Applying Retrieval Practice in Climate Change Education: How Retrieval-based Learning May Enhance High School Students’ Conceptual Understanding on Climate Change Topics

Mengyu Wang
University of Missouri-St. Louis, mwh79@umsystem.edu

Follow this and additional works at: https://irl.umsl.edu/dissertation

Part of the Curriculum and Instruction Commons, Educational Assessment, Evaluation, and Research Commons, Educational Methods Commons, and the Educational Technology Commons

Recommended Citation
https://irl.umsl.edu/dissertation/1282

This Dissertation is brought to you for free and open access by the UMSL Graduate Works at IRL @ UMSL. It has been accepted for inclusion in Dissertations by an authorized administrator of IRL @ UMSL. For more information, please contact marvinh@umsl.edu.
Applying Retrieval Practice in Climate Change Education: How Retrieval-based Learning May Enhance High School Students’ Conceptual Understanding on Climate Change Topics

Mengyu Wang

M.Ed., Theory of Curriculum and Instruction, Beijing Normal University, 2017
B.S., Geography Science, Fujian Normal University, 2014

A Dissertation Submitted to The Graduate School at the University of Missouri-St. Louis in partial fulfillment of the requirements for the degree Doctor of Philosophy in Education with an emphasis in Teaching and Learning Processes

December 2022

Advisory Committee

William C. Kyle, Jr. Ph.D.
Chairperson

Cody Ding, Ph.D.

Charles Granger, Ph.D.

Keith Miller, Ph.D.

Copyright © Mengyu Wang, 2022
Abstract

This dissertation explored ways to apply retrieval practice strategy on climate change learning at the high school level. Three studies presented that retrieval practice could be used broadly on different topics of climate change and in different learning scenarios. The first study focused on the effect of question placement on learning when students studied sea level rise from an educational video lecture. The results suggested that applying retrieval practice in a style of embedded short-answer questions in videos has clear advantage over restudying the video lecture in both immediate and delayed tests, but no difference in students' performance was found between the two common question embedding practices - dispersed and stacked question styles. The second study expanded the application of retrieval practice by combining it with learner-generated drawing strategy. We explored the effect of retrieval-based sketching method with different levels of retrieval support when students studied the Earth's energy budget from a video lecture. Results did not show any significant difference in students' performance between the "no support", "50% support", and "100% support" drawing conditions on either the immediate or the delayed test. Building on the findings of the second study, the third study focused on the effect of retrieval practice combined with different drawing strategies (concept mapping or sketching) on carbon cycle learning. Results demonstrated that retrieval practice by paragraphing was more beneficial on learning than concept mapping or sketching with open access to the text learning material, or simply restudying the material. However, the results also revealed no difference between the two drawing types on learning, regardless of whether they were implemented with retrieval practice. Generally, retrieval practice could engage learners to deepen the learning processes by
reconstructing their knowledge after initial learning. We have demonstrated retrieval practice can be implemented at the high school level to facilitate the learning of various climate science topics.
Acknowledgements

People often say the dissertation journey is like climbing a high mountain, which tests our mettle and challenges our resolve. I am so fortunate that I have received a great deal of help, support, encouragement and love during this journey. I would like to thank the following people, without whom I would not have been able to reach the top of this journey.

I would like to thank my advisor, Dr. William C. Kyle, Jr., who provided me with the valuable opportunity to pursue my doctoral degree in the United States. He has devoted so much time, effort and wisdom for my doctoral program. Without his guidance and support, I would’ve not been able to complete this challenging article-based dissertation with three independent but interrelated studies.

I would like to thank my dissertation committee members. They provided a lot of insightful ideas and suggestions on my research. Dr. Cody Ding provided me with lots of guidance and help on the quantitative data analysis. With his instruction on various quantitative methods, I was able to establish the logic of hypothesis testing and master relevant statistical methods for my study. Dr. Charles Granger provided a lot of background information on the current STEM education by sharing insightful ideas and valuable articles, which helped me develop an overview of my research area. Dr. Keith Miller encouraged me to try both qualitative and quantitative methods for my research. I built a deep understanding of the advantages and disadvantages of both methods with his guidance.

I would like to thank Ms. Meredith Jacques, who is the bioscience director of Sparks program in the Parkway School District - St. Louis. She offered me opportunities
to visit her class on a regular basis. I was inspired a lot by her active instructional strategies and her personal professional experience on teaching science. She also shared a lot of insightful ideas and helped edit the research materials for my dissertation. In addition to this, with her help, I could complete data collection for the first and the second studies of the dissertation. I am also really excited that we made a presentation on our collaborative study together in the ISTE 2021 conference.

I would like to thank Mr. Ming Yang, who is a geography teacher from Huipu Middle school in Zhejiang, China. With his help, I could complete the data collection on the last study of the dissertation.

I would like to thank all my friends in China and the U.S. Thank you all for encouraging me when I wanted to give up.

I would like to thank my dear parents. With your continuous support, I had the courage to start my study journey in the U.S.

I would also like to thank my dear husband, Tianlong. I am fortunate to have met you on this journey and thank you for being with me all the time especially when I was frustrated.

Finally, I would like to thank my dear baby boy, Dundun. You are my best gift and reward on the journey!
Table of Contents

Abstract ...................................................................................................................... I
Acknowledgements .................................................................................................. III
List of Figures .......................................................................................................... VII
List of Tables ........................................................................................................... IX

Chapter 1: Introduction .......................................................................................... 1
References ................................................................................................................ 6

Chapter 2: Effect of question placement on learning from sea level rise online video lectures in high school environmental science class ......................................................... 9
Abstract .................................................................................................................. 9
Introduction ............................................................................................................. 10
Theoretical background .......................................................................................... 12
Method .................................................................................................................... 21
Results .................................................................................................................... 26
Discussion ............................................................................................................... 31
Appendix A ............................................................................................................ 36
References .............................................................................................................. 39

Chapter 3: Effect of different levels of retrieval-based drawing on high school students’ conceptual understanding of the Earth’s energy budget ...................................................... 50
Abstract .................................................................................................................. 50
Introduction ............................................................................................................. 52
Theoretical background .......................................................................................... 55
Method .................................................................................................................... 62
Results .................................................................................................................... 67
Discussion ............................................................................................................... 74
Appendix B ............................................................................................................ 79
References .............................................................................................................. 87

Chapter 4: Effect of retrieval practice and drawing on high school students’ conceptual understanding of the carbon cycle .............................................................................. 96
Abstract .................................................................................................................. 96
Introduction ............................................................................................................. 97
Theoretical background .......................................................................................... 101
Method .................................................................................................................... 111
Results .................................................................................................................... 116
Discussion ............................................................................................................... 124
Appendix C ............................................................................................................................ 129
References ............................................................................................................................ 141
Chapter 5: Summary ............................................................................................................. 152
List of Figures

Figure 1  Schematic procedure of the design ................................................................. 25
Figure 2  Mean score proportion of students for short-answer embedded questions during video learning ................................................................. 27
Figure 3  Mean score of students for all 16 items in the SLRS on the immediate test ............ 28
Figure 4  Mean score of students for both the near transfer and far transfer questions in the SLRS on the immediate test ................................................. 28
Figure 5  Mean score of students for all 16 items in the SLRS on the one-week delayed test ...... 29
Figure 6  Mean score of students for both the near transfer and far transfer questions in the SLR survey on the one-week delayed test ............................................... 31
Figure 7  Visual framework for the generative theory of drawing construction ..................... 56
Figure 8  Scheme of the design ..................................................................................... 66
Figure 9  Students’ performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means. Condition 1 represents the non-support retrieval-based sketching condition, condition 2 represents the half-support retrieval-based sketching condition and condition 3 represents the restudy condition .... 68
Figure 10 Students’ drawing task’s performance (proportion correct) on the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means .......... 69
Figure 11 Students’ factual knowledge performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means ..................................................................................... 72
Figure 12 Students’ application knowledge performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means ..................................................................................... 72
Figure 13 Students’ performance (proportion correct) on the immediate and the one-week delayed tests as a function of the question type. Error bars represent standard errors of the means ..................................................................................... 73
Figure 14 Students’ performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means. GS = generative sketching, GC = generative concept mapping, RS = retrieval sketching, RC = retrieval concept mapping ..................................................................................... 121
Figure 15 Students’ factual knowledge performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means. GS = generative sketching, GC = generative concept mapping, RS = retrieval sketching, RC = retrieval concept mapping ..................................................................................... 123
Figure 16 Students’ application knowledge performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors
of the means. GS = generative sketching, GC = generative concept mapping, RS = retrieval sketching, RC = retrieval concept mapping

124
List of Tables

Table 1 Mean Score of Students for Different Questions from the Three Conditions on Both Immediate and Final Tests. .................................................................................................................. 26

Table 2 Mean Score of Students for Different Questions from the Three Conditions on Both Immediate and Final Tests. .................................................................................................................. 68

Table 3 Students’ Ratings of Enjoyment, Interestingness and Difficulty of the learning activity 74

Table 4 Introduction of the six conditions .................................................................................................................. 112

Table 5 Procedure of the study design .................................................................................................................. 116

Table 6 Mean Score of Students for Different Questions from the Six Conditions on Both Immediate and One-week delayed Tests............................................................................................................. 117

Table 7 Mean Score of Students for the number of idea units covered in the initial learning activity .................................................................................................................. 119
Chapter 1: Introduction

In January 2020, a group of more than 11,000 scientists from 153 countries declared that climate change is a clear and unequivocally "emergency" that requires decision-makers and all of humanity promptly to respond (Kyle, 2020a,b; Ripple et al., 2020). According to the Intergovernmental Panel on Climate Change (IPCC) recent special reports, there is strong evidence that climate change is more severe than anticipated, which is not only affecting weather patterns and ecosystems but is aggravating existing risks to biodiversity, food system, economy, human-health and individuals’ life safety (IPCC, 2018). In fact, we have experienced unprecedented and devastating natural disasters since 2020 (Kyle, 2020a,b). Apart from the pandemic of Covid-19, record-breaking wildfires had been raging along not only the coast of Australia but also the western United States. The Atlantic hurricane season broke all-time record with 30 named storms and 12 landfalling storms in the continental United States, leaving the Gulf Coast, the eastern coastline of United States and even Central America battered (NOAA, 2020; Samenow et al., 2020). Moreover, Eastern Africa faced the worst locust outbreak in the past 25 years, which threatened food security throughout Africa, where 51.2% of people face moderate or severe acute food insecurity (FAO, 2020). One of the main factors causing these afore-mentioned climate catastrophes is the rising global temperature.

It is no doubt that the global warming is inevitable due to industrialization and human activities. Human activities modify the Earth’s landscape and change the atmospheric composition of greenhouse gases, causing global temperatures to rise, which in turn results in the Earth’s climates to change (IPCC, 2013, 2019; EPA, 2012). Thus, it
is imperative that young generation understands the causes, mechanisms and implications of climate change in order to become informed citizens who engage in democratic decision making regarding a collective future (Shepardson et al., 2017).

Education is an essential element of the global response to climate change (UNESCO, 2013). To help young people understand and address the impact of the global warming, and to encourage them to change their attitudes and behaviors for adapting to climate change, the United Nations Educational Scientific and Cultural Organization (UNESCO) developed two global programs -“Climate Change Education for Sustainable Development Program” and “The Global Action Program (GAP) on Education for Sustainable Development (ESD)”, aiming to make education a more central and visible part of the international response to climate change and sustainable development.

Unfortunately, many students lack a deep comprehension of global warming and climate change. Generally, secondary students have misunderstandings of basic concepts on climate change, including, they couldn’t differentiate greenhouse gases from non-greenhouse gases and weather from climate (Shepardson et al., 2014). Besides, they have difficulties understanding global warming on a basis of climate system, that is, they struggled to understand how climate change is a result of a change to the climate system and how changing one component of the climate system impacts other components of the climate system (Shepardson et al., 2012). Therefore, there is a need to improve students’ conceptualization of climate change, such that they could become better informed decision makers.

On the other hand, the UN General Assembly proposed 17 sustainable development goals (SDGs) to advocate people around the world to take actions to
“combat climate change and its impacts, improve education on this very topic, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning” (United Nations, 2015, p.25). Science education plays a crucial role in preparing youth to be in a knowledgeable and decision-making position to address climate change and sustainable development issues (Kyle, 2020b, c; Sharma, 2012). However, “science educators have often remained silent with respect to ensuring the goals of science education are linked intrinsically to the central tenets of sustainable development” (Kyle, 2020c, p.1). Therefore, it is also critical for science educators to work on developing instructional strategies to facilitate students’ fundamentals about global warming and climate change.

Knowledge regarding climate change is complex and interdisciplinary which requires creative instructional strategies. Herein, this dissertation explores the effectiveness of instructional strategies on high school students’ basic comprehension specifically over three important topics of climate change: sea level rise, the global carbon cycle and the Earth’s energy budget. Chapter 2 discusses how to facilitate students’ conceptual understanding on sea level rise in an online video learning environment. Sea level rise is one of the most visible and major impacts of climate change, while students hold incomplete understandings of the causes and mechanisms. Although video lectures with embedded questions provide students convenient ways to learn about sea level rise, how to optimize video-based learning on climate change with embedded questions remains an open issue.

Chapter 3 focuses on the Earth’s energy budget which is an important concept for high school students to get a deep understanding of the mechanisms of the greenhouse
effect and global warming. Many students struggle to understand the greenhouse effect and its connection to the Earth’s energy budget and global warming (Shepardson et al., 2011). For example, students do not understand that the greenhouse effect is the main driver of global warming, causing the Earth’s climate to change. This requires more thoughtful instructional methods in high school classroom teaching. Retrieval-based drawing has been found to be effective in facilitating students’ learning performance as well as long-term retention of knowledge in elementary and college level (Heideman et al., 2017; Karpicke & Blunt, 2011, 2014). However, there is no prior study investigating the effectiveness of this learning strategy at the high school level. This work aims to fill this gap by answering an interesting question which is whether students benefit more if provided with scaffolds when applying retrieval-based drawing to learn the Earth’s energy budget.

The carbon cycle is a typical subsystem of climate and it lies at the heart of climate change and sustainability. It is important for students to think systematically on the climate system and climate change through developing and testing their own models (Shepardson et al., 2017). **Chapter 4** describes how to improve students’ basic understanding of the global carbon cycle using different drawing activities. Drawing has been shown to be an effective learning tool to enhance learning (Ainsworth et al., 2011). However, little is known when drawing is combined with different learning strategies such as generative learning practice or retrieval practice. Thus, this work aims to investigate how an integration of different drawing methods with either generative or retrieval practice influences students’ mental model of the global carbon cycle.
These three studies focus on the issues and challenges to teaching and learning about climate change. Findings will be used for helping develop instructional materials and approaches that better improve students’ learning and prepare students to make informed decisions about the climate change.
References


Intergovernmental Panel on Climate Change (IPPC). (2013). Climate change 2013: The physical science basis. IPCC, Geneva, Switzerland.


https://doi.org/10.1016/j.jarmac.2014.07.008

Kyle Jr, W. C. (2020a). Youth are demanding action regarding climate change: will educators have the wisdom and courage to respond? *APeduC Revista-Investigação e Práticas em Educação em Ciências, Matemática e Tecnologia, 1*(1), 150-160.


https://doi.org/10.1093/biosci/biz088


Chapter 2: Effect of question placement on learning from sea level rise online video lectures in high school environmental science class

Abstract

Although embedded questions have been shown to facilitate understanding while learning from video lectures, little is known on how question placement could affect students’ learning outcomes, retention, and transfer of knowledge. In this study, we compared two different question placement strategies with a typical learning method (restudy) in terms of their impact on learning in an environmental science class. Sixty high school students participated in the study. The first group (stacked condition) watched a video lecture on sea level rise and answered a set of embedded relevant short-answer questions at the end of the video lecture. The second group (dispersed condition) watched the same video in eight segments and answered one embedded question after each segment. The third group (restudy condition) re-watched the video. All students completed a conceptual survey consisting of multiple-choice questions on sea level rise immediately after the initial learning. Students returned to class and worked on the same survey again after one week. The results demonstrated using embedding questions has a clear advantage over restudying on both immediate and delayed tests. No difference has been observed between the stacked and dispersed conditions even though the dispersed condition showed an advantage over the stacked condition in performance on the embedded short-answer questions during the initial learning. Moreover, both the dispersed and stacked conditions significantly outperformed the restudy condition on the near-transfer questions in the survey but not on the far-transfer questions.
Keywords: sea level rise, embedded questions, interactive video lectures, learning transfer, environmental science education

Introduction

Learning science with educational videos has become ubiquitous in both K-12 and higher education settings in recent years. Students could learn directly from recorded video lectures in current flipped classrooms, blended courses and MOOCs (Bond, 2020; de Koning et al., 2018; Guo et al., 2014; Holbrook & Dupont, 2011; Lo & Hew, 2017; McGarr, 2009; Schultz et al., 2014). Video materials such as demonstrations, documentaries, and educational TV shows can be used as supporting or supplementary materials in class (Mayer & Fiorella, 2020). Although most previous studies suggested that educational videos contribute to better learning experience and outcomes on science subjects compared to reading-based learning materials (Love et al., 2013; Mayer, 2009), there are some challenges faced by students when video is the main source of instructional content delivery. For example, previous research in higher education indicated that college students showed the tendency of mind wandering while watching video lectures (Jing et al., 2016; Risko et al., 2012; Szpunar et al., 2013). It has also been found that students tend to disengage with the courses in the MOOC environment (Davis et al., 2016, 2018; Yong & Lim, 2018). In addition, a handful of studies in K-12 also revealed that high school students have difficulty staying focused on watching video lessons and thought that learning from video is time-consuming and boring (Schultz et al., 2014; Snyder et al., 2014). Thus, there is a need for educators to consider how to optimize the use of educational videos to better engage students with video-based learning.
Recently, there is a body of studies examining that testing or quizzing could increase the effectiveness of learning from multimedia video materials. For example, researchers found that using embedded questions as a way of testing leads to a positive effect on learners’ attention and engagement (e.g., Cummins et al. 2016; Jing et al., 2016; Szpunar et al., 2013; Rose et al., 2016), motivation and self-efficacy (e.g., Buil et al. 2016; Mischel, 2019; van der Meij & Böckmann, 2020), and learning outcomes (e.g., Johnson & Mayer, 2009; Rose et al., 2016; Yong & Lim, 2018). Yet, studies exploring the effect of embedded questions during video learning have limited their research scope to higher education context and very few studies have focused specifically on science education in K-12 settings. This certainly requires more attention since online video lessons not only have been heavily adopted in higher education (Hung et al., 2018), but have become increasingly important in K-12 education, especially due to the outbreak of Covid-19 pandemic, where during the 2020-21 school year most classes were forced to be delivered in the virtual environment. It is inevitable that teachers will face the issue of disengagement from students when the course or a part of the course is delivered asynchronously such as using a flipped classroom format. To address this problem, different technology for delivering digital content remotely and facilitating engagement have been developed. For example, video-sharing platforms such as EDpuzzle, EduCanon, and Zaption offer ways to enhance video-based learning processes by adding interactive elements into educational video lessons (e.g., Kleftodimos & Evangelidis, 2016; Mischel, 2019; Rose et al., 2016). Amongst all interactive elements used, integrating embedded questions in video is a typical option for interactive methods.
However, how to best use embedded questions to improve high school students’ learning performance remains an open issue.

In this study, we explored the impact of two common patterns of embedded questions in video-based learning on high school students’ understanding of sea level rise. The two patterns used are 1) dispersing questions throughout video lectures and 2) stacking all questions at the end of video lectures. In particular, we planned to compare the benefits of the two different ways of embedding questions on students’ learning outcomes when learning materials are delivered in video format, and also examine how embedded questions as a way of testing impact students’ conceptual understanding and retention of knowledge on an important topic of climate change—sea level rise. Thus, the research questions are:

Q1: How do embedded questions impact high school students’ learning performance, retention of knowledge when learning from online video lectures?

Q2: How do different question placements (dispersed questions throughout the video lectures vs. stacked questions at the end of the video lectures) impact high school students’ conceptual understanding and learning transfer on sea level rise?

**Theoretical background**

**Sea level rise education and educational videos**

The current generation of youth is facing complex global challenges that will demand active response to the challenges and transformation of human’s lifestyle (Kyle, 2020a, b). According to the Intergovernmental Panel on Climate Change (IPCC) recent special reports, there is strong evidence that climate change is more severe than anticipated. It is not only affecting weather patterns and ecosystems but is aggravating
existing risks to people’s living environment. Sea level rise is one of the most visible and major impacts of climate change. In particular, according to IPCC, the sea level increased 3.6mm per year over the period of 2006-2015 (IPCC, 2019). Given that half of world’s population live along coastline and will be influenced by rising sea level if global temperature continues increasing (IPCC, 2019), the young generation should be provided with adequate opportunities to get deep comprehension of climate change and sea level rise.

Science education plays a major role in climate change education (McNeal, et al., 2014; Sharma, 2012). The Next Generation Science Standards (NGSS, 2013) has explicitly included the topic of climate change for K-12 science education. In particular, at the high school level, it emphasizes students should be able to construct evidence-based explanation on how climate change has influenced human activity; describe how variations in the flow of energy through the Earth’s systems result in changes in climate by constructing a model; and apply data and results from global climate models to make forecast of the current rate of climate change and the impacts to the Earth’s ecosystem (NGSS, 2013). However, sea level rise is not constructed as a whole component in any science education standards, and there are only a handful of studies examining learners’ climate change knowledge related to the topic of sea level rise (Breslyn et al., 2016).

Additionally, it has been recorded that some students hold misconceptions about sea level rise. For example, students might hold the view that sea level rises because of increased precipitation, or they might expect sea level to stay or decline the same as a result of evaporation (Shepardson et al., 2011). Other studies showed that middle school students have limited ability to provide sophisticated explanation for sea level rise, especially at
the atomic-molecular level (Breslyn et al., 2016). Therefore, there is a need for educational efforts that would facilitate learners’ understanding of the mechanism and impacts of sea level rise.

Multimedia instruction has a strong influence on the social, emotional, and cognitive development of learners, thus, plays an important role in science education (Clark & Mayer, 2016; Mayer, 2003; Mayer & Estrella, 2014). In the past decades, there has been tremendous growth in the popularity of the use of video for educational purposes in both formal and informal educational contexts (Bétrancourt & Benetos, 2018). The presentation format of educational videos is not limited to the direct recording of classroom teaching using a camera but expands to enhanced video of Power Point slides presented with an audio explanation, writing board screencasts with voice-over, and highly elaborate video post-production with picture-in-picture (Kay, 2012). Additionally, rapid technological advances in hardware (e.g., computers, smartphones, video cameras) and software (e.g., video recording apps, video-editing apps), as well as increasing access to the Internet allow educational videos to be created relatively easily and shared with others in online environments with minimal effort (van der Meij & van der Meij., 2013).

While learning from educational videos and online video-recorded courses (e.g., Coursera and edX) is not new in higher education, the emergence of video networking sites (e.g., YouTube) and a variety of online courses have expanded accessibility and use in K-12 settings in recent years (Kleftodimos & Evangelidis, 2016; Lo & Hew, 2017). Since climate change is a messy, complex scientific and social issue, educational videos can provide rich and vivid content delivery by integrating different sources and format of
visualization information into learning materials, which would help students better comprehend the complicated concepts of climate change. Thus, it is imperative to further explore how to design video lectures in climate change education.

**Interactive video lectures and embedded questions**

Interactive video lectures have evolved as the development of online educational environment and diverse needs of learners. The concept of interactive video can be traced back as early as 1980s, to the introduction of video-based delivery system with some level of microprocessor support, where learners can control the video presentation with a corresponding response by the delivery system that guarantees active participation in learning (Woolley, 1982). After decades of numerous improvements in video technology, video-based learning has evolved from passive instructive manner to an engaging interactive video experience for learners (Kay, 2012; Shephard, 2003). Compared with conventional educational video, interactive video incorporates the information management and decision-making capacity of computers with the capabilities of video (Schaffer & Hannafin, 1986; Woolley, 1982), which can support engagement by managing students’ attention. Specifically, interactive video allows learners to access to video content proactively and randomly based on queries or search targets (Zhang et al., 2006). For example, functions like playing, pausing, forwarding, or rewinding the video can provide adequate interactivity for users (Delen et al., 2014). Moreover, continuing advances in technology and theory-driven techniques provide interactive video lectures with rich-content interactive elements such as interactive learning activities (e.g., Hung et al., 2018), generative note-taking (e.g., Delen et al., 2014) and self-evaluation through use of interactive questions (e.g., Schacter & Szpunar, 2015). These interactive tools
allow learners to make their own choice and self-direction during video learning and therefore interactive video lectures are being gradually adopted in digital learning contexts.

Integrating embedded questions in online video lectures is a common strategy to stimulate more active or deeper learning (e.g., Christiansen et al. 2017; Cummins et al. 2016). van der Meij and Böckmann (2020) investigated the effectiveness of embedded open-ended questions in delivering social science content through online video lectures by comparing the test performance, self-efficacy and usability of college students who completed the embedded questions and those without such questions. Findings revealed that students from embedded questions group engaged significantly more and gained better learning performance than students who only watched video-recorded lectures (Meij and Böckmann, 2020). In another study, Szpunar et al. (2013) compared the test performance of college students who answered interpolated questions, students who did unrelated arithmetic questions and students who restudied the key content of the learning materials during a four-segmented video lecture on statistics, finding that the students in the interpolated question group had better learning outcomes on subsequent tests of the material and exhibited less task-irrelevant mind wandering. The results also suggested that interpolating video lectures with embedded questions as memory tests could help students sustain attention to lecture content in a manner that encourages task-relevant note-taking activities and reduce anxiety and subjective estimates of cognitive demand (Szpunar et al., 2013).

Although embedding questions in video lectures showed positive effect on learning, there still remains a question on what the best strategy is to place embedded
questions in video lectures. Prior studies focused on the effect of question placement showed mixed results when learning materials are in text format. For instance, Duchastel and Nungester (1984) suggested that inserting questions within a passage might be more beneficial than placing questions at the end when the passage is lengthier and more complex. However, Uner and Roediger (2018) compared the benefits of doing practice tests after each section and after the whole chapter when university students studied an introductory biology textbook chapter. The results showed that no difference was found on a final test which happened two days after the initial learning. Similar results were found in a study by Weinstein et al. (2016). In one experiment of their study, college students learned APA style online by viewing PowerPoint slides and answered one short-answer question on the materials either after 1-2 slides or after all questions together at the end of the presentation. They found that the students answering interspersed questions significantly had better performance than students who answered all at-the-end questions on the initial tests, but there was no difference between these two groups in a one-week delayed test.

In another study (Rice et al., 2019), university students completed learning from a set of three video lectures, with the first video having quiz questions embedded throughout, the second one having quiz questions added at the end, and the last one without any questions presented. The results showed that a quiz interpolated throughout a video generated significantly better learning outcomes compared to placing all quiz questions at the end of a video lecture. However, their study design did not counterbalance different treatments, such that it did not allow to rule out the possible confounding effects due to the temporal proximity of the tasks. That means the final
performance difference might be because the final test was taken one week after video quiz 3, two weeks after video quiz 2 and three weeks after video quiz 1. Thus, more evidence is needed to support the superiority of interpolated questions over stacked questions on learning from online video lectures.

**The Testing Effect and Transfer**

The explanation for the benefit of taking embedded questions on meaningful learning processes in online video lectures comes from the *testing effect* or the *retrieval practice effect*, which means that taking a memory test on previously presented learning information enhances learners’ later memory retention of those information (Carpenter, 2012; Karpicke, 2017; Roediger & Karpicke, 2006a, 2006b). In other words, engaging learners to take quizzes during study may stimulate them to recall or reconstruct information they have just learned and achieve more long-term benefits than others who adopt often-used learning strategies such as restudy or note-taking. While most previous research on the test-enhanced learning focused on retention of information measured through a final assessment that has similar questions or requirements to the initial test (e.g., Roediger & Karpicke, 2006a; Weinstein et al., 2016; Uner & Roediger, 2018), recent studies have started to pay more attention to the testing effect on “meaningful learning”, which refers to situations where the final assessment is considerably novel compared to the initial testing condition (e.g., Blunt & Karpicke, 2014; Butler, 2010; Hinze et al., 2013; Zu et al., 2019), since simple rote memorization of learning is not the ultimate and desired goal in education (Carpenter, 2012; Karpicke, 2017; Thomas et al., 2018).
Transfer of learning can be defined as being able to generalize previously learned information from one context to another (Barnett & Ceci, 2002), and is a critical goal in education (McDaniel et al., 2007; Pan & Rickard, 2018). Meaningful transfer occurs when students activate previously learned knowledge to make inferences or apply the knowledge in novel situations. Previous research examining the effects of testing-enhanced learning on transfer can be categorized in two general ways. The first approach is to assess final tests that include questions in different format compared to the initial testing but the context of questions in the final exam is almost similar to the original context in the initial testing. For example, McDaniel et al. (2007) examined testing effect in a university online course through either weekly quizzes or through restudy. They used fill-in-the-blank questions in weekly quizzes (e.g., “All preganglionic axons, whether sympathetic or parasympathetic, release __________ as a neurotransmitter”) and reorganized the questions by omitting different words of initial testing in final criterial tests (e.g., “All __________ axons, whether sympathetic or parasympathetic, release acetylcholine as a neurotransmitter”). They observed having students work on quizzes every week improved performance in relative to restudy even though the test formats were not identical between the initial testing (i.e., weekly quizzes) and the final exams.

The second general approach is to examine transfer by varying the types of questions in final exams compared to initial testing. In one experiment, Butler (2010) asked learners to learn from educational texts either through rereading the texts repeatedly or through taking repeated initial short-answer tests with feedback. One week later, learners were asked to complete a final test consisted of two types of questions: factual questions and conceptual questions, which required students to make inferences
rather simply recognize from learning materials. Most importantly, these questions in the final test were novel questions in relative to original questions in initial testing. Results indicated that the benefits of repeated testing are not limited to the originally practiced materials but rather extend to the transfer in a variety of contexts.

Although there has been a trend to examine transfer of test-enhanced learning in real educational settings, such studies have indicated mixed results (Glass 2009; Jensen et al., 2014; McDaniel et al., 2007, 2013; McDermott et al., 2014; Thomas et al., 2018). For example, McDaniel et al. (2013) found that middle school students demonstrated limited transfer if the final exam questions were different from the quiz questions during the initial learning in a science class. In their study, students completed multiple-choice questions based on either the definitional knowledge or application of the learned content. The findings showed that students who completed application questions during initial learning improved performance on definitional questions in the final exam relative to those who did not complete those application questions. Moreover, the use of quizzing with definitional knowledge did not facilitate students’ performance on the application questions on the final test even though they were testing the same concepts.

Contrary to this observation, results in a related study with college students revealed that the benefits of quizzing could be extended to different question types. In one study, Thomas et al. (2018) found that students’ performance on both factual and application exam questions improved if they took a quiz with either factual or application questions during initial learning. In addition, Glass (2009) investigated whether adopting a distributed presentation of different versions of a question during initial learning would help produce better learning outcomes on a new version of the question than using a
distributed presentation of the question of the same version. Results showed that students had better performance answering inference questions on the exam if they have taken a quiz with different versions of questions compared to those who took the same version of the question during initial learning phase. Given the above mixed results from previous studies, it also remains an open question on whether having embedded questions during video-based learning can enhance high school students’ transfer on a final exam.

**Method**

To address the research questions, this study employed three conditions. Two treatment conditions are based on two different ways of embedding questions during initial learning from educational videos on the sea level rise: 1) dispersing questions throughout the video lectures (dispersed condition) and 2) stacking all questions at the end of the video lectures (stacked condition). A restudy approach (restudy condition) was used as a control condition due to its popularity amongst students in their daily learning (Karpicke et al., 2009).

**Participants**

60 students from grade 11-12 who are concurrently enrolled in an environmental science class in an urban high school from a large Midwest city in the U.S participated in the study. All sessions of learning and tests took place during the regular class time. Students were randomly assigned into the three conditions with equal number of students per condition.

**Materials**

A 13-min video lecture introducing sea level rise is the main learning material. Eight short-answer questions related to sea level rise were used in both treatment
conditions. Most of the questions require only one answer, but the fourth question includes three sub-questions and the fifth question includes two sub-questions. The video and questions were designed by two experienced environmental science educators. The content was also independently inspected by a third science education researcher who agreed the content was suitable for the participants of the study. Feedback to each question was provided after each question. All materials are shown in the Appendix A.

The assessment tool\(^1\) for the immediate and the delayed test was developed and validated by the climate change learning science research group at the University of Maryland. The sea level rise survey (SLRS) consists of 16 multiple-choice questions which can be further sorted into 8 near-transfer questions and 8 far-transfer questions. The categorization of transfer is consistent with the work of Pan and Rickard (2018). According to their taxonomy, there are six major dimensions affecting the degree of transfer spanning from “near” to “far” transfer. The six factors include test formats, stimulus-response rearrangement, untested materials seen during initial study, application and inference questions, problem-solving skills, and mediator and related word cues. In the current study, we operationally defined the near transfer questions if the context is similar to the context of the embedded questions during initial learning but different in format (i.e. the embedded questions are in short-answer questions format while the survey questions are in multiple-choice questions format). The far transfer questions were different in both format and context than the embedded questions, so the students needed to make inference or apply their knowledge to respond these questions which are in novel contexts.

\(^1\) It can be downloaded from their website (http://www.climateedresearch.org).
Procedure

Participants were randomly assigned to the three conditions: dispersed ($n = 20$), stacked ($n = 20$) and restudy ($n = 20$). A schematic procedure can be found in Figure 1. Students from the dispersed condition watched the sea level rise video lecture which was divided into 8 clips and completed related short-answer questions after each video clip. In addition, students in the dispersed condition received immediate feedback and were asked to summarize what they have learned from the previous video clip. Students from the stacked condition watched the undivided video lecture and answered all 8 short-answer questions at the end. Students from the stacked condition also received immediate feedback to each question and were required to summarize what they have learned from the whole video in the end. Students in the restudy condition watched the undivided video lecture twice and were prompted to take notes the second time watching the video. For both treatment conditions, feedback to embedded questions was provided since immediate feedback allows students to correct memorial errors and facilitates comprehension (Hattie & Timperley, 2007). In addition, feedback was found to be quite effective in classroom (Kulik and Kulik 1988) and could help learners correct the metacognitive error that occurs when they could provide correct responses but with low confidence in their response (Butler et al., 2008).

After the initial learning session, all students took an immediate final test with the SLRS. To evaluate the retention of knowledge, students completed the survey again in one week. All sessions were completed by students on the Qualtrics platform. The total amount of learning time as recorded by Qualtrics platform was not significantly different among the three conditions ($p > .05$), indicating the time on task was equivalent in the
three conditions. The total time spent completing initial learning, immediate final test, and the one week-delayed test was about 70 mins spanning over one week.
Scoring and data analysis

We analyzed students’ answers to the short-answer embedded questions during the initial learning, SLRS questions in the immediate and delayed final tests. The short-answer questions were graded by determining the number of core ideas students included in their answers. There were 11 core ideas in the whole sea level rise video lecture. 6 of the 8 questions were based on a single core idea. Question 4 examined 3 core ideas and questions 5 examined 2 core ideas. Students received 1 point for providing any core idea, 0.5 point for partially correct core idea and 0 point for incorrect or vague answers. Thus, the score ranged from 0 to a maximum of 11 points for the embedded questions. For the SLRS, since there was only one correct answer to each multiple-choice question, students got 1 point for a correct answer and 0 point for an incorrect answer. Students’ scores on the SLRS ranged from 0 to 16 points. Students’ final scores on each test were converted to percentages for further analysis.

One-way ANOVA was conducted to compare students’ performance on the short-answer questions during initial learning between the two treatment conditions, on the immediate and the delayed test between the three conditions, and on the on the 8 near
transfer and the 8 far transfer questions in the SLRS separately. We also compared the GPA of students from different conditions to show if the three conditions were equivalent in terms of their academic baselines. All statistical analyses were conducted at a .05 alpha level.

Results

The ANOVA analysis test on GPA score revealed no significant difference among the three conditions at the .05 level, indicating the participants assigned to different conditions were equivalent satisfying the randomization assumption. Table 1 shows the mean proportion correctness and standard deviation of each condition on the overall SLRS, on the 8 near-transfer and the 8 far-transfer questions respectively. The proportion correctness score for the overall test was determined by dividing the total score of each participant by the highest possible score on the test, which is 16. The proportion correctness score for the near (far) transfer problems was determined by dividing the score of each participant on the 8 near (far) transfer questions by the maximum possible score of 8.

Table 1

*Mean Score of Students for Different Questions from the Three Conditions on Both Immediate and Final Tests.*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Overall survey 16-questions</th>
<th>8-near-transfer questions</th>
<th>8-far-transfer questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Immediate final test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dispersed(N=20)</td>
<td>0.84</td>
<td>0.17</td>
<td>0.89</td>
</tr>
<tr>
<td>Stacked(N=20)</td>
<td>0.80</td>
<td>0.21</td>
<td>0.83</td>
</tr>
<tr>
<td>Restudy(N=20)</td>
<td>0.72</td>
<td>0.24</td>
<td>0.74</td>
</tr>
<tr>
<td>One-week delayed test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dispersed(N=20)</td>
<td>0.86</td>
<td>0.14</td>
<td>0.91</td>
</tr>
<tr>
<td>Stacked(N=20)</td>
<td>0.86</td>
<td>0.11</td>
<td>0.91</td>
</tr>
<tr>
<td>Restudy(N=20)</td>
<td>0.74</td>
<td>0.18</td>
<td>0.81</td>
</tr>
</tbody>
</table>
Performance on the short-answer embedded questions

We first compared the performance of students on the short-answer embedded questions during the initial learning for the two treatment conditions (see Figure 2). We expected that the dispersed condition would show a benefit relative to the stacked condition when answering the short-answer embedded questions, due to separate questions with immediate feedback after learning each video clip. A one-way between-subject ANOVA revealed a significant main effect of condition on the short-answer questions performance, $F(1, 38) = 4.653, p = .037, \eta^2_p = 0.109$, such that students in the dispersed condition significantly outperformed those in the stacked condition.

Figure 2

Mean score proportion of students for short-answer embedded questions during video learning

Performance on the immediate final test

Next, we examined the performance of students from the three conditions on the immediate final test. A one-way ANOVA revealed no significant difference in
performance for the overall immediate test, $F(2, 57) = 1.961, p = .15, \eta^2_p = .064$. A parallel ANOVA analysis reported no significant difference in performance on the near-transfer questions among three conditions, $F(2, 57) = 2.709, p = .075, \eta^2_p = .087$. For the far-transfer questions, the difference among the three conditions did not reach significance either, $F(2, 57) = .961, p = .389, \eta^2_p = .033$. Further analysis on the trend lines found that for the scores on the whole survey (see Figure 3), the 8 near transfer questions and 8 far transfer questions separately (see Figure 4), the dispersed condition performed the best, followed by the stacked condition, which in turn did slightly better than the restudy condition.

**Figure 3**

*Mean score of students for all 16 items in the SLRS on the immediate test*

![Mean score of students for all 16 items in the SLRS on the immediate test](image)

**Figure 4**

*Mean score of students for both the near transfer and far transfer questions in the SLRS on the immediate test*

![Mean score of students for both the near transfer and far transfer questions in the SLRS on the immediate test](image)
Performance on the one-week delayed test

We then examined the performance of all three conditions on the one-week delayed test. Overall, a one-way ANOVA revealed a significant main effect of condition on the overall one-week delayed test performance, $F(2, 57) = 4.129, p = .021, \eta^2_p = .127$. Post hoc pairwise comparison analysis with a Bonferroni adjustment revealed that the dispersed condition significantly outperformed the restudy condition, ($M = .86$ vs. $M = .74$), $p = .047$, $d = .71$, and the stacked condition also scored significantly higher than the restudy condition, ($M = .86$ vs. $M = .74$), $p = .047$, $d = .78$, whereas the difference between the dispersed and stacked condition did not reach significance, $p = 1.0$ (see Figure 5).

Figure 5

*Mean score of students for all 16 items in the SLRS on the one-week delayed test*

We also conducted further analyses for the near-transfer and far-transfer questions performance on the delayed test among the three conditions (see Figure 6). For the near-transfer questions, a one-way ANOVA showed a main effect of condition, $F(2, 57) = 3.525, p = .036, \eta^2_p = .11$. Follow-up pairwise comparisons revealed that the dispersed condition marginally outperformed the restudy condition, ($M = .91$ vs. $M = .81$), $p = .091$, $d = .69$, and the stacked condition also marginally outperformed the restudy condition, ($M$
= .91 vs. \( M = .81 \), \( p = .064 \), \( d = .73 \). Again, the difference between the dispersed and stacked conditions did not reach significance, \( p = 1.0 \). For the far-transfer questions, a similar pattern of better performance for both the dispersed and stacked conditions than for the restudy condition was observed. However, data analysis revealed there was only marginally significant difference among the three conditions, \( p = .06 \).
Discussion

The outcome of the study reinforces that the integration of embedded questions into online video lectures could help increase students’ learning outcomes at the high school level, which extends the recent research highlighting the effectiveness of testing to promote video-based learning in both laboratory and authentic educational settings at university level. Specifically, when the time on task is the same, answering embedded questions led to greater learning than a common learning strategy - restudy where learners watch the video lecture again and take notes which provides evidence supporting the effectiveness of the testing effect in a video learning setting. In addition, we also found that whether spreading the embedded questions within a video lecture or stacking all questions at the end does not differentially affect students’ final performance or retention of knowledge in a week’s time. Further, embedded questions with feedback in a video lecture could promote students’ learning outcomes on the near-transfer questions but with limited influence on the far-transfer questions.
The primary interest of this study was to identify a more effective question placement strategy when learning materials are educational videos in high school environmental science class. By comparing the two common patterns of embedded questions in authentic video-based learning contexts, we found that despite the initial advantage of answering dispersed questions throughout the video learning, the dispersed and stacked conditions provided comparable benefits on both the immediate and delayed tests. Our findings confirmed and complemented prior studies showing equal benefits of interpolated questions and at-the-end questions using authentic learning materials, with normal lectures and lecture slides (Weinstein et al., 2016) and textbook chapters (Uner & Roediger, 2018).

On the other hand, our outcome also extended previous studies in which researchers focused on the effect of interpolated questions in video-based learning. Many educators accept and acknowledge the benefit to pause and interpolate questions during video learning (i.e., adding clicker questions or pop-up questions), which stimulated prior researchers to focus on the effect of interpolated questions to improve video-based learning (e.g., Jing et al., 2016; Szpunar et al., 2013). In one study, Jing et al (2016) illustrated that students working on interpolated recall questions during a video lecture outperformed those who only finished a single recall question for the last portion of the lecture on both the last recall question and the final cumulative assessment, indicating potential benefits of interspersing questions during video learning. However, they provided neither equivalent number of recall questions for each group nor feedback at the end of each question in their study.
The outcome of our study may seem counterintuitive since students in the dispersed condition did not outperform those in the stacked condition like they did on the embedded questions during initial learning, especially in the long run. One possible explanation is that both treatment conditions received detailed feedback to the embedded questions of which the corrective functions to students’ memorial errors during initial learning might have elevated the performance of those from the stacked condition to the same level as those from the dispersed condition on the following exams (Uner & Roediger, 2018). Another possible explanation is based on the theoretical framework of desirable difficulties or retrieval effort (Bjork, 1994; Bjork & Bjork, 2011), of which the general idea is that more effortful retrieval practices during initial learning could lead to better long-term retention. Based on this perspective, a delayed recall requires more effort than an immediate recall, which in turn, could result in better retention performance. In the current study, the retrieval practice conducted by students from the stacked condition was more effortful since they did not answer those questions until they completed watching all the video segments. As such they had a significant lower accuracy compared to those from the dispersed condition in answering the embedded questions due to more effortful retrieval. Also, since they had a relatively low initial success, the benefit of more effortful retrieval practice was eliminated in the delayed test, which is based on the idea that harvesting the benefit of retrieval practice requires high initial correctness (Karpicke, 2017).

Another important finding of this study is that providing embedded questions along with feedback during online video learning could provide a greater boost to the learning transfer for high school students. Quizzing with short-answer questions, no
matter interpolated throughout or stacked at the end of the video, improved performance on a set of multiple-choice questions in later exams. This finding replicated prior research showing testing could facilitate transfer of learning on questions with different test formats in authentic educational settings, beyond simply improving rote memorization of practiced information (Thomas et al., 2018). Yet, we did not find any significant superiority of quizzing with feedback specifically on facilitating performance on far-transfer knowledge (i.e., application questions) compared to the restudy strategy. This result is consistent with prior research by McDaniel et al. (2013). In their study, researchers revealed that middle school students who practiced definition questions during the initial learning did not confer benefits to answering the application questions on the final exam. However, Thomas et al. (2018) found the evidence of transfer to application questions when college-level students were quizzed with factual content. One potential explanation is that the age of students and their corresponding stage of cognitive development may be a boundary condition for determining the utility of test-enhanced learning for promoting extensive transfer compared to the restudy method. Another account comes from the theory of the transfer-appropriate processing (Morris et al. 1977), where the final test performance is determined by the overlap in the processes required at the final test and the processes engaged in during initial retrieval practice. According to the transfer-appropriate processing theory, the greater the similarity with the type of processing between initial quizzing and final test, the better learners’ performance ought to be. Indeed, in the current study, the testing effect is significant on the near-transfer questions that were adapted as embedded questions while there was no difference between doing embedded questions and restudying on the far-transfer questions of which
the contexts were not seen during initial learning. Together, this study shows that doing embedded questions with feedback benefit transfer of learning in part. Specifically, for high school students, the benefits of embedded questions with feedback occur regardless of the format of questions (i.e., either short-answer questions or multiple-choice questions) under similar question context, while confer less benefits for questions with more complex and real-life related context on later exams. Future work may explore the extent to which specific measures or activities incorporated with embedded questions to improve students’ transfer to complex contexts in video-based learning environment.
Appendix A

EQ 1: This graph shows projections of sea level rise by the year 2100 based on different models. Can we be absolutely certain how much the sea level will rise based on these projections? Why or why not?


- Feedback: No. Although the trend of sea level rise is increasing based on available data, the value of sea level projections may actually be lower or higher than predicted.

EQ 2: The picture was taken in West Antarctica. Sea level rise is projected to rise between 10 cm and 70 cm by the year 2100. What do you see in the picture that could contribute to the rising sea levels?

( http://giphygifs.s3.amazonaws.com/media/nI3XFL8eMnlUk/giphy.gif )

- Feedback: An increase in global temperature is causing ice on land to melt, which adds to the volume of water in the sea.

EQ 3: Do floating ice chunks in the sea lead to sea level rise? Please describe your answer.

Feedback: No. This is because floating ice chunks are already in the water. The volume of water they displace as ice is the same as the volume of water they add to the ocean when they melt. As a result, sea level does not rise when sea ice melts.

EQ 4: a. What are the two main reasons for sea level rise?
b. How does thermal expansion contribute to sea level rise and what will happen to water molecules in the sea as global temperature increases?

c. Do the size and the number of water molecules change when temperature changes?

• Feedback: Ice on land melting and thermal expansion of sea water. An increased temperature causes water molecules to spread out, which causes them to occupy more space, so the water volume increases. No. It only changes the space between water molecules and leads to more volume in the same mass of water.

EQ 5: a. Why might sea level rise be different in different geographic locations?

b. How does land sinkage affect the local environment in a coastal area?

• Feedback: The land in some areas is sinking, but it is rising in other geographic areas. It depends on the changes in the relative elevation of the land. For example, where land is rising, sea level rise will have less extreme impact. It will cause the land to be more vulnerable to sea level rise and some of lands will be completely covered by sea water. It also will increase the risk of flooding in cities and inhabited islands etc.

EQ 6: What would happen to animals and plants living in the coastal area if sea level were to rise above 4 feet (1.2m)?

• Feedback: Many animals and plants would die off but some would be able to cope or move to other places.

EQ 7: How would people who live in coastal areas be impacted if the sea level rose 4 feet (1.2m)?

• Feedback: Flooding would erode cities, damage cropland and infrastructure, causing the loss of property. Due to flooding, people will have to move and relocate further inland.

EQ 8: Large populations of people live in coastal areas. According to the image, if sea level were to rise dramatically, what would happen during storms to both inland and coastal areas?

• Feedback: Both coastal areas and inland would experience increased flooding and erosion.
References


Davis, D., Chen, G., Van der Zee, T., Hauff, C., & Houben, G. J. (2016, September). Retrieval practice and study planning in MOOCs: Exploring classroom-based self-
regulated learning strategies at scale. In European conference on technology enhanced learning (pp. 57-71). Springer, Cham. https://doi.org/10.1007/978-3-319-45153-4_5


conference on Learning@ scale conference (pp. 41-50).

https://doi.org/10.1145/2556325.2566239


Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded


https://doi.org/10.1080/0020739X.2013.822582


https://doi.org/10.1016/S0959-4752(02)00016-6


https://doi.org/10.1016/j.learninstruc.2014.02.004
https://doi.org/10.1007/s11423-020-09749-6

https://doi.org/10.1080/09541440701326154

https://doi.org/10.1002/acp.2914

https://doi.org/10.1037/xap0000004

https://doi.org/10.14742/ajet.1136

https://doi.org/10.5408/1089-9995-62.3.291
https://doi.org/10.1177/2379298118773418


https://doi.org/10.1007/s11528-019-00374-6


https://doi.org/10.1111/j.1467-9280.2006.01693.x


Chapter 3: Effect of different levels of retrieval-based drawing on high school students’ conceptual understanding of the Earth’s energy budget

Abstract

Previous studies have found that learner-generated drawing and retrieval practice methods could effectively enhance science learning independently. In the current study, we explored how to use sketching in combination with different levels of retrieval practice to help high school students to improve their understanding on one important climate change topic: the Earth energy budget. 55 high school students were assigned randomly into three conditions: a non-support retrieval-based sketching condition, a half-support retrieval-based sketching condition and a restudy condition. Results revealed no significant difference in students’ test performance between the three conditions, and this pattern held even if further analyses were implemented for factual and application questions. However, the trend indicated a slightly higher mean score for the non-support retrieval-based sketching condition on a one-week delayed test. Additionally, students from all conditions rated similarly on the enjoyment and interestingness surveys for the drawing activities, respectively. However, students from the half-support retrieval-based sketching condition rated their drawing task as more difficult compared to the other two conditions. The article offers evidence that retrieval-based drawing could be used for learning science at the high school level and there is a continuing need for more research on how to better combine retrieval practice and other highly effective learning strategies to maximize efficiency of science learning.

*Keywords:* learner-generated drawing, retrieval practice, climate change education, the Earth’s energy budget
Introduction

Science education plays a crucial role in fostering new generations to become aware of global environmental issues and engage actively in addressing global challenges (Kyle, 2020). With wide ranging effects such as global warming, sea level rise, drought and severe weather, climate change has become one of the most serious and challenging problems facing today’s global society (Glenn & Florescu, 2015; IPCC 2013; IPCC 2014). To understand climate change, students need to develop robust models of the Earth’s energy budget and use their models to get deep understandings of climate change, for example, being able to explain the greenhouse effect scientifically (Shepardson et al., 2017).

Yet, at the middle and high school level, many students have misconceptions of the components and function of atmosphere when connecting to climate change. For example, most students believe carbon dioxide is the only greenhouse gas (Shepardson et al., 2011), and ozone depletion is caused by global warming (Andersson & Wallin, 2000; Shepardson et al., 2017). Furthermore, most students have trouble understanding processes of energy transfer in the earth-climate system. For example, they are often confused between solar radiation and terrestrial radiation (Koulaidis & Christidou, 1999; Shepardson et al., 2009, 2011, 2017), and they do not differentiate that light of different wavelengths could be transmitted, reflected, or absorbed differently in the earth-climate system (Shepardson et al., 2017). Although previous studies contributed to designing curricular interventions to support students in building mental models of the greenhouse effect, few studies offer approaches to help learners develop “a richer model about the physics of the greenhouse effect” (Shepardson et al., 2017, p. 96), that is, the main
processes of the Earth’s energy budget, including “the absorption, reflection, and radiation of the sun’s energy and the differentiation between solar radiation and terrestrial radiation” (Shepardson et al., 2017, p. 96).

This study applied retrieval practice with drawing methods to facilitate high school students’ conceptual understanding of the Earth’s energy budget. The retrieval practice, also known as retrieval-based learning, derives from memory studies in educational psychology and has been shown to promote meaningful learning (Roediger & Karpicke, 2006a, 2006b). The essence of retrieval-based learning is to strengthen learners’ memory routes of knowledge by actively conducting a mental search of concepts and reconstructing learners’ knowledge structure (Carpenter et al., 2009; Rowland, 2014). Retrieval practice in this study was combined with drawing methods because drawing directly reflects one’s construction of an internal, nonverbal representation and enables visual model-based reasoning processes (Quillin & Thomas, 2015; van Meter & Garner, 2005). Specifically, we were interested in whether different levels of retrieval based-drawing activities could be applied to help students get deep understandings of the greenhouse effect. Thus, this study proposed three main research questions:

1. How do different levels of retrieval-based drawing influence students’ learning on the Earth’s energy budget; for example, sketching with partial support and sketching with non-support?

2. In what ways do different levels of retrieval-based drawing influence students’ learning outcomes with respect to different levels of conceptual knowledge acquisitions on climate change; for example, factual and application knowledge?
3. How do different levels of retrieval-based drawing influence students’ short-term and long-term knowledge retention?
Theoretical background

Drawing-to-learn

Drawing is a useful strategy that can potentially promote generative learning (Fiorella & Mayer, 2016). The drawing-to-learn method is grounded in learner-generated drawing in which readers construct drawings to represent learning materials to improve learning from content area text (Alesandrini, 1981; Hall et al., 1997; van Meter et al., 2006). From the perspective of the Select-Organize-Integrate (SOI) model of generative learning (Fiorella & Mayer, 2016; Mayer, 2014), there are four main steps to complete drawing construction. In the first step, learners select the key information when they are reading from a text. In the second step, learners organize the selected key information and establish an internal verbal representation of the text information. In the third step, learners construct an internal visual representation of the text information and link it to the internal verbal representation. In the final step, learners draw an external representation which is based on the integration of the internal verbal and the internal visual representations. Understanding the process underlying drawing construction is important because it provides a framework to show that drawing can be seen as an active cognitive strategy, which enables learners to construct their own interpretations and draw inferences on them (Fiorella & Mayer, 2016). In addition, van Meter and Garner (2005) pointed out drawing activities could foster metacognitive processes: “Attempts at constructing the nonverbal representation can send learners back to either the verbal representation or the text as difficulties building the internal image are encountered” (van Meter & Garner, 2005, p. 317). Namely, learner-generated drawing is not a simple linear process, but it can be iterative and involves metacognitive processes such as self-
monitoring and self-regulation during drawing processes (Quillin & Thomas, 2015; Schmeck et al., 2014; van Meter, 2001; van Meter & Garner, 2005; van Meter et al., 2006). Figure 7 provides a visual framework for the generative theory of drawing construction (Quillin & Thomas, 2015).

**Figure 7**

*Visual framework for the generative theory of drawing construction*


Although van Meter and Garner (2005) provided a systematic framework to show how learner-generated drawing encourages learners to be involved actively in generative learning processes, some boundaries and prerequisites need to be considered. First, learner-generated drawing is often considered to be a type of representational drawings (Carney & Levin, 2002; van Meter & Garner, 2005), which means the final external representations are intended to show what depicted objects look like or to depict a physical resemblance with the objects. Second, learner-generated drawing focuses on how verbal representations transform to nonverbal representations (Leutner et al., 2009;
Leopold & Leutner, 2012; Schmeck et al., 2014; van Meter, 2001; van Meter & Garner, 2005; van Meter et al., 2006), and the basic process is that learners receive a text to read and draw pictures to reflect the content of the text. Based on the two main characteristics of the learner-generated drawing strategy, previous studies found mixed results on the effectiveness of drawing in which some studies pointed out positive effects of drawing on text comprehension (e.g., Leopold & Leutner, 2012; Schmeck et al., 2014; van Meter, 2001; van Meter et al., 2006), whereas others proposed contrary results (e.g., Hall et al., 1997; Leutner et al., 2009).

There are some important findings when applying learner-generated drawing as a learning 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 reflect 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 are more effective than photographs especially when learning materials are complex with rich contents (Quillin & Thomas, 2015; Rennie & Jarvis, 1995). Another important finding is that learner-generated drawing sometimes leads to lower learning performance. 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 students 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).
On the other hand, the diversity of learning resources has been a trend in today’s educational settings where students can get access to learning materials presented in all sorts of format, such as audio, visual and audiovisual, and therefore learning materials in a text format could not be a limitation when talking about learner-generated drawing strategies. Drawing can 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). In the current study, drawing methods are identified as learning activities that engage learners to create any type of external visual representations by themselves from any type of learning materials.

Recently, there is an emerging line of research suggesting that drawing can serve as a tool to improve learning in science education. Specifically, drawing can foster observation skills (Baldwin & Crawford, 2010; Quillin & Thomas, 2015), facilitate thinking and communicating skills (Ainsworth et al., 2011; Roam, 2008), and deepen understanding of science processes (Ainsworth et al., 2011; Lehrer & Schauble, 2006). Moreover, drawing could be a good way to enhance model-based reasoning when learning science (Quillin & Thomas, 2015; Schwarz et al., 2009). Model-based reasoning is a sophisticated process that enables learners to solve problems depending on working memory to construct a mental model – an analogue representation with highly personal, dynamic and format-diversified characteristics (Harrison & Treagust, 2000; Nersessian, 2002). A scientific model is commonly “an abstract, simplified, representation of a system of phenomena that makes its central features explicit and visible and can be used to generate explanations and predictions” (Schwarz et al., 2009, p. 633). 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-generative visual models (Greca & Moreira, 2000; Shepardson et al., 2017).

However, in most previous studies, especially in environmental science education, drawing is mainly applied to help develop or evaluate students’ model-based reasoning process on a complex environmental topic (e.g., Assaraf & Orion, 2005; Shepardson et al., 2011, 2017; Zangori et al., 2017), rather than being designed as a specific tool to facilitate their learning outcomes or knowledge acquisition. Therefore, although drawing shows promise as a learning tool for students, a general challenge is to develop interventions that could motivate students to improve learning with drawing methods (Heideman et al., 2017; Quillin & Thomas, 2015).

**Retrieval practice and retrieval-based drawing**

Recent work in educational psychology has pointed out that retrieval practice could facilitate learning in educational settings (Karpicke & Blunt, 2011; Roediger & Karpicke, 2006a, 2006b; Karpicke & Roediger, 2010). This approach involves having learners recall contents they just learned without access to learning materials. By practicing retrieval repeatedly, learners facilitate encoding of information from short term memory into long term memory (Karpicke, 2015). Importantly, retrieval-based learning is shown to be efficient in enhancing both knowledge and knowledge transfer (Butler, 2010; Carpenter, 2012; Jensen et al., 2014; Smith et al., 2016).
The benefits of retrieval-based learning are grounded in basic memory research (e.g., Allen et al., 1969), which often involved memory of word lists or paired associates. In recent years, researchers have investigated the effectiveness of retrieval practice with more educational relevant contents, such as science-related topics (e.g., Johnson & Mayer 2009; Karpicke & Blunt, 2011); more practical learning outcomes, such as transfer (e.g., Butler, 2010; Johnson & Mayer, 2009); and more diverse student populations, such as K-12 students (e.g., Agarwal et al., 2014). One important finding is that retrieval practice is also effective when combined with drawing activities, such as concept mapping and sketching (Heideman et al., 2017; Blunt and Karpicke, 2014; Karpicke & Blunt, 2011). For example, Blunt and Karpicke (2014) showed concept mapping could work as an effective tool when it was incorporated into a retrieval-based learning activity. In their study, students created concept maps about the material they just learned without viewing the material. The results showed that the retrieval-based concept mapping activity facilitated long-term retention to the same extent as recalling materials by paragraphing (Blunt and Karpicke, 2014; Karpicke & Blunt, 2011; Karpicke & Roediger, 2010). In addition, Heidman et al., (2017) developed a drawing practice called Minute Sketches in Folded Lists (MSFL) tool for college students to self-assess their recall and problem-solving on biology concepts. The MSFL involving retrieval practice led to better knowledge retention and performance compared to the common preferred visual review (VR) method in eight months after initial learning (Heidman et al., 2017). This demonstrated incorporating drawing into retrieved-based learning could be useful for students to improve both lower-order (e.g., recall facts and basic concepts) and higher-
order (e.g., solve problems in new situations) knowledge and skills (Heideman et al., 2017).

Although prior studies have developed a few types of drawing strategies combining drawing with retrieved practice and found evidence supporting their effectiveness for college students to learn science content, it is necessary to explore if these findings can be extended from college to K-12 classroom settings. One study investigated using adapted retrieval-based concept mapping techniques with children between ages 9 and 11 at the elementary school level (Karpicke et al., 2014). Researchers found it was not effective for young children when working with retrieval-based concept mapping tasks compared to college students, so they modified the retrieval-based learning activities by providing some supports. There were two levels of support in the concept mapping activity: 1) more support, where much of the map was filled in and 4 concepts were required to be completed by children; and 2) less support, where less of the map was completed and children were required to complete 13-14 concepts. The results indicated adapted retrieval-based concept mapping with support might be more effective than complete retrieval-based concept mapping without any cues, but elementary students with less support was more successful than students who received more support (Karpicke et al., 2014).

Although Karpicke et al. (2014) point out elementary students would require more support than college students when working on retrieval-based drawing activities to get better learning results, little is known about the impact of retrieved-based drawing with high school students in the context of environmental science. That is, it is an open issue
whether retrieval-based drawing methods can be adopted by high school students to learn environmental science concepts and what kind of scaffolds are effective for them.

**Method**

To address the research questions, we first conceptualized drawing in this study as an annotated sketching activity. Instead of using concept mapping, we chose annotated sketching as the basic drawing method because sketching could depict the processes of the Earth’s energy budget vividly by creating an external visuo-spatial representation of a to-be-learned content (Johnson & Reynolds, 2005; Scheiter et al., 2017). By incorporating annotations during drawing the processes of the Earth’s energy budget, students could identify features and characterize the relationship between components of the Earth’s energy budget. Furthermore, to reduce the influence of limitations on drawing abilities and minimize cognitive load of students when they participate in drawing activities (Fiorella & Mayer, 2016; Quillin & Thomas, 2015), symbols representing different components of the earth’s system were provided for students to select and use. As such, this annotated sketching method enabled students to present their intuitive model of what they learned or observed in a concrete way.

To explore how to better use retrieval-based drawing to learn the Earth’s energy budget, three types of retrieval support were provided: 1) non-support, where students drew the whole model of Earth’s energy budget without any hints, 2) half (50%) support, where half of the drawing was provided and students completed the rest by sketching; and 3) complete (100%) support, where students just copied a pre-completed diagram of the Earth’s energy budget. Finally, we included an immediate test and an identical one-week delayed test after initial learning activities to see whether scaffolded retrieval-based
drawing could promote better knowledge retention since both drawing and retrieval practice were demonstrated to foster memorization of work pairs in vocabulary studies (Wammes et al., 2015, 2018; van Meter et al., 2006).

**Study design and participants**

A between-subject design was adopted with the learning condition manipulated as a between-subject factor and the retention interval as a within-subject factor. Three conditions corresponded to the three types of retrieval-based drawing activities, including a “non-support retrieval-based sketching” condition, a “half-support retrieval-based sketching” condition and a restudy condition as control (i.e., students just copy the pre-completed diagram of the Earth’s energy budget and review what they learned on the topic). 55 students in grade 10-12, from a Midwest high school in the U.S participated in the study. They were assigned randomly into the three learning conditions. The study took place across two weeks: all students completed the learning session with drawing activities and then completed an immediate test. After one week, they completed the delayed test. We calculated students’ scores on both tests.
Learning and test materials

The learning material used in this study was a 7min video lecture introducing the Earth’s energy budget, which was developed by the South Central Climate Science Center\(^2\). The video lecture was selected by the author and reviewed by a science teacher who instructs the environmental science class in the high school to confirm the content of the video was appropriate for the students. In addition, three different drawing tasks were developed by the author in accordance with the three learning conditions. Students from different conditions completed the respective tasks after learning from the video. The content of the drawing tasks was related to the model of the Earth's energy budget.

The test material was designed to assess students’ conceptual understanding of the Earth’s energy budget after completing the drawing tasks. There were two types of questions in the test: 1) factual questions; and 2) application questions. The factual questions assessed students on their understanding of basic knowledge such as concepts and processes of the Earth’s energy budget. The application questions assessed students’ capability of applying what they learned to solve novel problems. An example of a factual knowledge question is: “What is the primary source of energy for the earth?” All the factual questions were developed directly from the script of the video lecture. An example of an application question is: “Can you explain why the day and night temperature of the moon's surface changes so much?” The application questions were adapted from the General High School Curriculum Standard Test Textbook-The First Compulsory Geography (Institute of Curriculum and Teaching Materials of People’s

\(^2\) https://www.youtube.com/watch?v=el7ygWztQqA
Education Press, 2004). These questions were tested by a pilot study and verified to be proper for high school students (details can be found in the Appendix B).

**Procedure**

There were three phases in the initial learning session of the study. In the first phase, all the students watched the video for 7 min. In the second phase, all students drew a model of the Earth’s energy budget on a sheet of paper depending on the condition to which they had been assigned. Specifically, in the non-support retrieval-based sketching condition, students were required to sketch and annotate the whole model of the Earth’s energy budget processes. In the half-support retrieval-based sketching condition, half of the Earth’s energy budget processes in the task were completed, students were only required to draw and annotate the remaining half to complete the whole model. For the restudy condition, students only needed to copy the pre-completed model diagram of the Earth’s energy budget as a practice. All the drawing tasks for the three conditions were provided with a corresponding instruction (details can be found in the Appendix B), and all the students took 10 min to complete their own drawing task. In the third phase, students in the first two conditions were provided feedback which is identical to the completed Earth’s energy budget model diagram. In this phase, they took 5 min to correct their drawings with access to feedback. After that, they completed a free recall task to summarize what they learned for 2 min. Whereas students in the restudy condition just re-watched the video lecture for another 7 min as a review.

After the learning session, all the students took around 15 min to complete the immediate test. Before the test, they were asked to rate on several aspects of their drawing task, including how they enjoyed the “annotated sketches” task, how interesting
and difficult they found the task. The students made their ratings on a 0-100 scale where 0 was the lowest rating and 100 was the highest rating. One week later, they were asked to complete the delayed test including the same questions as they answered in the immediate test. Apart from these questions, an additional drawing task was developed to investigate how much knowledge students had retained on the Earth’s energy budget processes. Therefore, the one-week delayed test lasted 20 min. The time of the learning session and the test session was equal among the three conditions (See Figure 8).

**Figure 8**

*Scheme of the design*

![Scheme of the design](image)

*Note.* Timeline for condition 1, 2 and 3: Non-support retrieval-based sketching condition vs. Half-support retrieval-based sketching condition vs. Restudy condition.

**Coding of tests**

Rubrics were developed for the test and the drawing task separately. For the test, a correct response to each question was worth one point, and a partially correct answer to a question was rewarded 0.5 points, with a maximum of 16 points. Students did not receive any credit for any other situations. For the drawing task, the Earth’s energy budget model was divided into 9 idea units representing the main processes of the model. A correct drawing for each main process was rewarded one point. Zero point was given to any
incorrect drawing. 20 students’ responses to both the test and the drawing task were independently graded by two raters. The Pearson correlations (r) between the two sets of scores were .98 and .95 for the test and the drawing task, respectively. Given the high levels of interrater reliability, the remaining responses to the test and the drawing task were scored by one of the raters.

**Results**

50 students completed both the learning and test sessions, with 16 students in the “non-support retrieval-based sketching” condition, 16 students in the “half-support retrieval-based sketching” condition and the 18 students in the restudy condition.

**Performance on the immediate and one-week delayed tests**

Table 2 shows the mean proportion correctness and standard deviation of each condition on the immediate and the one-week delayed test. First, to evaluate the main effects of learning condition, a 3 (learning condition: non-support retrieval-based sketching vs. half-support retrieval-based sketching vs. restudy condition) × 2 (retention interval: immediate test vs. one-week delayed test) mixed ANOVA was carried out. Overall, the results showed that there was a significant main effect of retention interval, \( F(1, 47) = 14.519, p < .001, \eta^2_p = .236 \), indicating a significant knowledge forgetting from the immediate to the delayed test. We did not find either main effects of learning condition, \( F(2, 47) = 1.086, p = .346, \eta^2_p = .044 \), or significant interaction between the retention interval and the learning condition, \( F(2, 47) = .883, p = .42, \eta^2_p = .036 \) (see Figure 9).
Table 2

Mean Score of Students for Different Questions from the Three Conditions on Both Immediate and Final Tests.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Overall questions</th>
<th>Factual questions</th>
<th>Application questions</th>
<th>Final drawing task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Immediate test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition 1 (N=16)</td>
<td>0.58</td>
<td>0.23</td>
<td>0.60</td>
<td>0.24</td>
</tr>
<tr>
<td>Condition 2 (N=16)</td>
<td>0.48</td>
<td>0.21</td>
<td>0.47</td>
<td>0.20</td>
</tr>
<tr>
<td>Condition 3 (N=18)</td>
<td>0.58</td>
<td>0.19</td>
<td>0.62</td>
<td>0.21</td>
</tr>
<tr>
<td>One-week delayed test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition 1 (N=16)</td>
<td>0.51</td>
<td>0.25</td>
<td>0.52</td>
<td>0.30</td>
</tr>
<tr>
<td>Condition 2 (N=16)</td>
<td>0.41</td>
<td>0.21</td>
<td>0.39</td>
<td>0.21</td>
</tr>
<tr>
<td>Condition 3 (N=18)</td>
<td>0.45</td>
<td>0.23</td>
<td>0.44</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Note: Condition 1 is the non-support retrieval-based sketching condition, condition 2 is the half-support retrieval-based sketching condition and condition 3 is the restudy condition.

Figure 9

Students’ performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means.

Condition 1 represents the non-support retrieval-based sketching condition, condition 2
represents the half-support retrieval-based sketching condition and condition 3 represents the restudy condition

We next analyzed students’ performance on the drawing task in the one-week delayed test. Again, a one-way ANOVA revealed there was not any significant difference between all three learning conditions, $F(2, 47) = 2.083, p = .136, \eta^2_p = .083$ (See Figure 10). However, the trend indicates a slightly higher mean for the non-support condition, although not significantly.

**Figure 10**

Students’ drawing task’s performance (proportion correct) on the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means.
To further examine whether retrieval-based sketching has any effects on facilitating the answering of different type of questions, we added the type of question as the third factor to further conduct a 3 (learning condition: non-support retrieval-based sketching vs. half-support retrieval-based sketching vs. restudy condition) × 2 (retention interval: immediate test vs. one-week delayed test) × 2 (question type: factual questions vs. application questions) mixed ANOVA, with learning condition as a between-subject factor, and retention interval and question type as within-subject factors. Again, results revealed that there was a significant main effect of retention interval, $F(1, 47) = 9.204, p = .004, \eta^2_p = .164$, but no difference was found between all three learning conditions, $F(2, 47) = .721, p = .492, \eta^2_p = .03$. The main effect of question type was not significant, $F(1, 47) = .105, p = .747, \eta^2_p = .002$.

In addition, these effects were not qualified by a significant Learning Condition × Retention Interval interaction, or Question Type × Retention Interval interaction, $F(1, 47) = .410, p = .666, \eta^2_p = .017$ and $F(2, 47) = .653, p = .525, \eta^2_p = .027$, respectively (See Figure 11 and 12). However, we found a marginally significant interaction of retention interval and question type, $F(1, 47) = 3.406, p = .071, \eta^2_p = .068$. Planned comparisons confirmed that students scored higher on factual questions for the immediate test compared to the delayed test ($M = .57$ vs. $M = .45$), $t(49) = 4.395, p < .001, d = .49$, 95% CI [0.063, 0.168], but no significant difference was found between the two tests on the application questions ($M = .52$ vs. $M = .48$), $t(49) = 1.103, p = .138, d = .14$, 95% CI [-0.033, 0.113] (See Figure 13). This interaction between retention interval and question type indicated that the students kept better retention on the application questions compared to the factual questions. Finally, no significant interaction between learning
condition, question type and retention interval was found, $F(2, 47) = .730, p = .487$, 
$\eta_p^2 = .030$, such that the retrieval practice did not produce an apparent effect on students’ 
overall learning outcomes on the Earth energy budget in this study.
Figure 11

*Students’ factual knowledge performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means.*

![Figure 11](image1)

Figure 12

*Students’ application knowledge performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means.*

![Figure 12](image2)
Figure 13

Students’ performance (proportion correct) on the immediate and the one-week delayed tests as a function of the question type. Error bars represent standard errors of the means.

Subjective rating performance

Table 3 shows students’ subjective ratings (means and standard deviations) of their experiences during the learning activity. A one-way ANOVA found that there was no significant difference in either students’ ratings of enjoyment between the three conditions, $F(2, 47) = .189, p = .828, \eta_p^2 = .008$, or ratings of the interestingness of the tasks, $F(2, 47) = 1.561, p = .221, \eta_p^2 = .062$. However, there was a significant difference in the ratings of difficulty between the three learning conditions, $F(2, 47) = 6.332, p = .004, \eta_p^2 = .212$. Post hoc analysis with a Bonferroni correction revealed that students in the half-support retrieval-based sketching condition ($M = 46.88$) reported a higher level of difficulty on the drawing task than did the restudy condition ($M = 17.22$), $p = .003, d = 1.17, 95\% \text{ CI} [8.58, 50.72]$. Students in the non-support retrieval-based sketching condition ($M = 36.25$) also reported a higher level of difficulty on the drawing
task than did the restudy condition \((M = 17.22)\), but this difference was close to marginally significance, \(p = .09, d = .96, 95\% \text{ CI } [-2.04, 40.1]\). Finally, the difference between the non-scaffolded retrieval-based drawing condition and the 50% scaffolded retrieval-based drawing condition did not reach significance, \(p = .69\).

**Table 3**

*Students’ Ratings of Enjoyment, Interestingness and Difficulty of the learning activity*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Enjoyment</th>
<th>Interestingness</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M)</td>
<td>(SD)</td>
<td>(M)</td>
</tr>
<tr>
<td>Condition 1(N=16)</td>
<td>60.00</td>
<td>23.38</td>
<td>61.88</td>
</tr>
<tr>
<td>Condition 2(N=16)</td>
<td>55.00</td>
<td>28.52</td>
<td>54.38</td>
</tr>
<tr>
<td>Condition 3(N=18)</td>
<td>60.00</td>
<td>28.08</td>
<td>47.78</td>
</tr>
</tbody>
</table>

**Discussion**

The main purpose of the current study was to examine the effectiveness of retrieval-based drawing with high school students when learning the Earth energy budget. Specifically, we investigated the effects of prompting high school students to either drawing a model of the Earth energy budget by recalling without any support or with some support, or just copying a pre-completed model diagram of the Earth energy budget followed by restudying the video lecture. As only watching a video is a passive learning activity, these drawing-to-learn activities were expected to increase students’ knowledge gaining and retention since it involves two active learning strategies: drawing and retrieval practice. Based on the retrieval practice literature, we further hypothesized that asking students to sketch a diagram of the learning topic by recalling with some or without any support would benefit learning more than merely copying an answer and restudying the learning materials. Interestingly, we did not find any apparent benefits of
retrieval practice over the restudy condition since no significant difference on the performance was found between treatments and the control condition. Furthermore, the same pattern holds when the analysis was conducted for the factual and application questions separately.

Why does the current study fail to provide evidence for the effect of retrieval practice? One possible explanation comes from the complex nature of learning materials used in this study. van Gog and Sweller (2015) proposed a boundary situation for the effect of retrieval-based learning, suggesting that “testing effect” could be reduced or even eliminated if learning materials are complex and high in element interactivity. Based on the theory of cognitive load, van Gog and Sweller suggested that when learning materials are complex, the content contains various information elements that are related and therefore impose a heavy working memory load to learners because these elements have to be considered simultaneously (Sweller et al., 2011; van Gog & Sweller, 2015). From this perspective, the learning material of the current study can be defined as highly complex since the learning material is an educational video introducing the processes of the Earth’s energy budget, the content of which is about different atmospheric components and interactions between these components in the earth-climate system. Therefore, when students practiced retrieval after initial learning, the effectiveness of retrieval-based learning was reduced or eliminated relative to restudy strategy because the advantages of retrieval (i.e., affording organizational/relational processing for long-term retention) were no longer present.

We also do not rule out the possibility that the effectiveness of restudy was increased. In this situation, restudy strategy might avoid overloading the working
memory since students had access to the solutions directly, which provided an opportunity to deeply process relations between elements and maintain students’ motivation for engaging in restudy (van Gog & Sweller, 2015). Complexity of learning materials could offer one such explanation. Nevertheless, other researchers provided counterarguments supporting the effectiveness of retrieval-based learning over complicated, educationally relevant materials (Karpicke & Aue, 2015).

A second account of the null results might be due to a small sample size in this study. Further results of data analysis showed a trend of lower forgetting rate (i.e. (scores of the immediate test – scores of the delayed test) / scores of the immediate test) for both the non-support retrieval-based sketching and the half-support retrieval-based sketching condition in relative to the restudy condition (11.13% vs. 21.97%, 9.71% vs. 21.97%, respectively). We might have seen the difference to reach significance between both the retrieval-based drawing conditions and the restudy condition if a larger sample size had been used. This begs for a future study with more participants.

Another interesting finding of this study is that applying drawing activities after educational video learning seems to benefit more on knowledge retention of the application questions than the factual questions. This result is consistent with prior findings that drawing was more beneficial on higher-level assessments (van Meter & Garner, 2001, 2006). The current study find evidence supporting drawing is still effective even when it is used in combination with other learning strategies such as retrieval practice. Furthermore, prior literature indicated that drawing could foster learners’ memorization of knowledge (Quillin & Thomas, 2015; Wammes et al., 2015, 2018), but less revealed is the level of knowledge that would be enhanced the most in a long run.
This study provides some initial evidence demonstrating drawing as a learning strategy could foster high school students’ retention of higher-order but not of lower-order knowledge.

The last interesting finding of this study revealed that high school students rated differently on the difficulty of the drawing activities. Specifically, students in the restudy condition rated the task they engaged in was the least difficult, presumably because they just copied the completed model solutions. Surprisingly, students from the half-support retrieval-based sketching condition felt their drawing activity was more difficult than those in the non-support retrieval-based sketching condition. This may seem counterintuitive because more supports should help lower the difficulty of learning activities. This outcome also contradicts the results of a study conducted by Karpicke et al (2014), where they found elementary students rated groups with less support during a retrieval concept map activity as more difficult than groups with more support.

One possible explanation for this unexpected result is it could be caused by a various levels of cognitive load perceived by students during the drawing session. A Likert scale question asking for the level of difficulty has been used for measuring cognitive load since Fred Paas first introduced it in the literature (Paas, 1992). From the cognitive load perspective, students from the non-support condition had full control of freely recalling the information without any constraints, which does not require them to process different pieces of information simultaneously. Students from the restudy condition had access to all the given answers and were asked to restudy the materials without necessarily imposing too much effort. Thus, students from both non-support and restudy conditions would not report high cognitive load as measured by the difficulty
question. On the contrary, students from half-support condition were asked to complete the blanks with partially filled solutions. Students might have felt the need to provide solutions similar to the given, as such, they need to process higher number of pieces of information at the same time, which results in higher reported cognitive load. On the other hand, later studies have found that the difficulty level questions can be used to measure one of the three sub-types of cognitive load: germane load which is an indicator of the amount of energy devoted to processing relevant information in working memory (Leppink et al, 2013; Zu et al., 2021). From this perspective, students from the half-support condition might have devoted more energy into constructing their answers compared to those in condition 1 and 3. This begs for future studies with more solid measures of cognitive load.

Overall, the current study provided evidence that retrieval-based drawing could be used for learning science at the high school level and there is a continuing need for more examinations of active learning strategies like retrieval practice and drawing so that students and teachers can draw evidence-based conclusions about their effectiveness.
Appendix B

Task (a):

After watching the video lecture, you are going to complete an activity named “Annotated Sketches” on Earth’s energy budget.

- Copy the diagram below on your paper sheet.
- In the box below "Different-wave radiation", label the type of radiation each line represents (e.g. Solid line can represent ____ wave radiation, and dashed line can represent ____ wave radiation).
- Use the lines you labeled to draw the main processes of the Earth’s energy budget as described in the video.
- Then, add in short descriptions of what is happening for each of the lines that you drew in the diagram (label each process of the Earth’s energy budget).

You’ll have 10 min to complete it then you will be advanced automatically to the next page.

(Note: O3 is ozone; CO2 is carbon dioxide; CH4 is methane; H2O is water vapor; The Earth’s surface includes different landscapes such as deserts and rainforests.)
Task (b):

After watching the video lecture, you are going to complete an activity named "Annotated Sketches" on Earth's energy budget.

- Copy the diagram and descriptions below on your paper sheet.
- Some of the processes of Earth's energy budget are already drawn in the diagram. Notice that the lines represent different wavelengths of radiation. Complete the diagram by adding in the rest of the processes of Earth's energy budget as described in video.
- Then, add in descriptions of what is happening for each of the lines that you added.

You'll have 10 min to complete it then you will be advanced automatically to the next page.

(Note: O3 is ozone; CO2 is carbon dioxide; CH4 is methane; H2O is water vapor; the earth's surface includes different landscapes such as deserts and rainforests.)
Task (c):

The diagram below shows the main processes of Earth’s energy budget, please read it carefully and then complete your “Annotated Sketches” on the Earth’s energy budget.

- Copy the diagram and descriptions below on your paper sheet.

You'll have 10 min to complete it, then you will be advanced automatically to the next page.

(Note: O3 is ozone; CO2 is carbon dioxide; CH4 is methane; H2O is water vapor; the earth’s surface includes different landscapes such as deserts and rainforests.)
Assessment of the Earth energy budget

Q1: What is the primary source of energy for the earth?

- The Sun (1pt).

Q2: The Sun emits _________-wave radiation and the Earth emits _________-wave radiation. (Fill in the wavelength of radiation.)

- Short (1pt); Long (1pt).

Q3: What is the effect of the atmosphere and clouds on different wavelengths of radiation and energy? (Discuss your answers for both short-wave radiation and long-wave radiation separately. There are several different effects, please answer as many as you can).

- 1) Primarily transmit incoming short-wave solar radiation (1pt).
- 2) Reflect some incoming short-wave solar radiation (1pt).
- 3) (High clouds) absorb some of the outgoing long-wave radiation emitted by the Earth and trap the heat (1pt).
- 4) Emit some long-wave radiation and energy to the outer space (1pt).

Q4: What is the effect of greenhouse gases on different wavelengths of radiation and energy? (Describe how short-wave radiation and long-wave radiation interact with atmosphere and clouds separately.)

- 1) Primarily transmit incoming short-wave solar radiation (1pt).
- 2) Absorb the amount of long-wave radiation and trap the heat, re-emit it back to the earth’s surface (1pt).

Q5. Which of the following is not a common greenhouse gas? (1pt). And what is the function of the one you chose? (1pt).

A. Methane (CH₄)
B. Carbon Dioxide (CO₂)
C. Ozone (O₃)
D. Water vapor (H₂O)

- The main function of the ozone layer is to absorb the Sun's ultraviolet radiation, hence protecting the Earth from its harmful effects.

Q6. The diagram below indicates a simplified model of the earth’s energy budget processes.
6-1: If “a-i” represents different processes, which of the following is the correct order for how the atmosphere is warmed? (1pt).

A. a — d — e  
B. a — c — e  
C. a — b — g  
D. a — d — g

6-2: Which process of the following is the correct statement: (1pt)

A. The process “d” is influenced by the current weather conditions  
B. Process “a” causes the temperature of the atmosphere to rise the most  
C. The clouds are warmed mainly by the process “c”  
D. The process “h” is the result of the greenhouse effect

6-3: If “a-i” represents different radiation and energy, which of the following is the correct statement below: (1pt)

A. The energy of “b” is totally absorbed by the earth’s surface  
B. The energy of “e” is absorbed by ozone  
C. The radiation wavelength of “a” is shorter than that of “h”  
D. “f” contains the equal amount of energy with “i”

Q7: The pictures below indicate the typical weather conditions in forest and desert of the same latitude (which means these places receive the same amount of incoming solar radiation).

Which of the situations has the smallest change in temperature from day to night? (1pt)
Q8. Temperatures on the moon are very hot in the daytime, and can reach 260 °F. At night, the lunar surface gets very cold, down to minus 280 °F. Can you explain why the day and night temperature of the moon's surface changes so much? (1pt)

(The diagram task for the one-week delayed test)

According to the principle of Earth’s energy budget, can you draw a model and explain why large amount of greenhouse gas emissions affect global warming? (9pt)
References


https://doi.org/10.1037/0022-0663.89.4.677

https://doi.org/10.1080/095006900416884

https://doi.org/10.1187/cbe.16-03-0116
https://doi.org/10.1080/095006900289976


https://doi.org/10.1007/s10648-013-9248-9


Kyle, W. C. (2020). Expanding our views of science education to address sustainable
development, empowerment, and social transformation. *Disciplinary and
Interdisciplinary Science Education Research, 2*(1), 1-9.


selection, and summarizing as learning strategies. *Learning and
Instruction, 22*(1), 16-26. https://doi.org/10.1016/j.learninstruc.2011.05.005

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J.
(2013). Development of an instrument for measuring different types of cognitive
https://doi.org/10.3758/s13428-013-0334-1

comprehension: Effects of drawing and mentally imagining text

learner-generated drawing and imagination in comprehending a science text. *The


Chapter 4: Effect of retrieval practice and drawing on high school students’ conceptual understanding of the carbon cycle

Abstract

Both learner-generated drawing and retrieval practice methods are effective to enhance science learning. To compare the impact of combing different drawing methods (representational drawing vs. abstract drawing) with retrieval practice on the carbon cycle learning, 136 Chinese high school students enrolled in geography classes were assigned randomly to six learning conditions: students built their mental model of the carbon cycle by either generating sketches with or without access to the learning material (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 learnt 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 an immediate and a one-week delayed test. 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 environmental science learning at the high school level.
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 the components and their relationship of the carbon cycle 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 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 the carbon cycle 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; Rosengrant et al., 2009; Schwarz et al., 2009; NRC, 2012). 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 (van Meter & Garner, 2005; Schwarz et al., 2009; Ainsworth et al., 2011; Quillin and Thomas, 2015; Wammes et al., 2016). Studies have revealed that college students who were asked to draw pictures while reading a scientific text had better performance than students who did self-written explanations (Schmeck et al., 2014; Scheiter et al., 2017). However, the benefits of drawing by learners themselves were only found on higher-order versus lower-order knowledge elements (Van Meter, 2001; Van Meter & Garner, 2005). Thus, how to harness the benefit of drawing to solve the second challenge is an open question. We report herein how to use retrieval practice to support drawing to enhance students’ understanding of the carbon cycle.

Retrieval practice can be described as a learning strategy of recalling just learned knowledge without viewing the 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, 2006b). 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 the combination of retrieval practice and drawing methods could produce better learning outcomes than adopting either one alone at the high school level.
Prior studies described drawing on a spectrum from representational which is more concrete to abstract (Quillin & Thomas, 2014). According to Quillin and Thomas (2014), 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, 2014). In the current study, we focused on two main drawing methods in science class, a representational drawing method: sketching; and an abstract drawing method: concept mapping. In particular, sketching is conceptualized as a technique that uses the 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 methods 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 methods (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?
Theoretical background

The carbon cycle and model-based learning

The carbon cycle is a major topic of climate change and sustainability, which describes the movement of carbon through atmosphere, biosphere, geosphere, and hydrosphere (You et al., 2018). One of the major carriers of carbon in the cycle is carbon dioxide. Carbon dioxide as a main greenhouse gas affects the movement and transformation of solar energy through the Earth system, leading to the change of global climate (IPCC, 2014).

In science education, the carbon cycle has become a key central construct of the climate change topic and has been emphasized in the K-12 science education curriculum (NGSS, 2013). In particular, at the high school level, students are required to develop a robust model to describe the carbon cycle across the Earth system (HS-ESS2-6). Models, especially scientific models, are often referred to as simplified abstraction that represents a system of phenomena and can be both external and internal representations (Gilbert, 2005; Schwarz et al., 2009; Sorensen et al., 2016). Internal representations of scientific models are often outlined as mental models of scientific concepts and ideas, which are a kind of dynamic and complex memory structures that can be used to store spatial, physical, and conceptual features of external experiences. Mental models are useful for retrieval in the service of problem solving, inference generation and decision making (Rapp, 2005). While external representations are based on accepted knowledge shared by a community of practitioners (Greca & Moreira, 2000), and can be any types of "seen" analogous and physical visualizations of scientific phenomena or processes (e.g. mathematical equations, dynamic simulations, molecular and anatomical models).
respect to drawing, external representations are often represented as static visualizations such as sketches, maps, flowcharts, graphs and diagrams.

In the current environmental science classrooms, students are encouraged to make their thoughts and reasoning visible when learning on climate change topics. Effective learning occurs when students continually use, evaluate, and revise the explanatory of their own models on climate change (Zangori & Forbes, 2016; Zangori et al., 2017). One study focused on the development of middle school students’ mental models of the greenhouse effect (Shepardson et al., 2017). Students in that study were required to do a draw-and-explain task to represent their understandings of the greenhouse effect. Specifically, the task was designed to assess students’ mental model of the greenhouse effect by three steps: drawing, labeling and writing a paragraph to explain their drawing. This task was used in students’ pretest and posttest to help researchers clarify any changes in students’ mental models. Therefore, drawing is an important “media” that connects with both external and internal representations in learning on climate change.

However, although prior studies highlighted the significance of developing model-based learning of the carbon cycle and climate change (e.g., Jacobson et al., 2016; Mohan et al., 2008, 2009), only a handful of studies have specifically investigated student-generated drawing of carbon cycle models. Zangori et al. (2017) developed a model-based curriculum focusing on socio-scientific 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 the carbon cycle. The results of the study indicated that it is important for students to hold conceptual understanding of causal mechanisms for
transformation and transfer of carbon, and once students have developed robust models of carbon cycling, their reasoning could shift in complexity to understand the relationship between the carbon cycle and climate change (Zangori et al., 2017). Moreover, Düsing et al. (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 drawing and writing explanations of the carbon cycle, researchers found that students had difficulties identifying carbon compounds and the processes where carbon compounds are transformed (Düsing et al., 2019). Prior studies revealed how student-created models of the carbon cycle can be an effective way for teachers to assess their learning processes and misconceptions, however, few studies investigated how drawing itself enhances students’ learning outcomes on the carbon cycle.

**Learner-generated drawing**

It is widely known that interpretations of others’ visual information is critical to science learning (e.g. interpreting charts and graphs), however, becoming proficient in science also requires students to develop their own representational skills (Ainsworth et al., 2011). *Learner-generated drawing* can be an effective way to facilitate learning processes in science learning, especially when students learn from expository text materials (e.g. Roelle & Nückles, 2019; Scheiter et al., 2017; Schmeck et al., 2014; Schwamborn et al., 2010, 2011; van Meter, 2001; van Meter et al., 2006). According to van Meter and Garner’s (2005) learner-generated drawing, which is based on Mayer’s (2005) framework of multimedia learning, students benefit from drawing since this strategy encourages them to engage in generative learning processes. The core idea of the learner-generated drawing is that drawing itself fosters deep cognitive processing on
three interactive components: (a) selecting key information from text materials, (b) organizing the selected information to build up internal verbal representations of the text materials, (c) constructing internal visual representations of the text materials by connecting with the internal verbal representations and relevant existing knowledge. When drawing external representations during learning, learners rely mainly on the integration of internal verbal and visual representations (van Meter & Garner, 2005). Previous studies revealed that learner-generated drawing is beneficial to learners’ higher-order knowledge such as comprehension and transfer, and problem solving (Leopold & Leutner, 2012; Schmeck et al., 2014; Schwamborn et al., 2010, 2011; van Meter, 2001; van Meter et al., 2006). In an exemplary study, researchers used a posttest including lower-level recognition questions and higher-level free recall questions on system thinking to assess fifth and sixth grade students’ learning outcomes of reading a science text (van Meter, 2001). The results showed that students who constructed drawing with support gained better scores on higher-level questions than that of students in a non-drawing group, but there was no positive effect of drawing compared to the non-drawing group on recognition questions. Additionally, Schwamborn et al. (2010) asked 9th-grade students to study a scientific text explaining the chemical process of doing laundry with soap and water by either generating drawings during learning or only reading the text. Results indicated that students from the drawing group outperformed students from the reading group on a subsequent posttest consisting of retention, drawing and transfer questions. Moreover, this effect was stronger for students who created high-accuracy drawing relative to students who created low-accuracy drawing.
Although previous studies showed benefits of learner-generated drawing activities, many of these studies implemented only tests that followed immediately after an initial learning phase, evidence concerning beneficial effects of drawing tasks in the long run is relatively scarce. In addition, literature review on linguistics studies (e.g. learning new words) examined that drawing could improve memory compared to other learning methods such as writing (Wammes et al., 2016), but few studies have investigated whether drawing can be beneficial for knowledge retention when to-be-learned materials are complex as science topics and are in a passage or essay format. Therefore, this study aims to fill this gap by investigating if learner-generated drawing could lead to long-term retention of knowledge.

**Representational drawing and abstract drawing**

There are different terms representing drawing, such as external representations, external models, visualizations, illustrations and pictures. However, 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 and 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, that is, the drawing is intended to look-like the appearance and the structure of scientific phenomena (Alesandrini, 1984; van Meter & Garner, 2005). One typical representational drawing is sketching or diagraming which individuals use the
minimum number of lines and symbols to rapidly represent what they have learned. Different from other external representational modeling, sketching serves as a useful learning strategy to support learners to construct visuo-spatial external models of scientific phenomena especially when learning from expository text materials (Scheiter et al., 2017). Specifically, sketching could help students to get a deep understanding of spatial and causal knowledge of scientific domains (Goel, 1995; Suwa & Tversky, 1997). For example, the phenomenon of plate tectonics is not easy to explain only by verbal 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. Apart from the feature of sketching, prior research also found some boundary conditions that affect the effectiveness of using sketching as an aid to learn from text materials. One main factor is whether providing learners with a clear specification of what to draw before learning. For example, Leutner et al. (2009) revealed that 10th grade students didn’t get better performance by sketching pictorial representations of water molecules than students who only mentally imagined in the process. Sketching without any additional instructions or scaffolds could impede learners’ comprehension of learning text materials, due to increased extraneous cognitive loads caused by the logistics of managing drawing activities (Schwamborn et al., 2011). By contrast, 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). In addition, students also benefit from receiving feedback on their drawing (van Meter & Garner, 2005). After initial sketching, providing students with timely feedback could enable them
to reflect on drawing and help them self-regulate their interactions with learning contents (Wu & Rau, 2019). Therefore, teachers ought to provide students opportunities to construct their external pictures as well as modifying their creations. The carbon cycle is a complex, large-scale process that occurs over a long period of time, which requires students to establish spatial thinking skills to understand the interactions among the components of the climate system. Regarding the features and boundary conditions of sketching, students can engage in well-designed activities of sketching to identify and get deep understandings of the spatial configurations of different carbon components and causal knowledge of matter and energy transfer in the carbon cycle.

Abstract drawing, on the other hand, is much more analogical than representational drawing. This type of external representations is mostly consisted of words, numbers, lines, and/or arrows, but with spatial relationships, such as in flowcharts, graphs and phylogenetic trees (Quillin & Thomas, 2015). Concept mapping is a typical example of abstract drawing which has played a significant role in science education since the 1970s. Novak and Gowin (1984) indicated that concept mapping is a meaningful learning tool that enables learners to externalize their ideas of scientific phenomena and depicts suggested relationships that link these ideas. 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). This graphic organizer often includes concepts enclosed in boxes or circles, and relationships between concepts indicated by labeled arrows or lines. Previous studies showed that concept mapping activities are effective on both assessment and instruction (Novak & Gowin, 1984; Willerman & Mac Harg, 1991; Novak & Cañas, 2008). In particular, doing concept
mapping can facilitate meaningful learning processes by incorporating new knowledge into prior knowledge (Novak & Cañas, 2008). In an exemplary study, 8th grade students from four science classes were asked to either draw concept maps (experimental group) or received a series of guided questions at the beginning of each class (Willerman & Mac Harg, 1991). After two weeks, students who participated in drawing concept maps outperformed those who received guided questions. This result was explained by concept maps that serve as external scaffolds to help learners organize and structure different 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 learning (Chularut & DeBacker, 2004), but concept mapping is much more concise, and emphasizes the processes of different related aspects into coherent structure, which in turn is beneficial for deep scientific reasoning. Given the carbon cycle is a typical complicated subsystem in the climate system, involving many highly interconnected, imperceptible and invisible interactions among the components, creating concept maps may be effective to enhance students’ system thinking on the carbon cycle.

Overall, although different external representations share some similar effects on learning, only a limited amount of research has been conducted regarding the application to promote the learning of the carbon cycle. Additionally, either sketching or concept mapping potentially plays a major role in enhancing the comprehension of the global carbon cycle. Yet, we were not sure what differences exist between these two drawing methods in terms of their effectiveness on enhancing students’ construction of a
conceptual model of the carbon cycle. Thus, another goal of the study was to compare the effectiveness between sketching and concept mapping for learning on the carbon cycle.

**Drawing with retrieval practice**

Retrieval practice is a learning strategy that enhances meaningful learning and long-term retention of memory (Karpicke, 2012, 2017; Roediger & Karpicke, 2006b; Roediger et al., 2010). By practicing retrieval, learners frequently recall what they have learned by reactivating information from long term memory (Glover, 1989; Gates, 1917; Spitzer 1939; Tulving, 1967). This activation process is accompanied by actively constructing and storing knowledge (Karpicke, 2014). Research of 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. Studies using educationally relevant materials on this topic have grown intensively over the past decade (see Pan & Rickard, 2018; Roediger et al., 2010; Rowland, 2014), and the general experimental procedure of retrieval practice has three main phases - an initial learning phase, an interventional testing phase, and a final assessment phase. In the initial learning phase, learners study to-be-learned materials. After that, they are required to retrieve what they have just learned without access to the original materials. The types of interventional activities are usually tests or quizzes, which can be also referred to as self-testing activities. As a comparison to the intervention, there is often a control condition where learners typically do not engage in any additional activities or just restudy the materials. Finally, learners in all conditions take a final assessment which can be from a few minutes (Rowland &
DeLosh, 2015; Smith et al., 2013) after interventional activities to several weeks later (Carpenter et al., 2009; Larsen et al., 2013).

Recently, researchers found that combining retrieval practice with drawing could improve the effectiveness of science learning. Heideman et al. (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 contents. In their study, they found students had better performance as well as better recall of contents with the MSFL relative to students who only restudied the material by 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 conceptual and inferential questions. However, these studies did not include immediate tests, that is, we do not know the effect of different drawings on students’ immediate performance. It is important to know the immediate benefit as well as the long-term benefit of certain interventions in different educational settings.

Together, both drawing and the retrieval practice methods 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’ learning outcomes on the carbon cycle. Therefore, the current study aims to
investigate how different types of drawing methods combined with retrieval practice impact high school students’ conceptual understanding and learning outcomes on the carbon cycle.

**Method**

**Overview of the study design**

In the current study, high-school students learned the knowledge of the carbon cycle by reading an introductory essay. A between-subject design was applied with students randomly assigned to six different learning conditions (see Table 4). Of the six conditions, four were retrieval-based drawing treatments in which students were asked to visualize their mental models of the main processes of the carbon cycle to re-consolidate their understandings of this topic. To examine if additional benefits existed when drawing methods combining with retrieval practice, we added a retrieval practice only condition—recalling by paragraphing and a restudy condition as control. The retrieval and the restudy conditions are typical conditions used by other researchers studying the impact of retrieval practice on learning (Blunt & Karpicke, 2014; Karpicke & Blunt, 2011). After this initial learning phase, students’ conceptual understanding of the related concepts was assessed by an immediate test, and their retention of knowledge was assessed by the same test in a week.
Table 4

Introduction of the six conditions

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

Participants

The study was conducted at an urban high school in a city located in east China. N = 186 tenth-grade students enrolling in geography class were assigned randomly to the six conditions. 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.

Learning and test materials

The learning material used in the current study was an expository text of the global carbon cycle designed by the author, with a total of 659 words. From this material, students learned the basic components of the carbon cycle and the processes of carbon transfer and transformation across atmosphere, hydrosphere, biosphere and lithosphere.
An experienced Chinese high school geography teacher reviewed the text and edited it to be appropriate for the students to use as a class reading material.

Additionally, six different learning tasks were developed by the author in accordance with the six learning conditions. That is, students in both RS and GS conditions were provided a sheet of paper containing a background picture and a couple of figures of the carbon cycle, along with an example of sketching on the urban heat island. Students in both RC and GC conditions were provided with an example of concept maps on the Earth’s energy budget. Students in both retrieval practice condition and restudy condition were provided a sheet of blank paper. Moreover, students in all conditions received necessary instructions and completed their respective learning tasks independently (see details in the next session).

The immediate test and the one-week delayed test were essentially the same which covered the concepts of the carbon cycle. Three different types of tasks were included in the tests: five fill-in-the-blank questions, a diagram task and a flow-chart task. The fill-in-the-blank questions and the diagram task were designed to assess students’ factual knowledge on the carbon cycle, and the flow-chart task was used to assess how they applied what they had learned on the carbon cycle to a new context. Specifically, the flow-chart task was adapted from a box-diagram assessment tool by Sibley et al. (2002), and the other question items were designed by the author. A pilot study was conducted with a different group of students from the same school to check if the instructions were clear to students.

For scoring purposes of the tests, a rubric was developed by the author beforehand and the pilot study was conducted to validate the clarity of the instructions. Specifically,
students received one point for each fill-in-the-blank question if they gave the correct answer (totally 5 points). Similarly, they were given one point when they correctly answered each item of both 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 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 author. All materials mentioned above can be found in the Appendix C.

Procedure

Before the learning session, all students were taught 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. Also, since the carbon cycle is an interdisciplinary topic, 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. In addition to learning and reviewing basic knowledge, students were shown 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 (2006). Students were not provided training sessions on sketching since the sketching tasks were simpler and more straightforward compared to the concept mapping method. Besides, students were
provided with necessary symbols and figures in the sketching tasks’ instruction to reduce
the influence of limitations on students’ drawing abilities.

After being equipped with background knowledge and drawing skills, students
participated in the learning session. Table 5 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. That is, 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 of the
current study showed that the time was enough for students to complete each learning
phase under 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 study. In all conditions, students were given 30 min totally to complete their learning tasks.

At the end of the learning session, students from all six conditions took the immediate test, the content of which is 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 test again one week later in 20 min (i.e., the delayed test).

Table 5

*Procedure of the study design*

<table>
<thead>
<tr>
<th>Learning condition</th>
<th>Learning session</th>
<th>Testing session</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Retrieval Sketching</td>
<td>Study (5min)</td>
<td>Sketching (10min)</td>
</tr>
<tr>
<td>Retrieval Concept mapping</td>
<td>Study (5min)</td>
<td>Concept mapping (10min)</td>
</tr>
<tr>
<td>Generative Sketching</td>
<td>Sketching while reading the text (30 min)</td>
<td>Concept mapping while reading the text (30 min)</td>
</tr>
<tr>
<td>Generative Concept mapping</td>
<td>Study (5min)</td>
<td>Recall (10min)</td>
</tr>
<tr>
<td>Retrieval only</td>
<td>Study and restudy by taking notes (30min)</td>
<td>Study (5min)</td>
</tr>
</tbody>
</table>

**Results**

A one-way ANOVA test on the GPA scores of all students revealed no significant difference among the six conditions at the .05 level, indicating the participants assigned to different conditions were equivalent satisfying the randomization assumption. Table 6
shows the mean proportion correctness and standard deviation of each condition on the immediate and the one-week delayed test. They are also displayed separately for the overall carbon cycle test, the factual knowledge questions, and the application questions. The proportion correctness score for the overall test is determined by dividing the total score of each participant by the highest possible score (26 points) on the test. The proportion correctness score for the factual knowledge questions is 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. The proportion correctness score for the application knowledge is determined by dividing the score of each participant on the flow-chart task by the maximum possible score of 8.

**Table 6**

*Mean Score of Students for Different Questions from the Six Conditions on Both Immediate and One-week delayed Tests*

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

*Note: Standard deviations are reported in parentheses.*

**Initial learning activities**
We first investigated the number of idea units covered in each condition (except the restudy condition) during the learning session (see Table 7). The global carbon cycle expository text was divided into 23 idea units for scoring purposes. Specifically, students’ artifacts were scored by the following grading protocol: one point is rewarded for each correctly presented idea unit in all five conditions rendering a maximum 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 a significant difference between the five conditions, $F(4, 109) = 2.336, p = .06, \eta^2_p = .079$. Students in the retrieval practice condition produced relatively more idea units during study, but there was no difference between the rest of the four drawing conditions.
Table 7

Mean Score of Students for the number of idea units covered in the initial learning activity.

<table>
<thead>
<tr>
<th>Condition</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generative Sketching (N=23)</td>
<td>13.96</td>
<td>3.76</td>
</tr>
<tr>
<td>Generative Concept mapping (N=23)</td>
<td>12.74</td>
<td>3.86</td>
</tr>
<tr>
<td>Retrieval Sketching (N=23)</td>
<td>13.30</td>
<td>3.37</td>
</tr>
<tr>
<td>Retrieval Concept mapping (N=22)</td>
<td>12.73</td>
<td>2.66</td>
</tr>
<tr>
<td>Retrieval Practice (N=23)</td>
<td>15.30</td>
<td>2.97</td>
</tr>
<tr>
<td>Restudy (N=22)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Overall performance on the immediate and one-week delayed test

We examined the performance of students from all six conditions on the immediate and the one-week delayed tests (see Figure 14). 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, $F(1, 130) = 2.104, p < .001, \eta^2_p = .106$, indicating forgetting occurred during one week. We also found a main effect of learning condition, $F(5, 130) = 2.358, p = .044, \eta^2_p = .083$. However, the main effects were qualified by a marginally significant Learning Condition × Retention Interval interaction, $F(5, 130) = 2.168, p = .062, \eta^2_p = .077$.

Subsequent analysis showed no significant difference between all 6 conditions on the immediate test performance, $F(5, 130) = 1.214, p = .306, \eta^2_p = .045$. However, for the one-week delayed test performance, there was a significant main effect on condition, $F(5, 130) = 3.389, p = .007, \eta^2_p = .115$. 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, (.58 vs. .51), $t(44) = 2.25, p = .03, d = .11, 95\% \text{CI} [ .01, .14]$, as did the GC condition, (.58 vs. .50), $t(44) = 2.94, p = .005, d = .09, 95\% \text{CI} [ .03, .14]$, and did the restudy condition, (.58 vs. .44), $t(43) = 3.27, p = .002, d = .14, 95\% \text{CI} [ .05, .22]$, respectively. In addition, students in the RC condition had better performance than students in the restudy condition, (.55 vs. .44), $t(42) = 2.68, p = .005, d = .13, 95\% \text{CI} [ .03, .19]$, 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, (.54 vs. .44), $t(43) = 1.96, p = .06, d = .16, 95\% \text{CI} [ -.00, .18]$. 

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, $F(3, 87) = 1.245, p = .30, \eta^2_p = .041$, 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, (.58 vs. .55), $t(43) = .78, p = .44, d = .10, 95\% \text{CI} [ -.04, .09]$, and to the RS condition (.58 vs. .54), $t(44) = 1.08, p = .29, d = .13, 95\% \text{CI} [ -.04, .12]$. 

120
Figure 14

Students’ performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means. GS = generative sketching, GC = generative concept mapping, RS = retrieval sketching, RC = retrieval concept mapping.

![Bar chart showing performance](chart.png)

Factual knowledge performance

Figure 15 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, $F(1, 130) = 5.094, p = .026, \eta_p^2 = .038$. There was also a main effect of learning condition, $F(5, 130) = 2.436, p = .038, \eta_p^2 = .086$. However, there was no statistically significant interaction effect, $F(5, 130) = 1.077, p = .376, \eta_p^2 = .04$.

Subsequent one-way ANOVA revealed a non-significant result for all 6 conditions on the immediate factual knowledge performance, $F(5, 130) = 1.245, p = .292, \eta_p^2 = .046$. There was a significant main effect of the condition on the one-week delayed performance, $F(5, 130) = 2.436, p = .038, \eta_p^2 = .086$. However, there was no statistically significant interaction effect, $F(5, 130) = 1.077, p = .376, \eta_p^2 = .04$. 
factual performance, \( F(5, 130) = 3.21, p = .009, \eta_p^2 = .11 \). 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 scores than the GS condition, ( .57 vs. .48), \( t (44) = 2.43, p = .02, d = .12, 95\% CI [.01, .16], \) as did the GC condition, ( .57 vs. .46), \( t (44) = 3.22, p = .002, d = .11, 95\% CI [.04, .17], \) and did the restudy condition, ( .57 vs. .44), \( t (43) = 3.32, p = .002, d = .13, 95\% CI [.05, .20], \) respectively. In addition, both the RC condition and the RS condition outperformed the restudy condition, ( .52 vs. .44), \( t (42) = 2.19, p = .035, d = .13, 95\% CI [.00, .16], \) and ( .53 vs. .44), \( t (43) = 2.09, p = .042, d = .15, 95\% CI [.00, .19]. \)

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, \( F(3, 87) = 1.664, p = .181, \eta_p^2 = .054 \). In addition, we also found that the retrieval practice condition produced essentially equivalent outcomes as the RC conditions, ( .57 vs. .52), \( t (43) = 1.35, p = .183, d = .11, 95\% CI [-.02, .11], \) and as the RS condition ( .57 vs. .53), \( t (44) = .793, p = .432, d = .13, 95\% CI [-.05, .11]. \)
**Figure 15**

Students’ factual knowledge performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means. GS = generative sketching, GC = generative concept mapping, RS = retrieval sketching, RC = retrieval concept mapping.

**Application knowledge performance**

Figure 16 shows students’ performance on the application knowledge questions for both the immediate and the one-week delayed test. Again, a $6 \times 2$ mixed ANOVA revealed a statistically significant main effect of retention interval, $F(1, 130) = 9.63, p = .002, \eta_p^2 = .069$. However, there was no statistically main effect of learning condition, $F(5, 130) = 1.21, p = .308, \eta_p^2 = .044$. We also did not find significant interaction effect, $F(5, 130) = 1.42, p = .221, \eta_p^2 = .052$. Thus, for the application knowledge performance, we did not find any difference between all the conditions on either the immediate or the delayed tests.
Students’ application knowledge performance (proportion correct) on the immediate and the one-week delayed tests as a function of the condition. Error bars represent standard errors of the means. GS = generative sketching, GC = generative concept mapping, RS = retrieval sketching, RC = retrieval concept mapping

Discussion

The first goal of this study was to investigate whether retrieval practice could enhance high school students’ learning of the concepts related to the carbon cycle. The results replicated 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). Importantly, this study also demonstrated that retrieval practice not only could be adopted as a stand-alone learning strategy, but also could be implemented in combination with other well-established elaborative learning methods 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 global carbon cycle. When students generated their models of carbon cycle, similar levels of learning benefits were produced by creating either abstract drawing - the concept mapping or the representational drawing – sketching. This 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. The present study illustrated that high school students could very well master different types of representational skills for the learning of environmental science. One of the things this study did not examine, but future studies may investigate is the subjective evaluation of students on different drawing methods when applying them in environmental science learning contexts.

The last goal of this study focuses on whether different drawing methods 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 with 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 method (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. Also, 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). They 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 practice retrieval either by paragraphing or by creating 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. However, the authors did not compare the closed-book concept mapping condition with the open-book concept mapping condition, therefore they did not report there was actually 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 method begs for an explanation. One possible explanation is the typical cognitive process required by retrieval practice is disturbed when students work on drawing tasks. 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.

Another possible account could be generated from the findings of Roelle and Nückles (2019). In their study, they 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. 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.

Research based on the effect of combining different educational activities remains quite small, especially in environmental sciences. There is considerable room for further exploration of various approaches to integrate retrieval practice with other effective learning strategies in environmental science learning. This current study sheds some initial light on the effect of using two different drawing methods synergistically with retrieval practice to promote students’ understanding of the global carbon cycle. We found that retrieval practice combined with drawing methods such as sketching or concept mapping has a positive impact on the long-term retention of knowledge for high school students. The benefits are strongest when compared with the commonly adopted restudying method by students. The implication for high school science teaching is that teachers should be encouraged to adopt retrieval practice in their everyday teaching to reap its long-term mnemonic benefits for students. In the meanwhile, it also opens the possibility for future research on exploring ways to broadly incorporate retrieval practice with other highly effective learning strategies to maximize learning efficiency in STEM education.
Appendix C

Reading text (translated from Chinese):

Global Carbon Cycle

Carbon, a common chemical element, is the basic building block of life and can be found as forms of carbon compounds everywhere on Earth. A carbon atom can spend millions of years traveling and exchanging between the biosphere, hydrosphere, atmosphere, and lithosphere of Earth in a complex cycle, which we call the carbon cycle. According to the carbon cycling period, we can divide the carbon cycle into the biological carbon cycle (short-term) and the geological carbon cycle (long-term).

In the atmosphere, carbon is mainly present in the form of carbon dioxide, the key component of the carbon cycle. On land, plants absorb carbon dioxide from the atmosphere through the process of photosynthesis, and then the carbon dioxide is converted into glucose molecules ($C_6H_{12}O_6$). When plants and animals respire, they release the carbon dioxide into the atmosphere through the process of cellular respiration.

Meanwhile, when animals eat plants, carbon atoms are passed through food chains (webs) from producers to consumers. When plants and animals die, a part of their remains gets buried deep into the soil and eventually converts into fossil fuels after a long period of time through geological processes. The other part containing organic matter starts to decay and is decomposed by soil microbes through the process of decomposition. This process finally releases carbon dioxide back into the atmosphere. The aforementioned content is about the carbon cycle between terrestrial organisms and the atmosphere.

The carbon dioxide can be released from Earth into the atmosphere directly through volcanic eruptions. Earth’s land and ocean surfaces sit on several moving crustal plates. When the plates collide, one sinks beneath the other, and the rock it carries melts under the extreme heat and pressure. The heated rock recombines into silicate minerals, releasing carbon dioxide.

The ocean is another significant component of the carbon cycle. The limestone in the ocean is the largest carbon reservoir on Earth, also known as the carbon sink. The ocean absorbs carbon dioxide through physical and biological processes.

At the ocean’s surface, carbon dioxide from the atmosphere dissolves into the water and evaporates when the water temperature changes. Tiny marine plants called phytoplankton use this carbon dioxide for photosynthesis. Phytoplankton are the base of the marine food web. After marine animals eat the phytoplankton, they breathe out the carbon dioxide or pass it up the food chains. Sometimes, these phytoplankton and animals die and decompose, releasing carbon dioxide. While a part of their remains sinks to the ocean floor, carrying carbon as they descend. Over long time scales, this process has made the ocean floor the largest reservoir of carbon on the planet.
After millions of years of geological processes, marine sediments convert into fossil fuels stored in the Earth's crust under extreme temperature and pressure conditions. Most of the ocean’s nutrients are in cold deep water. In a process called upwelling, currents bring nutrients and carbon up to the surface and release carbon dioxide back to the atmosphere.

The movement of carbon from the atmosphere to the lithosphere (rocks) begins with rain. Atmospheric carbon combines with water to form a weak acid—carbonic acid—that falls to the surface in rain. The acid dissolves rocks—a process called chemical weathering—and releases bicarbonate, calcium and magnesium ions etc. Rivers carry these ions to the ocean. Due to the alkaline nature of ocean water, bicarbonate ions combine with calcium ions to form calcium carbonate (CaCO$_3$) through a chemical reaction-neutralization, eventually sinking to the ocean floor.

Chemical equation (Weathering): $\text{CaCO}_3 + \text{H}^+ + \text{HCO}_3^- \rightarrow \text{Ca}^{2+} + 2\text{HCO}_3^-$

Human activities have had a tremendous impact on the carbon cycle in recent years. Burning fossil fuels such as oil, coal and natural gas, and changing land use by deforestation all transfer significant quantities of carbon dioxide into the atmosphere. As a result, the amount of carbon dioxide in the atmosphere is rapidly rising, breaking the balance of the original carbon cycle, causing the greenhouse effect and aggravating global warming.
Task a
Instructions: Please read the “Global carbon cycle” text and sketch your model of the global carbon cycle in the diagram. After that, label each process of the carbon movement or the name of each chemical/physical reaction in your drawing.

*You can choose any figures below heat island)  
*The example of annotating sketches (the urban
**Task b**
Instructions: Please read the “Global carbon cycle” and create your concept map of the global carbon cycle.

*The example of concept maps (Earth’s energy budget)*
Task c
Instructions: Please read the “Global carbon cycle” text. Then turn to the next page to sketch your model of the global carbon cycle in the diagram. After that, label each process of the carbon movement or the name of each chemical/physical reaction in your drawing. Please follow the instructions of your teacher on each step.

*You can choose any figures below heat island)  

*The example of annotating sketches (the urban
Task d
Instructions: Please read the “Global carbon cycle” text. Then turn to another page to create your concept map of the global carbon cycle. Please follow the instructions of your teacher on each step.

*The example of concept maps (Earth’s energy budget)
Tasks e
Instructions: Please read the “Global carbon cycle” text. Then turn to another page to write down as much information as you could remember from the text you just read and learned. Please follow the instructions of your teacher on each step.

Tasks f
Instructions: Please read the “Global carbon cycle” text carefully. Then you can go back to review the text and take some notes on any important information you may think on the text.
Assessment of the global carbon cycle

Name: ___________________ Class: ___________________ Date: ________________

**Fill-in-the-blank questions:**

Item 1) The carbon cycle can be categorized into the __________(short-term) and the_________ (long-term) carbon cycle.

Item 2) The carbon compounds are passed along the ______________ from producers to consumers.

Item 3) In the ocean, ___________ (A kind of plants) can take in carbon dioxide through photosynthesis.

Item 4) The limestone in the ocean is the largest carbon reservoir on Earth, also known as the________.

Scoring rubric:

<table>
<thead>
<tr>
<th>Item 1: Biological (1pt)</th>
<th>Geological (1pt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 2: Either food chains or food webs (1pt)</td>
<td></td>
</tr>
<tr>
<td>Item 3: Phytoplankton (1pt)</td>
<td></td>
</tr>
<tr>
<td>Item 4: Carbon sink (1pt)</td>
<td></td>
</tr>
</tbody>
</table>

**Free-respond questions:**

The diagram below represents a part of the global carbon cycle.

Item 5) Arrows and uppercase letters “A-D” in the diagram represent processes of carbon movement. The lowercase letter “a” represents a carbon matter after reactions of the extreme temperature and pressure. Please write down the name of each letter represents.
Item 6) What other processes of carbon dioxide transfer do not show up in the diagram?

Scoring rubric:

<table>
<thead>
<tr>
<th>Item 5</th>
<th>A: Human activities (1pt)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B: Respiration (1pt)</td>
</tr>
<tr>
<td></td>
<td>C: Photosynthesis (1pt)</td>
</tr>
<tr>
<td></td>
<td>D: Upwellings (1pt)</td>
</tr>
<tr>
<td>a:</td>
<td>Fossil fuels (1pt)</td>
</tr>
</tbody>
</table>

| Item 6 | Decomposition (1pt), Weathering (1pt), Volcanic eruptions (1pt), Evaporation in the ocean (1pt), Dissolving in the ocean (1pt), Evaporation from the rivers (1pt), Dissolving in the rivers (1pt), Dissolving in the processes of precipitation (1pt) |

The Figure 1 below shows me dead and buried on a hill. But do not worry, as a geologist I’m happy to know that someday some of my carbon atoms may end up in a limestone beneath the ocean.

The flow-chart represents one of processes on the movement of carbon atoms from “me” to a limestone. Each solid box is labeled as substance (upper portion) and location/reservoirs (lower portion) and dashed boxes are labeled processes of carbon movement. Please complete the flow-chart.
Figure 1

Item 7)

Scoring rubric:

1. **CaH$_2$O$_6$**
   - Remains/Soil
   - Decomposition (1pt)

2. **CO$_2$**
   - The atmosphere/The air (2 pts)
   - Raining/Precipitation (1pt)
   - Rocks on land (1pt)
   - Weathering

3. **HCO$_3^-$**
   - Rivers (1pt)
   - Erosion (1pt)
   - The ocean (1pt)
   - Neutralization (1pt)

4. **H$_2$CO$_3$**
   - Limestone (CaCO$_3$) The ocean
References


https://doi.org/10.1187/cbe.16-03-0116


http://dx.doi.org/10.5751/ES-08626-210330


https://doi.org/10.1177/0963721412443552


https://doi.org/10.1126/science.1199327


https://doi.org/10.1016/j.jml.2012.02.004


Schmeck, A., Mayer, R. E., Opfermann, M., Pfeiffer, V., & Leutner, D. (2014). Drawing pictures during learning from scientific text: Testing the generative drawing effect


Chapter 5: Summary

The goal of this dissertation was to explore ways to apply retrieval practice strategy on climate change learning at the high school level. Three studies presented that retrieval practice could be used broadly on different topics of climate change and in different learning scenarios. Generally, retrieval practice engages learners to deepen the learning processes by reconstructing their knowledge after initial learning. During the process of retrieval-based learning, learners were able to improve their comprehension of knowledge, rendering a longer period of retention of what they have learned versus just studying once or repeatedly reviewing the learning materials.

The first study of the dissertation focused on the effect of question placement when students studied sea level rise from an educational video lecture. In this study, retrieval practice was used in a style of short-answer quiz questions which were embedded in the video lectures. The results suggested that using embedded questions in videos has a clear advantage over restudying the video lecture in both immediate and the delayed tests. No difference in performance has been observed between two common practices of embedding question -- dispersed and stacked question styles. Nevertheless, we found the dispersed condition demonstrated an advantage over the stacked condition in explaining the embedded questions during the video learning.

The second study of the dissertation expanded the application of retrieval practice by combining with another useful learning strategy - learner-generated drawing. Specifically, we explored the effect of the retrieval-based sketching method with different level of retrieval support when students studied the Earth’s energy budget from a video lecture. Results did not show any significant difference in performance between the “no
support”, “50% support”, and “100% support” drawing conditions on both the immediate and delayed tests. However, students showed different perceived difficulty of the retrieval-based sketching tasks. Very interestingly, students from the non-support retrieval based-sketching condition reported their sketching task was less difficult than those from the half-support retrieval based-sketching condition reported on their sketching tasks. This counter-intuitive finding begs for more studies in the future.

Building on the findings of the second study, the third study focused on the effect of retrieval practice combined with different drawing strategies on carbon cycle learning. Concept mapping and sketching are two commonly adopted drawing strategies used to promote model-based reasoning in science learning. The results demonstrated that retrieval practice by paragraphing was more beneficial on learning than any concept mapping or sketching with access to the learning materials or restudying the learning materials. But there was no difference between the two drawing types on learning, regardless of whether they were retrieval-based or not.

This dissertation examined that retrieval practice could be used and implemented as common teaching and learning strategies appropriate for high school students in the field of environmental science. However, retrieval-based learning in the dissertation was only found effective for better retention of factual knowledge instead of inferential or application knowledge which requires higher-order skills. It solicits future studies to further explore how to better develop effective retrieval-based learning strategies and settings to improve students’ problem solving and argumentation skills in STEM education.
Finally, it should be noted that all three studies in the dissertation were conducted only once with relatively small sample sizes in a controlled lab-like environment. Future environmental science learning studies could design science courses integrating retrieval-based learning with more participants for a longer duration, say, at least one semester to examine its effectiveness in real educational settings. In addition, besides quantitative data collection, future studies could also integrate qualitative methods, such as structured or semi-structured interviewing and/or field observations to further investigate students’ attitudes or experiences towards retrieval-based learning. Thus, such results could contribute to our understandings of how to maximize the effectiveness of retrieval-based learning for science teaching and learning.