Wang *et al.*
Disciplinary and Interdisciplinary Science Education Research (2023) 5:15
https://doi.org/10.1186/s43031-023-00083-4

Disciplinary
and Interdisciplinary Science
Education Research

RESEARCH Open Access

Effect of retrieval practice and drawing on high school students' conceptual understanding of the carbon cycle

Mengyu Wang[1*], Ming Yang[2] and William C. Kyle Jr.[1]

Abstract
Both learner-generated drawing and retrieval practice methods are effective to enhance science learning. To compare the impact of combining different drawing methods (representational drawing vs. abstract drawing) with retrieval practice on the carbon cycle learning, 136 Chinese high school students enrolled in a geography course were assigned randomly to six learning conditions: students built their mental models of the carbon cycle by either generating sketches with or without access to the text learning material introducing the carbon cycle (i.e., generative sketching vs. retrieval sketching), or by creating concept maps with or without access to the learning material (i.e., generative concept mapping vs. retrieval concept mapping), or students just freely recalled on what they have learned from the learning material by paragraphing (i.e., retrieval practice), or restudied the learning material with note-taking (i.e., restudy). Students' learning outcomes were assessed by immediate and one-week delayed tests. Results revealed that no difference was found between the six conditions on the immediate test, whereas students in the retrieval practice condition with paragraphing significantly outperformed those who did not practice retrieval on the one-week delayed test. However, there was no difference between the two drawing conditions regardless of whether they were adopted with or without retrieval practice. Furthermore, the same pattern was found on the factual knowledge questions in both tests, but no main effect of condition was found on both the immediate and the delayed tests for the application questions. We conclude that retrieval-based drawing could be adopted for climate change education at the high school level.

Keywords Retrieval practice, Learner-generated drawing, Climate change education, Carbon cycling

Introduction

The carbon cycle is an important disciplinary core idea in climate change education, which could help learners construct a deep understanding of global climate change (NGSS, 2013). The Next Generation Science Standards has emphasized the significance of developing and using models to illustrate components of the carbon cycle and their relationships in carbon cycling among the biosphere, atmosphere, hydrosphere, and geosphere (NGSS, 2013). Quite a few previous studies have demonstrated that students have difficulties when learning phenomena associated with the carbon cycle. For instance, students hold incomplete or inaccurate conceptual models of the carbon flow (e.g., Düsing et al., 2019; Hartley et al., 2011; Mohan et al., 2009; Zangori et al., 2017), and some students are confused about the relationship between carbon dioxide and global warming (e.g., McNeill & Vaughn, 2012; Shepardson et al., 2012). Further, students have

---

*Correspondence:
Mengyu Wang
mengyuwang0909@gmail.com
[1]College of Education, University of Missouri - St. Louis, St. Louis, MO 63121-4499, USA
[2]Huipu Middle School, Xuepu Rd, Linhai, Taizhou, Zhejiang 317000, China

difficulty visualizing the carbon cycle on the global scale (e.g., Hartley et al., 2011; Mohan et al., 2009; Zangori & Koontz, 2016). Therefore, it is imperative to enhance students' knowledge regarding carbon cycling and enhance their basic understandings of climate change.

The afore-mentioned challenges of learning the carbon cycle can be categorized in two main types: (i) developing visualizing and modeling skills, and (ii) enhancing conceptual understanding of basic knowledge. To overcome barriers in visualizing and modeling on the carbon cycle, prior studies suggested that learner-generated drawing could enhance the model-based reasoning process across science disciplines (Grosslight et al., 1991; Kindfield, 1994; NRC, 2012; Rosengrant et al., 2009; Schwarz et al., 2009). Specifically, drawing has numerous benefits to learning, such as facilitating students' thinking and communicating processes, generating predictions and explanations, developing memory for core content, and enhancing comprehension of the scientific processes (Ainsworth et al., 2011; Quillin & Thomas, 2015; Schwarz et al., 2009; van Meter & Garner, 2005; Wammes et al., 2016). For example, Zangori and colleagues (2017) developed a model-based curriculum focusing on socioscientific issues for high school students. In their study, students deepened their understanding and reasoning of the relationship between the carbon cycle and climate change through developing, using, evaluating and revising their own models of carbon cycling. Yet, Düsing and colleagues (2019) investigated middle and high school students' conceptions of the carbon cycle and the relationship between the carbon cycle and its components. From students' schematic drawings and writing explanations of the carbon cycle, the researchers found that students had difficulties identifying carbon compounds and the processes where carbon compounds are transformed. Thus, how to harness the benefits of drawing to solve the second challenge -improving students' basic knowledge on the carbon cycle is an open question. We report herein how to use retrieval practice to support drawing for enhancing students' understanding of carbon cycling.

Retrieval practice can be described as a learning strategy of recalling previous learned knowledge without viewing learning materials. Prior studies indicated that retrieval practice could produce robust knowledge gains, especially with respect to the benefits for learners' long-term retention of knowledge (Carpenter, 2009; Karpicke, 2017; Roediger & Karpicke, 2006a, b). Research at the college level suggested that applying retrieval practice in terms of drawing tasks led to better learning outcomes compared to repeatedly studying materials (Karpicke & Blunt, 2011, 2014; Heideman et al., 2017). However, little is known whether a combination of retrieval practice and drawing methods could produce better learning outcomes than adopting either one of them alone at the high school level.

Prior studies described drawing on a spectrum from representational which is more concrete to abstract (Quillin & Thomas, 2015). According to Quillin and Thomas (2015), representational drawing could foster learners' active learning, observational skills, memorization, understandings of spatial relationships and enjoyment of learning, while abstract drawing could facilitate learners' motivation of learning. Also, abstract drawing could foster learners' development of mental models and help learners enhance acquisition of content knowledge and problem-solving processes (Quillin & Thomas, 2015). In the current study, we focused on two main drawing strategies in science class, a representational drawing strategy: sketching; and an abstract drawing strategy: concept mapping. In particular, sketching is conceptualized as a technique that uses a minimum number of lines and symbols to rapidly represent knowledge, while concept mapping is often described as a diagram that depicts suggested relationships between concepts, which is a graphical tool for learners to organize and structure knowledge (Novak, 2005). Although a body of research has applied drawing strategies in science class (e.g., Johnson & Reynolds, 2005; Rennie & Jarvis, 1995; Smith & Bermea, 2012), few studies focused on how to improve the efficiency of different types of drawing by combining other highly effective learning strategies, say, retrieval practice. Therefore, the goal of the present study was to explore how different drawing strategies (representational vs. abstract) in terms of retrieval practice facilitate high school students' conceptual understanding of the carbon cycle and climate change. Specifically, the research questions are:

1. Does retrieval practice lead to better performance than restudy in the learning of the carbon cycle at the high school level?
2. How do different types of drawing (representational drawings vs. abstract drawings) in combination with retrieval practice influence high school students' conceptual understanding and learning outcomes on the carbon cycle?

Literature Review

Learner-generated drawing

It is widely known that interpretation of others' visual information is critical to science learning (e.g., interpreting charts and graphs).However, students are also required to develop many representational skills to become proficient in science (Ainsworth et al., 2011). The *learner-generated drawing* strategy, which is grounded in generative learning theory (Mayer et al., 1995), engages students to construct their own visual representations to get better understanding of complex scientific expository text materials (van Meter, 2001; van Meter et al., 2006). From the perspective of the Select-Organize-Integrate (SOI) model of generative learning theory (Fiorella & Mayer, 2016; Mayer, 2014), there are three main steps to complete drawing construction. In the first step, learners select key information when reading from a text for processing in the working memory. In the second step, learners organize this selected key information and establish internal verbal representations of the text. Meanwhile, learners construct internal visual representations of the text and link them to the internal verbal representations. In the final step, learners draw an external representation of the text content which is based on the integration of the internal verbal and the internal visual representations (van Meter & Garner, 2005). This construction process supports learners to store and represent a mental model of knowledge that is credited with improving their problem-solving abilities and conceptual understanding (Ainsworth et al., 2011; van Meter & Garner, 2005). In addition, learner-generated drawing is an iterative process and involves metacognitive processes such as self-monitoring and self-regulation (Fiorella & Mayer, 2016; Mayer, 2014; Quillin & Thomas, 2015; Schmeck et al., 2014; van Meter, 2001; van Meter & Garner, 2005; van Meter et al., 2006). Thus, drawing as an active learning strategy could potentially enhance students' learning on sophisticated scientific phenomena such as carbon cycling.

There are some important findings when applying leaner-generated drawing strategy. First, the benefits of learner-generated drawing are relevant to the accuracy and the quality of drawings. Prior studies found that learners had deeper comprehension of text learning materials if their drawings accurately and completely reflected what was described in the materials (Schmeck et al., 2014; van Meter, 2001; van Meter & Garner, 2005; van Meter et al., 2006). While some researchers pointed out that simple line drawings were more effective than photographs, especially when learning materials were complex with rich content (Quillin & Thomas, 2015; Rennie & Jarvis, 1995). Second, learner-generated drawing sometimes leads to lower learning outcomes. One reason is that the drawing processes may cause higher extraneous cognitive load on learners in a way that hinders learning (Leutner et al., 2009). When learners have little experience with drawing or do not have any drawing skills, it might be difficult for them to draw what they really want to express (Quillin & Thomas, 2015).

Another important finding is that drawing strategy could potentially enhance model-based reasoning (Quillin & Thomas, 2015; Schwarz et al., 2009). Model-based reasoning is a sophisticated process that enables learners to solve problems via constructing a mental model – an analogue representation with highly personal, dynamic, and format-diversified characteristics (Harrison & Treagust, 2000; Schwarz et al., 2009). Given the goal proposed by the National Research Council (NRC) of developing and using models for K-12 science education, especially on the item of "construct drawings or diagrams as representations of events or system" (NRC, 2012, p. 58), students should be encouraged to represent and explain phenomena, make predictions, and solve different scientific problems through constructing their self-generated visual models (Greca & Moreira, 2000; Shepardson et al., 2017, Zhai et al., 2022).

In current environmental science classrooms, students often make their thoughts and reasoning visible while studying topics related to climate change. Effective learning occurs when students consistently employ, evaluate, and refine their own models with respect to climate change (Zangori & Forbes, 2016; Zangori et al., 2017). For example, Shepardson et al. (2017) investigated how middle school students enhance their comprehension of the greenhouse effect. They designed a draw-and-explain task that prompted students to create, label, and compose a paragraph explaining their mental models of the greenhouse effect. The researchers found that students developed more sophisticated mental models of the greenhouse effect after making certain modifications to their drawings during the process of completing the learning activity. Therefore, drawing could serve as a significant medium that connects with both internal and external representations to facilitate analyzing, reasoning, and synthesizing scientific knowledge for a better understanding of the carbon cycle.

Representational drawing and abstract drawing

When considering drawing as a learning strategy to enhance model-based learning, it could be broadly defined as "a learner-generated external visual representation depicting any type of content, whether structure, relationship, or process, created in static two dimensions in any medium." (Quillin & Thomas, 2015, p. 2). All visualizations of drawing are analogical and cannot truly represent the real world, but they could be varied in the extent to which they are representational or abstract.

Representational drawing shares a physical resemblance with the objects that the drawing depicts (Alesandrini, 1984; van Meter & Garner, 2005). One typical representational drawing is sketching. Different from other external representational modeling, sketching allows learners to construct visuo-spatial external models of scientific phenomena especially when learning from expository text materials (Scheiter et al., 2017), which could help students gain a deep understanding of spatial and causal knowledge of scientific domains (Suwa & Tversky, 1997). For example, the phenomenon of plate tectonics is not easy to explain verbally due to the complex spatial relations of the Earth's layered structure (Smith & Bermea, 2012). One good solution could be asking students to sketch the structure to identify the main features of plate tectonics such as rift, oceanic crust, and lithosphere.

While sketching offers notable advantages, prior research has identified certain boundary conditions that could limit the effectiveness of using sketching as an aid to learn from text materials. One such factor is whether learners should be provided with a clear specification regarding what to depict during the learning process (Leutner et al., 2009). Offering instructional scaffolds such as guided questions, labels or cut-out figures could improve the outcomes of sketching activities (Quillin & Thomas, 2015; van Meter, 2001; van Meter et al., 2006). By contrast, sketching without any additional instructions or scaffolds could impede learners' comprehension of to-be-learned materials, due to an increased extraneous cognitive load caused by logistics of managing drawing activities (Schwamborn et al., 2011). In addition, learners also benefit from receiving feedback on their drawings (van Meter & Garner, 2005). Providing learners with timely feedback could enable them to reflect on their drawings and facilitate them to self-regulate their interactions with learning content (Wu & Rau, 2019). Therefore, teachers ought to provide students with opportunities to construct their external pictures as well as modify their creations.

Abstract drawing, on the other hand, is much more analogical than representational drawing. This type of external representation mostly consists of words, numbers, lines, and/or arrows, such as flowcharts, graphs, and phylogenetic trees (Quillin & Thomas, 2015). Concept mapping is a typical example of abstract drawing that has played a significant role in science education since the 1970s. This graphic organizer often includes concepts enclosed in boxes or circles, and relationships between concepts indicated by labeled arrows or lines. Novak and Gowin (1984) indicated that concept mapping was a meaningful learning tool that enables learners to externalize their ideas of scientific phenomena and depict suggested relationships between concepts in an organized way. In general, "a concept map is a schematic device for representing a set of the concept meanings embedded in a framework of propositions." (Novak & Gowin, 1984, p. 15). Doing concept mapping could facilitate meaningful learning processes by incorporating new knowledge into prior knowledge (Novak & Cañas, 2008). In addition, concept mapping could serve as an external scaffold to support learners to organize and structure information during learning, thereby they had opportunities to enhance the comprehension of knowledge (Novak, 1990; Novak & Wandersee, 1991). Similar to learning through sketching, learners need to know how to create concept maps prior to engaging in the learning process (Chularut & DeBacker, 2004), but concept mapping is much more concise and emphasizes the processes of placing different related aspects into a coherent structure, which in turn is beneficial for deep scientific reasoning.

Retrieval practice

Retrieval practice is a powerful learning strategy that could foster long-lasting learning (Karpicke, 2012, 2017). Through retrieval practice, learners could promote their long-term retention of knowledge by recalling what they have learned previously (e.g., Roediger & Karpicke, 2006a, b; Roediger et al., 2010). Research on retrieval practice can be traced back to a series of memory experiments on testing effect, which has been shown to be a robust and replicable phenomenon in different educational settings including laboratory, physical and online classrooms (e.g., Carpenter, 2009; McDaniel et al., 2013; Millet et al., 2021; Zu et al., 2019). The general experimental procedure of retrieval practice involves three main phases: an initial learning phase where learners first study to-be-learned materials; an interventional testing phase where learners recall what they have just learned without access to the original materials through quizzing typically; and a final assessment phase where students take a summative assessment, which can be from a few minutes after interventional activities (Rowland & DeLosh, 2015; Smith et al., 2013) to several weeks later (Carpenter, 2009; Larsen et al., 2013). As a comparison to the retrieval intervention, a commonly adopted learning strategy where students restudy the materials with notetaking is often applied as a control condition. Retrieval practice was demonstrated to be superior to the restudy strategy during the same amount of time (e.g., Roediger & Karpicke, 2006a, b; Roediger et al., 2010).

Two prominent accounts describe why retrieval practice benefits long-lasting retention of knowledge: the elaborative retrieval account and the episodic context account. The general idea of the elaborative retrieval account is that learners enhance their memory of knowledge by generating several knowledge items that are semantically relevant to the retrieval cue provided, then incorporating these items with the targeted knowledge items to form an elaborated memory trace for a future search (Carpenter, 2009, 2011). By contrast, the episodic context account proposes that memory performance is consolidated by reinstating and updating the episodic context in which knowledge items are coded and stored. Specifically, learners first attempt to reinstate the episodic context associated with a knowledge item as part of a memory search process. When an item is successfully retrieved, the mental representation of knowledge items is updated by adding features from the current retrieval context to the features from the original study context.

Finally, when future retrieval attempts occur, these updated contexts can be used as additional cues to access the knowledge items such that memory is strengthened via not only the knowledge items but the addition of context features (Karpicke et al., 2014; Karpicke, 2017).

Most previous applied research on retrieval practice focused on low-level recognition tasks and revealed a robust effect on factual knowledge performance (e.g., Carpenter, 2009; Johnson & Mayer, 2009; Smith et al., 2013). Only a handful of studies have begun to explore if retrieval practice could also benefit more meaningful learning such as transfer and comprehension of complex content (e.g., Butler, 2010; Jensen et al., 2014, 2020; McDaniel et al., 2013; Zu et al., 2019). In such studies, students worked on quizzes via answering inferential or application questions as retrieval practice after initially learning the materials. The results showed that the application of a concept in the retrieval quizzing phase facilitated application of the same concept in new situations on a final assessment (Butler, 2010; Jensen et al., 2014, 2020; McDaniel et al., 2013; Zu et al., 2019). While retrieval practice produced benefits on higher-order meaningful assessment in the above-mentioned studies, it is important to note that some limitations exist while using testing as a way of practicing retrieval. One such limit is that questions used between the initial quizzing and the final summative assessment should be similar or at the same level to allow successful transfer of learning (Karpicke, 2017; Karpicke et al., 2014). It brings challenges for the application in real classrooms since it may cause high pressure for students if they work on high-level questions just after the initial exposure to learning materials, which in turn results in less or no effect of retrieval practice. Another limitation is that when using question items with "high element interactivity", the testing effect may decrease or even disappear (van Gog & Sweller, 2015). For example, prior studies showed that no testing effect was presented on a one-week delayed assessment as students worked on complex worked examples questions in the initial learning phase (Leahy et al., 2015; van Gog et al., 2015). Given the limitations of testing as a way of practicing retrieval, it is crucial to innovate approaches that effectively harvest the benefits of retrieval practice in real classroom settings.

Theoretical Framework

Applying retrieval-based drawing to learn carbon cycling

It is undoubtful that involving students in meaningful learning is one of the most effective ways to achieve goals of science education (Novak, 1990). Meaningful learning refers to engaging students to actively construct new knowledge with their relevant prior knowledge through an organized, coherent, and integrated ways that allow them to make inferences and apply their knowledge in novel situations in real life (Ausubel, 1968). Such processes emphasize that both construction and consolidation of knowledge information are equally important during meaningful learning (Roelle et al., 2022). Drawing as a typical generative learning strategy engages students to actively construct their own visual representations during learning, thereby it is reasonable to assume that drawing could promote construction of knowledge. Yet, the benefits of drawing by learners themselves were only found on higher-order instead of lower-order factual knowledge items (Van Meter, 2001; Van Meter & Garner, 2005). Moreover, many of drawing studies implemented posttests that followed immediately after an initial learning phase with drawing (for an exception, see Wammes et al., 2016), while evidence concerning long-term benefits (i.e., the consolidation of knowledge) of drawing tasks is relatively scarce. In contrast, retrieval practice revealed a robust effect on factual knowledge gains as well as on long-lasting instead of rote, transient learning outcomes (Karpicke, 2012). Thus, it is plausible to combine retrieval practice with drawing to exploit the function of consolidating knowledge at all cognition levels to optimize learning goals.

In the current study, we adopted retrieval-based drawing strategy to explore if the combination of retrieval practice and drawing could be a promising approach to foster students' meaningful learning. We only found a handful of studies that investigated the effectiveness of strategies integrating retrieval and drawing (Blunt & Karpicke, 2014; Heideman et al., 2017; Karpicke & Blunt, 2011). Heideman and colleagues (2017) designed a learning tool called Minute Sketches in Folded Lists (MSFL) for college students to self-assess their knowledge retention and problem solving on introductory biology concepts. They indicated that students using MSFL had better learning outcomes relative to students who only restudied learning materials via visually reviewing. Moreover, Blunt and Karpicke (2014) found that university students benefited more from learning science texts when using concept mapping as a retrieval practice activity (i.e., creating concept maps in a closed-book format) than when using it as a generative learning activity (i.e., creating concept maps in an open-book format). They compared students' performance in a one-week delayed test including both factual and inferential questions. The results revealed that drawing concept maps in the form of retrieval practice led to better knowledge retention on both factual and inferential questions. Based on these landmark findings, it is necessary to explore if the benefits of retrieval-based drawing can be extended from college to high school classroom settings.

In addition, both sketching and concept mapping have the potential to significantly contribute to the comprehension of the carbon cycle. Yet, we were not sure what differences exist between these two drawing strategies in terms of their effectiveness in enhancing students' construction of a conceptual model of carbon cycling. Therefore, one of the goals in the current study is about comparing the effectiveness between sketching and concept mapping for learning the carbon cycle.

Together, both learner-generated drawing and the retrieval practice strategies could improve meaningful learning processes and outcomes, but less is known whether drawing in combination with retrieval practice would be better suited to optimizing high school students' comprehension on the carbon cycle. Therefore, we combined retrieval practice with drawing in the current study to explore if any additional benefits can be produced on high school students' conceptual understanding of carbon cycling.

Method

Overview of the current study

We designed an independent course where a group of Chinese high-school students learned carbon cycling through reading an introductory essay and employing retrieval-based drawing to construct their own models of the carbon cycle.

There have been numerous studies conducted to examine Chinese students' performance on the learning of science via using various international assessment frameworks. These studies have provided valuable insights into the academic achievements and capabilities of Chinese students in comparison to their international counterparts. It has been revealed that Asian Chinese students have the highest performance on science international tests such as Program for International Student Assessment (PISA) (Cheng & Wan, 2016; Schleicher, 2019), although they tend to report lower self-efficacy than their western peers (Lau & Ho, 2020). Some other studies revealed more nuanced results. For example, researchers found that even though Chinese students did significantly better on well-validated physics conceptual surveys than their U.S. counterparts, they did not demonstrate different performance on the Lawson classroom test of scientific reasoning (Bao et al., 2009). These studies have contributed to our understanding of the learning achievements and strengths of Chinese students within international assessment frameworks, specifically in the context of science education. This current work also aims at contributing to the discussion of Chinese students' learning performance on learning of environmental science.

This course was a part of Chinese high school geography curriculum focusing on earth science and physical geography. Prior to the study, the students commonly adopted learning strategies such as re-reading learning materials or relying on existing models provided by teachers (i.e., instructor-provided models). Thus, one of objectives of the course is to transform teacher-centered instructions via engaging high school students to actively construct their own models to get deep understandings on the spatial configurations of different carbon components, causal knowledge of matter, and energy transfer in the carbon cycle. Moreover, after the course, students would know if retrieval-based drawing could be an optional learning strategy when learning other cyclic nature of an ecosystem.

A between-subject design was employed in this study. Student participants were randomly assigned to six different learning conditions (see Table 1). Four of the six conditions were based on a factorial combination of retrieval practice and the two drawing strategies. Students in these four conditions were instructed to visualize their mental models of the main processes of carbon cycling. To examine whether additional benefits exist when drawing methods were combined with retrieval practice, two control conditions were included: a retrieval-only condition, where students recalled information by paragraphing, and a restudy condition, rendering six conditions in total. The retrieval by paragraphing and the restudy conditions are commonly employed in retrieval practice research (e.g., Blunt & Karpicke, 2014; Karpicke & Blunt, 2011).

We examined the effect of retrieval-based drawing on students' conceptual understanding of the carbon cycle with an immediate test. Their retention of knowledge was assessed by the same test in a week (referred to as the delayed test). The test consisted of factual questions and application questions. We hypothesized that students practicing retrieval in a paragraphing format would outperform all other open-book conditions on factual questions due to the robust effect of retrieval practice on learning factual knowledge (e.g., Carpenter, 2009; Johnson & Mayer, 2009; Smith et al., 2013). We also hypothesized that students in generative sketching and generative concept mapping conditions would outperform those in the restudy condition on application questions since drawing benefits more on higher-order learning performance (i.e., questions on applying problem-solving skills, analysis, and application skills) than on lower-order recognition performance (van Meter, 2001; van Meter & Garner, 2005). We also anticipated that retrieval-based sketching and retrieval-based concept mapping conditions would outperform other conditions on both factual and application questions since the combination of the two methods might work synergistically with each other to benefit both factual and application knowledge.

Participants

The study was conducted at an urban high school with a student population of approximately 3000 in a mid-sized city located in East China. We invited 186 tenth-grade

Table 1 Introduction of the six conditions

Learning condition	Learning activities	Rationale of the design
Generative sketching (GS)	Students drew sketches of the carbon cycle while reading the text material.	Representational drawing in an open-book style
Generative Concept mapping (GC)	Students drew concept maps of the carbon cycle while reading the text material.	Abstract drawing in an open-book style
Retrieval Sketching (RS)	Students read the text first and then drew sketches of the carbon cycle without access to the text material.	Representational drawing in a closed-book style
Retrieval Concept mapping (RC)	Students read the text first and then drew concept maps for the carbon cycle without access to the text material.	Abstract drawing in a closed-book style
Retrieval practice	Students read the text first and then recalled as much of the information as they could by paragraphing.	Control 1
Restudy	Students read the text and then restudied it by taking notes	Control 2

students (94 males and 92 females) for participation. These students, at the time, were enrolled in the geography course taught by a geography teacher who had over 11 years of teaching experience at the high school level.

Before the study, all the students and their guardians were introduced the study and signed their consent forms to grant their consent for students' participation. All procedures performed in the study involving the participants were in accordance with the Institutional Review Board of the university and ethical standards of the Institutional and National Research Committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Out of the 186 participants, 136 students eventually completed all learning activities required of the study such that we had 23 students in the generative sketching (GS) condition, the generative concept mapping (GC) condition, the retrieval sketching (RS) condition and the retrieval practice condition, respectively; and 22 students in the retrieval concept mapping (RC) and the restudy condition, respectively.

Materials

The students were required to learn an expository text on the topic of the global carbon cycle by themselves. The text material had a total of 659 words which covered basic components of carbon cycling, the processes of carbon transfer and transformation across atmosphere, hydrosphere, biosphere, and lithosphere (see Appendix A).

Additionally, six different learning tasks were developed in accordance with the six learning conditions. Specifically, students in both RS and GS conditions were provided with a sheet of paper containing a background image and several figures possibly representing the different components of the carbon cycle, along with an example sketch of the urban heat island. Students in both RC and GC conditions were provided with a sheet of paper including an example of concept maps on the Earth's energy budget. Students in the retrieval practice and the restudy conditions were provided with a sheet of blank paper (details are provided in Appendix B). Moreover, students in all conditions received necessary instructions and completed their respective learning tasks independently.

The same test material was used in both the immediate and one-week delayed tests, which covered knowledge of the carbon cycle. There were three types of tasks on the test: five fill-in-the-blank questions, a diagram task, and a flow-chart task (see Appendix C). The fill-in-the-blank questions and the diagram task were designed to assess students' factual knowledge on the carbon cycle. The flow-chart task was adapted from a box-diagram assessment tool by Sibley and colleagues (2007), and it was used to assess the degree to which students could apply the knowledge learned in a new context. All the learning and test materials were first designed by the first author, and reviewed and edited by the geography teacher to make sure they were appropriate for the students.

Procedure

Prior to the formal study, a pilot study was conducted to test the validity of both the materials and design used in the study. Twelve students in the school were randomly selected to participate in the pilot study, with two in each learning condition. They were from a different class taught by the same geography teacher. Minor adjustments were made accordingly after the pilot study.

In the beginning of the formal study, all the students were instructed the background knowledge of the carbon cycle, including atomic-molecular models of different carbon compounds, the concept and/or chemical equations of different processes relevant to the carbon movement, such as photosynthesis and cellular respiration. In addition, the students were required to review relevant prior knowledge they learned in middle school, including concepts of evaporation and condensation, acid and bases, combustion, and food web. Additionally, the students were instructed on how to create concept maps. Specifically, the geography teacher guided the students to review the water cycle by constructing a concept map with reference to the guideline proposed by Novak and Cañas (2008). The students did not receive training on sketching since it was a familiar and straightforward learning technique to them compared to the concept mapping strategy. Besides, to reduce the influence of limitations on students' drawing abilities, the students

Table 2 Procedure of the study design

Learning condition	Learning session				Testing session	
	1	2	3	4	Immediate	Delayed
Retrieval Sketching	Study (5 min)	Sketching (10 min)	Restudy (5 min)	Sketching (10 min)	20 min	20 min
Retrieval Concept mapping	Study (5 min)	Concept mapping (10 min)	Restudy (5 min)	Concept mapping (10 min)		
Generative Sketching	Sketching while reading the text (30 min)					
Generative Concept mapping	Concept mapping while reading the text (30 min)					
Retrieval only	Study (5 min)	Recall (10 min)	Study (5 min)	Recall (10 min)		
Restudy	Study and restudy by taking notes (30 min)					

were provided with necessary symbols and figures in the sketching tasks' instruction.

After being equipped with background knowledge and drawing skills, the students participated in the learning session. Table 2 provides an overview of the process of the learning and testing sessions for the six conditions. During the learning session, students in the RS and the RC conditions had four consecutive task phases. First, students in both conditions studied the text material on a paper sheet, in an initial 5-min study period. After the initial study phase, students in the RS condition drew their mental models of the carbon cycle by annotated sketching without access to the text material. Namely, students in this condition practiced recalling with representative drawing strategy in a closed-book style. Besides drawing, students were required to label each process of the carbon cycle while they were drawing. Students in the RC condition constructed their carbon cycle models by creating concept maps. Likewise, they were not allowed to access the text material while creating concept maps, thereby they freely recalled the abstract drawing method in a closed-book style. The drawing phase for both retrieval practice drawing conditions lasted about 10 min. After the drawing phase, students in both conditions had 5 min to check and restudy the text material. During the last phase, students from both conditions took 10 min to edit their drawings without access to the text material. The schedule for time spent on each task step was designed to be consistent with previous studies (i.e., Blunt & Karpicke, 2014; Karpicke & Blunt, 2011). Pilot testing showed that the time was enough for students to complete each learning phase in these conditions.

For the GS and the GC conditions, students either generated their annotated sketching or concept maps according to their conditions with access to the text material (i.e., in an open-book style). For the retrieval practice condition, students were required to recall and write down as much information of the text as they could. For the restudy condition, students were required to restudy the text and take notes by themselves after the initial learning. In all conditions, students were given 30 min to complete their learning tasks.

Table 3 Mean Score of Students for the number of idea units covered in the initial learning task

Condition	*M*	*SD*
Generative Sketching (N = 23)	13.96	3.76
Generative Concept mapping (N = 23)	12.74	3.86
Retrieval Sketching (N = 23)	13.30	3.37
Retrieval Concept mapping (N = 22)	12.73	2.66
Retrieval Practice (N = 23)	15.30	2.97
Restudy (N = 22)		

At the end of the learning session, students from all six conditions took the immediate test, the content of which was relevant to the text learning material on the carbon cycle. The time allowed for completing the immediate test was 20 min. Students took the same delayed test in another 20 min one week later. The details of the scoring process can be found in the results section below.

Results

GPA comparison

We first conducted a one-way ANOVA test on the GPA scores of all students and found no significant difference among the six conditions at the level of $p = .05$, indicating the participants assigned to different conditions were equivalent satisfying the randomization assumption.

Initial learning activities

We investigated the number of idea units covered in each condition (except the restudy condition) during the learning session (see Table 3). The global carbon cycle expository text was divided into 23 idea units. Specifically, students' artifacts were scored by the following grading protocol: one point is rewarded for each correctly presented idea unit (through either paragraphing or drawing) in all five conditions, rendering a maximum of 23 points. 0.5 point was rewarded if any idea unit was partially recalled or drawn (Karpicke & Blunt, 2011). A one-way ANOVA suggested that there was no significant difference between the five conditions, $F(4, 109) = 2.336$, $p = .06$, $\eta_p^2 = 0.079$, although students in the retrieval practice condition produced relatively more idea units during study.

Overall performance on the immediate and one-week delayed test

For purposes of grading the tests, a rubric was developed by the first author beforehand. Specifically, students received one point for each fill-in-the-blank question if they gave the correct answer (for a total of 5 points). Similarly, they were given one point when they correctly answered each item in the diagram and the flow-chart task, yielding a maximum score of 26 points. Participants did not receive any credit for vague or partially correct answers but could receive credit for correct answers that were worded differently than the text materials.

Initially, 30 immediate tests were independently graded by the first author and the geography teacher with 5 from each of the six conditions, and the interrater correlation of Pearson between scores was 96%. Given the high interrater reliability, the remaining tests were scored by the first author.

Table 4 presents learning performance of the students in each condition on both the immediate and one-week delayed tests. The learning results are also displayed separately for the overall carbon cycle test, the factual knowledge questions, and the application questions. The performance for the overall test was determined by dividing the total score of each participant by the highest possible score (26 points) on the test. The performance for the factual knowledge questions was determined by dividing the score of each participant on the fill-in-the-blank questions and the diagram task by the maximum possible score of 18 points. The performance for the application knowledge was determined by dividing the score of each participant on the flow-chart task by the maximum possible score of 8 points.

We examined the performance of students from all six conditions on the immediate and the one-week delayed tests (see Fig. 1). The results were submitted to a 6 × 2 mixed measures ANOVA, with *learning condition* as the between-subject independent variable and *retention interval* as the within-subject independent variable. Overall, there was a significant effect of retention interval, $F(1, 130)=2.104$, $p<.001$, $\eta_p^2 = 0.106$, indicating forgetting occurred during one week. We also found a main effect of learning condition, $F(5, 130) = 2.358$, $p = .044$, $\eta_p^2 = 0.083$. However, these effects were not qualified by Learning Condition × Retention Interval interaction, $F(5, 130) = 2.168$, $p = .062$, $\eta_p^2 = 0.077$.

Table 4 Mean Score of Students for Different Questions from the Six Conditions on Both Immediate and One-week delayed Tests[1]

Condition	Overall test	Factual knowledge questions	Application knowledge questions
Immediate test			
Generative Sketching (N = 23)	0.55 (0.10)	0.50 (0.13)	0.66 (0.12)
Generative Concept mapping (N = 23)	0.52 (0.08)	0.49 (0.12)	0.58 (0.18)
Retrieval Sketching (N = 23)	0.54 (0.17)	0.54 (0.17)	0.56 (0.26)
Retrieval Concept mapping (N = 22)	0.56 (0.10)	0.53 (0.09)	0.64 (0.18)
Retrieval Practice (N = 23)	0.59 (0.11)	0.56 (0.08)	0.65 (0.23)
Restudy (N = 22)	0.52 (0.13)	0.49 (0.14)	0.57 (0.23)
One-week delayed test			
Generative Sketching (N = 23)	0.51 (0.11)	0.48 (0.13)	0.56 (0.20)
Generative Concept mapping (N = 23)	0.50 (0.10)	0.46 (0.11)	0.57 (0.13)
Retrieval Sketching (N = 23)	0.54 (0.15)	0.53 (0.16)	0.54 (0.27)
Retrieval Concept mapping (N = 22)	0.55 (0.10)	0.52 (0.10)	0.62 (0.19)
Retrieval Practice (N = 23)	0.58 (0.11)	0.57 (0.11)	0.60 (0.20)
Restudy (N = 22)	0.44 (0.16)	0.44 (0.15)	0.45 (0.25)

Note: Standard deviations are reported in parentheses

[1] Students' anonymized raw data could be provided at the request of any interested reader. The authors could also provide a reasonable level of translation help

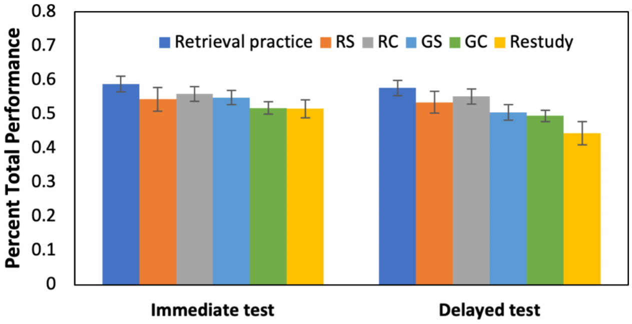

Fig. 1 Students' performance on the immediate and the one-week delayed tests

Subsequent analysis showed no significant difference between all 6 conditions on the immediate test performance, $F(5, 130)=1.214$, $p=.306$, $\eta_p^2 = 0.045$. However, for the one-week delayed test performance, there was a significant main effect on condition, $F(5, 130) = 3.389$, $p = .007$, $\eta_p^2 = 0.115$. Post hoc analyses revealed that students in the retrieval practice condition had better performance than students who did not practice retrieval during the initial learning. Specifically, the retrieval practice condition outperformed the GS condition, (0.58 vs. 0.51), $t\ (44) = 2.25$, $p = .03$, $d = 0.11$, 95% CI [0.01, 0.14], as did the GC condition, (0.58 vs. 0.50), $t\ (44) = 2.94$, $p = .005$, $d = 0.09$, 95% CI [0.03, 0.14], and did the restudy condition, (0.58 vs. 0.44), $t\ (43) = 3.27$, $p = .002$, $d = 0.14$, 95% CI [0.05, 0.22], respectively. In addition, students in the RC condition had better performance than students in the restudy condition, (0.55 vs. 0.44), $t\ (42) = 2.68$, $p = .005$, $d = 0.13$, 95%CI [0.03, 0.19], suggesting that retrieval practice was also beneficial when students drew concept map during the carbon cycle learning. Meanwhile, we found the RS condition outperformed the restudy condition, although the difference between these two conditions was marginally significant, (0.54 vs. 0.44), t (43) = 1.96, p = .06, d = 0.16, 95%CI [−0.00, 0.18].

To further investigate how different types of drawing (sketching vs. concept mapping) influence students' learning outcomes on the carbon cycle, we conducted a one-way ANOVA to examine if there was any significant difference between the four drawing conditions on the delayed performance. Similar to the results of the immediate test, students from the four drawing conditions had equivalent learning outcomes on the delayed test, $F(3, 87)=1.245$, p=.30, $\eta_p^2 = 0.041$, indicating that students could either adopt sketching or concept mapping to construct their carbon cycle model for learning, and whether combining with retrieval practice did not matter. The data analysis also showed that performance in the retrieval practice condition was essentially equivalent to the RC conditions, (0.58 vs. 0.55), t (43) = 0.78, p = .44, d = 0.10, 95%CI [−0.04, 0.09], and to the RS condition (0.58 vs. 0.54), t (44) = 1.08, p = .29, d = 0.13, 95%CI [−0.04, 0.12].

Factual knowledge performance

Figure 2 shows performance of students from different conditions on the factual questions for both the immediate and the one-week delayed test. The results were submitted to a 6 × 2 mixed ANOVA, with *learning condition* as the between-subject independent variable and *retention interval* as the within-subject independent variable. Similar to the overall performance, we found a statistically significant effect of retention interval, $F(1, 130)=5.094$, p=.026, $\eta_p^2 = 0.038$. There was also a main effect of learning condition, $F(5, 130) = 2.436$, p = .038, $\eta_p^2 = 0.086$. However, there was no statistically significant interaction effect, $F(5, 130) = 1.077$, p = .376, $\eta_p^2 = 0.04$.

Subsequent one-way ANOVA revealed a non-significant result for all 6 conditions on the immediate factual knowledge performance, $F(5, 130)=1.245$, p=.292, $\eta_p^2 = 0.046$. There was a significant main effect of the condition on the one-week delayed factual performance, $F(5, 130) = 3.21$, p = .009, $\eta_p^2 = 0.11$. Post hoc analyses revealed that students in the retrieval practice condition had better factual performance than students who did not practiced retrieval during the initial learning. Specifically, retrieval practice condition reflected a higher score than the GS condition, (0.57 vs. 0.48), t (44) = 2.43, p = .02, d = 0.12, 95% CI [0.01, 0.16], as did the GC condition, (0.57 vs. 0.46), t (44) = 3.22, p = .002, d = 0.11, 95% CI [0.04, 0.17], and did the restudy condition, (0.57 vs. 0.44), t (43) = 3.32, p = .002, d = 0.13, 95% CI [0.05, 0.20], respectively. In addition, both the RC condition and the RS condition outperformed the restudy condition, (0.52 vs. 0.44), t (42) = 2.19, p = .035, d = 0.13, 95%CI [0.00, 0.16], and (0.53 vs. 0.44), t (43) = 2.09, p = .042, d = 0.15, 95%CI [0.00, 0.19].

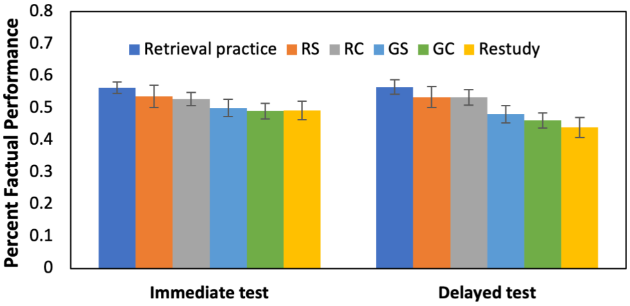

Fig. 2 Students' factual knowledge performance on the immediate and the one-week delayed tests

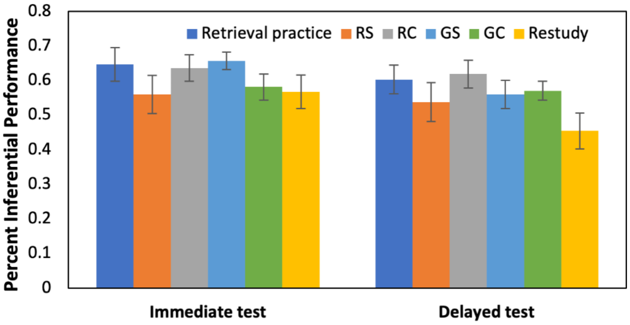

Fig. 3 Students' application knowledge performance on the immediate and the one-week delayed tests

We further conducted a one-way ANOVA to examine if there was any significant difference between the four drawing conditions on the delayed factual performance. Again, no difference was found on the four drawing conditions, $F(3, 87)=1.664$, p=.181, $\eta_p^2 = 0.054$. In addition, we also found that the retrieval practice condition produced essentially equivalent outcomes as the RC conditions, (0.57 vs. 0.52), t (43) = 1.35, p = .183, d = 0.11, 95%CI [−0.02, 0.11], and as the RS condition (0.57 vs. 0.53), t (44) = 0.793, p = .432, d = 0.13, 95%CI [−0.05, 0.11].

Application knowledge performance

Figure 3 shows students' performance on the application knowledge questions for both the immediate and the one-week delayed test. Again, a 6 × 2 mixed ANOVA revealed a statistically significant main effect of retention interval, $F(1, 130)=9.63$, p=.002, $\eta_p^2 = 0.069$. However, there was no statistically main effect of learning condition, $F(5, 130) = 1.21$, p = .308, $\eta_p^2 = 0.044$. We also did not find significant interaction effect, $F(5, 130) = 1.42$, p = .221, $\eta_p^2 = 0.052$. Thus, for the application knowledge performance, we did not find any difference between all the conditions on either the immediate or the delayed tests.

Discussion

The first goal of this study was to investigate whether retrieval practice could enhance high school students' learning of the carbon cycle. The results are consistent with findings from prior research which discovered retrieval practice could produce more benefits to long term retention of knowledge and conceptual understanding than the typically adopted restudy method (Karpicke, 2012, 2017; Roediger & Karpicke, 2006b; Roediger et al., 2010). More importantly, the current study also demonstrated that retrieval practice could be not only adopted as a stand-alone learning technique but implemented in combination with other well-established elaborative learning strategies such as sketching and concept mapping in the context of learning the carbon cycle topic at the high school level.

The second goal of the study was to compare the effect of two commonly used drawing strategies on students' learning outcomes of the carbon cycle. When students generated their models of the carbon cycle, similar levels of learning benefits were produced by creating either abstract drawing - the concept mapping or the representational drawing – sketching. The result pattern is unaffected regardless of whether they were implemented in combination with retrieval practice or not. Additionally, the concept mapping had comparable benefits for students' performance on both factual and application knowledge compared to the sketching. From a theoretical perspective, sketching and concept mapping could both follow the Select-Organize-Integrate procedure in generative learning theory for constructing a model (Fiorella & Mayer, 2016; Mayer, 2014). Also, the current study illustrated that high school students could very well master different types of representational skills for the learning of environmental science.

The last goal of this study focuses on whether different drawing strategies combined with retrieval practice could further enhance students' conceptual understanding of the carbon cycle. We found benefits when students created either concept maps or sketches with retrieval practice in the long run, especially on the factual knowledge, but this pattern was not found for the application questions (i.e., the flow-chart task) which requires students to transfer their learned knowledge to solve a problem in a new context. In addition, we were interested in whether drawing combined with retrieval practice (i.e., drawing in a closed-book format) could benefit students' learning outcomes compared to the stand-alone drawing strategy (i.e., drawing in an open-book format). In the present study, however, we did not find any additional benefits of combining retrieval practice with drawing over drawing with access to the learning materials. This null result was surprising since retrieval practice is often thought to be a powerful learning tool (Karpicke, 2012, 2017; Roediger & Karpicke, 2006b; Roediger et al., 2010). Drawing operationalized in retrieval practice should merge the benefit from both methods and the process should appear to encourage students to build a more robust mental model of the carbon cycle in the long run. Additionally, we did not find any difference between retrieval practice in the typical paragraphing format and retrieval practice combined with both drawing methods.

These null results are consistent with sparse prior studies which also failed to find additional benefits of combining retrieval practice with other learning strategies. One such study examined the effects of combining retrieval practice with an imagery-based elaborative keyword method (Karpicke & Smith, 2012). The researchers found this elaborative strategy did not produce any further learning when it occurred after several successful retrieval practices. Furthermore, Blunt and Karpicke (2014) asked students to practice retrieval either by paragraphing or by creating a concept map when learning a science expository text. On a one-week final test, they found either paragraphing or concept mapping in a closed-book style produced comparable benefits. Additionally, they also examined practicing retrieval by paragraphing is more effective than creating concept maps in an open-book style (Blunt & Karpicke, 2014). However, the authors did not compare the closed-book concept mapping condition with the open-book concept mapping condition, therefore they did not report if there was no difference between these two styles. In another study, O'Day and Karpicke (2020) had college students either create a concept map prior to practice retrieval by freely recalling or complete these two strategies separately when learning an educational material. The results revealed that when students worked on both activities, their outcomes failed to produce any additional benefits over practicing retrieval alone.

The lack of additional benefits when retrieval practice was implemented simultaneously with either concept-mapping or sketching begs for an explanation. One possible explanation is the typical cognitive process required by retrieval practice is disturbed when students work on drawing tasks. From the cognitive load perspective (Sweller, 2010; Sweller et al., 2019), engaging in both retrieval and drawing may overwhelm the limited working memory capacity (Cowan, 2010). When students sketched or created concept maps, they might distribute a portion of their cognitive energy on thinking of building graphical representations. As a result, they would not be able to spend all their cognitive resources at recalling information from the text materials as they would if they engaged in retrieval in its pure form. However, this explanation seems to contradict our discovery that there was no difference in the number of idea units about the carbon cycle text identified by students between the generative concept mapping and retrieval concept mapping conditions, or between the generative sketching and retrieval sketching conditions during initial learning. More future studies are needed to investigate the possibility.

Another possible account could be related to the cohesion and elaboration level of the materials adopted. This explanation is generated from the findings of Roelle and Nückles (2019). In their study, the authors compared generative learning and retrieval practice by using expository text of either high or low levels of cohesion and elaboration. Specifically, learners who are in the generative condition were asked to highlight the main content items of the expository text and then to illustrate the main content items of the text with access to the text. Learners in the retrieval condition, however, were prompted to recall as much information as possible from the expository text. The authors reported that engaging learners in retrieval practice was beneficial when the expository text was of high cohesion and elaboration, whereas engaging learners in generative activities was not. By contrast, when the learning material was of low cohesion and elaboration, only engaging learners in generative learning activities was beneficial (Roelle & Nückles, 2019). In the current study, the cohesion and elaboration level of the carbon cycle text might be positioned as "medium", thus the advantage of either generative drawings or drawing with retrieval practices might be "compromised" or "neutralized". This possibility certainly needs future research before such a conclusive assertion is offered.

Conclusions

Research exploring effect of combining different educational strategies remains quite scarce, especially within the context of environmental science. There is considerable room for future research on integrating retrieval practice with other effective learning strategies. This study sheds some initial light on the effect of using two different drawing methods synergistically with retrieval practice to promote students' understanding of the carbon cycle. The implication for climate change education is that high school students should be encouraged to adopt retrieval practice in their learning experience to reap its long-term mnemonic benefits. In the meanwhile, most previous applied studies on retrieval practice have been conducted with student population from the North American and Western Europe countries (Agarwal et al., 2021). This study expands the population sample by the inclusion of high school students from an East Asian country which has a different cultural background and educational practices. It provides further evidence supporting retrieval practice as a universally effective learning strategy.

Despite the encouraging findings, there are some limitations in this study. It should be noted that the current study was conducted only once with relatively small sample sizes. In addition, the measurement instrument adopted in the study is required to be tested with more student population to increase its validity and reliability. Future studies could design courses within the context of environmental science through integrating retrieval-based learning with more participants for a longer duration (e.g., one semester) to examine its effectiveness. Additionally, besides quantitative data collection, qualitative methods such as structured or semi-structured interviewing and/or field observations could be employed to further investigate students' attitudes and experiences towards retrieval-based learning. In general, our study opens the possibility for future research on exploring ways to broadly incorporate retrieval practice with other highly effective learning strategies to maximize the learning efficiency in climate change education.

List of abbreviations

NGSS	Next Generation Science Standards
NRC	National Research Council
MSFL	Minute Sketches in Folded Lists
PISA	Program for International Student Assessment
GS	Generative Sketching
GC	Generative Concept mapping
RS	Retrieval Sketching
RC	Retrieval Concept mapping

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43031-023-00083-4.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

The authors acknowledge E. Desmond Lee Scholars Program for the scholarship granted.

Authors' contributions

MW contributed to the design of the study, the analysis of data and drafted the manuscript. MY contributed to the improvement on the assessment tool for the study and implemented both the pilot and formal study. WK contributed to the development and improvement of the research questions, the theoretical framework and revised the manuscript.

Funding

This study was supported by the E. Desmond Lee Family Professor of Science Education endowment fund at the University of Missouri - St. Louis.

Availability of data and material

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

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 18 March 2023 / Accepted: 30 October 2023

References

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.

Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science*, *333*(6046), 1096–1097. https://doi.org/10.1126/science.1204153.

Alesandrini, K. L. (1984). Pictures and adult learning. *Instructional Science*, *13*(1), 63–77.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. Holt.

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, *106*(3), 849–858. https://doi.org/10.1037/a0035934.

Butler, A. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*, 1118–1133. https://doi.org/10.1037/a0019902.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(6), 1563–1569. https://doi.org/10.1037/a0017021.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(6), 1547. https://doi.org/10.1037/a0024140.

Cheng, M. H. M., & Wan, Z. H. (2016). Unpacking the paradox of Chinese science learners: Insights from research into Asian Chinese school students' attitudes towards learning science, science learning strategies, and scientific epistemological views. *Studies in Science Education*, *52*(1), 29–62. https://doi.org/10.1080/03057267.2015.1112471.

Chularut, P., & DeBacker, T. K. (2004). The influence of concept mapping on achievement, self-regulation, and self-efficacy in students of English as a second language. *Contemporary Educational Psychology*, *29*(3), 248–263. https://doi.org/10.1016/j.cedpsych.2003.09.001.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277.

Düsing, K., Asshoff, R., & Hammann, M. (2019). Students' conceptions of the carbon cycle: Identifying and interrelating components of the carbon cycle and tracing carbon atoms across the levels of biological organization. *Journal of Biological Education*, *53*(1), 110–125. https://doi.org/10.1080/00219266.2018.1447002.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9.

Greca, I. M., & Moreira, M. A. (2000). Mental models, conceptual models, and modelling. *International Journal of Science Education*, *22*(1), 1–11. https://doi.org/10.1007/1-4020-3613-2_2.

Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, *28*(9), 799–822. https://doi.org/10.1002/tea.3660280907.

Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education*, *22*(9), 1011–1026. https://doi.org/10.1080/095006900416884.

Hartley, L. M., Wilke, B. J., Schramm, J. W., D'Avanzo, C., & Anderson, C. W. (2011). College students' understanding of the carbon cycle: Contrasting principle-based and informal reasoning. *Bioscience*, *61*(1), 65–75. https://doi.org/10.1525/bio.2011.61.1.12.

Heideman, P. D., Flores, K. A., Sevier, L. M., & Trouton, K. E. (2017). Effectiveness and adoption of a drawing-to-learn study tool for recall and problem solving: Minute sketches with folded lists. *CBE—Life Sciences Education*, *16*(2), ar28. https://doi.org/10.1187/cbe.16-03-0116.

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test… or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, *26*, 307–329. https://doi.org/10.1007/s10648-013-9248-9.

Jensen, J. L., McDaniel, M. A., Kummer, T. A., Godoy, P. D., & Clair, S., B (2020). Testing effect on high-level cognitive skills. *CBE—Life Sciences Education*, *19*(3), ar39. https://doi.org/10.1187/cbe.19-10-0193.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, *101*(3), 621. https://doi.org/10.1037/a0015183.

Johnson, J. K., & Reynolds, S. J. (2005). Concept sketches - using student-and instructor-generated, annotated sketches for learning, teaching, and assessment in geology courses. *Journal of Geoscience Education*, *53*(1), 85–95.

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.

Karpicke, J. D. (2017). *Retrieval-based learning: A decade of Progress*. Grantee Submission.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775. https://doi.org/10.1126/science.1199327.

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, *67*(1), 17–29. https://doi.org/10.1016/j.jml.2012.02.004.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of learning and motivation* (Vol. 61, pp. 237–284). Academic Press.

Kindfield, A. C. (1994). Biology diagrams: Tools to think with. *The Journal of the Learning Sciences*, *3*(1), 1–36. https://doi.org/10.1207/s15327809jls0301_1.

Larsen, D. P., Butler, A. C., & Roediger, I. I. I., H. L (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education*, *47*(7), 674–682. https://doi.org/10.1111/medu.12141.

Lau, K. C., & Ho, S. C. E. (2020). Attitudes towards science, teaching practices, and science performance in PISA 2015: Multilevel analysis of the Chinese and western top performers. *Research in Science Education*, 1–12. https://doi.org/10.1007/s11165-020-09954-6.

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*, 291–304. https://doi.org/10.1007/s10648-015-9296-4.

Leutner, D., Leopold, C., & Sumfleth, E. (2009). Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. *Computers in Human Behavior*, *25*(2), 284–289. https://doi.org/10.1016/j.chb.2008.12.010.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). Cambridge University Press. Second Editionth ed.https://doi.org/10.1017/CBO9781139547369.

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*, *43*, 31–41. https://doi.org/10.1007/BF02300480.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, *27*(3), 360–372. https://doi.org/10.1002/acp.2914.

McNeill, K. L., & Vaughn, M. H. (2012). Urban high school students' critical science agency: Conceptual understandings and environmental actions around climate change. *Research in Science Education*, *42*(2), 373–399. https://doi.org/10.1007/s11165-010-9202-5.

Millet, A., Turcotte, N., & Yan, S. (2021). Retrieval Practice and Online Learning. *International perspectives in Online instruction* (Vol. 40, pp. 95–112). Emerald Publishing Limited.

Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching*, *46*(6), 675–698. https://doi.org/10.1002/tea.20314.

National Research Council. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. National Academies Press.

NGSS Lead States. (2013). *The next generation science standards: For states, by states*. National Academies Press.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, *27*(10), 937–949. https://doi.org/10.1002/tea.3660271003.

Novak, J. D. (2005). Results and implications of a 12-year longitudinal study of science concept learning. *Research in Science Education*, *35*(1), 23–40. https://doi.org/10.1007/s11165-004-3431-4.

Novak, J. D., & Cañas, A. J. (2008). The theory underlying concept maps and how to construct and use them. Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, 2008, available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge University Press.

Novak, J. D., & Wandersee, J. (1991). Coeditors, special issue on concept mapping. *Journal of Research in Science Teaching*, *28*(10).

O'Day, G. M., & Karpicke, J. D. (2020). Comparing and combining retrieval practice and concept mapping. *Journal of Educational Psychology*, *113*(5), 986–997. https://doi.org/10.1037/edu0000486.

Quillin, K., & Thomas, S. (2015). Drawing-to-learn: A framework for using drawings to promote model-based reasoning in biology. *CBE—Life Sciences Education*, *14*(1), es2. https://doi.org/10.1187/cbe.14-08-0128.

Rennie, L. J., & Jarvis, T. (1995). Children's choice of drawings to communicate their ideas about technology. *Research in Science Education*, *25*, 239–252. https://doi.org/10.1007/BF02357399.

Roediger, I. I. I., H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x.

Roediger, I. I. I., H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x.

Roediger, I. I. I., Agarwal, H. L., Kang, P. K., S. H., & Marsh, E. J. (2010). Benefits of testing memory: Best practices and boundary conditions. *Current issues in applied memory research* (pp. 27–63). Psychology Press.

Roelle, J., & Nückles, M. (2019). Generative learning versus retrieval practice in learning from text: The cohesion and elaboration of the text matters. *Journal of Educational Psychology*, *111*(8), 1341–1361. https://doi.org/10.1037/edu0000345.

Roelle, J., Froese, L., Krebs, R., Obergassel, N., & Waldeyer, J. (2022). Sequence matters! Retrieval practice before generative learning is more effective than the reverse order. *Learning and Instruction*, *80*, 101634. https://doi.org/10.1016/j.learninstruc.2022.101634.

Rosengrant, D., Van Heuvelen, A., & Etkina, E. (2009). Do students use and understand free-body diagrams? *Physical Review Special Topics-Physics Education Research*, *5*(1), 010108. https://doi.org/10.1103/PhysRevSTPER.5.010108.

Rowland, C. A., & DeLosh, E. L. (2015). Mnemonic benefits of retrieval practice at short retention intervals. *Memory (Hove, England)*, *23*(3), 403–419. https://doi.org/10.1080/09658211.2014.889710.

Scheiter, K., Schleinschok, K., & Ainsworth, S. (2017). Why sketching may aid learning from science texts: Contrasting sketching with written explanations. *Topics in Cognitive Science*, *9*(4), 866–882. https://doi.org/10.1111/tops.12261.

Schleicher, A. (2019). *PISA 2018: Insights and interpretations*. oecd Publishing.

Schmeck, A., Mayer, R. E., Opfermann, M., Pfeiffer, V., & Leutner, D. (2014). Drawing pictures during learning from scientific text: Testing the generative drawing effect and the prognostic drawing effect. *Contemporary Educational Psychology*, *39*(4), 275–286. https://doi.org/10.1016/j.cedpsych.2014.07.003.

Schwamborn, A., Thillmann, H., Opfermann, M., & Leutner, D. (2011). Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *Computers in Human Behavior*, *27*(1), 89–93. https://doi.org/10.1016/j.chb.2010.05.028.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *46*(6), 632–654. https://doi.org/10.1002/tea.20311.

Shepardson, D. P., Niyogi, D., Roychoudhury, A., & Hirsch, A. (2012). Conceptualizing climate change in the context of a climate system: Implications for climate and environmental education. *Environmental Education Research*, *18*(3), 323–352. https://doi.org/10.1080/13504622.2011.622839.

Shepardson, D. P., Roychoudhury, A., & Hirsch, A. (2017). Climate change as an issue for socio-scientific issues teaching and learning. In D. P. Shepardson, A. Roychoudhury, & A. S. Hirsch (Eds.), *Using conceptual and physical models to develop students' Mental models of the Greenhouse Effect. In teaching and learning about Climate Change* (pp. 85–105). Routledge.

Sibley, D. F., Anderson, C. W., Heidemann, M., Merrill, J. E., Parker, J. M., & Szymanski, D. W. (2007). Box diagrams to assess students' systems thinking about the rock, water and carbon cycles. *Journal of Geoscience Education*, *55*(2), 138–146.

Smith, G. A., & Bermea, S. B. (2012). Using students' sketches to recognize alternative conceptions about plate tectonics persisting from prior instruction. *Journal of Geoscience Education*, *60*(4), 350–359. https://doi.org/10.5408/11-251.1.

Smith, M. A., Roediger, I. I. I., H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(6), 1712. https://doi.org/10.1037/a0033569.

Suwa, M., & Tversky, B. (1997). What do architects and students perceive in their design sketches? A protocol analysis.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*, 123–138. https://doi.org/10.1007/s10648-010-9128-5.

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*, 261–292. https://doi.org/10.1007/s10648-019-09465-5.

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*, 247–264. https://doi.org/10.1007/s10648-015-9310-x.

van Gog, T., Kester, L., Dirkx, K., Hoogerheide, V., Boerboom, J., & Verkoeijen, P. P. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review*, *27*, 265–289. https://doi.org/10.1007/s10648-015-9297-3.

van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology*, *93*(1), 129–140. https://doi.org/10.1037/0022-0663.93.1.129.

van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, *17*(4), 285–325. https://doi.org/10.1007/s10648-005-8136-3.

van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology*, *31*(2), 142–166. https://doi.org/10.1016/j.cedpsych.2005.04.001.

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *The Quarterly Journal of Experimental Psychology*, *69*(9), 1752–1776. https://doi.org/10.1080/17470218.2015.1094494.

Wu, S. P., & Rau, M. A. (2019). How students learn content in science, technology, engineering, and mathematics (STEM) through drawing activities. *Educational Psychology Review*, *31*(1), 87–120. https://doi.org/10.1007/s10648-019-09467-3.

Zangori, L., & Koontz, J. A. (2016). Supporting upper-level undergraduate students in building a systems perspective in a botany course. *Journal of Biological Education*, *51*(4), 399–411. https://doi.org/10.1080/00219266.2016.1257502.

Zangori, L., Peel, A., Kinslow, A., Friedrichsen, P., & Sadler, T. D. (2017). Student development of model-based reasoning about carbon cycling and climate change in a socio-scientific issues unit. *Journal of Research in Science Teaching*, *54*(10), 1249–1273. https://doi.org/10.1002/tea.21404.

Zhai, X., He, P., & Krajcik, J. (2022). Applying machine learning to automatically assess scientific models. *Journal of Research in Science Teaching*, *59*(10), 1765–1794. https://doi.org/10.1002/tea.21773.

Zu, T., Munsell, J., & Rebello, N. S. (2019). Comparing retrieval-based practice and peer instruction in physics learning. *Physical Review Physics Education Research*, *15*(1), 010105. https://doi.org/10.1103/PhysRevPhysEducRes.15.010105.

Publisher's Note