, U. (2000). The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal*, 42, 17–25.
- Chiu, T., & Churchill, D. (2016). Design of learning objects for concept learning: Effects of multimedia learning principles and an instructional approach. *Interactive Learning Environments*, 24(6), 1355–1370.
- De Witte, K., & Rogge, N. (2014). Does ICT matter for effectiveness and efficiency in mathematics education? *Computers & Education*, 75, 173–184.
- Freudenthal, H. (1987). *Mathematics starting and staying in reality*. In I. Wirsup & R. Street (Eds.), *Proceedings of the USCMP Conference on Mathematics Education on Development in School Mathematics around the World*. NCTM, Reston, VA.
- Gastwirth, J. L., Gel, Y. R., Wallace, W. L., Lyubchich, V., Miao, W., & Noguchi, K. (2017). lawstat: Tools for biostatistics, public policy, and law. *R package version 3.1*.
- Granberg, C., & Olsson, J. (2015). ICT-supported problem solving and collaborative creative reasoning: Exploring linear functions using dynamic mathematics software. *The Journal of Mathematical Behavior*, 37, 48–62.

- Gross, J., & Ligges, U. (2015). nortest: Tests for normality. R package version 1.0–3.
- Hiebert, J. (2003). What research says about the NCTM standards. In J. Kilpatrick, G. Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. *Second handbook of research on mathematics teaching and learning*, 1, 371–404.
- Hohenwarter, J., Hohenwarter, M., & Lavicza, Z. (2008). Introducing dynamic mathematics software to secondary school teachers: The case of GeoGebra. *Journal of Computers in Mathematics and Science Teaching*, 28(2), 135–146.
- Jäder, J., Sidenvall, J., & Sumpter, L. (2017). Students' mathematical reasoning and beliefs in non-routine task solving. *International Journal of Science and Mathematics Education*, 15(4), 759–776. <https://doi.org/10.1007/s10763-016-9712-3>.
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85.
- Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science*, 39(4), 561–579.
- Konietschke, F., Friedrich, S., Brunner, E., & Pauly, M. (2016). rankFD: Rank-based tests for general factorial designs. R package version 0.0.1.
- Lampert, M. (1988). What can research on teacher education tell us about improving quality in mathematics education? *Teaching and Teacher Education*, 4(2), 157–170.
- Leung, A. (2008). Dragging in a dynamic geometry environment through the lens of variation. *International Journal of Computers for Mathematical Learning*, 13(2), 135–157.
- Lithner, J. (1998). Mathematical reasoning and familiar procedures. *International journal in science and technologies*, 1, 83–95.
- Lithner, J. (2000). Mathematical reasoning in task solving. *Educational Studies in Mathematics*, 41, 165–190.
- Lithner, J. (2004). Mathematical reasoning in calculus textbook exercises. *Journal of Mathematical Behavior*, 23(4), 405–427.
- Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics*, 67(3), 255–276.
- Lithner, J. (2015). Learning mathematics by creative or imitative reasoning. In *Selected Regular Lectures from the 12th International Congress on Mathematical Education* (pp. 487–506). Cham: Springer.
- Mincu, M. E. (2009). Myth, rhetoric and ideology in eastern European education. *European Education*, 41(1), 55–78.
- Munzel, U., & Brunner, E. (2002). An exact paired rank test. *Biometrical Journal*, 5, 584–593.
- National Council of Teachers of Mathematics. (2014). *Principles to actions*. Reston, VA: National Council of Teachers of Mathematics.
- Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, 50(12), 1–23.
- 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., Qwillbaird, T., & Holm, L. (2019). Investigating algorithmic and creative reasoning strategies by eye tracking. *The Journal of Mathematical Behavior*, 55, 100701.
- Organization for Economic Cooperation and Development. (2019). *PISA 2018 results (volume I): What students know and can do, PISA*. Paris: OECD Publishing. <https://doi.org/10.1787/5f07c754-en>.
- Olsson, J. (2017). The contribution of reasoning to the utilization of feedback from software when solving mathematical problems. *International Journal of Science and Mathematics Education*, 1–21.
- Olsson, J. (2018). Relations between task design and students' utilization of GeoGebra. *Digital Experiences in Mathematics Education*, 1–29.
- Palm, T., Boesen, J., & Lithner, J. (2005). The requirements of mathematical reasoning in upper secondary level assessments. *Research Reports in Mathematics Education, Dept. of Mathematics, Umeå Universitv*.
- Pepin, B., Choppin, J., Ruthven, K., & Sinclair, N. (2017). Digital curriculum resources in mathematics education: Foundations for change. *ZDM - Mathematics Education*, 49(5), 645–661. <https://doi.org/10.1007/s11858-017-0879-z>.
- Radišić, J., & Baucal, A. (2018). Teachers' reflection on PISA items and why they are so hard for students in Serbia. *European Journal of Psychology of Education*, 33(3), 445–466.
- Radović, S., Radojčić, M., Veljković, K., & Marić, M. (2020). Examining the effects of Geogebra applets on mathematics learning using interactive mathematics textbook. *Interactive Learning Environments*, 28(1), 32–49.



- Sumpter, L. (2015). Investigating upper secondary school teachers' conceptions: Is mathematical reasoning considered gendered? *International Journal of Science and Mathematics Education*. <https://doi.org/10.1007/s10763015-9634-5>.
- Takači, D., Stankov, G., & Milanovic, I. (2015). Efficiency of learning environment using GeoGebra when calculus contents are learned in collaborative groups. *Computers & Education*, 82, 421–431.

## Affiliations

**Marija Kaplar<sup>1</sup> · Slaviša Radović<sup>2</sup> · Kristina Veljković<sup>3</sup> · Ksenija Simić-Muller<sup>4</sup> · Miroslav Marić<sup>5</sup>**

Slaviša Radović  
slavisa.radovic@ou.nl

Kristina Veljković  
kristina.veljkovic@fri.uni-lj.si

Ksenija Simić-Muller  
simicmka@plu.edu

Miroslav Marić  
maricm@matf.bg.ac.rs

<sup>1</sup> Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, Novi Sad 21000, Serbia

<sup>2</sup> Faculty of Educational Sciences, Open University of the Netherlands, Valkenburgerweg 177, 6419 AT Heerlen, the Netherlands

<sup>3</sup> Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113, Ljubljana 1000, Slovenia

<sup>4</sup> Mathematics Department, Pacific Lutheran University, 1010 122nd St S, Tacoma, WA 98447, USA

<sup>5</sup> Faculty of Mathematics, University of Belgrade, Studentski trg 16, Belgrade 11000, Serbia