

Retrieval practice and test-potentiated learning: A comparison of high and low prior topic knowledge students in terms of procedural fluency and conceptual understanding



Author:
Bruce M. May¹ 

Affiliation:
¹School of Science and Mathematics Education, Faculty of Education, University of the Western Cape, Cape Town, South Africa

Corresponding author:
Bruce May,
bmay@uwc.ac.za

Dates:
Received: 07 May 2023
Accepted: 28 Feb. 2024
Published: 29 Apr. 2024

How to cite this article:
May, B.M. (2024). Retrieval practice and test-potentiated learning: A comparison of high and low prior topic knowledge students in terms of procedural fluency and conceptual understanding. *Pythagoras*, 45(1), a755. <https://doi.org/10.4102/pythagoras.v45i1.755>

Copyright:
© 2024. The Authors.
Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Read online:



Scan this QR code with your smart phone or mobile device to read online.

A cohort of pre-service mathematics students was exposed to a teaching strategy based on retrieval practice and test-potentiated learning. The aim of the study was to determine how high and low prior topic knowledge study participants compare in terms of their procedural fluency and conceptual understanding after exposure to the teaching strategy. A pre-test and post-test repeated measures design was employed in the study to compare within groups. A revised taxonomy table based on Bloom's taxonomy was utilised to categorise test items. Findings indicate significant differences between pre-test and post-test scores within groups. Results from the independent samples *t*-test show a significant difference between the two groups. Outcomes confirm that the benefits of retrieval practice are greatest for unfamiliar content. Findings indicate that for low prior topic knowledge students, procedural fluency is enhanced and retained more than conceptual understanding whereas for the high prior topic knowledge students it was the reverse. The strategy was not as effective for improving conceptual understanding.

Contribution: How a teaching strategy based on retrieval practice and test-potentiated learning affects the mathematical competencies of procedural fluency and conceptual understanding has not been researched. There is also a dearth of studies that set out to investigate how retrieval practice and test-potentiated learning affect research participants with different levels of prior knowledge. The contribution of this study therefore is to contribute to understanding of how retrieval practice and test-potentiated learning can be utilised to improve learning and teaching of mathematics at the school level.

Keywords: pre-service mathematics teachers; retrieval practice; test-potentiated learning; prior topic knowledge

Introduction

In South Africa, high school students are required to study either Mathematics or Mathematical Literacy from Grades 10 through 12. The Curriculum and Assessment Policy Statement (CAPS) (DBE, 2011, p. 8) defines Mathematical Literacy as a subject that does not concentrate on abstract mathematical ideas but is grounded in basic mathematical content. It is emphasised that calculations in the subject should only require a basic four-function calculator. Furthermore, it is required that all content should be context based since the focus of Mathematical Literacy is to make sense of real-life contexts. The development of procedural knowledge is therefore not a priority in Mathematical Literacy. Consequently, in this subject the development of the skill of algebraic manipulation does not enjoy a high priority. Given that the emphasis is on basic mathematical content without a focus on the growth of mathematical conceptual comprehension, it is anticipated that students of Mathematical Literacy will not possess the same degree of conceptual and procedural understanding (pertaining to the topics of the high school curriculum in South Africa) as those who pursued Mathematics. Accordingly, the expectation is that Mathematical Literacy students do not have well-developed conceptual understanding and procedural knowledge of the mathematical topics that form part of school-level mathematics. Students who enter university mathematics content courses with Mathematical Literacy therefore are perceived to have low prior knowledge as compared to students who enter with Mathematics.

At some universities in South Africa entrance criteria for pre-service mathematics teachers allow students to enter with Mathematical Literacy. Upon graduation, these students are

permitted to teach Mathematics in the Senior Phase (Grades 7 to 9). To provide high-quality education, teachers at this stage require a robust understanding of both the procedures and concepts related to the mathematical topics covered in Grades 7 through 9. Since Mathematical Literacy students for example are not exposed to algebraic concepts beyond Grade 9 there is a high possibility that their understanding of these concepts is limited. It is therefore crucial that the curricula of university-level mathematics courses include in their objectives the development of reasoning, and procedural and conceptual knowledge of school-level topics. Therefore, educators of these prospective mathematics teachers need to devise teaching methods that would boost mathematical reasoning, procedural and conceptual understanding, as well as memory retention.

Literature review

Retention of knowledge is of central importance in mathematics learning. Therefore, since retention of knowledge plays such a crucial role in mathematics learning it is incumbent on mathematics educators to make certain that they expose students to learning and teaching strategies that enhance retention. Studies have indicated that retrieval practice is a learning technique that improves the preservation of knowledge (for instance, Abott, 1909; Glover, 1989; Roediger & Karpicke, 2006). The concept that the process of remembering information aids in its preservation has been recognised for quite some time (Abott, 1909; Roediger & Butler, 2011). This process of recalling information from memory without relying on external resources (such as during exams or while studying) is referred to as retrieval practice.

Retrieval practice can also have an implicit or facilitating effect on learning. Evidence suggests that efforts to recall test items can enhance subsequent re-study of those items, even if the initial retrieval attempt was unsuccessful and no feedback was provided (Arnold & McDermot, 2013; Grimaldi & Karpicke, 2012; Izawa, 1966; Wissman & Rawson, 2018).

Izawa (1966) was the first researcher to identify this phenomenon (known as test-potentiated learning). Arnold and McDermot (2013) were subsequently the pioneering researchers who first identified the intermediary impact of testing on re-studying. They argue that test items that were successfully recalled in the test and after re-studying could have gained advantages from both the effect of testing and the enhancement of learning potentiated by testing. In contrast, items that were not successfully recalled in the initial test but were successfully retrieved in a subsequent test (after re-studying) could have only gained from the enhancing effect of the initial test. Research findings indicate that conducting more practice tests, as opposed to fewer, amplifies the efforts of re-studying (also known as test-potentiated learning) (Wissman & Rawson, 2018). In the current study, there was no effort made to differentiate between the indirect and direct impacts of retrieval practice.

Put differently, the impacts of both the testing effect and test-potentiated learning were assessed collectively. This is because in real-world mathematics education scenarios, the same material is often tested multiple times (through tests and examinations), and as a result, student performance is typically influenced by both the testing effect and test-potentiated learning.

Research evidence suggests that certain learning strategies may work well for some students, while they may not be as beneficial for others (Dunlosky et al., 2013; Xiaofeng et al., 2016).

Consequently, it is important to know how individual differences such as fluid intelligence, working memory capacity, level of prior topic knowledge, etc. influence the effectiveness of mnemonic enhancing strategies such as retrieval practice. The primary focus of this study is to evaluate the efficacy of retrieval practice and test-potentiated learning as strategies for enhancing memory, taking into account the influence of pre-existing topic knowledge. The research question for this study is: How do students with high and low pre-existing topic knowledge differ in their mathematical competencies, specifically procedural fluency and conceptual understanding, after being exposed to a teaching method that incorporates retrieval practice and test-potentiated learning? Data for the study were provided by research done for a doctoral thesis (May, 2017).

Prior topic knowledge and learning

Prior knowledge plays a major role in learning in practically all learning areas. It is therefore important for educators to understand how different learning and teaching strategies are affected by levels of prior knowledge. There is evidence that effectiveness of learning strategies is moderated by level of prior knowledge. The argument is made that the same learning materials and teaching strategy could yield different learning outcomes for students, depending on their varying levels of prior knowledge on the topic (Xiaofeng et al., 2016). Furthermore, students who already have a substantial understanding of the topic are more capable of grasping new content (Dunlosky et al., 2013), and they can retain more information compared to those with less pre-existing knowledge (Thompson & Zamboanga, 2003).

Prior topic knowledge delays forgetting of novel subject material since it can be integrated into a pre-existing knowledge framework (Dunlosky et al., 2013). Since it is integrated in existing knowledge structures with different information clues more hooks for retrieval are provided and hence retrieval is enhanced. Generally, it has been found that greater pre-existing knowledge in a domain enhances student learning and achievement (Thomson & Zamboanga, 2004). The literature however is silent on which types of knowledge are retained better and understood better. Understanding this is important for the design of effective teaching strategies.

Prior topic knowledge and retrieval practice

Studies indicate that the practice of retrieval not only improves the retention of knowledge, but also aids in learning (Butler, 2010; Chan et al., 2018; Karpicke, 2012). It is therefore also important for mathematics educators to understand the relationship between level of prior topic knowledge and effectiveness of retrieval practice. This is because an essential goal of teaching is the enhancement of knowledge. A number of studies set out to investigate how levels of prior topic knowledge influence the effectiveness of retrieval practice. The following is a discussion of some of these studies.

Findings of some studies indicate that levels of prior topic knowledge have little or no influence on the effectiveness of retrieval practice (for e.g. Carroll et al., 2007; Cogliano et al., 2019; Xiaofeng et al., 2016) but benefits of retrieval practice are greatest for unfamiliar content (i.e. low prior knowledge) (Cogliano et al., 2019).

Learning materials used by some of the studies (on how level of prior knowledge influences the effectiveness of retrieval practice) included topics in educational psychology (Cogliano et al., 2019), texts from abnormal psychology (Carroll et al., 2007), and key psychology terms (Xiaofeng et al., 2016). Study participants in these studies included undergraduate educational psychology students (Cogliano et al., 2019), undergraduate and advanced psychology majors (Carroll et al., 2007), and undergraduate psychology students (Xiaofeng et al., 2016).

Outcome measures in these studies included responses to multiple-choice questions based on education psychology course content (Cogliano et al., 2019), student responses to short-answer tests based on content from abnormal psychology (Carroll et al., 2007), and selection from a wordlist based on key psychology terms (Xiaofeng et al., 2016).

The fact that most of these studies were done in the domain of psychology is not surprising since most research regarding learning is done by cognitive psychologists. However, what is missing in all of the aforementioned research is how exposure to retrieval practice affects the different knowledge and reasoning types. Moreover, there is a dearth of research that utilised mathematics learning material and whose outcomes measure included knowledge and reasoning type. The author is not aware of any studies that included knowledge type and reasoning in their outcome measures.

Despite the fact that retrieval practice and its effects on learning have been studied extensively there are some areas where very few investigations have been done. An area of interest is how the impact of retrieval practice might be influenced by individual variances, such as pre-existing knowledge on the topic (Dunlosky et al., 2013). Dunlosky et al. (2013) argue that this issue is significant because studies have demonstrated that learning strategies do not have a uniform effect on all students. This inconsistency in the literature is what the present study aims to resolve.

Theoretical framework

Procedural fluency and conceptual understanding

This study posits that it is crucial for future teachers to master the five types of mathematical competencies defined by the National Research Council (2001), as they are fundamental to successful mathematics learning. These competencies include: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and a productive disposition. It is recognised that these five competencies are interconnected and mutually dependent.

Conceptual understanding is defined as the grasp of mathematical concepts, operations, and relations. Procedural fluency, on the other hand, is the ability to execute procedures in a flexible, accurate, efficient, and appropriate manner (National Research Council, 2001). Studies have indicated that these two aspects, namely conceptual understanding and procedural fluency, are crucial for the development of mathematical proficiency (Hiebert & Grouws, 2007). As a result, there is a demand for research in pre-service mathematics education, particularly in the South African context, that explores which teaching strategies could enhance these competencies in pre-service mathematics education students for school-level mathematics topics.

There exist four theoretical perspectives regarding the relationship between procedural and conceptual knowledge (Rittle-Johnson et al., 2015; Schneider et al., 2011). Each theory is backed by certain empirical evidence, but at the same time, other evidence contradicts them. The concepts-first theory posits that students initially gain conceptual knowledge in a field and then use this understanding to formulate procedures for problem-solving. Conversely, the procedures-first theory suggests that students initially learn problem-solving procedures in a field and then gradually build conceptual knowledge through repeated problem-solving (Rittle-Johnson et al., 2001).

This study aligns with the iterative model as suggested by Rittle-Johnson et al. (2001). A key tenet of this theory is that either procedural or conceptual knowledge can develop first, but it is not a rule that one type of knowledge always precedes the other. The authors maintain that it is common for a specific type of knowledge to be incomplete (for a topic). In other words, one type of knowledge might be more advanced at a certain moment, but this doesn't mean that the other type of knowledge is entirely lacking. As a rule, initial knowledge in any field is quite limited. The argument is that the levels of pre-existing knowledge in a field determine which type of knowledge will surface first, thereby initiating the learning process.

The argument is made that conceptual knowledge, being general and abstract, can be applied to new types of problems (Schneider et al., 2011). Procedural knowledge, on the other hand, is believed to be more specific to a particular problem

type, as procedures are typically practised with a specific problem type. Therefore, it is suggested that if students have some pre-existing knowledge of the material to be learned, then conceptual knowledge might have a larger role in the development of procedural knowledge, and not the other way around. Students with minimal prior knowledge in a domain tend to first develop procedural knowledge, which then aids in the development of conceptual knowledge. For instance, when students are introduced to the topic of linear functions, they often first learn the procedures (which includes some initial conceptual knowledge), and then, through exposure to different problem types in this domain, they develop their conceptual understanding. Hence, according to the iterative model, improved conceptual knowledge leads to improved procedural knowledge. This improved procedural knowledge, in turn, leads to improved conceptual knowledge, and this cycle continues.

As previously stated, when students first encounter a mathematical topic, their initial understanding of the topic is usually quite limited. As a result, it is common for students to have some knowledge about a topic, but not a complete understanding. Conceptual knowledge is crucial for the creation, selection, and proper application of procedures. However, practising established procedures is believed to assist students in developing and deepening their understanding of concepts. The primary argument is that both types of knowledge are necessary for effective learning in mathematics and, over time, each type of knowledge needs to reinforce the other. Furthermore, if a student's conceptual knowledge about a mathematical topic hasn't yet matured, they will likely rely heavily on this conceptual knowledge when devising a solution procedure for a given mathematical task. Therefore, mathematics instruction should aim to develop both procedural and conceptual knowledge.

A student with limited prior knowledge in a mathematical field typically possesses disjointed knowledge and lacks the ability to perceive how procedures and concepts interrelate within the field. As the knowledge within a domain expands, so does the ability to merge the pieces of conceptual and procedural knowledge into a unified knowledge structure (Eds. Baroody & Dowker, 2003; Linn, 2006; Schneider & Stern, 2009; Schneider et al., 2011). Theories commonly used to elucidate retrieval practice involve the concepts of storage and retrieval strength. Storage strength pertains to the permanency of knowledge, while retrieval strength indicates the ease or difficulty of recalling knowledge. A model grounded in these theories posits that retrieval strength has a negative correlation with increases in storage strength (Bjork & Bjork, 1992). This suggests that strenuous retrieval (low retrieval strength) bolsters storage strength and fosters long-term learning, whereas easy retrieval is linked with lower levels of storage strength. 'Desirable difficulties', such as extending the interval between retrieval practice sessions, are thought to exemplify instances of strenuous retrieval. In the present study intervals between retrieval practice sessions were approximately three weeks. Long retention

intervals such as this introduce desirable difficulties since they require effortful retrieval.

The taxonomy table

A taxonomy table was employed to classify test and examination items based on their perceived level of difficulty. The table was initially based on Bloom's original taxonomy. This taxonomy comprised six primary categories: knowledge, comprehension, application, analysis, synthesis, and evaluation. In this original taxonomy, the categories were arranged from the simplest to the most complex, and from the most concrete to the most abstract. The underlying assumption was that the mastery of learning objectives followed a hierarchical pattern. That is, understanding each simpler category was a necessary step before grasping the next, more complex category (Krathwohl, 2002).

In the original taxonomy, statements of learning objectives typically comprised a noun or noun phrase, and a verb or verb phrase (Krathwohl, 2002). These noun and verb elements were linked to the knowledge category in Bloom's original taxonomy. This meant that the knowledge category had a dual nature, setting it apart from the other categories. To eliminate this inconsistency, the noun and verb elements were divided into two dimensions in Bloom's revised taxonomy. In this revised version, the noun formed the foundation for the knowledge dimension, while the verb established the basis for the cognitive process dimension (Anderson et al., 2001). Using these two dimensions, a two-dimensional table, referred to as the taxonomy table, was created. In this table, the knowledge dimension constituted the vertical axis, while the cognitive process dimension made up the horizontal axis. The cells of the table were formed by the intersections of the knowledge and cognitive process categories.

The taxonomy table's knowledge dimension primarily consists of four categories: Factual Knowledge, Conceptual Knowledge, Procedural Knowledge, and Meta-cognitive Knowledge. Conversely, the Cognitive Process dimension is mainly composed of: Remember, Understand, Apply, Analyse, Evaluate, and Create. It's crucial to note that within the taxonomy table, cognitive processes are perceived to function on knowledge structures during cognitive processing. For instance, one might refer to 'understanding' as it pertains to 'conceptual knowledge' when using the taxonomy table.

The categories of the Knowledge and the Cognitive Process labelled in the revised taxonomy table did not fully meet the requirements of the current study. As a result, modifications were made to the table to better align with the study's needs. In my adapted version of the taxonomy table, I retained some of the Knowledge dimension categories, but eliminated most of the Cognitive Process dimension categories. The reason for this was to concentrate on the categories of cognition that are prevalent in mathematical cognition. Similarly, the knowledge categories I included are those that are most common in

mathematics. In my version, sub-categories are not included in the Knowledge dimension, but they are incorporated in the Cognitive Process dimension. Moreover, since this study focuses on the teaching and learning of mathematics, the knowledge categories and cognitive process categories are viewed from a mathematical perspective. The categories of my revised Knowledge dimension include: Factual Knowledge, Procedural Knowledge, Flexible Procedural Knowledge, and Conceptual Knowledge. Each category will be discussed in more detail in the following sections. Factual Knowledge encompasses understanding of terminology, the structure and syntax of symbol representation, permissible operations, and so on (Eisenhart et al., 1993).

Flexible procedural knowledge refers to profound procedural knowledge that enables a student who has it to apply the most appropriate mathematical procedures to a given familiar or new problem scenario (Star, 2005).

In the reasoning structure proposed by Lithner (2008), reasoning is characterised as the thought process necessary to formulate assertions and draw conclusions in the resolution of mathematical tasks. Within this structure, reasoning pertains to the cognitive processes, the outcome of these processes, or both. Thinking can be classified as reasoning within this framework, even if it is incorrect, as long as it is logical to the thinker.

Lithner's (2008) framework identifies two distinct types of reasoning. Imitative Reasoning (IR) is when a student generates a solution procedure that they have memorised. Conversely, Creative Reasoning (CR) is a type of reasoning marked by flexibility and innovative approaches to solving mathematical problems (Bergqvist, 2007).

Lithner (2008) distinguishes between two primary categories of IR: Memorised Imitative Reasoning and Imitative Algorithmic Reasoning (AR). For a reasoning to be classified as memorised, it must meet two conditions: the reasoning should be based on the recall of a complete answer, and the implementation strategy should involve only writing down the answer.

If a reasoning sequence is based on recalling an algorithm, it falls under Imitative Algorithmic Reasoning (AR). To be classified as AR, it must meet two conditions: the strategic choice for the reasoning should involve recalling an algorithm as a solution, and no other reasoning should be required except to implement the algorithm.

Algorithmic Reasoning is further divided into two sub-categories: Familiar Algorithmic Reasoning (FAR) and Delimiting Algorithmic Reasoning (DAR). If a mathematical task is identified as a familiar type and can be solved using a known algorithm, then it requires FAR. Conversely, if a task requires the student to select from a set of algorithms, then the reasoning required is DAR.

Creative Reasoning (CR) is defined as a type of reasoning that is not restricted by fixation, but rather is marked by flexibility and innovative methods for solving mathematical problems (Bergqvist, 2007). If a task primarily requires CR, the reasoning involved is categorised as Global Creative Reasoning (GCR). On the other hand, if a task is almost solvable using IR and only needs CR to adjust an algorithm, then the reasoning required is termed Local Creative Reasoning (LCR). The taxonomy table, which is based on the aforementioned discussion, is presented in Table 1.

Method

This study employed convenience sampling to choose its research participants. The participants were a group of second-year pre-service teachers from the 2014 cohort who were enrolled in a mathematics course taught by the researcher. The research was conducted over both semesters of 2014. Initially, 88 students were chosen to participate, but ultimately, only 63 students were included in the study. Data from students who missed tests and examinations were excluded. The study covered various topics, including but not limited to: analytical geometry, functions, remainder and factor theorems, Euclidean geometry, and matrices. With the exception of matrices all of the topics are high school-level mathematics topics in South Africa.

The research participants had a wide range of mathematical knowledge and skills upon leaving school. A portion of the participants had completed Mathematics up to Grade 12.

These students ($n = 44$) were perceived to have high prior topic knowledge whereas students who did Mathematical Literacy up to Grade 12 were considered to have low prior topic knowledge ($n = 19$). The sample was composed of 30 male and 33 female participants, with an average age of 23. The statistical analysis only included students who had completed all the assessment components. So for the statistical analysis high prior topic knowledge participants numbered 40 and low prior topic knowledge participants numbered 18.

TABLE 1: The taxonomy table.

The knowledge dimension (KD)	The cognitive process dimension (CPD)				
	1. Imitative reasoning (IM)		2. Creative mathematically founded reasoning		
	(a.) Memorised reasoning (MR)	(b.) Algorithmic reasoning (AR)		(a.) Local creative reasoning (LCR)	(b.) Global creative reasoning (GCR)
		(i) Familiar (FAR)	(ii) Delimiting (DAR)		
A.) Factual knowledge	-	-	-	-	
B.) Procedural knowledge	-	-	-	-	
C.) Flexible procedural knowledge	-	-	-	-	
D.) Conceptual knowledge	-	-	-	-	

The study's quantitative data were derived from test and examination scores. The research followed a quasi-experimental method, meaning that the independent variable was not altered, and there was no random group assignment (Johnson & Christensen, 2012). While the quasi-experimental method is seen as providing less robust evidence of a cause-and-effect relationship between variables, it remains crucial in educational research. This is because many educational research issues are not suitable for strict experimental methods (Kerlinger, 1986).

The study utilised a design that included pre-tests and post-tests with repeated measures. Four class tests administered per semester collectively constituted a pre-test, while the examinations (two per semester) written at the end of each semester were considered a post-test. The examinations were written about three and a half months after the first test. The first test covered only the content that had been taught prior to the test, while the second test included both the content taught before it and the content covered in the first test. Similarly, the third test covered the content from the first and second tests, as well as the content taught prior to it, and so on. As a result, the content taught before the first test was tested four times, while the content taught after the second test was tested three times, and so on. This indicates that the retrieval practice in the study was repeated. Given that the interval between tests was approximately three weeks, it suggests that the retrieval practice was also spaced (distributed). The fact that research participants had to revisit content covered in previous tests implies a high likelihood of test-potentiated learning occurring.

Questions from tests and examinations were sorted using a revised taxonomy table, which is grounded in Bloom's taxonomy (Table 1). This revised taxonomy table served as the tool for measuring the knowledge and reasoning proficiency of the research participants. The classification of problems was influenced by the number of previous practice sessions, the timing, and the perceived primary knowledge and reasoning requirements of the problem. To clarify the above statement, consider this: a problem might initially be classified as DAR based on Flexible Procedural Knowledge (C1bii), but after students have encountered it multiple times, the knowledge requirements shift to FAR, and the reasoning type becomes Ordinary Procedural Knowledge. As a result, the new classification is FAR based on Procedural Knowledge (B1bi), a category that is perceived to be less challenging.

After the assessment pieces were marked, each participating student received a score in each category for every individual class test and examination. Subsequently, all the individual scores for each category were summed to determine a total. This summing process was carried out separately for class tests and examinations. For instance, all individual scores for the category FAR based on Procedural Knowledge (B1bi) were added together to yield a total sum for this category per student. Following this, the sums for the categories Memorised Reasoning based on Factual knowledge (A1a), Memorised Reasoning based on Procedural Knowledge

(B1a), FAR based on Procedural Knowledge (B1bi), FAR based on Flexible Procedural Knowledge (C1bi), and DAR based on Flexible Procedural Knowledge (C1bii) were added sequentially to provide a grand total. This was done separately for pre-tests and post-tests. This grand total was deemed to represent measures of the mathematical competency procedural fluency. In the statistical analysis, the grand total for the pre-test was labelled as SKILLPRE and for the post-test, it was labelled as SKILLPOST.

In a similar manner, the total scores for the categories such as Memorised Reasoning based on Conceptual Knowledge (D1a), FAR based on Conceptual Knowledge (D1bi), DAR based on Conceptual Knowledge (D1bii), and Local CR based on Conceptual Knowledge (D2a) were combined to yield a new overall sum. This overall sum was then further added to generate a final sum total. This process was carried out separately for both pre-tests and post-tests, similar to the previous case. This final sum total was interpreted as a measure of the mathematical competency known as Conceptual Understanding. In the statistical analysis, the final sum total for the pre-test was labelled as CONCPRE and for the post-test, it was labelled as CONCPOST.

It is imperative that the difficulty level of test items in the post-test matches or exceeds that of the pre-test. If this is not the case then a false measurement of improvement may be determined. As indicated, the difficulty level of all questions in tests and exams was determined using the taxonomy table. The researcher used the taxonomy table to set standards for tests and exams. For example, 15,55% of test items of the pre-test of the first semester consisted of the category of question DAR based on Conceptual Knowledge (D1bii) whereas for the post-test the percentage was 34%. This category is perceived to be of a higher difficulty level. Post-test items on average were set at a higher difficulty level than for pre-tests to mitigate for the possibility of false measurement of improvement.

The following exemplars based on the topic of matrices (linear algebra) are used to illustrate the arguments regarding item difficulty level:

- An example of an item categorised as FAR based on Procedural Knowledge (perceived to be less difficult):

$$\text{Given the following system of linear equations: } \begin{cases} x + y - z = 2 \\ 2x + 3y - z = 7 \\ 3x - 2y + z = 9 \end{cases}$$

- Write an augmented matrix for the system.
- Use Gauss-Jordan elimination to solve the system of equations.

- An example of a more difficult item based on the same topic. This item would be categorised as DAR based on Flexible Procedural Reasoning:

Use a system of equations to find the quadratic function $f(x) = ax^2 + bx + c$ that satisfies the equations: $f(1) = 1$; $f(2) = -1$; $f(3) = -5$

- An example from the category DAR based on Conceptual Knowledge (perceived to be more difficult since it requires well-developed conceptual knowledge):

State whether the following statement is true or false. Provide a justification for your answer.

If two columns of a square matrix are the same, then the determinant of the matrix will be zero.

The teaching strategy

In this section the teaching strategy that was employed is discussed. Arguments are presented to motivate why this particular teaching strategy was employed.

Yeo and Fazio (2019) argue that the best learning approach is contingent on both the type of knowledge being acquired (whether it's stable facts or adaptable procedures) and the pertinent learning processes (whether it's schema induction or memory and fluency enhancement). The most effective learning strategy is determined by the learning goals.

These include learning to retain new information (memory building), learning a new problem-solving strategy (schema induction), learning to solve analogous problems (fluency building), and inducing a mathematical concept (schema induction).

It was the intention with the design of the teaching strategy to optimise learning opportunities for participants. As mentioned previously, the second test included content that was tested in the first test and the third test included content that was tested in the second and first tests, etc. This way of doing things allowed for the inclusion of analogous problems in tests which facilitated fluency building (i.e. enhancing procedural fluency). Testing on the same content repeatedly also offered the opportunity to participants to learn from previous tests (test-potentiated learning). For example, if a student could not provide correct responses to a test item in a test the student would be compelled to determine what the correct response was in preparation for the following test since test items would be included on the same content. The presentation of analogous problems requires the repeated retrieval of schemata regarding the same procedural and conceptual knowledge. Such cognitive exercises enhance retention since memory is 'refreshed' with each retrieval attempt. In addition, since the problems do not include exactly the same information, it will be stored in a different way. For example, students were exposed to the remainder and factor theorems of polynomial functions. The factor theorem they were exposed to was given as follows:

A polynomial $f(x)$ has a factor $(x - h)$ if and only if $f(h) = 0$

In classroom exercises they were required to solve problems based on the factor and remainder theorems. This included problems such as the following two examples:

Show that $x - 3$ is a factor of $f(x) = 2x^2 + 3x^2 - 11x - 6$

For which values of m will $x - 3$ be a factor of $x^3 + m^2x^3 - 11x - 15m$?

In a subsequent test students were presented with the problem:

Find p if $k + 2$ is a factor of $k^{50} - p^{25}$

The solution to this problem requires an understanding that since the factor theorem is an *if-and-only-if* mathematical statement it consists of two converse statements. The one statement being:

If a polynomial has a factor of $x - h$, then $f(h) = 0$

Its converse is:

If $f(h) = 0$, then $x - h$ is a factor of polynomial $f(x)$

Both of these statements are required for the solution. That is, since $k + 2$ is a factor, -2 should be substituted:

$$f(-2) = (-2)^{50} - p^{25}.$$

Next, the function should be equated to 0. Using the other statement:

$$(-2)^{50} - p^{25} = 0$$

The remainder of the solution requires an understanding of exponential laws. Now if a student provided a correct response to the problem it would imply that the student made a connection with the implicit conceptual information provided in the factor theorem. That is, that the theorem consists of two converse statements both of which are required for the solution procedure. The next part of the solution procedure requires a connection with exponential laws and their manipulation. This in turn implies that the student cognitively connected information concerning the factor theorem in a new way. In other words, schemata concerning the factor theorem were changed (schema induction). That is, the test item has resulted in learning. This is contrary to the conventional belief that tests are only used to assess learning.

Now let us suppose a student could not provide a solution. The student therefore would have identified a lack of knowledge regarding the solution procedure to such problems in themselves. It is highly probable that the student would then proceed to seek help to find a solution. This implies that test-potentiated learning has taken place.

Statistical analysis

IBM SPSS version 28.0 was utilised for the statistical analysis which follows. The focus of this study, as mentioned earlier, is to evaluate the differences in knowledge and reasoning types between students with high and low prior knowledge on the topic, after they have been subjected to a teaching method that emphasises retrieval practice and test-enhanced learning.

Statistical analysis was done as follows:

The comparison of scores from pre-tests and post-tests was conducted to assess the impact of the teaching method on

students' knowledge and reasoning abilities. This was done separately for low and high prior topic knowledge participants. As previously mentioned, all class tests were collectively used as the pre-test, while the end-of-semester exams were combined to form the post-test. The updated taxonomy table was employed to classify items from the tests and exams. The categories within this taxonomy table served as the primary measurement tool in this study. The dependent variable in this study is the scores from the pre-tests and post-tests, while the independent variable is the teaching approach that is grounded in retrieval practice and test-enhanced learning.

Initially, descriptive statistics were employed to scrutinise the data and identify any breaches of the assumptions inherent in statistical tests. Subsequently, a paired samples *t*-test was conducted to ascertain if significant disparities existed between the pre-test and post-test scores. This was done separately for participants with low and high prior knowledge of the topic, and also separately for procedural fluency and conceptual understanding.

An independent samples *t*-test was carried out to compare the average scores of participants with low and high prior knowledge on the post-test. This comparison was made separately for both procedural fluency and conceptual understanding. The purpose of this was to identify if a significant variance exists between the scores of participants with low and high prior topic knowledge after they were exposed to the teaching method.

Paired samples *t*-test (low prior topic knowledge)

As mentioned previously, to determine if there is a significant difference between low and high prior topic knowledge scores after exposure to the teaching strategy a paired samples test (*t* statistic) that is based on the overall mean difference (μ_D) was utilised.

The null hypothesis is that there is no significant difference after exposure to the teaching strategy. In other words, the mean difference of the pre-test and post-test scores for the population is zero:

$$H_0 : \mu_D = 0$$

TABLE 3: Paired samples test (low prior topic knowledge participants).

Pair	Variable	Paired differences					<i>t</i>	<i>df</i>	Significance	
		Mean	Standard deviation	Standard error mean	95% confidence interval of the difference				One-sided <i>p</i>	Two-sided <i>p</i>
					Lower	Upper				
1	SKILLPOST – SKILLSPRE	15.41	20.84	4.91	5.05	25.78	3.139	17	0.003	0.006
2	CONCPOST– CONCPRE	8.39	9.79	2.31	3.52	13.26	3.634	17	0.001	0.002

TABLE 4: Paired samples effect sizes (low prior topic knowledge participants).

Pair	Variable	Measure	Standardiser†	Point estimate	95% confidence interval	
					Lower	Upper
1	SKILLPOST – SKILLSPRE	Cohen's <i>d</i> ‡	20.84	0.74	0.208	1.26
		Hedges' correction §	21.82	0.71	0.198	1.199
2	CONCPOST– CONCPRE	Cohen's <i>d</i>	9.79	0.86	0.304	1.39
		Hedges' correction	10.25	0.82	0.291	1.33

†, The denominator used in estimating the effect sizes; ‡, Cohen's *d* uses the sample standard deviation of the mean difference; §, Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

The alternative hypothesis is that the intervention caused the post-test scores to be higher or lower than the pre-test scores. In other words, the mean difference is not zero:

$$H_1 : \mu_D \neq 0$$

The level of significance is set at $\alpha = 0.05$ for a two-tailed test.

The paired samples statistics for the SKILL and CONC variables (low prior topic knowledge participants) is presented in Tables 2, 3 and 4.

A paired samples *t*-test was conducted to evaluate the impact on low prior topic knowledge participant scores of a teaching intervention based on retrieval practice and test-potentiated learning. There is a statistically significant increase in scores for the variable SKILL (procedural fluency) from pre-test ($M = 53.32$, $SD = 19.65$) to post-test ($M = 68.73$, $SD = 11.7$) $t(18) = 3.139$, $p = 0.006$ (two-tailed). The mean increase in scores for the skill variable is 15.41 with a 95% confidence interval ranging from 5.05 to 25.78. The value for Cohen's *d* is 0.74 which is close to a large effect size. The null hypothesis for the SKILL variable is rejected. In other words $H_1 : \mu_D \neq 0$.

Next we discuss the statistics for the competency conceptual understanding. There is a statistically significant increase in scores for the variable CONC (conceptual understanding) from pre-test ($M = 35.90$, $SD = 12.81$) to post-test ($M = 44.29$, $SD = 13.56$), $t(18) = 3.634$, $p = 0.002$ (two-tailed). The mean increase in scores for the CONC variable is 8.39 with a 95% confidence interval ranging from 3.52 to 13.26. The value for Cohen's *d* is 0.86 which is a large effect size. The null hypothesis for the CONC variable is rejected. That is $H_1 : \mu_D \neq 0$.

TABLE 2: Paired samples statistics (low prior topic knowledge participants) ($N = 18$).

Pair	Variable	Mean	Standard deviation	Standard error mean
1	SKILLPOST	68.74	11.70	2.76
	SKILLSPRE	53.32	19.65	4.63
2	CONCPOST	44.29	13.56	3.195
	CONCPRE	35.90	12.81	3.02

Paired samples *t*-test (high prior topic knowledge)

The paired samples statistics for the high prior topic knowledge group is presented in Tables 5, 6 and 7.

A paired samples *t*-test was conducted to evaluate the impact on high prior topic knowledge participant scores of a teaching intervention based on retrieval practice and test-potentiated learning. There is a statistically significant increase in scores for the SKILL variable (procedural fluency) from pre-test ($M = 73.38, SD = 18.56$) to post-test ($M = 84.71, SD = 13.29$), $t(40) = 5.87, p < 0.001$ (two-tailed). The mean increase in scores for the skill variable is 11.33 with a 95% confidence interval ranging from 7.43 to 15.25. The value for Cohen’s *d* is 0.93 which is a large effect size. The null hypothesis for the SKILL variable is rejected. In other words $H_1 : \mu_D \neq 0$.

The statistics for the CONC variable are as follows. There is a statistically significant increase in scores for the variable CONC (conceptual understanding) from pre-test ($M = 56.36, SD = 17.05$) to post-test ($M = 69.09, SD = 13.98$), $t(40) = 7.14, p < 0.001$ (two-tailed). The mean increase in scores for the CONC variable is 12.66 with a 95% confidence interval ranging from 9.08 to 16.25. The value for Cohen’s *d* is 1.13 which is a large effect size. The null hypothesis for the CONC variable is rejected. That is, $H_1 : \mu_D \neq 0$.

TABLE 5: Paired samples statistics (high prior topic knowledge) ($N = 40$).

Pair	Variable	Mean	Standard deviation	Standard error mean
1	SKILLPOST	84.71	13.29	2.100
	SKILLSPRE	73.38	18.56	2.940
2	CONCPOST	69.03	13.78	2.180
	CONCPRE	56.36	17.05	2.695

TABLE 6: Paired samples test (high prior topic knowledge).

Pair	Variable	Paired differences					<i>t</i>	<i>df</i>	Significance	
		Mean	Standard deviation	Standard error mean	95% confidence interval of the difference				One-sided <i>p</i>	Two-sided <i>p</i>
					Lower	Upper				
1	SKILLPOST – SKILLSPRE	11.33	12.22	1.93	7.43	15.24	5.87	39	< 0.001	< 0.001
2	CONCPOST– CONCPRE	12.66	11.21	1.77	9.08	16.25	7.14	39	< 0.001	< 0.001

TABLE 7: Paired samples effect sizes (high prior topic knowledge).

Pair	Variable	Measure	Standardiser †	Point estimate	95% confidence interval	
					Lower	Upper
					1	SKILLPOST – SKILLSPRE
Hedges’ correction §	12.46	0.91	0.541	1.270		
2	CONCPOST– CONCPRE	Cohen’s <i>d</i>	11.21	1.13	0.727	1.523
		Hedges’ correction	11.43	1.11	0.713	1.493

†, The denominator used in estimating the effect sizes; ‡, Cohen’s *d* uses the sample standard deviation of the mean difference; §, Hedges’ correction uses the sample standard deviation of the mean difference, plus a correction factor.

TABLE 9: Independent samples test.

Variable	Equal variances	Levene’s test for equality of variances		<i>t</i>	<i>df</i>	<i>t</i> -test for equality of means					
		<i>F</i>	Significance			Significance		Mean difference	Standard error difference	95% confidence interval of the difference	
						One-sided <i>p</i>	Two-sided <i>p</i>			Lower	Upper
SKILLPOST	Assumed	0.042	0.84	4.39	56.00	< 0.001	< 0.001	15.98	3.64	8.68	23.270
	Not assumed	-	-	4.61	37.02	< 0.001	< 0.001	15.98	3.47	8.95	23.000
CONCPOST	Assumed	0.408	0.53	6.36	56.00	< 0.001	< 0.001	24.73	3.89	16.94	32.530
	Not assumed	-	-	6.395	33.34	< 0.001	< 0.001	24.73	3.87	16.87	32.597

Independent samples *t*-test

As indicated, an independent samples *t*-test was performed to compare the mean scores of the low and high prior knowledge participants’ post-tests. This was done separately for the competencies procedural fluency and conceptual understanding. The statistics are shown in Tables 8 and 9.

For the independent samples *t*-test the null hypothesis is that there is no significant difference after exposure to the teaching strategy. In other words, the mean difference of the low and high prior topic knowledge groups is zero:

$$H_0 : \mu_D = 0$$

The alternative hypothesis is that the intervention caused the mean difference to be non-zero:

$$H_1 : \mu_D \neq 0$$

The level of significance is set at $\alpha = 0.05$ for a two-tailed test.

Levene’s test for equality of variances tests whether the variance of scores for the two groups is the same. The outcome of this test determines which of the *t* values is the correct one to use. If the significance value for Levene’s test is larger than 0.05 the first row is used, otherwise the second row is used.

TABLE 8: Group statistics.

Variable	V5	<i>N</i>	Mean	Standard deviation	Standard error mean
SKILLPOST	1	40	84.71	13.29	2.100
	2	18	68.74	11.70	2.760
CONCPOST	1	40	69.03	13.78	2.800
	2	18	44.29	13.56	3.195

An independent samples *t*-test was conducted to compare SKILL scores (after exposure to the teaching strategy) between low and high prior topic knowledge groups. There was a significant difference in scores for low prior topic knowledge participants ($M = 63.74$, $SD = 11.70$) and high prior topic knowledge participants ($M = 84.71$, $SD = 13.29$); $t(56) = 4.39$, $p < 0.001$ (two-tailed).

The magnitude of the differences in the means is 15.98 with a 95% confidence interval of 8.68–23.27. Therefore, the null hypothesis is rejected. That is $H_1 : \mu_D \neq 0$.

An independent samples *t*-test was conducted to compare CONC scores (scores after exposure to the teaching strategy) between low and high prior topic knowledge groups. There was a significant difference in scores for low prior topic knowledge participants ($M = 44.29$, $SD = 13.56$) and high prior topic knowledge participants ($M = 69.03$, $SD = 13.78$); $t(56) = 6.36$, $p < 0.001$ (two-tailed). The magnitude of the differences in the means is 24.73 with a 95% confidence interval of 8.68–23.27. Consequently, the null hypothesis is rejected. In other words, $H_1 : \mu_D \neq 0$.

It is possible that the data might be conflated in terms of those topics that were new to both groups. To investigate this possibility, data for only those topics that were new to both high and low prior knowledge groups were analysed separately and were then compared with an analysis of data for the ‘familiar’ topics. A paired samples test was utilised for this purpose. It should be noted that not all the statistics are presented in the tables other than those that are relevant for the discussion. Table 10 shows the analysis for the low prior knowledge group whereas Table 11 shows the analysis for the high prior knowledge group and Table 12 shows the analysis for all participants.

A comparison of Tables 10, 11 and 12 indicates that there was a net increase for the unfamiliar topics. Likewise, if the mean differences of the unfamiliar topics are compared to those of the familiar topics, then the increase for the unfamiliar topics was greater. Moreover, the low prior topic knowledge

TABLE 10: Low prior topic knowledge paired samples statistics.

Topics	Mean		Mean difference
	Pre	Post	
Unfamiliar	67.644	68.596	-0.932
Familiar	47.89	50.556	-2.666

TABLE 11: High prior topic knowledge paired samples statistics.

Topics	Mean		Mean difference
	Pre	Post	
Unfamiliar	63.808	59.859	3.9495
Familiar	63.937	59.410	4.5270

TABLE 12: All participants paired samples statistics.

Topics	Mean		Mean difference
	Pre	Post	
Unfamiliar	64.687	61.850	2.8375
Familiar	59.626	57.032	2.5944

participants had a higher mean score than their high prior knowledge counterparts for the unfamiliar topics. This holds true for both the pre-test and the post-test. However, for the familiar topics the reverse is true.

Another possible confounding concerns the amount of retrieval practice for the different topics. The concern is that topics covered near the beginning of the study would enjoy more retrieval practice than topics dealt with near the end of the study. The topics at the end would therefore not be influenced as much by the retrieval practice. To investigate this issue, only data from topics that were tested two or more times were used in an additional statistical analysis. Table 13 shows the analysis for the low prior knowledge participants whereas Table 14 shows the statistics for high prior knowledge participants. It should be noted that not all the statistics are presented, only those that are relevant to the discussion.

A comparison of the statistics of the new analysis with the previous analysis shows that for the low prior knowledge group there were slight increases for both SKILL and CONC variables. A similar scenario holds for the high prior knowledge group. The exception being a small decrease for the SKILLPOST variable. Since there is no major difference when the analysis is done without the data where less retrieval practice was done, one can conclude that the analysis was not conflated to the extent that all previous arguments based on the original analysis must be discarded.

Discussion and conclusion

The statistical analysis clearly shows that students with low prior topic knowledge started the course with a significantly low base (53%) in the competency of procedural fluency, compared to their high prior topic knowledge counterparts where the competency was at 73% pre-intervention (a 20% difference). For the competency conceptual understanding the low prior topic knowledge students entered with an even lower base at 36% compared with 56% (a 20% difference) for the high prior topic knowledge students.

Post-intervention procedural fluency improved for both low and high prior topic knowledge students. There was an increase from 53% to 69% (30 percentage points) for the former and an increase from 73% to 85% (16 percentage

TABLE 13: Low prior topic knowledge paired samples statistics ($N = 18$).

Pair	Variable	Mean	Standard deviation	Standard error mean
1	SKILLPOST total (%)	71.308	8.6072	2.0287
	SKILLSPRE total (%)	60.925	14.9772	3.5301
2	CONCPOST total (%)	44.444	15.1339	3.5671
	CONCPRE total (%)	35.022	12.5881	2.9670

TABLE 14: High prior topic knowledge paired samples statistics ($N = 40$).

Pair	Variable	Mean	Standard deviation	Standard error mean
1	SKILLPOST total (%)	82.342	15.7214	2.4858
	SKILLSPRE total (%)	78.912	12.1798	1.9258
2	CONCPOST total (%)	75.059	15.1699	2.3986
	CONCPRE total %	60.320	17.0735	2.6996

points) for the latter. For conceptual understanding post-intervention there were increases from 36% to 44% (22 percentage points) and from 56% to 69% (23 percentage points). Post-intervention for the competency procedural fluency the difference between the two groups decreased to 16% whereas for the competency conceptual understanding it increased to 24%.

It is not easy to attribute causality using a pre-test post-test quasi-experimental design. Also as argued earlier since the independent variable was not manipulated and there was no random assignment to groups it is not easy to attribute causality. However, if all possible rival explanations can be eliminated then one can make a case for causality. One rival explanation for the findings could be that difficulty level of post-test items was lower than that for pre-test items. However, it was indicated previously that the percentage of more difficult items was more for the post-test than for the pre-test. Hence the improvement in scores cannot be attributed to difficulty level. Furthermore, if a student performs badly in a test then there is a chance to make up in the next test. However, should a student perform badly in the exam there is no second chance. It could be argued therefore that anxiety for the exam would be higher. Since the exam covers all the work and the difficulty level was higher the expectation was that post-test scores would be lower than that of the pre-tests. Since this was not the case there has to be some variable that is attributable to the improvements.

Another possible rival explanation could be good studying methods. Since tests covered less material and were closer to when it was presented the expectation is that performance would be better in the pre-tests. Since performance in post-tests was better this argument is rejected.

It can also be argued that exceptionally good lecturing caused the improvement. However, the argument can again be presented that performance in tests should be better if this was the case. Since the exams took place long after some of the lecturing sessions a major amount would be forgotten. This argument therefore is also rejected.

When the data regarding content that were tested only once (or not at all) was ignored, the scores improved. There was therefore weaker performance in those tests where there was less retrieval practice. Since there are no plausible rival explanations for the findings, the contention is that the improvements are attributable to retrieval practice and test-potentiated learning.

The research question for the study was how do high and low prior topic knowledge students compare in terms of the mathematical competencies of procedural fluency and conceptual understanding after exposure to a teaching strategy based on retrieval practice and test-potentiated learning? The findings show that the teaching strategy enhanced procedural fluency for both groups but the low prior topic knowledge group showed a higher increase.

This finding is consistent with the literature. That is, that benefits of retrieval practice are greatest for unfamiliar content (Cogliano et al., 2019). The Mathematical Literacy students entered the course with weakly developed procedural knowledge and hence would have found much of the course content unfamiliar.

The findings also indicate that there was significant enhancement of the competency of conceptual understanding for both groups post-intervention. The low prior topic knowledge participants however still had a mean below the pass cut-off (44%) whereas the high prior topic knowledge group improved to close to 70%. The difference between the groups also increased for this competency signifying that the low prior topic knowledge group was still at an early developmental stage for this competency.

The iterative model for the development of procedural and conceptual knowledge, proposed by Rittle-Johnson et al. in 2001, suggests that students with low prior knowledge in a particular field are likely to first develop procedural knowledge, which then aids in the cultivation of conceptual knowledge. The findings of this study support this proposition. That is, the findings show that the development of procedural knowledge of low prior topic knowledge students is ahead of their development of conceptual knowledge. This can be seen in the fact that the improvement for procedural fluency is nearly double that of their conceptual understanding (15% versus 8%). This shows that for the low prior topic knowledge students the ability to integrate conceptual understanding and procedural fluency is at an early developmental stage. Evidence for this is the low base conceptual understanding. The fact that it is increasing is evidence that it is developing. This aligns with existing literature which suggests that as competencies in procedural fluency and conceptual understanding improve, the capacity to merge conceptual and procedural knowledge structures into a unified knowledge structure also enhances (Eds. Baroody & Dowker, 2003; Linn, 2006; Schneider & Stern, 2009; Schneider et al., 2011). Hence although the development is still low it is moving in the right direction. This was the ultimate objective of the study: the development of procedural fluency and conceptual understanding of participants.

As indicated previously, findings of some research indicates that students with different levels of prior topic knowledge can have different learning outcomes after exposure to the same learning materials and teaching strategy (Xiaofeng et al., 2016). There is also evidence that students with high prior topic knowledge can develop better understanding of new content (Dunlosky et al., 2013) and are able to remember more than individuals with less prior knowledge (Thompson & Zamboanga, 2003). Findings from the present study corroborate these findings. The low and high prior topic knowledge students had different levels of development of procedural fluency and conceptual understanding after exposure to the teaching strategy.

The high post-intervention mean scores for the high prior topic knowledge students are evidence that these students are able to develop better understanding of new content and are able to remember more.

The foregoing discussion provides answers to the questions as to which type of knowledge is retained better and understood better after exposure to a teaching strategy based on retrieval practice and test-potentiated learning. The statistical results show that for low prior topic knowledge students procedural fluency is enhanced more and retained more than conceptual understanding. As indicated previously the competency of procedural fluency is predicated on Memorised Reasoning based on Factual Knowledge, Memorised Reasoning based on Procedural Knowledge, FAR based on Procedural Knowledge, FAR based on Flexible Procedural Knowledge, and DAR based on Flexible Procedural Knowledge. Therefore, these categories of reasoning and knowledge were privileged by the teaching strategy for the low prior topic knowledge group. The findings were similar for the high prior topic knowledge group.

Likewise as indicated previously, the competency of conceptual understanding is based on Memorised Reasoning based on Conceptual Knowledge, FAR based on Conceptual Knowledge, DAR based on Conceptual Knowledge and Local CR based on Conceptual Knowledge. Although both groups showed improvements in these categories, it was much higher for the high prior topic knowledge group. The improvement for these categories for the high prior topic knowledge group was also higher than their procedural fluency categories, whereas it was the reverse for the low prior topic knowledge group (procedural fluency greater than conceptual understanding). The question is: what can be possible explanations for these findings?

Present understanding of conceptual and procedural knowledge indicates that these two types of knowledge are not held as completely separate systems (Hiebert & Lefevre, 1986; Rittle-Johnson et al., 2001). It's possible that one type of knowledge may be more developed at a certain point in time, but this doesn't mean that the other type of knowledge is entirely lacking (Rittle-Johnson et al., 2001). A well-developed mathematical knowledge base includes important and fundamental cognitive links between conceptual and procedural knowledge. Importantly, the formation of suitable cognitive links between conceptual and procedural knowledge is believed to aid in efficient memory storage and successful retrieval of procedures in relevant situations (Hiebert & Lefevre, 1986).

Procedural knowledge is generally believed to be more associated with a particular type of problem, as procedures are typically practised in relation to a specific problem type. Hence, if students already have some familiarity with the material to be learned, conceptual knowledge

might have a greater influence on the development of procedural knowledge, and not the other way around. On the other hand, students who have little prior knowledge in a field are likely to first develop procedural knowledge, which then aids in the cultivation of conceptual knowledge (Schneider et al., 2011). This argument is an explanation of why the low prior knowledge participants of the current study showed a lower increase in the development of their conceptual knowledge. That is these students entered with low levels of procedural knowledge and hence the development of their conceptual knowledge was slower than that of their high prior topic knowledge counterparts.

Hiebert and Lefevre (1986) propose several reasons for the more successful storage and retrieval of procedures when they are linked to conceptual knowledge. When procedures are tied to conceptual knowledge, they become part of an information network bound by semantic relationships, which are less likely to degrade over time as memory tends to last longer for meaningful connections. Since the procedures are integrated into a knowledge network, a variety of cognitive links can be employed to access and retrieve them. It is believed that conceptual knowledge also serves an executive control function, as it is used not only to oversee the selection and application of a procedure, but also to evaluate the appropriateness of the procedural result.

Familiar procedures can lessen the cognitive load involved in problem-solving, thereby improving the ability to tackle more intricate problems. The rationale for this is that automated procedures free up cognitive resources. These resources can then be used, for instance, to identify connections between new aspects of problems or to apply pertinent conceptual knowledge (Hiebert & Lefevre, 1986). This argument provides explanations why the low prior topic knowledge participants' improvement of conceptual knowledge is much lower than that of the high prior topic knowledge participants. Since the low prior topic knowledge students were still developing their procedural knowledge their cognitive resources were occupied with applying their procedural knowledge in problem-solving. This effort inhibited their ability to identify relationships between novel features of problems and hence they could not provide correct solutions to the more complex problems where conceptual knowledge was required.

The theories of retrieval and storage strength (Bjork & Bjork, 1992) provide explanations for the findings based on the implemented teaching strategy. As previously stated, storage strength pertains to the durability of knowledge, while retrieval strength signifies the ease or challenge of recalling that knowledge. Additionally, there is a negative correlation between retrieval strength and increases in storage strength. Given that the gap between tests (or retrieval practice) was roughly 3 weeks, it necessitated strenuous retrieval as the retained knowledge would have significantly decayed over

this period. The effortful retrieval in turn enhanced storage strength. That is long-term retention of knowledge was promoted.

The fact that analogous test items were presented to participants in each subsequent test meant that the same knowledge would be retrieved more than once. Moreover, since analogous items were not exactly the same in order to produce a solution, one needed to use and view stored knowledge in new ways. That is the retrieved knowledge was changed. In other words the participant learnt from the test. More complex problems enhanced the development of conceptual understanding. This is these kinds of problems required that connections are made between disparate pieces of information. Making connections between pieces of knowledge implies that conceptual knowledge is enhanced. This is of cardinal importance in the learning of mathematics.

The fact that South African students struggle with mathematics is well documented. Therefore research is required to determine which teaching and learning strategies could help to alleviate the problem. This study is a contribution in this regard.

Possible confounding

For parametric techniques it is assumed that the population from which the samples are taken have normally distributed scores. For large samples (e.g. 30+) the violation of this assumption should not cause any major problems. Since the number of participants for the high prior topic knowledge group was greater than 30 it does not present a problem for them. It could however be a problem for the low prior topic knowledge students since they numbered only 18. However, the data were checked for severe violations of the normality criteria. Histograms together with skewness and kurtosis parameters indicated that the normality requirement was not violated too severely.

In quantitative research, the four primary forms of validity typically examined are internal, external, construct, and statistical conclusion. In a design that involves a single group undergoing pre-testing and post-testing, factors such as history, maturation, testing, instrumentation, and regression could potentially impact internal validity.

In this research, there were no historical events (known to the investigator) that influenced the results. Maturation is defined as any physical or psychological transformations that might happen within the participants of the study, which could potentially impact their performance on the dependent variable. Given that the majority of the participants in this study were mature individuals, no maturation-related threats were anticipated.

In this research, the pre-test and post-test scores were derived from multiple tests, and each subsequent test had different

items. Also since students wrote tests individually it negated the possibility of intergroup influence. Therefore, the risk to internal validity was diminished. Since the same instrument was utilised in all parts of the study and test items were equivalent, no instrumentation validity threat is expected. The researcher therefore contends that since no rival plausible explanation was found that explained the findings it is attributed to the independent variable.

External validity is the extent to which a study's results can be extrapolated to other groups, environments, and situations. Achieving external validity in quantitative research involves two steps: first, identifying the target population, and second, randomly selecting a sample from that population. However, due to various practical constraints, these steps may not always be feasible. Often in educational research, an accessible population is used in place of the target population. A random sample is then chosen from this accessible population. In this study, the students enrolled in the author's class were chosen as the accessible population. Rather than randomly selecting from this group, the entire class was included in the study. The participation of the whole class increased the likelihood that the accessible population is a good representation of the target population.

Acknowledgements

Competing interests

The author declares that they have no financial or personal relationship(s) that may have inappropriately influenced them in writing this article.

Authors' contributions

B.M.M. declare that they are the sole author.

Ethical considerations

Ethical clearance to conduct this study was obtained from the University of the Western Cape Senate Research Committee (No. 14/1/5). Informed consent was obtained from all participants.

Funding information

This research received no specific grant from any agency in the public, commercial or not-for-profit sectors.

Data availability

The data that support the findings of this study are available from the corresponding author, B.M.M., on reasonable request.

Disclaimer

The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of any affiliated agency of the author.

References

- Abott, E.E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements*, 11(1), 159. <https://doi.org/10.1037/h0093018>
- Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives (complete edition)*. Longman.
- 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. <https://doi.org/10.1037/a0029199>
- Baroody, A.J., & Dowker, A. (Eds.). (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise*. New York, NY: Routledge.
- Bergqvist, E. (2007). Types of reasoning required in university exams in mathematics. *Journal of Mathematical Behavior*, 26, 348–370. <https://doi.org/10.1016/j.jmathb.2007.11.001>
- Bjork R.A. & Bjork, E.L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kasslyn, & R. Shiffrin (Eds), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum
- Butler, A.C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118. <https://doi.org/10.1037/a0019902>
- Carroll, M., Campbell-Ratcliffe, J., Murnane, H., & Perfect, T. (2007). Retrieval-induced forgetting in educational contexts: Monitoring, expertise, text integration, and test format. *European Journal of Cognitive Psychology*, 19(4–5), 580–606. <https://doi.org/10.1080/09541440701326071>
- Chan, J.C., Meissner, C.A., & Davis, S.D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, 144(11), 1111–1146. <https://doi.org/10.1037/bul0000166>
- Cogliano, M., Kardash, C.M., & Bernacki, M.L. (2019). The effects of retrieval practice and prior topic knowledge on test performance and confidence judgments. *Contemporary Educational Psychology*, 56, 117–129. <https://doi.org/10.1016/j.cedpsych.2018.12.001>
- Department of Basic Education. (2011). *National curriculum statements Grades R-12*. Retrieved from <http://www.education.gov.za>
- Dunlosky, J., Rawson, K.A., Marsh, E.J., Nathan, M.J., & Willingham, D.T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Eisenhart, M., Borko, H., Underhill, R., Brown, C., Jones D., & Agard, P. (1993). Conceptual knowledge falls through the cracks: Complexities of learning to teach mathematics for understanding. *Journal for Research in Mathematics Education*, 24(1), 8–40. <https://doi.org/10.5951/jresmetheduc.24.1.0008>
- Glover, J.A. (1989). The 'testing' phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81(3), 392. <https://doi.org/10.1037//0022-0663.81.3.392>
- 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>
- Hiebert, J., & Grouws, D.A. (2007). The effects of classroom mathematics teaching on students' learning. In F.K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Information Age Publishing.
- Hiebert, J., & Lefevre, P. (1986). *Conceptual and procedural knowledge in mathematics: The case of mathematics: The case of mathematics* (pp. 1–27). Lawrence Erlbaum Associates.
- Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports*, 18(3), 879–919. <https://doi.org/10.1037/e469452008-011>
- Johnson, B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approaches*. SAGE.
- Karpicke, J.D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, 21(3), 157–163. <https://doi.org/10.1177/0963721412443552>
- Kerlinger, F.N. (1986). *Foundations of behavioural research* (3rd ed.). New York, NY: Holt, Rinehart & Winston.
- Krathwohl, D.R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2
- Linn, M.C. (2006). The knowledge integration perspective on learning and instruction. In R.K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 243–264). Cambridge: Cambridge University Press.
- Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics*, 67(3), 255–276. <https://doi.org/10.1007/s10649-007-9104-2>
- National Research Council. (2001). Adding it up: Helping children learn mathematics. In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), *Mathematics learning study committee, center for education, division of behavioral and social sciences and education*. National Academy Press, pp. 5–6
- Rittle-Johnson, B., Schneider, M., & Star, J.R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, 27, 587–597. <https://doi.org/10.1007/s10648-015-9302-x>
- Rittle-Johnson, B., Siegler, R.S. & Alibali, M.W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, 93(2), 346–362. <https://doi.org/10.1037/0022-0663.93.2.346>
- Roediger, H., & Karpicke, D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210.
- Roediger III, H.L., & Butler, A.C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–27. <https://doi.org/10.1016/j.tics.2010.09.003>
- Schneider, M., Rittle-Johnson, B., & Star, J.R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology*, 47(6), 1525. <https://doi.org/10.1037/a0024997>
- Schneider, M., & Stern, E. (2009). The inverse relation of addition and subtraction: A knowledge integration perspective. *Mathematical Thinking and Learning*, 11, 92–101. <https://doi.org/10.1080/10986060802584012>
- Star, J.R. (2005). Reconceptualising procedural knowledge. *Journal for Research in Mathematics Education*, 36(5), 404–411.
- Thompson, R.A., & Zamboanga, B.L. (2003). Prior knowledge and its relevance to student achievement in introduction to psychology. *Teaching of Psychology*, 30(2), 96–101.
- Thompson, R.A., & Zamboanga, B.L. (2004). Academic aptitude and prior knowledge as predictors of student achievement in introduction to psychology. *Journal of Educational Psychology*, 96(4), 778.
- Wissman, K.T., & Rawson, K.A. (2018). Test-potentiated learning: Three independent replications, a disconfirmed hypothesis, and an unexpected boundary condition. *Memory*, 26(4), 406–414. <https://doi.org/10.1080/09658211.2017.1350717>
- Xiaofeng, M., Xiao-e, Y., Yanru, L., & AiBao, Z. (2016). Prior knowledge level dissociates effects of retrieval practice and elaboration. *Learning and Individual Differences*, 51, 210–214. <https://doi.org/10.1016/j.lindif.2016.09.012>
- 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. <https://doi.org/10.1037/edu0000268>