




How do testing and test-potentiated learning versus worked example method affect medium- and long-term knowledge in abstract algebra for pre-service mathematics teachers?

Anna Muzsnay^{1,2}  · Csilla Zámbo^{2,3} · Janka Szeibert^{3,4} · László Bernáth^{2,5} · Brigitta Szilágyi^{2,6} · Csaba Szabó^{7,8}

Received: 9 January 2024 / Revised: 5 June 2024 / Accepted: 10 June 2024
© The Author(s) 2024

Abstract

The retention of foundational knowledge is crucial in learning and teaching mathematics. However, a significant part of university students do not achieve long-term knowledge and problem-solving skills. A possible tool to increase further retention is testing, the strategic use of retrieval to enhance memory. In this study, the effect of a special kind of testing versus worked examples was investigated in an authentic educational setting, in an algebra course for pre-service mathematics teachers. The potential benefits of using tests versus showing students worked examples at the end of each practice session during a semester were examined. According to the results, there was no difference between the effectiveness of the two methods in the medium term—on the midterm that students took on the 6th week and the final that students took on the 13th week of the semester, the testing group performed the same as the worked example group. However, testing was more beneficial regarding long-term retention in studying and solving problems in abstract mathematics. Analyzing the results of the post-test that students took five months after their final test, the authors found that the improvement of those students who learned the material with testing was significantly larger than that of the worked example group. These findings suggest that testing can have a meaningful effect on abstract algebra knowledge and a long-lasting impact on solving complex, abstract mathematical problems.

Keywords Abstract algebra · Retrieval practice · Testing effect · Test-potentiated learning · Complex material · Worked examples

Motivation for submitting our article to the journal.

The topic of this article—i.e., investigation of the application of retrieval practice in abstract algebra for pre-service teachers—can be viewed from several aspects: it is undergraduate mathematics education, if we focus even more: teacher training; it includes high-level abstract mathematics; it addresses educational psychological issues etc. While trying to grab the ideas and questions in their complexity, we chose to concentrate on the psychological point of view. We aimed to publish our results in a high-quality journal in order to gain visibility for our work among the international community of researchers in the topic and also to contribute actively to the common knowledge on these issues, that is why we considered choosing a high-impact journal of Springer Nature. Our decision to submit the article to the European Journal of Psychology of Education is based on consultation with a Submission Editorial Advisor who suggested this journal as one of the best fits.

Extended author information available on the last page of the article

Introduction

The retention of foundational knowledge is crucial in learning and teaching mathematics (Hopkins et al., 2016). Since mathematics teachers are required to explain their understanding of mathematical concepts and procedures to students during instruction, mathematics teachers need to have a well-connected deep understanding of fundamental mathematics (Ma, 2020). Advanced mathematics, like abstract algebra, serves to deepen and confirm more rigorously the specific mathematical ideas secondary teachers will teach (Wasserman, 2018). Pre-service mathematics teachers, therefore, have to be taught in a way that promotes understanding and long-term retention of mathematical concepts and the relationship between them. In other words, in university mathematics courses, the learning strategies applied should prevent forgetting and enhance understanding and the creation of long-term knowledge. Retrieval practice (the strategic use of retrieval to enhance memory) and worked examples (showing students a step-by-step demonstration of how to solve a problem) are both recommended as effective methods for improving learning (Adeniji & Baker, 2023; Dunlosky et al., 2013). However, they are based on different underlying cognitive processes. In this article, the effectiveness of worked examples and a specific type of retrieval practice were compared in an abstract algebra course concentrating on medium- and long-term retention. We conducted an experiment for second-semester pre-service mathematics teachers in the context of the “Algebra and Number Theory 2” course (ANT 2) in the 2017/18 spring term. During the semester, we divided the class into two groups of sizes 39 and 37. One of the groups studied the material using retrieval practice, while the other one studied with worked examples. They took two tests on the material they learned on the topic of polynomials: one on the 6th week of the semester and another one at the end of the semester. This way, we measured what students could retrieve in the medium-term. To measure the long-term effects of these two methods, students took a post-test five months after their final test, on the topic of polynomials. This test was taken by all the students who completed the course “Algebra and Number Theory 2” and who had enrolled for the continuation of the course, Algebra and Number Theory 3 (ANT 3). In this paper, we analyze the results of these three tests.

Literature review

Testing and test-potentiated learning

Unless the learned material is consciously reviewed, people forget about half of the newly gained knowledge within days or weeks (Averell & Heathcote, 2011; Ebbinghaus, 1913; Murre & Dros, 2015). Testing works against forgetting. Several studies showed that knowledge acquired by testing is more resistant to interference effects and leads to a lower forgetting rate (Kliegl & Bäuml, 2016; Racsomány & Keresztes, 2015; Szpunar et al., 2008). In addition, it has a promoting effect on learning: it produces better organization of the acquired knowledge, enhances its transfer to new contexts, and produces faster access to learned information (Chan et al., 2018; Jacoby et al., 2010; Racsomány et al., 2018; Zaromb & Roediger, 2010). Testing, also known as retrieval practice, is a learning technique that involves recalling to-be-remembered information from memory. It can refer to any activity (such as questions during class, quizzes, flashcards, brain dump, and examination questions) that requires retrieving information from memory without the help of any external

sources. Although it was more than 100 years ago that testing was first studied (Abbott, 1909), it is only in the last 20 years that the testing effect—the fact that active retrieval produces better retention than passive rereading—has become the focus of research (Rowland, 2014). The direct effects of retrieval can be seen when students study a set of materials and then practice retrieval without restudying or receiving feedback after retrieval (Karpicke, 2017). Any gains in learning from practicing retrieval, without restudy or feedback, represent the direct effects of retrieval processes on learning.

Also, testing can have an indirect effect on learning: unsuccessful retrieval attempts can enhance the effectiveness of subsequent restudy. The attempts to retrieve information from memory may improve later restudy items even in conditions when the retrieval attempt fails and feedback is not given (Arnold & McDermott, 2013; Grimaldi & Karpicke, 2012; Izawa, 1966; Wissman & Rawson, 2018). This effect is known as test-potentiated learning. Since in an authentic mathematics education context, it is difficult to separate (and we do not necessarily aim to distinguish) the indirect and direct effects of testing, we did not separate the two in this study but measured direct and indirect effects together.

Testing and test-potentiated learning have proved to be successful in several learning environments, such as text memorizing, foreign language vocabulary, general knowledge facts, learning materials that include visual or spatial information, and skill learning (Rowland, 2014) and have been demonstrated as powerful tools for improving learning not only in laboratory circumstances but in real educational environments as well (Dunlosky et al., 2013; McDermott et al., 2014; Roediger et al., 2011). However, more applied research is needed in mathematics learning (Agarwal et al., 2021). The literature on the application of retrieval practice in real mathematics educational environments, in particular in learning higher mathematics, is limited. Since several factors can affect retention, such as context, subject, and material complexity, it is legitimate to ask whether testing is an effective tool for learning higher mathematics.

Testing and test-potentiated learning when learning complex materials

Some researchers argue that retrieval practice is not beneficial for “complex” materials (e.g., Van Gog & Sweller, 2015). They suggest that the testing effect decreases or disappears as the complexity of learning material increases. By “complex material,” they mean “high in element interactivity, containing various information elements that are related and must therefore be processed simultaneously in working memory” (Van Gog & Sweller, 2015, p. 248). Also, the research of Van Gog and Kester (2012) points to the direction that in the field of science, the testing effect might not work in acquiring problem-solving skills from worked examples. On the other hand, a wealth of studies, such as the research of Butler (2010), Chan (2010), and McDaniel et al. (2009), emphasize the benefits of retrieval practice for learning complex materials. Although only a few studies investigated the testing effect on mathematical problem-solving in a classroom environment, recent studies suggest that increasing retrieval practice may be an effective way of learning (Fazio, 2019; Lyle & Crawford, 2011; Lyle et al., 2016, 2020; Szeibert et al., 2022). The study of Lyle and Crawford (2011) was carried out in a statistics for psychology course, where students had to answer a small set of questions at the end of each lecture to retrieve information learned the same day. This method significantly and substantially increased exam scores. Lyle et al. (2020) investigated spaced versus massed retrieval practice in a precalculus course for engineering students and their impact on long-term retention by independently manipulating the amount and the spacing of retrieval practice. The long-term retention of

the students on the material learned in the precalculus course was assessed one month after the end of the course. On this post-test, the memory was significantly better for precalculus objectives that had been the target of spaced quizzing versus massed quizzing. Also, increasing the number of quiz questions did not significantly affect retention. In similar research by Lyle et al. (2016), across-semester retention was measured less directly. In this study, researchers came to the same conclusion as in the abovementioned experiment. Finally, the study of May (2021) investigated spaced-retrieval practice in a mathematics course for second-year pre-service mathematics teachers. The topics of the course were analytical geometry, functions, remainder and factor theorems, Euclidean geometry, and matrices. He found that the intervention enhanced retention in those categories of knowledge and reasoning that were relatively similar to what was presented in class and at the same time, it required imitative reasoning and well-established procedural knowledge. In categories where flexible and creative use of conceptual and procedural knowledge was required, students' performance was less impressive. The results of May's study show that repeated and spaced retrieval was effective in enhancing near transfer but was not that effective for far transfer.

Worked examples

The effect of testing when learning complex material is not obvious. Numerous studies report that when learning flexible procedures, retrieval practice is no more effective than repeated studying (Leahy et al., 2015; Van Gog & Kester, 2012; Van Gog & Sweller, 2015; Van Gog et al., 2015). Moreover, it can even be suboptimal: some researchers found that in the case of math and science problem-solving, using worked examples instead of problem-solving tasks was more beneficial (Atkinson et al., 2000; Renkl, 2014; Van Gog & Rumel, 2010). By worked examples, we mean a step-by-step demonstration of how to perform a task or how to solve a problem (Clark et al., 2006). In terms of problem-solving, studying carefully selected worked examples can often be a more effective way of learning than solving further problems. This worked example advantage is often called the worked example effect. The effectiveness of using worked examples is usually explained by the cognitive load theory. It helps in reducing extraneous cognitive load and in increasing germane cognitive load in the initial phase of learning. In other words, it frees up working memory resources so that the learner can pay more attention to constructing and automating schemas (Kalyuga et al., 2010; Sweller, 1988, 2010). Typically, it is beneficial for novice learners, and it becomes less effective as expertise increases (Kalyuga et al., 2010).

Testing versus worked examples

Testing and studying worked examples can be both effective learning methods in the field of problem-solving. However, which one is more effective in the medium and long term when learning to solve abstract mathematical problems is still a question. Regarding mathematics education, one of the most relevant works on the topic of comparing testing and worked examples is the work of Yeo and Fazio (2019). In their research, the efficacy of retrieval practice versus studying worked examples was examined on the topic of Poisson distribution. They argue that both testing and studying worked examples are required to effectively learn complex material. Their experiment included a learning phase and a test phase. The test phase occurred either five minutes or one week after the learning phase. During the one-session learning phase, one part of the

participants learned the material using retrieval practice, while the other part learned it by (re)studying worked examples. They found that the optimal learning strategy depends on various factors. Namely, it depends on the learning goal, the retention interval, and the kind of knowledge being learned (stable facts or flexible procedures). The learning processes involved (schema induction vs. memory and fluency building) are also a major factor. They argue that learning objectives determine when it is better to test or study. When the learning goal was to remember stable facts in the text of a worked example, repeated testing resulted in higher recall performance than repeated studying one week after the intervention. When the learning goal was to learn a novel math procedure, the optimal learning strategy depended on the retention interval (five-minute or one-week delay) and the nature of the material (identical or nonidentical learning problems). Five minutes after the learning phase, repeated studying was more effective than repeated testing both with identical and nonidentical problems. One week after the learning phase, repeated testing was as effective as repeated studying with nonidentical learning problems and it was more effective than repeated studying with identical learning problems. Their research findings suggest that the testing effect can occur for flexible procedures as well.

Research focus and research questions

The “[Literature review](#)” section focused on studies that have not only investigated the retrieval practice in real educational environments but explicitly studied it within the subject of mathematics. In mathematics, problems often require developed problem-solving skills—the application of procedures and deep conceptual understanding are necessary. Also, mathematical problems themselves are usually complex. The present research was carried out in an abstract algebra course. We investigated prospective mathematics teachers’ retention of ANT 2 problem-solving skills. University algebra courses play a huge role in improving students’ abstraction ability, making connections within different parts of mathematics and science, and developing a deeper understanding of concepts, such as formal calculations, introducing algebraic structures, and understanding their behaviors. By the level of abstraction, we mean the degree of connection between pieces of information; it refers to the extent to which a unit of knowledge is bound to a given context (Hiebert & Lefevre, 1986). In other words, the level of abstraction increases as the knowledge moves away from a specific context. Even though abstract algebra is taught in university-level mathematics, it has numerous connections to primary and secondary school mathematics (Even, 2011; Wasserman, 2018.). Understanding concepts related to abstract algebra and uncovering its connections to primary and secondary school mathematics has a significant impact on mathematics teachers’ beliefs and their practices of teaching (Wasserman, 2014). We believe that understanding algebra at a higher level is essential for mathematics teachers. However, many studies showed that students have difficulties learning abstract algebra (Agustyaningrum et al., 2021; Veith et al., 2022; Wasserman, 2018). Even if students pass their algebra exam, they have trouble remembering the learning material later. The question arises as to whether continuous testing helps retain information in abstract algebra in the long term. It would be a big step forward for university mathematics education if it turned out that testing leverages to increase long-term retention in studying and solving problems in abstract mathematics. In this research, we aim to explore and gain a better understanding of how testing affects studying and solving problems in abstract algebra, if at all.

Research questions

In this study, we were curious about the medium-term (within-semester) effects and long-term (across-semester) effects of testing in solving complex problems in abstract mathematics.

The research questions were the following:

Do students who study abstract algebra in a testing way (described below) perform better on complex mathematical problems than those who study by worked examples (described below) on the midterm and the final test (in the medium term)?

Do students who study abstract algebra in a testing way (described below) perform better on complex mathematical problems than those who study by worked examples (described below) on the post-test five months after the final test (in the long term)?

To what extent do students forget abstract algebra in the long term—five months after the final test—studying it in a testing way (described below) versus studying by worked examples (described below)?

Methods

Sample and study design

The experiment consisted of two parts: in the first part, we investigated the medium-term effect of the interventions in an undergraduate abstract algebra course, while in the second part, we focused on the long-term effect.

The first part of the study was conducted in the spring semester of the academic year 2017/2018 at ELTE Eötvös Loránd University in the capital of Hungary. Participants in the study were $N=76$ prospective high-school teachers enrolled in the “Algebra and Number Theory 2” course (ANT 2), so they were students, who had already attended a semester at the university.

The lectures were presented by one of the researchers. He was an experienced university mathematics lecturer who taught this course many times. The course consisted of a 90-min lecture and a 90-min practice session and lasted for 13 weeks. The students attended the same lectures, while their practice sessions were taught in six groups of 15 students on average. All students learned the same material and solved the same types of problems during their practice sessions. Students were not allowed to miss more than two practice sessions; therefore, students attended almost all practice sessions.

If one wants to implement retrieval practice in the classroom, one needs to consider a few requirements: the first retrieval should take place within 24 h, and it must be an activity where students do not use any external help (for example, their books, notes...). Students should be involved in it and should not be exposed to a high level of stress. Also, it should not take a lot of time. Furthermore, the form of testing and the types of questions have to be properly chosen. Taking all these aspects into account, the design of the experiment was the following. There were two types of groups: a testing group and a worked example group. The testing group and the worked example group were selected from the different practice groups of the academic year 2017/18: half of the practice groups were chosen to be testing groups ($N=39$), and the other half of the practice groups were selected to

be worked example groups ($N=37$). Two practice groups were run by professors, two by younger members of the department, and two groups were taught by a Ph.D. student. One of each pair of groups was a testing group and the other one was a worked example group. This way, the teacher's effect was eliminated as much as possible. These groups did the same exercises during the practice sessions. However, there was a difference between the structures of the two types of groups' practice sessions in the last 5–10 min of each session.

At the end of the practice sessions, students from the testing group took a test on the material of the given practice session without any aid, or any external help. These end-of-class tests consisted of two open-ended questions related to the topic of the given lesson. This way, they had to retrieve immediately what they had just studied. These tests were corrected (in a simple way: correct or incorrect solution) and evaluated (0–1–2 points). Students did not get the evaluated tests back, and they did not get any feedback about how to solve the problems of the end-of-class tests unless they asked for it. During the semester, the corrected tests were shown to the students if they were curious about their end-of-class test results. However, this was a rather sporadic phenomenon. To make sure that students took these tests seriously, they needed to score 50 percent of the maximum points by the end of the semester to pass the course. It was not difficult to achieve 50 percent, all the students managed to achieve it. To give the reader an impression of these end-of-class tests, six questions are presented (see Table 1). We intended to ask questions that were neither too easy nor too hard for the students, they were of moderate difficulty. The solution of these tasks required the same knowledge content and ideas as the tasks of the midterm and final tests, but they were easier and not identical to them.

The same two problems that the testing group solved in the end-of-class tests were shown to the worked example group at the end of the practice sessions. However, in this group, the teacher solved these problems, not the students—in the worked example group, the solutions to these problems were presented step-by-step by the teacher. This way, students could concentrate on the solution pathway. Students' topic-related problem-solving skills were measured two times during the semester: they took a midterm and a final test on the 6th and the 13th week of the semester, in March and May 2018.

The second part of the study was carried out in the fall semester of the academic year 2018/2019 at ELTE Eötvös Loránd University in the capital of Hungary during the course ANT 3. In the teacher training program, ANT 3 can be taken in either in the third or the fifth semester. Participants of the study were those pre-service teachers who took the course ANT 3 in their third semester and participated in the first part of the study. For example, those who did not pass ANT 2 or completed ANT 2 and had another class instead of ANT 3 did not participate in this part of the study. In this part of the study, we investigated the long-term effect of the two methods. We measured the long-term retention of ANT 2 material by a post-test. Students took the post-test on the topic of polynomials in October 2018, five

Table 1 Sample questions of the end-of-class tests

1	Express the following polynomial by elementary symmetric polynomials: $a^2bc + ab^2c + abc^2$
2	Calculate the cyclotomic polynomial $\phi_{68}(x)$
3	Factor the polynomial $x^5 - 1$ over \mathbb{Z}_{11}
4	Is the polynomial $6x^7 + 52x^5 - 39x^3 + 52$ irreducible over \mathbb{Q} ?
5	What are the roots of the polynomial $x^{20} + x^{15} + x^{10} + x^5 + 1$ over \mathbb{C} and what are the orders of them?
6	Determine whether the polynomial $2x^3 + x^2 - x + 1$ can be factorized over \mathbb{Z}_3

months after they took their final exam in ANT 2. Altogether, 33 students took this post-test, $N=13$ studied ANT 2 in a testing way (testing group), and there were $N=20$ students from the worked example group. Students had been informed that they would take a test on what they had learned so far; however, they expected to be asked only about materials learned in ANT 3, not ANT 2. Tests on announced topics can influence student motivation because knowing about an upcoming test often leads students to increase their study efforts (Roediger & Karpicke, 2006). In that case, the effect of retrieval on learning is mediated by enhancements in subsequent restudy, which is why we did not explicitly tell students about the upcoming post-test. All the participants agreed to take the post-test and not doing well in the post-test did not have any negative consequences.

The material covered by the study

The topics covered by the course *Algebra and Number Theory 2* can be found in the Appendix. Topics for the study were a subset of the ones presented in the course, namely, Polynomials over fields; Equalities, main coefficient, normal polynomial; Operations with polynomials (addition, subtraction, multiplication, Euclidean division), the zero polynomial; Degree of a polynomial; Roots and factorizations of polynomials, Horner's method; The maximum number of roots; Identity theorem of polynomials over an infinite field; The relation between polynomial functions and polynomials; Multiplication of roots, root factor form; Lagrange interpolation; Rational roots of polynomials of integer coefficients; The relation between coefficients and roots of a polynomial; Fundamental theorem of symmetric polynomials, power sums, Viète's formulas; The fundamental theorem of algebra; Polynomials over \mathbb{Z} ; Gauss lemma; Reduction modulo p ; Polynomials over \mathbb{Z}_p ; Irreducible polynomials; Irreducibility criteria for polynomials; and Cyclotomic polynomials.

Instruments

In the first part of the study, students took two tests during the semester: one in March 2018 (in the 6th week of the semester) and another one in May 2018 (in the 13th week of the semester). We concentrated on their problem-solving abilities on polynomials. The first test—which students took in March—consisted of seven problems. Out of the seven problems, three tasks were on polynomials. On the second test—which the students completed in May—we asked students five problems, and all of them were on the topic of polynomials (see Appendix). The tests have essentially been the same every second year since 2003.

In the second part of the study, we were curious about students' long-term retention of the topic of polynomials. We measured it by a post-test at the end of October 2018 which consisted of three problems (see Appendix).

In the Appendix, we also highlight some of the most relevant problems and their solutions presented in the course ANT 2. We argue that all of these problems are considered to be complex. By the definition of Van Gog and Sweller by “complex,” we mean “high in element interactivity, containing various information elements that are related and must therefore be processed simultaneously in working memory” (p. 248). Most of these problems asked about factoring or reducibility of high-degree polynomials over various algebraic structures, \mathbb{Z} , \mathbb{R} , \mathbb{Q} , \mathbb{C} , etc.... If you look at Table 1, each problem requires different approaches and most of them can be approached from different directions. After choosing and applying an initial trick, in almost all cases, we obtain a lower-degree polynomial to factor, and so on.

Data analysis and results

To measure students' achievement and development, we used a combination of qualitative and quantitative research methods. To answer our first two research questions (RQ1 and RQ2) we used quantitative research methods. All statistical analyses were conducted using R 4.2.3 software (R Core Team, 2023). Qualitative research methods were used to answer the third research question (RQ3).

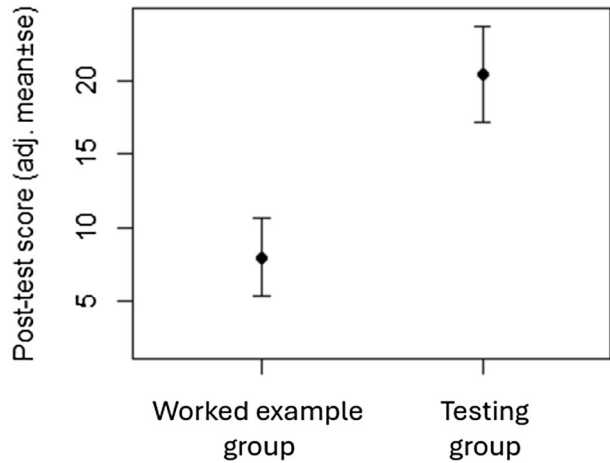
RQ1: Do students who study abstract algebra in a testing way perform better on complex mathematical problems than those who study by worked examples on the midterm and the final test (in the medium term)?

We investigated the medium-term effect of retrieval practice versus worked examples in learning abstract algebra in the first part of the study. Students took a midterm and a final test on the 6th and the 13th week of the semester. $N=76$ students took both tests, $N=39$ from the testing group and $N=37$ from the worked example group. To assess relationships between test scores, all scores were calculated as percentages. We used independent sample t -tests to compare the results of the midterm and the final test between the testing and the worked example groups. The normality of the distribution of test scores was checked on quartile-quartile (QQ) plots for the results of both groups on both tests. Assumptions for an equal variance were checked with F -tests for the results of the midterm ($F(1,38)=0.67, p=0.21$), and the final test ($F(1,38), p=0.057$). There was no significant difference between the midterm ($t(74)=1.19, p=0.21$) and final test ($t(74)=1.25, p=0.24$) scores of the two groups. It means that there was no significant difference between the scores of the testing group and the worked example group on either the first or the second test.

RQ2: Do students who study abstract algebra in a testing way perform better on complex mathematical problems than those who study by worked examples on the post-test five months after the final test (in the long term)?

In order to answer RQ2, in the second part of the study, we explored the long-term effect of retrieval practice in learning abstract algebra by a post-test. For this analysis, we chose to examine the scores of all four problems (two from the midterm and two from the final) that were asked again in the post-test. To obtain pre-test scores, we summed up the results of the four repeated tasks from the midterm and final. Participants of the study were those 33 pre-service teachers who completed all three tests: $N=13$ students from the testing group and $N=20$ from the worked example group. Test results were converted to percentages. Both pre- and post-test scores showed normal distribution in both groups on QQ-plots, so we used Pearson correlation to assess the relationship between pre- and post-test scores. Pre- and post-test scores moderately correlated ($r(31)=0.47, p=0.006$). We used one-way analysis of covariance (ANCOVA) to determine the effect of testing on post-test scores while controlling for pre-test scores. We validated our model and checked model residuals using the *DHARMA* package (Hartig, 2022). There was a significant difference between the testing and worked example group ($F(1,30)=8.9; p=0.006$). We performed a post hoc analysis using the *emmeans* package (Lenth, 2024) with pairwise comparisons applying Bonferroni adjustment. The testing group scored significantly higher (20.4 ± 3.3 ; adj. mean \pm std. error) in the post-test than the worked example group (7.9 ± 2.6), adj. $p=0.0057$ (see Fig. 1).

Fig. 1 Post-test results (adjusted mean \pm standard error) of students in the testing and the worked example group. The dots represent the adjusted mean values; the whiskers show the standard deviation (figure generated using R 4.2.3 software)



RQ3: To what extent do students forget abstract algebra in the long term—five months after the final test—studying it in a testing way versus studying by worked examples?

In order to answer RQ3, we investigated the post-test scores. We measured how well students remembered the topic of polynomials—which was a topic they covered in the ANT 2 course. Participants of the study were those 33 pre-service teachers who passed the course ANT 2 and took all the tests: the midterm, the final, and the post-test. Among the 33 students, $N=13$ studied ANT 2 in a testing way (testing group), and there were $N=20$ students from the worked example group.

On the post-test, students had to solve three problems related to polynomials. The three problems of the post-test resembled the ones that we asked for either in the midterm or in the final test. Each problem involved an initial trick or method that was typical in the theory of polynomials and necessary to solve the problem. In the first pair, there were two high-degree polynomials that students had to factor over Z_2 . The initial trick was that a sum can be squared elementwise over Z_2 , or one could observe that 1 is a root of the polynomial. The second problem was Problem 2 in both cases; the task was to factor the polynomial $p(x) = x^8 + x^4 + 1$ into a product of irreducible factors over \mathbb{R} . There are several ways to start the solution. The third problem was to factor a polynomial with integer coefficients. The first step in the solution was to realize that we had to apply the rational root test.

Those who managed to find the key idea of a problem got at least 1 point. Those who were not able to remember the key idea of a problem got 0 points. We define “forgetting” as achieving 0 points altogether on the post-test. In Table 2, we present the “forgetting rate” in the two groups. By forgetting rate, we mean the number of students who got 0 points on the post-test divided by the number of students. In the testing group ($N=13$), 23.08% of the students forgot the material. In the worked example group ($N=20$), 55.00% of the students got 0 points on the three problems of the post-test.

Table 2 Forgetting rate of the testing and the worked example group according to the post-test

	N	Forgetting rate
Testing group	13	0.2308
Worked example group	20	0.5500

Summary, discussion, and outlook

Summary and contribution to the field

The retention of foundational knowledge is crucial in learning and teaching mathematics. Retrieval practice, the strategic use of retrieval to enhance memory, and worked examples, showing students a step-by-step demonstration of how to solve a problem, are both recommended as effective methods for improving learning (Adeniji & Baker, 2023; Dunlosky et al., 2013). In this article, we compared the effectiveness of a particular way of using worked examples and a particular type of testing in an abstract algebra course concentrating on medium (within-semester) and long-term (across-semester) retention. We conducted an experiment in an authentic mathematics education context involving pre-service mathematics teachers in an “Algebra and Number Theory 2” course. We showed that in the medium-term there wasn’t any difference in problem-solving between the two groups—on the mid-term and the final, the testing group performed the same as the worked example group. On the other hand, we found that testing was more beneficial in the long term than studying with worked examples. The improvement of those students who studied ANT 2 with testing was significantly larger than that of the other group. At the same time, the forgetting rate was more than twice as high in the worked example group than in the testing group.

The method investigated in this study—students take a short test on the material of the given day without any aid, or any external help at the end of each lesson—does not need financial resources, it takes only a small amount of time, and it is simple; thus, it is easy to implement into classrooms. Quizzes and tests are already commonly used assessment tools in tertiary mathematics education and they can have a positive effect on students’ achievement. However, the impact of tests on students’ learning also depends on how educators (and students) use them (Feudel & Unger, 2022). The research results show that taking short tests that require students to retrieve the material studied on the given day can have a powerful effect in the long term even when learning abstract mathematics. In addition to these results, the advantage of the method is that it does not require teachers to reorganize their lessons to a large extent; they only have to spend the last few minutes of the lesson on the test and it does not require any special equipment. It is of high importance that educators do not consider the time of the tests as “lost time.” Instead, they should see taking a 5–10 min long test at the end of the lesson as a method that enriches long-term retention and highly reduces the forgetting rate.

In Hungary—where this research was conducted—students’ attitudes to learning are often such that if an assignment is not compulsory, or if there is no disadvantage in not completing it, students tend not to take it seriously. That is why in this experiment, students from the retrieval practice group had to reach at least 50 percent of the maximum points on the end-of-class tests by the end of the semester—it was a requirement in order to pass the course. This way, we made sure that students took these tests seriously. Still, we wanted to minimize the stress factor of the end-of-class tests: 50 percent was not that hard to achieve, and with a little effort, students could easily reach the minimum level to complete this criterion.

Limitations of the study and outlook for further research

In this study, we investigated the medium- and long-term effects of a certain type of test-enhanced learning in an abstract algebra course. Although this method seems to be an efficient way to be able to retain abstract mathematics in the long term, further research is

needed in this area to draw a far-reaching conclusion. We believe it would be important to test the method in different school environments: with different students, at different universities, and in other abstract mathematics courses as well.

In this experiment, the teacher corrected all the end-of-class tests. It would be practical to know how we can minimize teachers' tasks while maintaining the effectiveness of the method. For instance, is it important to give feedback to students about their performance? If the answer is yes, then what is the best way to give feedback? From whom and how does the feedback have to come from? Can it be a student, a group of students, or a machine who gives it? The question may also arise as to whether the same efficiency can be achieved if students are not tested in class—under controlled circumstances—but are allowed to take the test at home. It would also be useful to know whether the outcome depends on the scoring system we use (or not use at all). Achieving 50 percent was not too difficult; still, it may have had an impact on students' stress levels which might have affected their performance.

Furthermore, the form of the test should be considered as well. Several forms of testing can be applied, for instance, multiple choice questions, short answer questions, or true or false questions. The question is: are they as effective as open questions? We believe that true or false questions or multiple-choice questions do not provide the same results: we can imagine that if a question needs a more complex argumentation, but the final answer requires only a selection process (circling, matching, crossing, underlying), the student may simplify and not think deeply about the question.

Another limitation of our study may concern the decreased sample size for the post-test. We eliminated this issue by using ANCOVA instead of a *t*-test.

Last but not least, we cannot say that the beneficial effect of the retrieval practice method on long-term retention is entirely due to the retrieval itself. Retrieval practice has several indirect effects, as shown by McDaniel et al. (2015). For example, the end-of-class tests can give learners a fairly clear picture of what they can retrieve and apply in mathematical problem-solving, and what concepts, procedures, and skills they still need to acquire. Another indirect effect can be that taking tests by itself can help students develop test-taking skills which may improve performance on high-stakes exams (Adesope et al., 2017). One might also assume that there is a latent factor, namely that the method exerts its effect through the active engagement of students rather than via the retrieval itself. We have no valid information regarding students' higher attention.

The abovementioned limitations open up several pathways for future research. We believe it would be important to investigate how we can minimize teachers' tasks while maintaining the effectiveness of the testing method. Future research could explore if giving feedback to students should be a key element of the method. Does it have to come from the teacher, from a student, or from a machine? Changing the form of the end-of-class test from open questions to multiple choice questions, short answer questions, or true or false questions could also reduce teachers' work. However, we have to examine if we can maintain the same performance concerning complex learning material in the long term with these types of questions.

Another possible direction of future research is the separation of direct and indirect effects of testing in a similar context, for example, the "attention effect" and the "activation effect."

Also, it might be fruitful to explore more deeply whether there is a relationship between the level of abstraction of the material and the effectiveness of the method. In previous experiments, the authors examined the effect of end-of-class tests in learning high-school mathematics and in studying Number Theory problems at the university level as well (Szabó et al., 2023; Szeibert et al., 2022). The authors' experimental results so far

strengthen the positive effects of retrieval practice in abstract environments, and the more abstract the curriculum is, the later the testing effect becomes apparent. Finally, it would be interesting to explore to what extent the use of spaced or cumulative retrieval practice has a positive impact on students' long-term retention in mathematics and solving complex abstract problems.

Appendix

1. The topics covered by the course *Algebra and Number Theory 2*:

Systems of linear equations, Gaussian elimination. Matrices and operations. Complex numbers, trigonometric form, roots of unity. Polynomials over fields. Equalities, main coefficient, normal polynomial. Operations with polynomials (addition, subtraction, multiplication, Euclidean division), the zero polynomial. Degree of a polynomial. Roots and factorizations of polynomials, Horner's method. The maximum number of roots. Identity theorem of polynomials over an infinite field. The relation between polynomial functions and polynomials. Multiplication of roots, root factor form. Lagrange interpolation. Rational roots of polynomials of integer coefficients. The relation between coefficients and roots of a polynomial. Fundamental theorem of symmetric polynomials, power sums, Viète's formulas. The fundamental theorem of algebra. Polynomials over \mathbb{Z} . Gauss lemma. Reduction modulo p . Polynomials over \mathbb{Z}_p . Cyclotomic polynomials. The notion and rule of irreducible polynomials. The overview of irreducible polynomials above different fields. Equations of third and fourth degree.

2. The topics covered by the course *Algebra and Number Theory 3*:

Groups, examples from geometry, algebra, and number theory. Subgroup, normal subgroup, factor group, direct product, isomorphism. The fundamental theorem of finite abelian groups. Permutation groups and their applications. Lagrange's theorem, Cayley's theorem, and their importance. Vector spaces, independence, basis, dimension. Subspace, factor space, direct sum. Matrices and linear maps, image, kernel. rank and their connection with solutions of systems of linear equations. Determinants and their expansion, Cramer's rule. Vandermonde determinant and its connection with interpolation. Eigenvectors and eigenvalues, characteristic and minimal polynomial. Usage of computational software.

3. Midterm test:

For which value of $a \in \mathbb{Z}_7$ is there no solution to the following system of equations over \mathbb{Z}_7 ?

$$2x + 14y + 8z = 2$$

$$x + 4y + z = 1$$

$$3x + y + az = 5$$

Find the "exact" value of the following expression: $\left| \frac{(\sqrt{3}+3i)^{2018}}{(-\sqrt{2}-\sqrt{2}i)^{2018}} \right|$.

Factor the polynomial $x^8 + x^4 + 1$ to the product of irreducible polynomials over R .

Factor the polynomial $x^7 + 2x^6 + 125x + 250$ into linear factors over C .

Factor the polynomial $x^6 - 1$ into linear factors over \mathbb{Z}_{19} .

Where are those complex numbers located on the plane for which:

$$|z + 4 - 8i| = |z + 8i|$$

The wicked witch claims that there exists a system of equations with integer coefficients that has a solution over Z, Q, R, C in exactly two cases.

- a) Prove that she is lying.
- b) Show a system of equations with integer coefficients that has a solution over Z, Q, R, C in exactly three cases.

4. Final test:

1. Compute the polynomial $\phi_{36} \cdot \phi_{18} \cdot \phi_9$, where ϕ_n denotes the n^{th} cyclotomic polynomial.
2. Show an irreducible polynomial over

Z whose root is $\sqrt[3]{2} + 1$. (Do not forget to prove the irreducibility of the polynomial.)

3. Factor the polynomial

$x^{16} + x^{12} + x^8 + 1$ to a product of irreducible polynomials over Z_2 .

4. The wicked witch claims that the following 32-degree polynomial.

$x^{32} - 6x^{31} + 42x^{30} + 2x^{27} + \dots$ can be supplemented with such terms so that the polynomial will have 32 integer roots. Prove that she is lying.

Factor the polynomial $p(x) = x^8 + x^4 + 1$ to the product of irreducible polynomials over R .

5. Post-test

Factor the polynomial $x^{12} + x^8 + 1$ to the product of irreducible polynomials over Z_2 .

Prove that all roots of the polynomial $x^8 + x^4 + 1$ over C are roots of unity. Determine the order of these roots.

Find all integers d for which there exists a positive integer root of the polynomial $4x^6 + dx + 2$.

6. Most important problems and their solutions

The tests and the post-test consisted of similar problems to the ones below.

Problem 1: Factor the polynomial $x^{16} + x^{12} + x^8 + 1$ to a product of irreducible factors over Z_2 .

Solution to problem 1:

There is only one way to start to solve this problem, one needs to apply the “high-school dream.” The “high-school dream” is a theorem (and also an inside joke) in this course, namely: over Z_2 , $(a + b)^2 = a^2 + b^2$ holds. Thus, the first part of the solution is to rewrite the polynomial by using the “high-school dream” iteratively, so we get:

$$x^{16} + x^{12} + x^8 + 1 = (x^4 + x^3 + x^2 + 1)^4$$

This part itself can consist of several steps:

$$\begin{aligned} x^{16} + x^{12} + x^8 + 1 &= (x^8 + x^6)^2 + (x^4 + 1)^2 = (x^8 + x^6 + x^4 + 1)^2 \\ &= [(x^4 + x^3 + x^2 + 1)^2]^2 = (x^4 + x^3 + x^2 + 1)^4 \end{aligned}$$

Unfortunately, a non-negligible proportion of students stop here. However, we do not know at this point if we are ready, we still have to examine the degree 4 polynomial.

The next observation needed for the solution is the following: 1 is a root of this polynomial over Z_2 . It is realistic to expect that students make this observation. In the practice sessions, several possible strategies were discussed leading to this observation. Students know that the possible roots are 0 and 1 and they just substitute both elements into the polynomial. Or in particular, over Z_2 , it was a recurrent thread of thought that 1 is a root of every polynomial with even many terms.

Hence

$$x^4 + x^3 + x^2 + 1 = (x + 1)(x^3 + x + 1)$$

In order to continue, you need to apply the following theorem: “A polynomial of degree 3 is reducible if and only if it has a root.” Neither 0 nor 1 is a root of $x^3 + x + 1$, hence it is irreducible over Z_2 . Thus

$$x^{16} + x^{12} + x^8 + 1 = (x + 1)^4(x^3 + x + 1)^4$$

Problem 2: Factor the polynomial $p(x) = x^8 + x^4 + 1$ into a product of irreducible factors over \mathbb{R} .

Solution to problem 2:

A standard try for a student is to substitute $y = x^4$ and look for the roots of $y^2 + y + 1$. Then find the fourth roots of the two complex roots. Students starting this way usually stop at writing.

$x = \sqrt[4]{\left(-1/2 \pm \frac{\sqrt{3}}{2}i\right)}$. Those who continue either observe that these are the algebraic forms of the 3rd roots of unity or try to calculate $(a + bi)^4$. There were very few students who went in any of these directions.

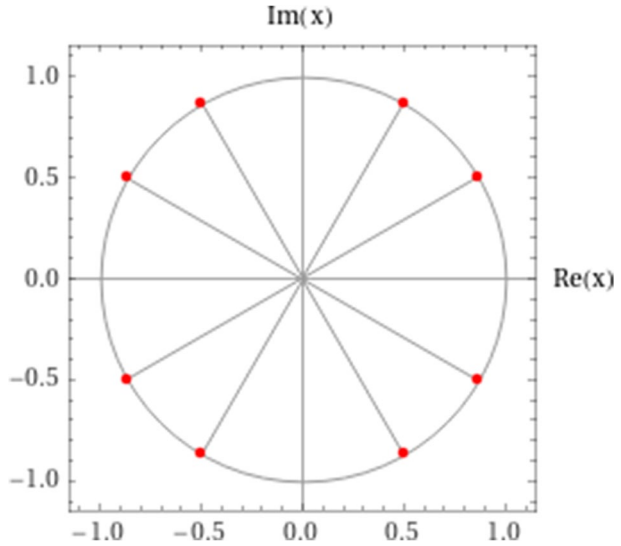
Another possible way to solve the problem is to observe (or remember) that the terms of $p(x)$ form a geometric progression, so $p(x) = (x^{12} - 1)/(x^4 - 1)$. From here, by observing the quotient and considering its algebraic meaning, we can see that the roots are the 12th roots of unity, except those that are roots of $x^4 - 1$. Figure 2.

Hence, the 8 roots are the 1st, 2nd, 4th, 5th, 7th, 8th, 10th, 11th, and the 12th among the 12th roots of unity. Then either based on the figure or even without a figure, only by using the formula for the order of a power, $o(\epsilon^i) = \frac{12}{(12,i)}$ we obtain that the orders are 12; 6; 3; 12; 12; 3; 6; 12 respectively.

Another, slightly easier way to solve the problem requires more knowledge. If we factor the polynomial into the products of cyclotomic polynomials, we get

$$x^8 + x^4 + 1 = \frac{x^{12} - 1}{x^4 - 1} = \frac{\phi_{12} \circ \phi_6 \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1}{\phi_4 \circ \phi_2 \circ \phi_1} = \phi_{12} \circ \phi_6 \circ \phi_3$$

Fig. 2 Roots of $p(x) = (x^{12} - 1)/(x^4 - 1)$ (figure generated using Wolframalpha)



Thus, the orders are 12; 6, and 3.

Problem 3: Factor the polynomial $x^7 - 3x^6 + 125x - 375$ into linear factors over C .

Solution to Problem 3:

Instructors of the ANT courses are convinced that looking at a polynomial with integer coefficients a student should remember the rational root test. In this case, there are quite a few candidates for an integer (rational) root, and indeed, in this particular case, 3 is obviously a root. The next step is to factor out $x - 3$. This can be easily done with this polynomial, but Horner’s method is also an option. Then we arrive at $(x - 3)(x^6 + 125)$. From here, one has to take the 6th roots of -125 . Then we obtain all the seven roots of the polynomial, and we can factor it into a product of linear terms.

Problem 4: Compute the polynomial $\phi_{36} \circ \phi_{18} \circ \phi_9$, where ϕ_n denotes the n^{th} cyclotomic polynomial.

Solution to problem 4:

The first approach could be to calculate ϕ_{36} , ϕ_{18} , and ϕ_9 and multiply them. As 9 is a power of a prime,

$$\phi_9(x) = \frac{x^9 - 1}{x^3 - 1} = x^6 + x^3 + 1$$

$$\phi_{18}(x) = \phi_9(-x) = x^6 - x^3 + 1$$

$$\phi_{36}(x) = \phi_{18}(x^2) = x^{12} - x^6 + 1$$

Multiplying these polynomials is neither easy, nor hard. Another approach is to skip the step-by-step multiplication and apply the pattern

$$(x^{2n} + x^n + 1)(x^{2n} - x^n + 1) = x^{4n} + x^{2n} + 1$$

twice to obtain:

$$\phi_{36} \circ \phi_{18} \circ \phi_9 = x^{24} + x^{12} + 1$$

As in most cases, there is a more straightforward way to solve this problem. Observe that the indices include exactly those divisors of 36 that are not divisors of 12. Thus

$$\phi_{36} \circ \phi_{18} \circ \phi_9 = \frac{\phi_{36} \circ \phi_{18} \circ \phi_{12} \circ \phi_9 \circ \phi_6 \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1}{\phi_{12} \circ \phi_6 \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1} = \frac{x^{36} - 1}{x^{12} - 1} = x^{24} + x^{12} + 1$$

Funding Open access funding provided by University of Debrecen. This work is supported by the Digital Education Development Competence Center at Eötvös Loránd University, Budapest, with project number 2022–1.1.1-KK-2022–00003.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbott, E. E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Monographs*, 11(1), 159–177. <https://doi.org/10.1037/h0093018>
- Adeniji, S. M., & Baker, P. (2023). Effects of worked example on students' learning outcomes in complex algebraic problems. *International Journal of Instruction*, 16(2), 229–246. <https://doi.org/10.29333/iji.2023.16214a>
- Adesope, O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701. <https://doi.org/10.3102/0034654316689306>
- Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review*, 33(4), 1409–1453. <https://doi.org/10.1007/s10648-021-09595-9>
- Agustyaningrun, N., Sari, R. N., Abadi, A. M., & Mahmudi, A. (2021). Dominant factors that cause students' difficulties in learning abstract algebra: A case study at a University in Indonesia. *International Journal of Instruction*, 14(1), 847–866. <https://doi.org/10.29333/iji.2021.14151a>
- Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 940–945. <https://doi.org/10.1037/a0029199>
- Atkinson, R. D., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, 70(2), 181–214. <https://doi.org/10.3102/00346543070002181>
- Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology*, 55(1), 25–35. <https://doi.org/10.1016/j.jmp.2010.08.009>
- Butler, A. J. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 36(5), 1118–1133. <https://doi.org/10.1037/a0019902>

- Chan, J. C. (2010). Long-term effects of testing on the recall of nontested materials. *Memory*, 18(1), 49–57. <https://doi.org/10.1080/09658210903405737>
- Chan, J. Y., Meissner, C. A., & Davis, S. L. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, 144(11), 1111–1146. <https://doi.org/10.1037/bul0000166>
- Clark, R. C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. Pfeiffer.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology. In *Teachers College Press eBooks*. <https://doi.org/10.1037/10011-000>
- Even, R. (2011). The relevance of advanced mathematics studies to expertise in secondary school mathematics teaching: Practitioners' views. *ZDM*, 43, 941–950. <https://doi.org/10.1007/s11858-011-0346-1>
- Fazio, L. K. (2019). Retrieval practice opportunities in middle school mathematics teachers' oral questions. *British Journal of Educational Psychology*, 89(4), 653–669. <https://doi.org/10.1111/bjep.12250>
- Feudel, F., Unger, A. (2022). Students' strategic usage of formative quizzes in an undergraduate course in abstract algebra. *International Journal of Research in Undergraduate Mathematics Education*<https://doi.org/10.1007/s40753-022-00194-9>
- Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, 40(4), 505–513. <https://doi.org/10.3758/s13421-011-0174-0>
- Hartig, F. (2022). DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models. R package version 0.4.6, <https://CRAN.R-project.org/package=DHARMa>
- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In: J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum Associate
- Hopkins, R. F., Lyle, K. B., Hieb, J. L., & Ralston, P. A. (2016). Spaced retrieval practice increases college students' short-and long-term retention of mathematics knowledge. *Educational Psychology Review*, 28(4), 853–873. <https://doi.org/10.1007/s10648-015-9349-8>
- Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports*, 18(3), 879–919. <https://doi.org/10.2466/pr0.1966.18.3.879>
- Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 36(6), 1441–1451. <https://doi.org/10.1037/a0020636>
- Kalyuga, S., Renkl, A., & Paas, F. (2010). Facilitating flexible problem solving: A cognitive load perspective. *Educational Psychology Review*, 22, 175–186. <https://doi.org/10.1007/s10648-010-9132-9>
- Karpicke, J. D. (2017). *Retrieval-Based Learning: A Decade of Progress*. <https://doi.org/10.1016/B978-0-12-809324-5.21055-9>
- Kliegl, O., & Bäuml, K. T. (2016). Retrieval practice can insulate items against intralist interference: Evidence from the list-length effect, output interference, and retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 42(2), 202–214. <https://doi.org/10.1037/xlm0000172>
- Leahy, W., Hanham, J., & Sweller, J. (2015). High element interactivity information during problem solving may lead to failure to obtain the testing effect. *Educational Psychology Review*, 27(2), 291–304. <https://doi.org/10.1007/s10648-015-9296-4>
- Lenth, R. (2024). Emmeans: Estimated marginal means, aka least-squares means. R package version 1.10.0, <https://CRAN.R-project.org/package=emmeans>
- Lyle, K. B., & Crawford, N. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology*, 38(2), 94–97. <https://doi.org/10.1177/0098628311401587>
- Lyle, K. B., Hopkins, R. F., Hieb, J. L., & Ralston, P. A. (2016). Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, 28(4), 853–873. <https://doi.org/10.1007/s10648-015-9349-8>
- Lyle, K. B., Bego, C. R., Hopkins, R. F., Hieb, J. L., & Ralston, P. A. (2020). How the amount and spacing of retrieval practice affect the short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, 32, 277–295. <https://doi.org/10.1007/s10648-019-09489-x>
- Ma, L. (2020). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States* (3rd ed.). Routledge. <https://doi.org/10.4324/9781003009443>

- May, B. M. (2021). Effects of spaced, repeated retrieval practice and test-potentiated learning on mathematical knowledge and reasoning. *International Journal of Mathematical Education in Science and Technology*, 53(1), 92–107. <https://doi.org/10.1080/0020739x.2021.1961034>
- McDaniel, M. A., Howard, D. L., & Einstein, G. O. (2009). The read-recite-review study strategy. *Psychological Science*, 20(4), 516–522. <https://doi.org/10.1111/j.1467-9280.2009.02325.x>
- McDaniel, M. A., Bugg, J. M., Liu, Y., & Brick, J. (2015). When does the test-study-test sequence optimize learning and retention? *Journal of Experimental Psychology: Applied*, 21(4), 370–382. <https://doi.org/10.1037/xap0000063>
- McDermott, K. B., Agarwal, P., D'Antonio, L., Roediger, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, 20(1), 3–21. <https://doi.org/10.1037/xap0000004>
- Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLoS ONE*, 10(7), e0120644. <https://doi.org/10.1371/journal.pone.0120644>
- R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org>
- Racsmány, M., Szöllösi, Á., & Bencze, D. (2018). Retrieval practice makes procedure from remembering: An automatization account of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(1), 157–166. <https://doi.org/10.1037/xlm0000423>
- Racsmány, M., & Keresztes, A. (2015). Initial retrieval shields against retrieval-induced forgetting. *Frontiers in Psychology*, 6, 657. <https://doi.org/10.3389/fpsyg.2015.00657>
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38(1), 1–37. <https://doi.org/10.1111/cogs.12086>
- Roediger, H. L. III, Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. P. Mestre & B. H. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (pp. 1–36). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00001-6>
- Roediger, H. L., 3rd., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 1(3), 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Rowland, C. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463. <https://doi.org/10.1037/a0037559>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- 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>
- Szabó, C., Bereczky-Zámbó, C., Muzsnay, A., Szeibert, J., & Bernáth, L. (2023). Investigación de la eficacia de práctica de recuperación en matemáticas universitarias. *Revista De Educación*, 401(1). <https://doi.org/10.4438/1988-592X-RE-2023-401-584>
- Szeibert, J., Muzsnay, A., Szabó, C., & Bereczky-Zámbó, C. (2022). A case study of using test-enhanced learning as a formative assessment in high school mathematics. *International Journal of Science and Mathematics Education*, 21(2), 623–643. <https://doi.org/10.1007/s10763-022-10264-8>
- Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34(6), 1392–1399. <https://doi.org/10.1037/a0013082>
- Van Gog, T., & Kester, L. (2012). A test of the testing effect: Acquiring problem-solving skills from worked examples. *Cognitive Science*, 36(8), 1532–1541. <https://doi.org/10.1111/cogs.12002>
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, 22(2), 155–174. <https://doi.org/10.1007/s10648-010-9134-7>
- Van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, 27(2), 247–264. <https://doi.org/10.1007/s10648-015-9310-x>
- Van Gog, T., Kester, L., Dirx, K., Hoogerheide, V., Boerboom, J., & Verkoijen, P. P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-015-9297-3>
- Veith, J. M., Bitzenbauer, P., & Girnath, B. (2022). Exploring learning difficulties in abstract algebra: The case of group theory. *Education Sciences*, 12(8), 516. <https://doi.org/10.3390/educsci12080516>
- Wasserman, N. H. (2014). Introducing algebraic structures through solving equations: Vertical content knowledge for K-12 mathematics teachers. *Primus*, 24(3), 191–214. <https://doi.org/10.1080/10511970.2013.857374>

- Wasserman, N. (Ed.). (2018). *Connecting abstract algebra to secondary mathematics, for secondary mathematics teachers*. Springer. <https://doi.org/10.1007/978-3-319-99214-3>
- Wissman, K. T., & Rawson, K. A. (2018). Test-potiated learning: Three independent replications, a disconfirmed hypothesis, and an unexpected boundary condition. *Memory (hove, England)*, 26(4), 406–414. <https://doi.org/10.1080/09658211.2017.1350717>
- Yeo, D. J., & Fazio, L. K. (2019). The optimal learning strategy depends on learning goals and processes: Retrieval practice versus worked examples. *Journal of Educational Psychology*, 111(1), 73–90. <https://doi.org/10.1037/edu0000268>
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38(8), 995–1008. <https://doi.org/10.3758/mc.38.8.995>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Current themes of research:

The authors of this article are members of the MTA-ELTE Theory of Learning Mathematics Research Group whose research areas encompass the examination of effective methods for presenting and teaching curriculum. Within this scope, we place particular emphasis on enhancing the efficiency of the mathematics learning process. Our primary research domains include investigating test-enhanced learning, the development of thinking skills through boardgames, gamification, the utilization of interleaved problem sets, problem posing, exploring geometric conceptual development, bridging knowledge gaps, and enhancing spatial skills.

Most relevant publications in the field of Psychology of Education:


Publications on testing:

- Szabó, C., Bereczky-Zámbó, C., Muzsnay, A., Szeibert, J., & Bernáth, L. (2023). Investigación de la eficacia de práctica de recuperación en matemáticas universitarias. *Revista De Educación*, 401(1). <https://doi.org/10.4438/1988-592X-RE-2023-401-584>.
- Szeibert, J., Muzsnay, A., Szabó, C., et al. (2022). A case study of using test-enhanced learning as a formative assessment in high school mathematics. *International Journal of Science and Mathematics Education* 21, 623–643. <https://doi.org/10.1007/s10763-022-10264-8>.
- Bereczky-Zámbó, Cs.; Muzsnay, A.; Szeibert, J.; Szabó, Cs.(2019). Test-enhanced learning in studying mathematics at secondary school and university. *In: Mellony, G.; Hamsa, V.; Anthony, A. E.; Pamela, V.: Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education (Vol 4)*. Pretoria, South Africa: PME. Pretoria, Dél-Afrikai Köztársaság: Minute Man Press. Paper: 4-124.
- Szabó, C., & Szeibert, J. (2018). Efficiency of test-enhanced learning in teaching elementary geometry. *In Proceedings Of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (pp. 296–296).
- Szabó, C., & Muzsnay, A. (2018). Efficiency of test-enhanced learning in teaching number theory. *In Proceedings Of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (p. 168).

Publications on other topics in Psychology of Education:

- Bereczky-Zámbó, Cs., Szabó, Cs., Muzsnay, A., Szeibert, J. (2022). Passing the exam and not mastering the material in geometry. *Annales Mathematicae et Informaticae*, 55, 189-195. ISSN 178.
- András, F. D., Katalin, F., & Csaba, S. (2021). The time spent on board games pays off : links between board game playing and competency motivation. *Teaching Mathematics and Computer Science*, 19(1), 119–131. <https://doi.org/10.5485/TMCS.2021.0523>.
- Szabó, C., Bereczky-Zámbó, C., Muzsnay, A., & Szeibert, J. (2020). Students' non-development in high school geometry. *Annales Mathematicae et Informaticae*, 52, 309–319. <https://doi.org/10.33039/ami.2020.12.004>.
- Dukán, A.F., Szabó, Cs., & Vásárhelyi, É. (2020). Logic in secondary school: From Tamás Varga's proposed curriculum to board games. *In Komplexer Mathematikunterricht* (pp. 143–156).
- Bereczky-Zámbó, C., Muzsnay, A., Szeibert, J., & Szabó, C. (2019). Testing the test: Rethinking the levels of understanding geometry based on maths students' van hiele test results. *In Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education (Vol 4)*. Pretoria, South Africa: PME.

Authors and Affiliations

Anna Muzsnay^{1,2}  · **Csilla Zámbo**^{2,3} · **Janka Szeibert**^{3,4} · **László Bernáth**^{2,5} · **Brigitta Szilágyi**^{2,6} · **Csaba Szabó**^{7,8}

✉ Anna Muzsnay
muzsnay.anna@science.unideb.hu

¹ Faculty of Science and Technology, University of Debrecen, Debrecen, Hungary

² MTA-ELTE Theory of Learning Mathematics Research Group, Budapest, Hungary

³ Faculty of Primary and Pre-School Education, ELTE Eötvös Loránd University, Budapest, Hungary

⁴ HUN-REN Alfréd Rényi Institute of Mathematics, Budapest, Hungary

⁵ Faculty of Education and Psychology, ELTE Eötvös Loránd University, Budapest, Hungary

⁶ Budapest University of Technology and Economics, Budapest, Hungary

⁷ Faculty of Science, ELTE Eötvös Loránd University, Budapest, Hungary

⁸ Edutus University, Tatabánya, Hungary