What role does causality play in comprehension of expository teaching materials and do cognitive load and individual differences impact that role?

by

Taylor Leigh Vargo

B.A., The University of Findlay, 2017

M.S., Kansas State University, 2021

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Abstract

This study investigates how cognitive load, causal connections, and individual differences influence comprehension of expository materials, particularly in online learning environments. This study focused on enhancing germane cognitive load by using graphic organizers—diagrams that visually represent concepts and the relationships between them to support student learning of classroom material. Participants engaged with educational content through self-generated and ready-made graphic organizers under explicit or implicit causality conditions and their comprehension for the content were tested across three delays ranging from 1 - 23 days. Results supported the Cognitive Load Hypothesis, revealing that self-generation tasks could hinder learning by imposing high cognitive demands. Explicit causal connections facilitated comprehension, but their effectiveness varied across different sessions, suggesting the importance of encoding accuracy and retrieval timing. Ready-made graphic organizers outperformed self-generated ones, particularly during delayed assessments, which was likely due to poor encoding of the material during the lecture (as evidenced by extremely low accuracy on the self-generated graphic organizer). These findings have significant implications for educational practices, especially in online settings. Instructors are encouraged to incorporate explicit causal structures and provide organized materials to reduce extraneous cognitive load and support retention. The study also highlights the role of individual differences—such as prior knowledge and motivation/engagement—in moderating learning outcomes, underscoring the need for tailored instructional strategies. By deepening our understanding of these factors, educators can better design interventions that enhance student learning and retention.

Key Words: Cognitive Load, Causality, Graphic Organizers, Mental Models, Online Learning

What role does causality play in comprehension of expository teaching materials and do cognitive load and individual differences impact that role?

by

Taylor Leigh Vargo

B.A., The University of Findlay, 2017

M.S., Kansas State University, 2021

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2025

Approved by:

Major Advisor
Dr. Heather Bailey

Copyright

Abstract

This study investigates how cognitive load, causal connections, and individual differences influence comprehension of expository materials, particularly in online learning environments. This study focused on enhancing germane cognitive load by using graphic organizers—diagrams that visually represent concepts and the relationships between them to support student learning of classroom material. Participants engaged with educational content through self-generated and ready-made graphic organizers under explicit or implicit causality conditions and their comprehension for the content were tested across three delays ranging from 1 - 23 days. Results supported the Cognitive Load Hypothesis, revealing that self-generation tasks could hinder learning by imposing high cognitive demands. Explicit causal connections facilitated comprehension, but their effectiveness varied across different sessions, suggesting the importance of encoding accuracy and retrieval timing. Ready-made graphic organizers outperformed self-generated ones, particularly during delayed assessments, which was likely due to poor encoding of the material during the lecture (as evidenced by extremely low accuracy on the self-generated graphic organizer). These findings have significant implications for educational practices, especially in online settings. Instructors are encouraged to incorporate explicit causal structures and provide organized materials to reduce extraneous cognitive load and support retention. The study also highlights the role of individual differences—such as prior knowledge and motivation/engagement—in moderating learning outcomes, underscoring the need for tailored instructional strategies. By deepening our understanding of these factors, educators can better design interventions that enhance student learning and retention.

Key Words: Cognitive Load, Causality, Graphic Organizers, Mental Models, Online Learning

Table of Contents

List of Figures

No table of figures entries found.

List of Tables

Acknowledgements

This journey would not have been possible without the support, guidance, and encouragement of so many people who walked alongside me.

To my advisor, Dr. Heather Bailey — thank you for shaping me into the researcher, teacher, and mentor I am today. Your thoughtful mentorship, steady guidance, and belief in my abilities have had a profound impact on both my work and who I've become through this process. I am deeply grateful for your time, your wisdom, and your trust.

To my past students — thank you for showing me that I was on the right path. You helped me realize that teaching and mentoring are not just part of what I do, but core to who I am. To my current students — you are the reason I kept going. Your energy, questions, growth, and ambition gave me the motivation to finish this degree and to keep striving to be better for you.

To my former colleagues and friends — thank you for being there during the hardest parts of this journey. You offered laughter, advice, support, and friendship when I needed it most. To my current colleagues and friends— thank you for showing me that there is life after the PhD. You remind me daily that the work we do matters, and that joy, community, and purpose belong in academia too. I'm so grateful to be part of this team with you.

And finally, to my husband, Seth — thank you for being encouraging, supportive, and yes, sometimes annoying, throughout this entire process. You've been my rock in the middle of anxiety, stress, and panic. Your unwavering support and love made this possible, and I'm endlessly thankful to have you by my side. You are deeply loved and appreciated.

Dedication

To me.

As a first-generation student who started this PhD journey eight years ago, I had no idea what I was getting into — I only knew where I wanted to end up. Along the way, I became something I never thought I'd be: a full-fledged researcher, a teacher who makes an impact, and a mentor who leads by example.

I've faced more than my fair share of obstacles, but I didn't just survive — I thrived. I had an incredible support system, but let's be clear: *I* did the work. Every long night, every tough choice, every deadline, and every milestone — that was all me.

I've accomplished more than I ever expected, and I dedicate this PhD to myself for doing exactly what I set out to do. You did it.

Chapter 1 - What Role does Causality Play in Comprehension of Expository Teaching Materials and do Cognitive Load and Individual Differences Impact that Role?

Almost 20 million students enroll in college courses each year with the hopes of passing classes and being ready to enter the workforce (Duffin, 2021). Yet, these students are not all equally successful in achieving these aspirations, which could be attributed to various factors. Most important for this experiment is a student's ability to properly encode new information into memory in a way that will increase the likelihood of later retrieval. This is no small feat, and even highly successful students experience memory failures. Researchers in both psychology and education domains have been working to better understand how students learn in the classroom and what can be done, from a pedagogical standpoint, to implement better educational methods and thus optimize learning.

Early foundational memory research by Ebbinghaus (1885) found that memory accuracy for new information significantly decreases immediately after learning, with the largest drop occurring in the first 20-30 minutes. This decline continues throughout the first day, with only about 30% of the material being retained. After 24 hours, memory performance levels off over the next 30 days, indicating that the first 24 hours are the most critical for memory retention. This original study has been replicated many times to include more participants (see Finkenbinder, 1913 for review; Heller et al., 1991; Murre & Dros, 2015), different presentation rates (Radossawljewitsch, 1907; Finkenbinder, 1913), and learning times (Radossawljewitsch, 1907), as well as using word stimuli (Shepard, 1967) and pictures (Intraub, 1980, 1981; Potter, 1976; Potter et al., 2002). Critically, more recent work using stories has evaluated memory for different levels of representation and found that long-term memory for event-level information (i.e., gist of the information conveyed) is remembered much better over time compared to text-based (i.e., paraphrase of words) and surface-form information (i.e., verbatim recall of words), which is forgotten more rapidly (Fisher & Radvansky, 2018). This suggests that emphasizing the gist in classroom settings may not be the best use of resources, since this information is well remembered over time. Rather, if instructors want students to be focused on text-based or surface-form information, more targeted interventions may be necessary to achieve this goal. Given that an instructor's main goal should be that students learn and remember course information over a long period of time, it is crucial to find ways to aid student retention for complex, educational materials over ecologically meaningful periods of time (i.e., longer than 5 minutes in a laboratory study).

Previous work has shown strategies such as rehearsal (e.g., Ebbinghaus, 1885), retrieval practice (e.g., Roediger & Karpicke, 2006), spaced study sessions (e.g., Cepeda et al., 2006) and self-reference (Rogers et al., 1977) can improve memory. Further, Bjork and Bjork (1992; 2020; Bjork, 1994) proposed the idea of *desirable difficulty*, which is a way to increase both encoding and retrieval processes for students, so they can achieve their optimal level of learning. Importantly, this concept focuses on two specific features: 1) increasing storage strength by increasing the extent to which information is integrated with existing items in memory, and 2) increasing retrieval strength, which helps the ease of access to an existing item in memory. Desirable difficulty can vary by individual, depending on their level of prior knowledge and the ease with which an individual can relate new information to this existing knowledge (McCarthy & McNamara, 2021). Integrating new information with existing knowledge is a powerful learning strategy that increases long-term retention of information. Linking new information to prior knowledge allows it to become part of the organized knowledge structure and is associated with better memory (Barlett, 1932). Highly organized schemas that make up prior knowledge help construct mental models, which are mental representations of experiences and textual content (Johnson-Laird, 1983).

Mental Model Construction

A mental model is an individual's mental representation of an ongoing event that is based on the external environment (i.e., what the individual is seeing, hearing, reading, etc.) as well as the individual's prior experiences, knowledge, and beliefs. Mental models help an individual understand the world around them by processing incoming information and making predictions about future events. When information in the environment changes, the mental model must be updated to reflect that change. Generally, there are three types of updating that can occur within a mental model: 1) new entities (e.g., a character or object) are added, 2) existing relationships amongst entities are updated to reflect the current situation, and 3) entities are removed when no longer relevant to the current situation (Radvansky & Copeland, 2001). When a mental model is updated, reading times slow down (e.g., Bohn-Gettler, 2014; Pettijohn & Radvansky, 2016; Zwann et al., 1995; Zwann et al., 1998), presumably because it takes time to integrate the changing information into the existing mental model.

The best explanation for why this occurs is that we integrate new information into our existing mental model instead of constructing a new model from scratch. This integration process is described by the Event Indexing Model (Zwaan et al., 1995; Zwaan & Radvansky, 1998), which describes readers' ability to track situational features within a text. These features include: 1) space, 2) entities, 3) motivations/goals, 4) time, and 5) causality. For instance, readers track the main characters in a text, and when the text focuses on a new character, readers must update their current mental model to reflect that change. While each of these situational features plays a crucial role in comprehension, causality is particularly significant because it helps readers understand the logical flow and connections between events (for review, Radvansky, 2012). The likelihood of detecting these changes and subsequently integrating them into an existing mental model can be increased by instructing people to attend to certain situational features (Bohn-Gettler et al., 2014).

In fact, causality interventions in educational settings have successfully increased learning (Cevasco & van den Broek, 2019). Encouraging students to generate (McMaster et al., 2012; Mokhtari & Reichard, 2002) or recognize causal connections (Linderholm et al., 2000; Vidal-Abarca et al., 2002) has been shown to improve reading comprehension and recall in younger children, high school, and college students. Causal connections also help in understanding spoken language (e.g., Cevasco 2010, Cevasco & Muller, 2009). Because of its importance in both laboratory and applied setting comprehension, causality will be the situational feature of primary focus in the current study.

Causality

According to the Event Indexing Model, readers represent the cause-and-effect relationships between events or ideas in a text (Zwann & Radvansky, 1998). Tracking and updating causality within our existing mental model is considered to be a natural and automatic process while reading narrative texts (Zwann et al., 1995). Specifically, people have better memory and higher ratings of importance when there are more causal connections between events depicted (Trabasso & van den Broek, 1985). Studies that evaluate the importance of causality have shown that reading times are related to its presence or absence (Radvansky & Copeland, 2004; Pettijohn et al., 2016; Zwaan et al., 1995; Bohn-Gettler et al., 2014; Radvansky, 2012). Specifically, when causal connections are absent from a text or when there are shifts in causality within an existing mental model (i.e., location changes or new characters are introduced; Bohn-Gettler et al., 2014) there is an increase in reading times.

Taken together, this work suggests that 1) causal relationships are automatically monitored and 2) doing so is important when building mental models. Importantly, though, this prior work did not use expository materials and focused primarily on narrative texts. In fact, most research evaluating causality involves narrative next, whereas the use of expository materials (i.e., educational materials, lectures) is less common. It is important to closely evaluate how people comprehend expository information because past work has shown that readers may use different comprehension strategies when reading narrative and expository text (Wolfe, 2005).

One model that can help guide testable predictions about how causal connections influence the comprehension of various situations, including narrative text (e.g., Radvansky et al., 2005) and expository text (e.g., Graesser et al., 2003; Clariana et al., 2014), is the Event Horizon Model (Pettijohn et al., 2016; Pettijohn & Radvansky, 2016; Radvansky, 2012; Radvansky & Zacks 2017). This model consists of 5 principles, but 2 of these are the most relevant for this paper. Principle 1 proposes that events are segmented and stored as smaller segments of information, and that these segments are influenced by changes in situational features. Principle 3 proposes that causal relations influence how events are segmented and are integrated into the memory representations of these events.

In sum, causality is an important index that people naturally track to help link information together to improve memory. When causal relationships are clear, comprehension is easier and reading times are faster (Zwann et al., 1995; Radvansky & Copeland, 2004). However, when there is a break in causality, reading times are slower as the current information is difficult to integrate with the preceding information. Thus, a person's mental model must be updated to reflect that change in causality. Given the importance of causal relationships, making these relationships more explicit should improve comprehension. To my knowledge, this idea has not been tested but will be one of the focuses of the current study. Specifically, it would be expected that if participants can easily track causal relationships in expository materials, they should have improved memory both when tested immediately and after a delay.

Expository Text Memory

Importantly, this idea of mental model updating, as discussed so far, has mostly been used in the realm of narrative texts. When focusing on the work done using expository texts, there is some reliance on this way of conceptualizing knowledge organization (e.g., Graesser et al., 2003; Clariana et al., 2014); however, it becomes more relevant to focus on discourse processing (e.g., Meyer 1975, 1982; Mayer & Freedle, 1984). In this line of work, the focus has been comprehension in an educational setting, rather than comprehension for reading for entertainment, which has been shown to lead to differences in comprehension over time (Narvaez et al., 1999). In this theory of comprehension, Meyer and colleagues (e.g., 1975, 1982) suggest that memory and comprehension depend on the level of organization within a text. Specifically, they propose 4 different types of discourse that exist in expository texts: comparisons (i.e., compare one method to a another method), problem/solution (i.e., describe a problem and then the steps to solve the problem), causation (i.e., what led to this outcome), and description (i.e., listing facts or relationships). Under this theory, none of the individual organizational structures, but rather it is the culmination of multiple organizational structures, enhances memory and comprehension.

When examining schema creation and comprehension, it's important to consider the organizational structure of a text. Descriptive expository texts, which rely on a single structure (i.e., grouping by association), tend to result in the poorest comprehension (Meyer & Freedle, 1984; Meyer & McConkie, 1973). Causation texts offer slightly better comprehension by incorporating additional structures such as temporal comparisons and causal relationships. Problem/solution texts tend to yield the strongest comprehension, as they integrate all of the above structures along with overlapping content that connects the cause of a problem to its potential solutions. Whereas comprehension of comparison texts varies depending on the topic and the number of similarities and differences presented. Further, the number of overlapping points is thought to lead to better encoding of information (i.e., stronger storage strength; Bjork & Bjork, 1992; 2020; Bjork, 1994) and, thus, an increased ease of retrieval (Bjork & Bjork, 1992; 2020; Bjork, 1994). According to the *desirable difficulties theory* proposed by Bjork and Bjork (1992; 2020), these are the two most important aspects of optimal learning.

Research has shown that this theory holds true in a classroom setting. Specifically, certain expository texts containing comparisons were remembered at higher rates than those containing descriptions only (Meyer & Freedle, 1984), and these differences hold up over a week-long delay. Importantly, information also seems to be stored in the way that it is presented to students. For example, when presenting a text in a descriptive manner, students are more likely to recall descriptive information. The same is true for causal and comparisons texts; however, the evidence is mixed for problem/solution texts. Meyer and Freedle (1984) suggested that this may have been due to students rejecting the solution that the text provided, indicating a role for top-down processing to play a role in the schema creation process.

This idea of top-down processes impacting the schema creation process was further shown when looking at different levels of reading comprehension in 4th-8th graders (Meyer et al., 2018). In this study, they looked at the effectiveness of an intervention that focused on enhancing the use of and understanding of signal words (e.g., “however”; i.e., causal connectives, Zwaan et al., 1995) in the classroom. Fourth grade students in the intervention group, especially those with low comprehension ability, had increased comprehension, which replicated past research (Vauras et al., 1994). Meyer et al. (2018) also found that 8^{th} graders in the intervention group showed an increase in the ability to recognize and comprehend difficult signal words. These effects indicate that students with certain levels of comprehension ability are more adept to learn and utilize these signal words, and that there is a learning curve/skill level required in some areas (Geva, 2007, as cited by Meyer et al., 2018); however, it is a skill that can be taught and reinforced in students. Some research suggests that these effects start to plateau around the 8^{th}-grade level (Cain & Nash, 2011), indicating a necessity to teach these skills as early as possible (Bartlett, 1978, as cited by Meyer & Freedle, 1984; Cook, 1983 as cited by Meyer & Freedle, 1984; Mayer, 1982).

However, if a reader is not yet skilled enough, then building a schema using these signal words will be less efficient (Crosson & Lesaux, 2013; Welie, Schoohen, Kulken, van den Bergh, 2017), because they do not yet know what these words signify in the text. Importantly, this struggle can be offset when teaching students to use and understand explicit signal words (e.g., “however”, “the solution is”, etc.) in an expository text (Meyer & Poon, 2001), suggesting further that causal language can be useful in the expository texts (e.g., Graesser et al., 2003; Clariana et al., 2014).

In sum, research suggests that having more organizational components will enhance memory, but it is rarely ever this straightforward when dealing with students in a classroom setting. First, discourse does not usually neatly fit into a single category (Meyer 1975, 1982), meaning it is normal to have elements of different types of expository texts within a single passage (e.g., describing the water cycle would have factual/descriptive information, but also causal information). Second, individuals can be taught to use signal words (Meyer et al., 2018) and how to use those signal words to better comprehend different types of passages (Mayer, 1982). Taken together, this suggests that there is a benefit to understanding these aspects of text and does lead to benefits in comprehension for individuals.

Cognitive Load

Underlying these ideas of mental model updating are theories related to cognitive load; specifically, Sweller's Cognitive Load Theory (i.e., Sweller, 1994; Sorden, 2005; Brunken et al., 2010; Plass & Kalyuga, 2018). In his theory, Sweller outlines various factors that negatively and positively impact learning and comprehension by increasing or decreasing cognitive load. Specifically, cognitive load that negatively affects comprehension could involve 1) a task that is unrelated to mental model creation and updating, 2) a task that requires learners to complete an activity in a short amount of time, and/or 3) a highly demanding task. Examples of this type of cognitive load would be instructional materials that use unrelated or decorative images, not allowing students to think through a response before providing an answer or having students complete multiple tasks at the same time. These types of situations could easily overwhelm a learner by using much of their available working memory (WM) resources and, thus, could hinder storage efficiency, leading to decreased learning and worse long-term memory.

Conversely, tasks that create a positive cognitive load are those that require learners to utilize newly learned information in some way beyond the processes that typically occur during the "learning phase". For instance, students could be asked to create examples of the new content, engage in a group discussion, create concept maps or mind maps, or apply the concepts to a real-world scenario. All of these examples help to increase comprehension of material by increasing storage strength (Bjork & Bjork, 1992; 2020; Bjork, 1994) by creating an elaborated understanding of the content (Colliot & colleagues, 2018b, 2019, 2020, 2021, 2022; Stull & Mayer, 2007). By making connections to content, cognitive demand on WM is reduced through reinforcement or reiteration (i.e., information is processed more deeply with some sort of learning activity; Craik & Lockhart, 1972) of what has already been presented.

While cognitive load can be positive and negative there are three different types of cognitive load: (1) *intrinsic cognitive load* (I-CL), (2) *extraneous cognitive load* (E-CL), and (3) *germane cognitive load* (G-CL). I-CL positively impacts a learner and is the information required to learn the material. I-CL is also based on the complexity of material and expertise of the learner. E-CL negatively impacts a learner and includes factors that are unnecessary for learning material. This may include internal factors (e.g., mental health, money struggles) and external factors (e.g., instructional presentation/format). G-CL positively impacts a learner and includes learners' efforts to enhance their comprehension. Specifically, this includes any activity or assessment meant to further a learners' understanding of material.

Generally, E-CL is most studied, because it can be easily manipulated by changing the presentation speed or increasing the amount of information presented at a single time (Mayer et al., 2019) or even manipulating the relevance of a picture with the content (Mayer, 2014). G-CL and I-CL are harder to measure because they vary across individual learners. I-CL is difficult to assess because it is the internal nature of the processes that would fully capture its impact on the learner. G-CL can more easily be manipulated by providing learners with different tasks related to their learning, such as labeling hierarchical structures depicting course concepts and the relationships within the content they had just learned (e.g., Colliot & Jamet, 2018a) or preparing to teach others new materials (Hoogerheide et al., 2019). The nature of G-CL and E-CL lead researchers to believe they are inversely related because optimal learning environments decrease E-CL and increase G-CL (Cierniack, et al., 2009). Thus, in an optimal learning environment, G-CL processes are being used, which should then activate and update any existing, relevant prior knowledge on the topic. For instance, imagine you know that mammals do not lay eggs, but then you learn about the platypus, which is a mammal that does lay eggs, and you now must update your “mammals” schema. Ideally, this process would then allow for more accurate and coherent representations of information by taking the established semantic information and integrating the newer information. If this is true, the process of engaging in a task that increases G-CL should strengthen the connections between new and prior knowledge, while also elaborating on these connections to form a more integrated and comprehensive understanding.

Researchers have measured cognitive load in numerous studies, but unfortunately, the methods used to do so are often subjective and/or indirect, rather than more cognitive-based, objective and/or direct measures (Plass & Kalyuga, 2018; Hoogerheide et al., 2019; Nourbakhsh et al., 2017; Brunken et al., 2010). Subjective rating scales, in which participants are asked to rate their perceived experience on a single question (e.g., how difficult was the task?’), are used most often (Brunken et al., 2010; Multu-Bayraktar et al., 2019). While these subjective methods are valuable, it is important to verify that they are related to observable and measurable behaviors, which is what makes objective measures so valuable.

Several researchers have recently begun to address this issue by using galvanic skin response (GSR; Nourbakhsh et al., 2017; Hoogerheide et al., 2019) and eye-tracking (Zu et al., 2020) to validate the commonly used self-report measures. For instance, Hoogerheide et al. (2019) had participants either practice math problems on their own or practice teaching the content to other people while being videotaped. Participants in the teaching condition rated the task as more difficult and had increased GSR. Such results had converging evidence from both indirect and direct measures and indicated that teaching increased G-CL. Similarly, Nourbakhsh et al. (2017) had students solve math problems that varied in difficulty and found that increased GSR was associated with increased subjective ratings of cognitive load. Further, Zu et al. (2020) used eye-tracking and WM to measure I-CL, E-CL and G-CL and found measures of attentional selection (e.g., saccades, fixations, etc.) that were independently related to each type of cognitive load. Together, these studies (Nourbakhsh et al., 2017; Hoogerheide et al., 2019; Zu et al., 2020) demonstrate that the commonly used subjective measures are reliable and valid measures of cognitive load. Furthermore, they confirm that the three types of cognitive load can be measured independently using both objective and subjective measures.

Graphic Organizers and Germane-CL

Some work has evaluated how cognitive load impacts comprehension using diagrams or graphic organizers that would both require an understanding of causal relationships to complete (i.e., rain must fall from the sky before it can be evaporated back into the atmosphere again). Graphic organizers are “spatial arrangements of words, indicating relations between concepts, that are intended to represent the conceptual organization of texts (Stull & Mayer, p. 810, 2007).” While there is evidence of comprehension-related benefits when using these visual representations (Cieniak et al., 2009; Van Merrienboer et al., 2010), some researchers have found that generating causal relationships can be cognitively demanding (Zwann & Radvansky, 1998; Sweller, 1994), whereas others have found it does not affect comprehension (McCrudden et al., 2007).

Most relevant to the current study is a series of studies by Colliot and colleagues (2018a, 2018b, 2019, 2020, 2021, 2022) who evaluated the efficacy of graphic organizers on comprehension. The basic paradigm across these studies shows participants a short PowerPoint presentation and instructs them to complete a task at the same time as viewing the presentation. This task requires participants to either read a fully complete (i.e., ready-made) graphic organizer or fill in the information (i.e., self-generate) representing the concepts and/or relationships amongst concepts. Across several experiments, Colliot and colleagues used this self-generation task, which required participants to use empty boxes (i.e., where they would later type the concepts from the lecture) and lines to create their own hierarchical depiction of what was discussed in the lecture. In some of the experiments, they included a third group that was labeled the *partially generated group*. In this group, participants received the hierarchical depiction (i.e., boxes and lines were in the correct places), but the content still needed to be generated by participants (2019). After learning the content and reviewing/completing the graphic organizer, all participants then answered questions about the content. The researchers expected to find that self-generating the information on the graphic organizer should increase G-CL, ideally then increasing comprehension of material, because it requires learners to elaborate on the new information.

In their first several experiments, Colliot and colleagues found no benefit of self-generation (Colliot & Jamet, 2018b; 2019; 2020). Instead, they found that participants who were given a ready-made graphic organizer remembered more information compared to those who had to fully self-generate their graphic organizers, who in turn performed similarly to those who partially-generated their graphic organizers. Colliot and colleagues claim that generating the entire graphic organizer may have resulted in high E-CL, which explains why self-generation did not produce comprehension benefits.

To address this issue, Colliot and Jamet (2020a) manipulated *when* participants viewed the graphic organizer: either during or after the lecture. They found that presenting the graphic organizer after the lecture produced better comprehension than presenting them simultaneously, specifically on transfer questions. Most important for the current study, when using this staggered design, they found that self-generation led to better comprehension compared to using a readymade graphic organizer or no graphic organizer at all (Colliot & Jamet, 2020b).

However, despite these benefits of self-generation, subsequent research by Colliot et al. (2022) revealed a different pattern of results. When evaluating the impact of graphic organizers over a delay, they found that the readymade condition consistently outperformed both the self-generation and no graphic organizer conditions on an immediate test and after a 2-day delay, replicating their earlier findings (2018b, 2019, 2020a). This apparent contradiction raises questions about why the readymade graphic organizer might offer an advantage, particularly in delayed testing situations. One potential explanation is that the readymade groups consistently outperform the self-generation groups because participants are presented with fully accurate, instructor-generated information. Notably, in their studies, Colliot et al. have not evaluated the quality of the self-generated responses, rather they use it as a threshold for inclusion in their study (i.e., 80% correct completion required for results to be considered). This suggests that the accuracy and completeness of the self-generated organizers may be key factors in understanding why readymade organizers often lead to better performance. Thus, there are several questions that remain left unanswered. First, the testing phase in most of these studies occurred immediately after learning. According to Bjork and Bjork (2020), it is important to differentiate between performance, which can be assessed at any time, and actual learning, which must be assessed after a delay. Although Colliot et al. (2022) found no benefit of self-generation after a delay, they did not stagger the presentation of the lecture and graphic organizer, which they had shown to be beneficial to the self-generation group previously (Colliot & Jamet, 2020b). Second, there was no consideration of the role of causality in the learning process even though it may improve long-term consolidation above and beyond G-CL tasks. Third, there was little to no consideration for the role of individual differences (i.e., O-SPAN, engagement/interest, prior knowledge, difficulty judgements), which are known to play a role in learning and comprehension (Mayer, 2014; Radvansky & Copeland, 2004; Mayer, 2005).

Individual Differences

Prior Knowledge

Several factors influence how cognitive load impacts learning, and one important factor is prior knowledge. In fact, an interesting result that is often observed is called the *expertise reversal effect* (Kalyuga et al., 2003) in which the effectiveness of instructional strategies differs depending on one's relevant prior knowledge. Specifically, people with low prior knowledge need more scaffolding by having information presented at a lower, more fine-grained level for improved comprehension and performance (i.e., more cohesion between sentences/ideas; McNamara & Kintsch, 1996). Conversely, people with high prior knowledge already have well-formed schemas for the content. Thus, presenting information at a fine-grained level is below their level of understanding and can be detrimental to their comprehension (Kalyuga et al., 2003). Thus, for these individuals, E-CL is lowest, and learning is highest when new but related information is presented at a higher level (i.e., more general and not as incremental or with less cohesion; McNamara & Kintsch, 1996). In sum, what increases helpful G-CL for one student may increase harmful E-CL for another student because they have differing amounts of prior knowledge.

Working Memory & Prior Knowledge

According to Cognitive Load Theory, researchers and/or instructors should aim to reduce negative cognitive load to then reduce the demand on WM at any given time (i.e., decrease negative E-CL or increase positive G-CL). However, the demand on WM at any given time will vary for everyone based on their WM capacity. For instance, individuals with higher WM capacity should be able to work with more new information at a given time and/or organize information more efficiently while learning new material, compared to those with lower WM capacity. Generally, WM is a good predictor of comprehension, but it does not explain all the variance in long-term memory performance (Unsworth & Engle, 2007). There is an important interaction between WM and knowledge that is critically important to learning processes. For instance, consider two people with equal WM capacity who encode the same information but have varying levels of prior knowledge. We would expect the high knowledge person to perform above the low knowledge person because they can leverage their prior knowledge to help them organize and integrate the new information into their existing knowledge structures. Keep in mind though, that even this effect may be influenced by the expertise reversal effect that was discussed above (Kalyuga et al., 2003).

Information Presentation Control

Another factor that Mayer and colleagues (2019) found to influence performance during learning was how much information was presented at any point in an online learning environment. In this study, participants were shown a PowerPoint lecture with either small segments (e.g., individual bullet-points) or larger segments (e.g., information taking up the full PowerPoint slide) and were asked to report the difficulty experienced while learning the material. They found that information presented in smaller segments was rated as less difficult, perceived as easier to comprehend and better remembered on a later transfer test compared to information presented in larger segments. Importantly, they did not measure the role of prior knowledge, which could play a further role in all these outcomes. It is possible that low-knowledge individuals may benefit from small segments, whereas high-knowledge individuals may benefit from larger segments (i.e., the expertise reversal effect; Kalyuga et al., 2003). Further, prior work has not evaluated the effect of segment sizes when causal connections are present or absent, and how prior knowledge influences this effect.

Chapter 2 - Overview of the Current Study

The aim of the current study is to further understand how cognitive load impacts encoding and retrieval processes. Specifically, this study will evaluate whether students' memory for expository, educational materials can be improved when implementing certain pedagogical techniques. Importantly, this will be done using theoretical ideas from the cognitive psychology literature by bringing together cognitive load theory (e.g., redundancy vs. self-generation effects; Sweller, 1994), event comprehension theory (Zwaan et al., 1995; Zwaan & Radvansky, 1998; Pettijohn et al., 2016; Pettijohn & Radvansky, 2016; Radvansky, 2012; Radvansky & Zacks 2017), and theories involving various individual differences variables (e.g., WM, prior knowledge; e.g., Unsworth & Engle, 2007; Kalyuga et al., 2003). This goal of improving student comprehension has become increasingly important since the pandemic forced many instructors to pivot to online teaching formats, with little-to-no preparation. Specifically, the use of multimedia presentations has increased; however, more studies are needed to evaluate how the designs of such presentations can help organize information and engage students in a meaningful way. The current study will advance the current literature in four main areas.

First, the research on self-generation strategies using expository materials is mixed. Some research has shown that self-generation is beneficial to comprehension of expository materials (Katayama, 1997; Colliot & Jamet, 2020b; Katayama & Robinson, 2000), whereas other research has shown no benefit of self-generation (i.e., the redundancy effect; Mayer et al., 2005; Sweller 1994). Thus, it is important to identify whether G-CL (i.e., in the form of a self-generation graphic organizer) is useful for comprehension of expository materials.

Second, it is important to expand the current research by evaluating the effects of causal connections on comprehension, by identifying whether causal connections help students comprehend educational, expository materials. Specifically, do explicit causal connections between content ideas help students create effective mental models as they do within narrative texts (Zwann et al., 1995; Pettijohn et al., 2016; Pettijohn & Radvansky, 2016; Radvansky, 2012; Radvansky & Zacks 2017)? Third, it is important to identify the long-term effects of self-generation and causal connections. Previous research has shown no benefit of self-generation (i.e., ready-made outperformed self-generated graphic organizers; Colliot & colleagues; 2018a, 2018b, 2019, 2020, 2021, 2022), which may be due to the short retention intervals that were used. It is possible that such effects are dependent upon memory consolidation that takes place over time (e.g., Muller & Pilzecker, 1900) and thus could be observed after longer delays that are like those experienced by students in the real world. Moreover, making causal connections very explicit to learners may improve the benefits of self-generated organizers, through an increased ability to successfully encode material within the existing mental model information, especially over longer delays.

Fourth, it is expected that individual differences, such as WM, prior knowledge, motivation/engagement, and perceived difficulty, will moderate the effects of self-generation, explicit causal connections and testing delays on learning and comprehension.

Hypotheses

The hypotheses for the proposed study are split into several smaller sections of alternative competing hypotheses. Each set of hypotheses will be discussed below.

Cognitive Load Hypotheses. There are two alternative competing hypotheses related to the cognitive load manipulation (i.e., G-CL Self-Generation vs. no G-CL Completed). The **Generation effect hypothesis** (Stull & Mayer, 2007) states that by actively engaging with material through the graphic organizer, the generative process will increase G-CL and, thus, will increase comprehension. If this hypothesis is supported, it is predicted that the Self-Generation group will outperform the Readymade group on the content quiz. Alternatively, the **Cognitive Load hypothesis** (Colliot & colleagues, 2018a, 2018b, 2019, 2020, 2021, 2022) states that generating responses on a graphic organizer will increase E-CL while learning and completing the graphic organizer. If this hypothesis is supported, it is predicted that the Readymade group will outperform the Self-Generation group, or there will be no difference in comprehension between groups. In a set of exploratory analyses, the quality of the graphic organizer will also be tested. That is, the quality of the graphic organizer produced by the Self-Generation group should predict quiz performance (i.e., more accurate graphic organizers should be associated with better memory).

Causality Hypotheses. The **Explicit Benefit hypothesis** states that when explicit causal connections are available, comprehension will increase. Conversely, the **Implicit Sufficient hypothesis** states that there will be no benefit to having explicit causal connections because implicit causal connections are sufficient for comprehension.

Delay-related Hypotheses. Overall, I expect to find that comprehension performance will be higher for the Immediate vs. Delayed condition (Ebbinghaus, 1885; see Finkenbinder, 1913 for review; Heller et al., 1991; Murre & Dros, 2015).

Interaction Effect Predictions. I expect to find an interaction between G-CL and Causality. Because causality and G-CL are hypothesized to have independent, and potentially additive, benefits on memory, I expect participants in the Self-Generated Graphic Organizer/Explicit Causality group will have the best comprehension. Additionally, I expect to observe interactions between Delay and both G-CL and Causality. That is, the benefits of self-generation and explicit causal connections may get stronger as the delay between learning and assessment increases because they help increase mental model strength by increasing the consolidation of material being learned. Importantly, the effects of self-generation and explicit causal connections may not be apparent until after a delay, when the other groups may see less effective consolidation of material (i.e., no long-term storage, but successful short-term storage).

A significant Graphic Organizer x Delay interaction could explain why previous research (Colliot & colleagues, 2018b, 2019, 2020, 2021) found no self-generation benefit. That is, the previous work did not include delay intervals, but the benefits of self-generation may not appear until a meaningful delay. Alternatively, if the Graphic Organizer x Delay interaction is not significant, then it could suggest that self-generation on a graphic organizer has little to no benefit on memory for expository materials, replicating Colliot and Jamet (2022).

Individual Differences to consider. Research has shown there are differences in how cognitive load impacts individuals based on their own prior knowledge (Kalyuga et al., 2003), WM (Unsworth & Engle, 2007), motivation (Multu-Bayraktar et al., 2019), content difficulty ratings (see Brunken et al., 2010; Nourbakhsh et al.,2017), and engagement (Mayer, 2014). In general, higher levels of most of these variables (i.e., prior knowledge, WM, engagement, motivation) may decrease the effect of cognitive load on a participants' comprehension and knowledge of material, regardless of whether causal connections are implicitly or explicitly available. Additionally, there could be more complicated interaction effects with individual differences variables and the manipulated, causal connections, delay times, and/or cognitive load. I predict these effects would be like that of the expertise reversal effects (Kalyuga et al., 2003), in which less information is better for comprehension for those with higher levels of prior knowledge due to the difficulty of integrating the smaller units of independent content into their existing schemas. It is expected that these types of relationships could be seen at higher levels of prior knowledge, WM, engagement/motivation, or difficulty ratings.

Chapter 3 - Methods

Participants

A power analysis was conducted in G*Power (version 3.1.9.2) using Cohen's $d = 0.71$ for the between-subjects effect of Self-Generation vs. Readymade graphic organizer manipulation from Colliot and Jamet (2020), power = .80 and alpha = .05. No prior work has evaluated the interaction between G-CL and causality, so Colliot and Jamet's (2020) effect was the most appropriate for the current study's power analysis. The power analysis determined that 52 participants were needed to detect the between-subjects effect of self-generation versus readymade graphic organizers. The goal was to collect a final sample of 60 participants in each session to ensure that the study was sufficiently powered.

A total of 103 participants were collected; 50 were from the Prolific online recruiting platform and 53 were from Simpson College Psychology classes. Prolific participants were paid $9 for the first session (~60 minutes), $10 for their second session (~20 minutes), and $11 for their final session (~10 minutes). This pay structure resulted in a high rate of retention among the first 5 participants and, thus, it was used for the subsequent participants. A small grant ($1,000) was funded through Kansas State's Arts, Humanities, and Social Sciences Small Grant Program to cover the costs of participant compensation. Simpson College students from several courses were offered extra credit for participation, and the amount of extra credit increased over each session (e.g., 5, 7, 8 points of extra credit for sessions 1, 2 and 3, respectively). Demographic information for both the online and in-person samples is provided in Table 1. In session 2, there were 62 returning participants in total (Prolific = 23; Simpson = 39) and in session 3 there were 53 returning participants (Prolific = 22; Simpson = 31). Two participants were excluded from data for being more than 3SD above the mean for their comprehension scores.

During data analyses, participant location was considered when doing model comparisons. However, when included as a random effect in models, location never resulted in a better-fitting model. Because of this, participants will not be considered as being from two different locations throughout the rest of the analyses.

Table 3-1. Participant demographics

Data Collection Location	Age (Range)	Education	Gender	Time Point
Prolific	38.5 years (21 - 63 years)	High school = 6 Associates = 11 Bachelors = 20 Masters-level = 10 PhD = 1 Professional = 1 Technical = 1	Female = 28 Male = 22 Non-Binary = 0	Session 1 =50 Session 2 = 23 Session 3 = 22
Simpson	19 years (18-22 years)	High school = 49 Associates = 4	Female = 31 Male = 20 Non-Binary = 2	Session 1 =53 Session 2 = 39 Session 3 =31

Materials

PowerPoint lecture

The lecture material was presented via 1 of 2 pre-recorded videos. Both recorded lectures were approximately 9 minutes in length. They were on a set of PowerPoint slides that were translated from French to English and modified from those used in a set of studies used by Colliot and colleagues (2018a, 2018b, 2019, 2020, 2021). The original lecture consisted of an introduction slide laying out what the participant was expected to complete during the learning phase, seven content slides that contain the to-be-learned material including text and corresponding images, and a closing slide indicating they have completed the learning phase of the study.

Modifications were made to the original slides in several ways, largely all allowing for clearer presentation of information. First, edits were made so that information was presented in smaller segments (i.e., bullet points added, sentences shortened, etc.). Second, causal connections were added throughout the PowerPoint slides, these are discussed further below. Importantly, these changes did not alter the content in any way. Third, more examples were included to further elaborate on the information (e.g., examples of each component of working memory) and to allow for more opportunities to insert causal connections. The slides had to also be translated from French to English. To do this, the information was uploaded into Google Translate. Then, a graduate student from the French department at Kansas State University (KSU) back translated the English version into French. This student identified several mistranslations, none of these impacted the overall meaning of the content, and adjusted the English version accordingly (e.g., 'submit' was changed to 'next' to match context; 'Frontal Lobe' was used instead of 'Lobe Frontal' to match English word order).

Causality Manipulations

Participants viewed one of two PowerPoint presentations: Explicit or Implicit Causality. In the Explicit Causality Group, participants were presented with the explicit relationships between the content topics (e.g., using words such as "thus" or "therefore"; see Table 2 below). In the Implicit Causality Group, participants were presented with similar PowerPoint slides, but now with only implied causal manipulations between each of the topics (see Table 2 below). In total there were 38 causal connections throughout the slides, with an average of 4-5 per slide.

Additionally, temporal connections (e.g., then, after, now) were removed from both the Explicit and Implicit groups to avoid any additive benefits of having those connections present (Zwann et al., 1995).

Table 3-2. Examples of Explicit and Implicit Causality Manipulation.

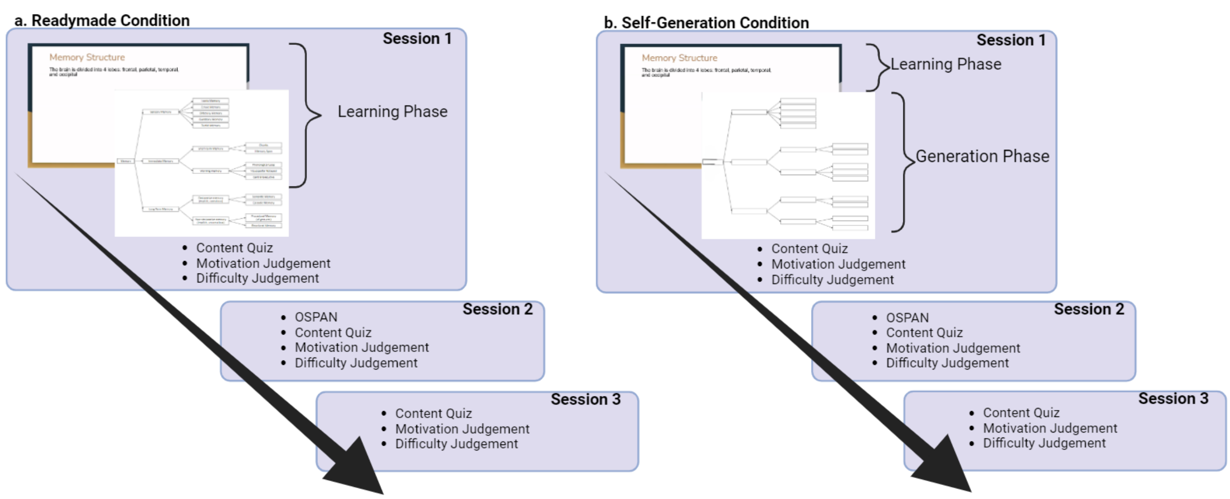
Group	Example from PowerPoint
Explicit Causality Group	*The brain is divided into 4 lobes: frontal, parietal, temporal, and occipital* *Each lobe is involved with memory in their own way, therefore there are different memory systems and subsystems* *Memory is the activity making it possible to:* *Encode, Store, and Restore information of various kinds (knowledge, know-how, etc.)* *Process information, thus creating an engram (memory trace)* *Memory is divided into 3 main types: Sensory, Immediate, & Long-term* *Therefore, we will discuss each type of memory on the following slides*
Implicit Causality Group	*The brain is divided into 4 lobes: frontal, parietal, temporal, and occipital* *Each lobe is involved in different memory systems and subsystems* *Memory is the activity making it possible to:* *Encode, Store, and Restore information of various kinds (knowledge, know-how, etc.)* *When information is processed, an engram (memory trace) is created* *Memory is divided into 3 main types: Sensory, Immediate, & Long-term* *We will discuss each type of memory on the following slides*

Note: *words in red are the causal manipulation changes and are only depicted in red for visualization purposes here. There are additional formatting changes that were made to the content to increase overall organization and structure of information being presented.*

Graphic Organizer

To manipulate G-CL, participants were assigned to one of two possible generation groups: Self-Generation (i.e., G-CL) or the Readymade (i.e., control) Group. These groups are used to largely replicate the methods used in Colliot and Jamet (2020, 2021). Immediately following the lecture, participants in the Self-Generation group completed a graphic organizer that visually depicted the relationships between the content (i.e., the generation phase). After the lecture, participants in the Readymade group were provided with a complete graphic organizer during their learning phase. See Figure 1 for a visual representation of the two groups.

Figure 3-1. Schematic for the (a) Readymade and (b) Self-Generation groups, across each of the 3 testing sessions.



Comprehension Measures (Outcome Measures)

The primary outcome variable of interest was comprehension accuracy for the PowerPoint content. The accuracy of the self-generated Graphic Organizers will also be treated as an outcome measure, but it is an exploratory variable of interest.

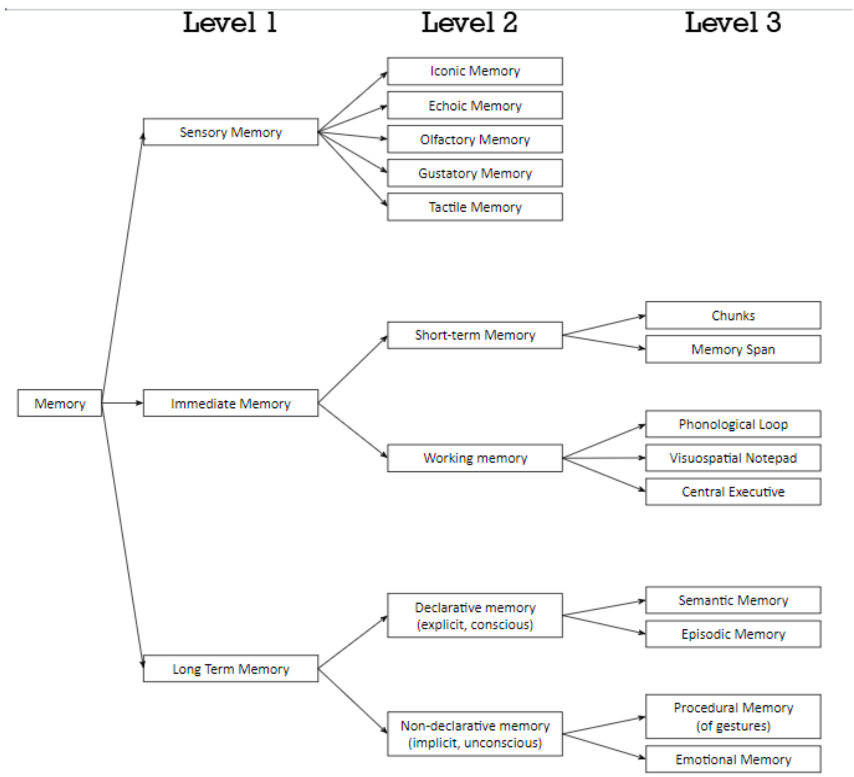
Content Quiz. Across 3 testing sessions, participants completed a quiz that assessed their overall learning of the PowerPoint content. This quiz was adapted from Colliot and Jamet (2018), which originally consisted of 45 transfer and retention questions, as well as a mix of question types (i.e., short answer, multiple choice). The final set of questions consisted of 45 questions (i.e., 24 multiple choice and 21 single-phrase short answer responses; see Table 3 for question type breakdown). These questions were either text-based (i.e., information was stated directly in the PPT) or situation-based (i.e., representation of the gist of what was presented, not the text itself; Radvansky & Copeland, 2004) and this was done to test multiple levels of representation that are important for comprehension. Questions were randomly split into three sets (i.e., 15 questions per set; groupings are depicted in Appendix A), for each of the three testing phases. Questions were then counterbalanced across participants, ensuring that different sets of questions were presented at different delays. All questions were scored as either correct or incorrect, and then a proportion correct was calculated separately for immediate (Session 1), mid-delay (Session 2) and long-delay (Session 3) testing conditions.

Table 3-3. The number of questions created from each major content area within the lecture, separated by the type of question and the level of representation.

	Text		Situation	
	Short Answer	Multiple Choice	Short Answer	Multiple Choice
Sensory Memory	2	3	2	2
Short-term/Working Memory	2	4	7	5
Long-term Memory	4	4	7	5
Memory Structure	1	2	0	0
Totals	9	13	16	12

Graphic Organizer. The relationships depicted in the Graphic Organizers were graded for accuracy. Points were earned through hierarchical relationships, such as understanding that “memory” encompasses other terms like “sensory” and “long-term” memory, and for the total number of relationships correctly depicted at each level, such as understanding that “long-term memory” and “sensory memory” are both on the same hierarchical level but are not equal to “chunking” in the hierarchy. The Graphic Organizer consisted of 21 labels: three at level 1, nine at level 2, and nine at level 3. Level 1 represents the top of the hierarchy (e.g., "long-term memory"), level 2 represents the subcategories of each memory type (e.g., "declarative memory"), and level 3 represents the subcategories of the level 2 terms (e.g., "semantic memory" and "episodic memory"). These were scored as a total proportion correct, either correct (1) or incorrect (0), for each participant in the self-generated group and then averaged for a single measure of accuracy on the graphic organizer. See Figure 2 for review of the tiers being discussed.

Figure 3-2. Completed Graphic Organizer that was shown to the Readymade group. The same hierarchical structure was shown to the Self-generation group, but the words were missing, and they were tasked with filling in the information.



Individual Differences

Participants were assessed on a range of individual difference variables described in detail below.

Working Memory. Participants completed the Operation Span (O-SPAN), which is a measure of working memory (WM; Turner & Engle, 1989). In this task, participants were presented with a series of math equations and words. First, the participants saw a math equation (e.g., "Does (3 + 2) x 3 = 14?") and their task was to evaluate whether the solution to the math equation is accurate. They had 4 seconds to respond ("yes, it is correct" or "no, it is not correct"). Afterwards, a word (e.g., DOG) was presented for 1 second, and their task was to remember it for a later memory test. After the word, a new math equation was presented, followed by another word, and so on until a prompt appeared that asked the participants to recall the words. The number of to-be-remembered words per set varied from 3 to 7 words.

Math scores were calculated as either correct (1) or incorrect (0). The proportion of correct math scores served as a single measure, typically used to determine a threshold for sufficient accuracy (e.g., 80%; see table 4). Math scores were calculated to ensure that participants were paying attention to this part of the task rather than only attending to memorizing the to-be-remembered letters. If participants' math scores were lower than 85%, then their data were removed from further analyses. Letters scores were assessed similarly, scored as either correct (1) or incorrect (0). Recalled letters were not considered correct if left unanswered or if they were out of order, reflecting a stricter coding method (see Table 4).

To foreshadow, performance on the math task was extremely low with 80% of participants scoring below the minimum cutoff of 85% correct. Therefore, this measure will not be assessed in the analyses since I cannot ensure that participants were 'mentally juggling' the two tasks, I have no way of saying that what I measured was working memory capacity. It is very possible that participants focused on only remembering the letter, thus inflating their working memory scores and skewing the results and the role that working memory could have on my data.

Table 3-4. Average OSPAN performance on both the Math and Letter portions.

Measure	M(SD)	Range
Math	.61 (.25)	0 - 0.92
Letters	.71 (.21)	0.05 - 0.99

Motivation and Cognitive Load Scales. Motivation is an often-overlooked measure when studying concepts related to comprehension, especially in the realm of education (Mayer, 2014). However, past work has shown that WM capacity is temporarily increased when motivation is high (Multu-Bayraktar et al., 2019). Engagement has been incorporated into the Cognitive Affective Theory of Multimedia Learning developed by Moreno (2007), which extends Mayer's CTML (Mayer & Moreno, 2005; Mayer, 2005) and Sweller's cognitive load theory (1994). This established measure was originally developed by Moreno (2007) and contained 2 subscales: the Affective Learning Scale and the Indirect measure of Cognitive Load (see Appendix C).

Motivation Scale. The original Affective Learning Scale (i.e., referred to as the motivation scale in the current manuscript) consisted of 6 questions. However, only 3 of the questions that assess general motivation for learning and engagement with the content were used in the current study. These questions were answered on a 5-point Likert-type scale, where higher scores meant higher levels of motivation. Participants provided these ratings on each of the 3 sessions, so a maximum of 3 sessions per person was collected.

Self-Reported G-CL Scale. Participants rated the subjective difficulty of the tasks on a scale from 1 (*extremely low mental effort*) to 5 (*extremely high mental effort*). This rating served as a subjective measure of G-CL and was used to evaluate whether we replicate previous work (Pass & Van Merriënboer, 1992; Pass et al., 2003; Zu et al., 2020) that has found participants report G-CL tasks as more difficult. In addition to this single question, the original Indirect Measure of Cognitive Load scale (i.e., referred to as the Self-Reported G-CL scale in this manuscript) included 2 questions to assess task difficulty broadly. A question was added to this scale to assess the amount of effort that participants put into the tasks they were completing, since that is an important feature for cognitive load. These 3 questions (i.e., 2 original questions and the 1 additional that was added for this study) were answered on a 5-point Likert-type scale, where higher scores meant more G-CL being experienced. Participants provided these ratings on each of the 3 sessions, so a maximum of 3 sessions per person was collected.

Prior Knowledge Assessment. Participants completed the prior knowledge assessment previously used by Colliot and Jamet (2019), which included questions that tap into the content being covered in the experiment. There were 5 short-answer questions, each scored as either correct or incorrect (see Appendix B), and then a proportion of correctly answered questions was calculated. Ideally, these questions tapped into the quality and quantity of participants' existing topic-based knowledge as defined by McCarthy and McNamara (2021).

Design and Procedure

This study used a 2 (Graphic Organizer: Readymade vs. Self-Generated) x 2 (Causality: Explicit vs. Implicit) x 3 (Delay: Immediate vs. Mid-delay vs. Long-delay) mixed experimental design. The Graphic Organizer and Causality manipulations are between-subjects variables, and the Delay manipulation is a within-subjects variable. The Graphic organizer manipulation served as a replication of previous work (Colliot & colleagues, 2018a, 2018b, 2019, 2020, 2021, 2022), whereas the Causality (Zwann et al., 1995; Radvansky & Copeland, 2004) and Delay manipulations are extensions of previous work (Colliot & Jamet, 2022).

The study was programmed in PsychoPy, hosted on Pavlovia, and the data was collected using the Prolific online data collection platform and at Simpson College. During Session 1, participants first read and signed an informed consent, approved by the KSU IRB. They then completed the prior knowledge assessment to gauge their knowledge of the target content. Participants were pseudo-randomly assigned to one of four groups (Readymade Graphic

Organizer/Explicit Causality; Readymade Graphic Organizer/Implicit Causality; Self-Generated Graphic Organizer/Explicit Causality; Self-Generated Graphic Organizer/Implicit Causality). In the *learning phase* of session 1, all participants were shown a video of the PowerPoint slides and asked to learn the information for later memory tests. If a participant was in the Readymade group, they were shown the PowerPoint slides, followed by a complete Graphic Organizer to study. If a participant was in the Self-Generate group, they completed a graphic organizer after the learning phase (see Figure 1). After participants in either group completed the learning phase, they were presented with motivation questions, and difficulty judgments, in that order.

Then, participants completed the *immediate testing phase*, which consisted of 15 comprehension quiz questions. Finally, they were thanked for their participation and reminded of their second test session. Participants received a reminder email about their Session 2 appointment, which occurred three days after Session 1. In Session 2, participants completed an additional 15 comprehension questions, a motivation assessment, and OSPAN. Participants received another email reminder for their Session 3 appointment, which occurred 1 week after Session 1. In Session 3, participants completed the remaining 15 comprehension quiz questions and the motivation assessment. Finally, they were thanked for their participation, debriefed and paid for their participation (i.e., Prolific participants were paid, Simpson students received extra credit).

Data Preparation

Motivation Scales

The Motivation Scale was created by averaging the three questions used to assess motivation, engagement, and interest levels in the content (Moreno, 2007; see Appendix C). Similarly, the Self-Reported G-CL Scale was created in the same way and used to assess the cognitive load experienced by the participants. See Tables 5 and 6 for averages of the Motivation Scale and Self-Reported G-CL Scale reports, broken up by condition and session.

Table 3-5. Motivation Assessment averages broken up by session

	Session	M (SD)
Overall	overall	3.23 (1.1)
	1	3.30 (1.10)
	2	3.06 (1.05)
	3	3.30 (1.11)
GO	overall	3.26 (1.1)
	1	3.34 (1.17)
	2	2.99 (1)
	3	3.43 (1.06)
Readymade	overall	3.21 (1.08)
	1	3.28 (1.06)
	2	3.12 (1.09)
	3	3.19 (1.15)

Table 3-6. Self-Reported G-CL Scale

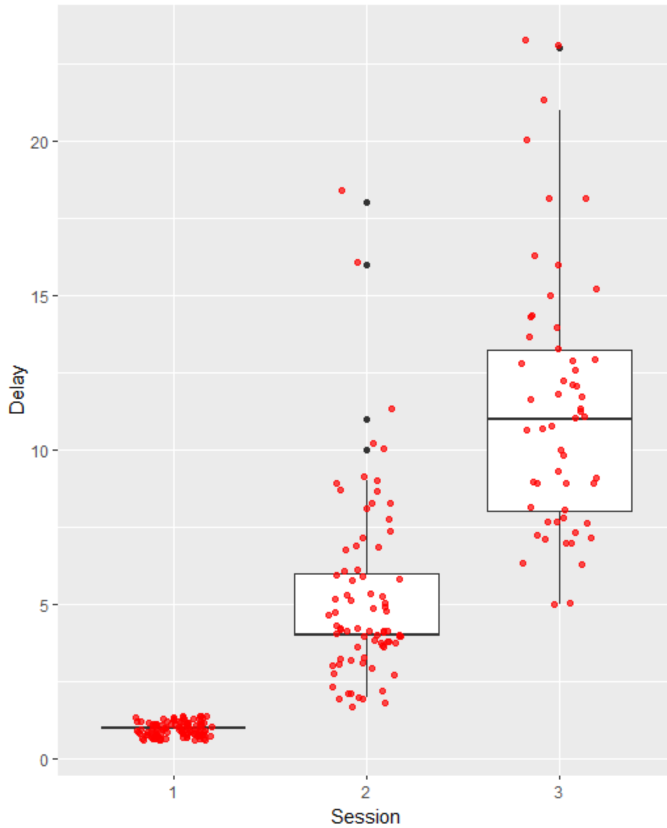
	Session	M(SD)
Overall	overall	3.74 (.78)
	1	3.83 (.78)
	2	3.62 (.79)
	3	3.69 (.74)

GO	overall	3.94 (.70)
	1	3.99 (.71)
	2	3.83 (.72)
	3	3.98 (.72)
Readymade	overall	3.58 (.79)
	1	3.71 (.81)
	2	3.48 (.81)
	3	3.46 (.72)

Delay

Naturally, this variable is a categorical variable: session 1, session 2, and session 3. However, this approach does not account for the fact that the time elapsed between each of the sessions varied by participant (see Figure 3). Specifically, all participants completed Session 1 on Day 1 (day of initial learning phase), but the delays varied for Sessions 2 (M = 5.29, SD = 2.94; range = 3-18 days) and Session 3 (M = 11.5, SD = 4.27; range = 6-23 days). To address this issue Delay was also treated as a continuous variable (e.g., Day 1-23). However, this variable does not include the specific session number (i.e., 1, 2, or 3) that participants were completing. Because each variable provides different information, they will both be investigated in the analyses below.

Figure 3-3. Each dot equals an individual participant's time that it took to participate in each session of the study. Session 2 was meant to be completed 4 days after completing Session 1 and Session 3 was meant to be completed 7 days after Session 1 was completed.



Chapter 4 - Results

First, I will evaluate whether comprehension is affected by the use of graphic organizers (G-CL), causal connectives, delays, and their interactions. Then, I will evaluate the accuracy of the participants' self-generated graphic organizers as well as whether that influenced final comprehension performance. Finally, I will conduct several exploratory analyses that evaluate whether various individual difference variables—such as motivation, prior knowledge, subjective cognitive load—moderate the relationships between G-CL, causality, and delay on comprehension.

Analyses were conducted using the R statistical software (version 3.1.1) or Jamovi statistical software (version 2.2). In R, I used the lme4 (Bates et al., 2014) library to run my mixed models, the Plyr (Wickham, 2011) and lattice (Sarkar, 2008) libraries were run to visualize data, the emmeans (Lenth, 2018) library to plot least squares predicted means from the model fit, the afex (Singmann et al., 2015) library to get parameter specific *p* values from the models, and the ggplot2 (Wickham, 2016) library for figure creation.

Generalized-multilevel models (GMLM) were used to analyze some data, as a way to treat the dependent variables appropriately for the non-normal distributions that they had, to better model repeated assessments and to utilize other measures (e.g., link functions). The square root link function from the gamma family was used for all model comparisons to handle the positive skew continuous outcome variables. Then random effect and fixed effect structures were compared to determine the best fitting model based on the lowest AIC and BIC (Schwarz, 1978) values. To streamline results, the best model in each case will be discussed in the main body of this document (see Appendix D and Appendix F for the model comparisons).

Comprehension Performance Scoring

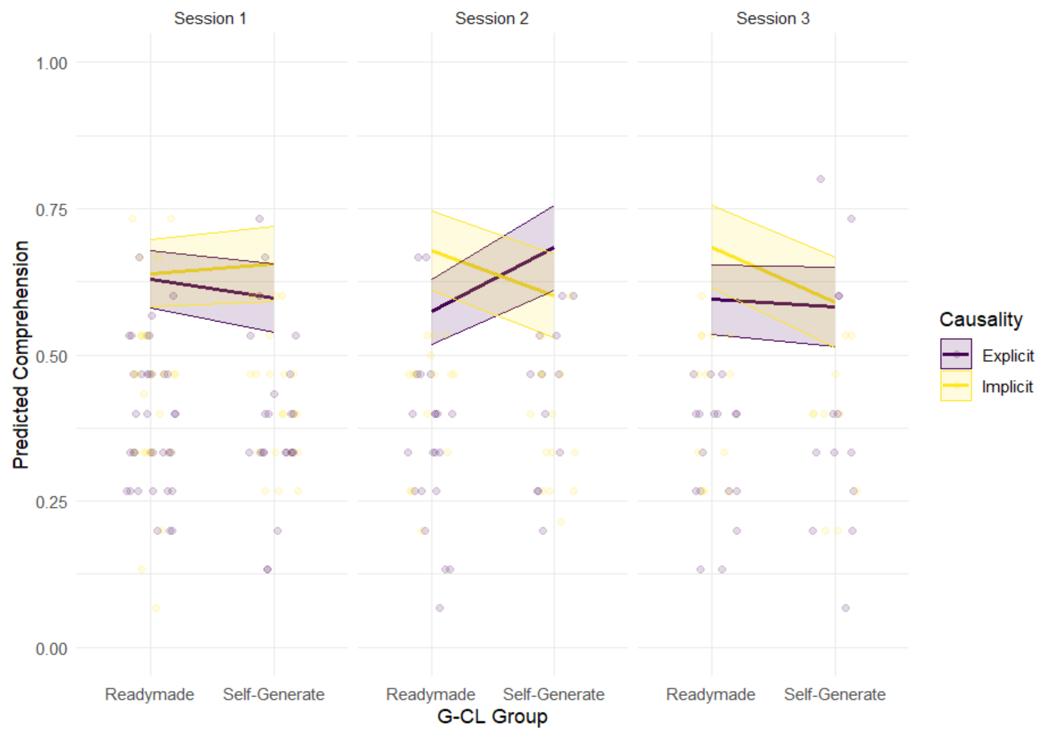
To evaluate final comprehension, performance on the multiple choice and short answer questions were scored. Multiple choice responses were each scored as correct or incorrect. Short answer responses were scored as correct or incorrect. To assess accuracy, two independent raters scored the responses from five participants, resulting in strong inter-rater reliability (Kappa = .816). Discrepancies were discussed and one rater completed the scoring for the remaining participants. Average performance for both question types (short answer, multiple choice) and both levels of representation (text-based, situation based) at each delay are reported in Table 7.

Table 4-1. Average comprehension scores for each type of question and each level of representation, at each time point.

		Average Correct (SD)			
Question Type	Questions level	Overall	Session 1	Session 2	Session 3
Overall Average		.40 (.49)	.41 (.49)	.39 (.49)	.38 (.49)
Multiple Choice	Situation-based	.45 (.49)	.45 (.49)	.47 (.50)	.42 (.49)
	Text-based	.37 (.48)	.39 (.49)	.35 (.48)	.34 (.48)
Short Answer	Situation-based	.35 (.48)	.35 (.48)	.34 (.47)	.37 (.48)
	Text-based	.42 (.49)	.45 (.50)	.40 (.49)	.38 (.49)

To evaluate whether G-CL (via graphic organizers), causal connectives, and delay affect comprehension performance, the best fit model was a GMLM with a square root link function predicting comprehension accuracy from the fixed effects of Causality group, Graphic Organizer group, delay and their interactions with the intercept of subject as a random effect and the Motivation Assessment as a covariate. Predicted comprehension performance based on this model is plotted in Figure 4 by G-CL, Causality and Delay. The main effect of G-CL was not significant, β = -0.031, t = -0.808, p = .418, supporting the *Cognitive Load hypothesis*. That is, overall comprehension performance did not differ between the Readymade group (M = 0.63, SD = 0.02) and the Self-Generate group (M = 0.62, SD = 0.02). Further, the main effect of Causality was not significant, β = 0.011, t = 0.288, p =.773: the Explicit (M = 0.61, SD = 0.02) and Implicit (M = 0.64, SD = 0.02) groups performed similarly, supporting the Implicit Sufficient hypothesis. The main effect of Delay was significant, (p = .035). Specifically, comprehension performance was significant for session 2, β = -0.055, t = -2.569, p = .01, but not session 3, β = -0.034, t = -1.449, p = .147, when compared to session 1. However, the pairwise comparisons showed no significant differences between each of the three comparisons (Session 1 - Session 2, p = .956; Session 2 - Session 3, p = .314; Session 1 - Session 3, p =.438). This may be because 1) the pairwise comparisons are more conservative in nature compared to the main model run or 2) the role that potential higher-level interactions are having on the data. Finally, the motivation scale also had a significant effect on comprehension, β = 0.021, t = 2.108, p = .035, such that the more motivated the participant reported being, the higher their comprehension performance was (see Appendix D for full model output).

Figure 4-1. Results of G-CL Condition X Causality X Delay on Comprehension



Note: The error bars in this figure are 95% Confidence Intervals. The data shown are predicted results from the model and the dots are raw observations.

The main effects are qualified by several significant interactions. First, the G-CL x Delay interaction was significant (p <.001), with significant effects of G-CL at Session 2, $\beta = 0.14$, $t = 3.74$, $p < .001$, but not at Session 3, $\beta = 0.019$, $t = 0.51$, $p = .609$, when compared to session 1. Surprisingly, pairwise comparisons indicated that the Readymade group ($M = 0.64$, $SD = 0.02$) significantly outperformed the Self-Generation group ($M = 0.59$, $SD = 0.026$) at Session 3 when compared to Session 2 ($p = .035$). Comprehension performance between these two groups did not differ between the other delays. Additionally, the Causality x Delay interaction was significant ($p < .001$), with significant effects at Session 2, $\beta = 0.094$, $t = 2.729$, $p < .001$, and Session 3, $\beta = 0.079$, $t = 2.123$, $p = .034$ when compared to Session 1. However, post-hoc analyses showed no significant differences of Causality at Session 1 (Explicit $M = 0.61$; $SD = 0.02$; Implicit $M = 0.65$; $SD = 0.02$; $p = .231$), Session 2 (Explicit $M = 0.63$; $SD = 0.02$; Implicit $M = 0.64$; $SD = 0.03$; $p = .738$) or Session 3 (Explicit $M = 0.59$; $SD = 0.02$; Implicit $M = 0.64$; $SD = 0..3$; $p = .169$). The G-CL x Causality, $\beta = 0.047$, $t = 0.798$, $p = .425$ was not significant.

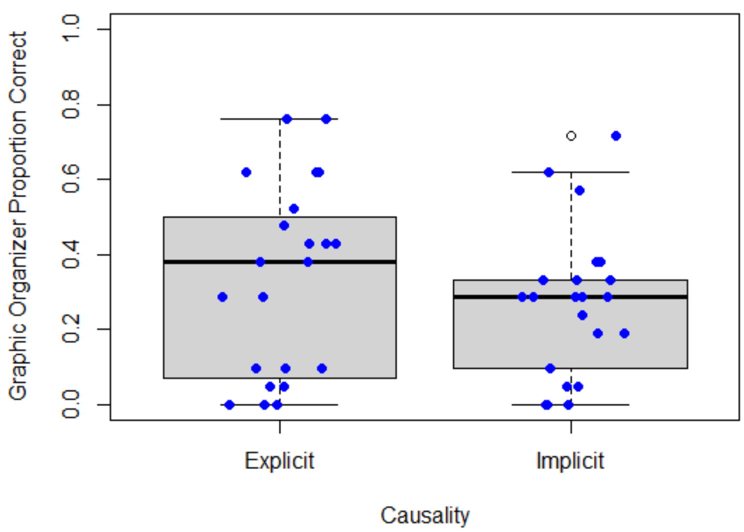
Importantly, the 3-way interaction between G-CL x Causality x Delay was significant (p < .001). The model showed that, when compared to the initial testing at Session 1, there was a significant G-CL x Causality interaction at Session 2, $\beta = -0.235$, $t = -4.332$, $p < .0001$, and at Session 3, $\beta = -0.13$, $t = -2.295$, $p = .022$. To unpack the 3-way interaction, post hoc analyses were conducted and showed that, when learners were given Explicit causal connections, those who Self-Generated their graphic organizers ($M = 0.68$, $SD = 0.04$) performed significantly better than those who were provided Readymade graphic organizers ($M = 0.57$, $SD = 0.03$) during Session 2 ($p = .019$). Interestingly, there is a trend that the Readymade group ($M = 0.69$, $SD = 0.04$) outperformed the Self-generate group ($M = 0.59$, $SD = 0.04$) during Session 3, especially for those who received implicit causal connectives ($p = .077$).

Graphic Organizer Accuracy

I expected to find that self-generating information on the graphic organizer would be beneficial for comprehension because G-CL is increased by requiring the learner to engage in effortful retrieval practice of the just-learned information (i.e., a main effect of G-CL). Such effortful processing has been shown to improve learning and comprehension (Stull & Mayer, 2007). However, in the analysis above, no benefit for self-generation was observed. However, one potential explanation for this result is that the Readymade group was able to encode accurate information twice (i.e., during the lecture and then again on the provided graphic organizer), whereas some learners in the Self-Generate may have retrieved the incorrect information.

To evaluate this explanation, I analyzed the accuracy of the Self-Generated graphic organizers. Accuracy was calculated as a proportion of correctly completed labels (out of 21). The average performance of the graphic organizer was quite low (overall $M = 0.29$, $SD = 0.23$). I also tested whether graphic organizer accuracy differed depending on whether learners received explicit or implicit causal connections. To do so, I conducted an independent samples t-test and found no significant effect of Causality, $t(43) = 0.763$, $p = .45$, such that the Explicit group generated 32% ($SD = 0.26$) correct labels and the Implicit group generated 29% ($SD = 0.2$) correct labels. Figure 5 shows graphic organizer accuracy separately for the explicit and implicit causality groups.

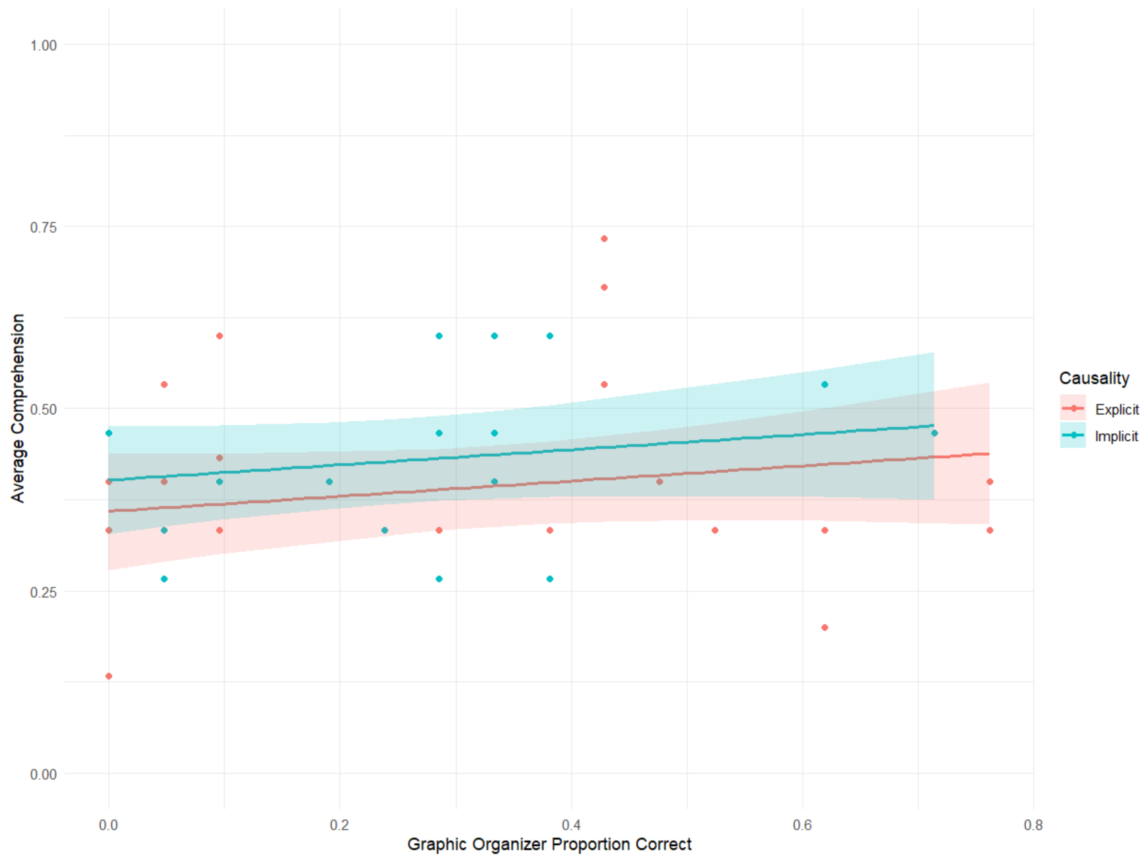
Figure 4-2. Proportion correct on Graphic Organizers, split by Causality condition



Evaluating Graphic Organizer Accuracy as a Predictor of Overall Comprehension Accuracy

Because graphic organizer accuracy was so low and varied quite a bit between learners in the Self-Generate group (range = 0-76% correct), it is possible that this low “immediate test” performance may explain why there was no G-CL benefit on comprehension (i.e., the Self-Generate group did not outperform the Readymade group). To evaluate this, I ran a general linear model looking at the role of causality and delay, their interaction, and Graphic Organizer accuracy to predict comprehension performance. This analysis only included those participants that were in the Self-Generate group (n = 44). Graphic Organizer accuracy was a significant predictor, $t(83) = 2.14$, $p = .035$, of comprehension performance (see Figure 6). There were no other significant effects in this model (see Appendix E for model output). That is, during initial learning, as assessed by graphic organizer accuracy, was a significant predictor of comprehension performance, it did not change the (lack of) effects of Causality, Delay and their interaction.

Figure 4-3. Figure showing the effect of Causality condition and GO accuracy to predict Comprehension



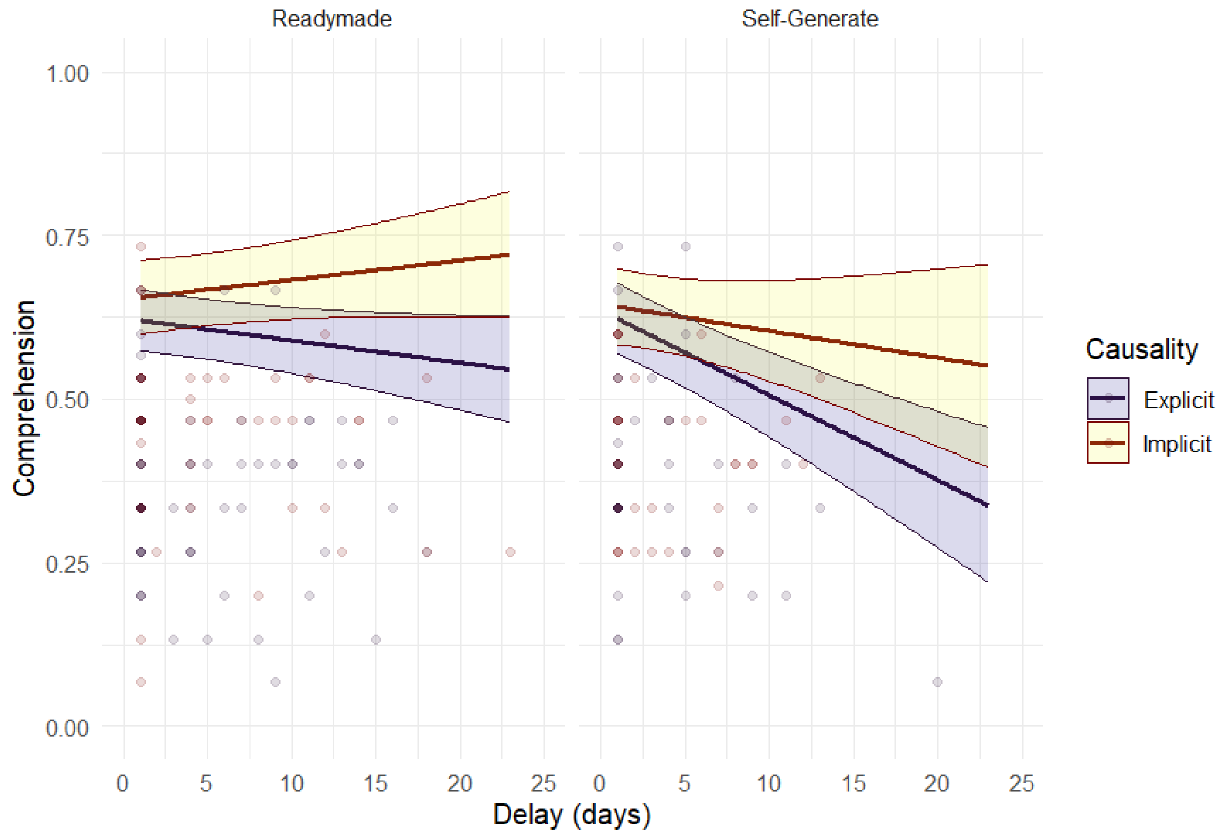
Continuous Time-based Model

To account for the wide range of delays in Sessions 2 (3-18 days) and Session 3 (6-23 days), I next treated delays as a continuous variable (i.e., instead of a categorical variable as in the previous model). This model evaluated whether G-CL, causal connectives, and delay (i.e., continuous time variable in this model) affect comprehension performance. To do so, the best fit model was a GMLM with a square root link function predicting comprehension accuracy from the fixed effects and full-factorial interactions for Causality group, Graphic Organizer group and delay, the intercept of subject as a random effect, and Self-reported G-CL scale as a covariate effect.

Comprehension performance based on G-CL, Causality and continuous Delay is plotted in Figure 7. Results from this model differed drastically from the original model (i.e., with Delay as a categorical variable). The one effect that was consistent across the analyses was that the main effect of G-CL was not significant, $\beta = 0.012$, $t = 0.34$, $p = .738$. That is, overall comprehension performance did not differ between the Readymade group ($M = 0.64$, $SD = 0.02$) and the Self-Generate group ($M = 0.61$, $SD = 0.02$), showing support for the *cognitive load hypothesis*. However, the remaining effects differed across models. Unlike in the original model, the main effect of Delay was trending towards significant, $\beta = -0.003$, $t = -1.91$, $p = .056$, but the main effect of Causality was not significant, $\beta = 0.029$, $t = 0.75$, $p = .451$, with the Explicit group ($M = 0.59$, $SD = 0.02$) having lower comprehension scores than the Implicit group ($M = 0.65$, $SD = 0.02$), suggesting support for the *implicit sufficient hypothesis*. Further, Self-report G-CL had no effect on comprehension, $\beta = -0.009$, $t = -0.819$, $p = .41$.

These main effects are qualified by several significant 2-way interactions. Specifically, the Causality x Delay, $\beta = 0.006$, $t = 2.266$, $p = .023$ and G-CL x Delay, $\beta = -0.01$, $t = -2.95$, $p = .003$, were both significant predictors in this model. However, the G-CL x Causality, $\beta = -0.019$, $t = 0.35$, $p = .728$, and surprisingly, the 3-way interaction, $\beta = 0.002$, $t = 0.46$, $p = .644$, were not significant (see Appendix F for full model output).

Figure 4-4. Figure showing effects of G-CL condition, causality, and delay (continuous) to predict comprehension performance



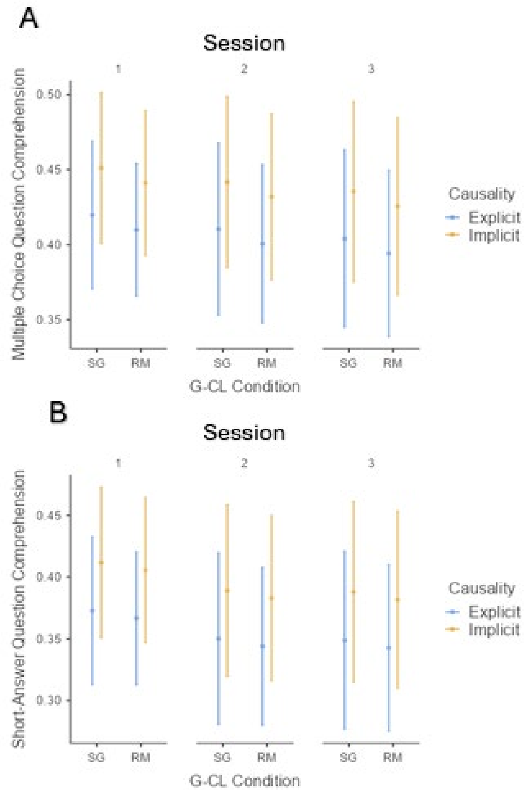
Chapter 5 - Exploratory Analyses

Exploring whether Question Type and Levels of Representation affected Comprehension

Type of Question

Because different question types elicit different retrieval processes—multiple choice involves recognition or identification of the correct response, whereas short answer involves recall of the information—the next analysis evaluated whether the influence of G-CL, causality and delay affected comprehension differently on multiple choice and short answer questions. Two 2 (Graphic Organizer: Readymade vs. Self-Generated) x 2 (Causality: Explicit vs. Implicit) x 3 (Delay: Session 1, Session 2, and Session 3) mixed ANOVAs were run separately for the multiple choice and short answer questions. For the Multiple Choice questions, the only significant effect was the G-CL x Causality interaction, $F(1, 11) = 5.56$, $p = .019$ (see appendix G). Specifically, a Tukey post-hoc comparison showed a significant difference between the Readymade Explicit group ($M = 0.38$, $SD = 0.17$) and the Readymade Implicit group ($M = 0.46$, $SD = 0.19$; $t = -2.54$, $p = .05$; see Figure 8). I ran a parallel analysis on the short-answer questions; however, there were no significant effects.

Figure 5-1. Average [A] Multiple Choice Comprehension Performance and [B] Short-Answer Comprehension Performance, both broken up by G-CL Condition, Session, and Causality



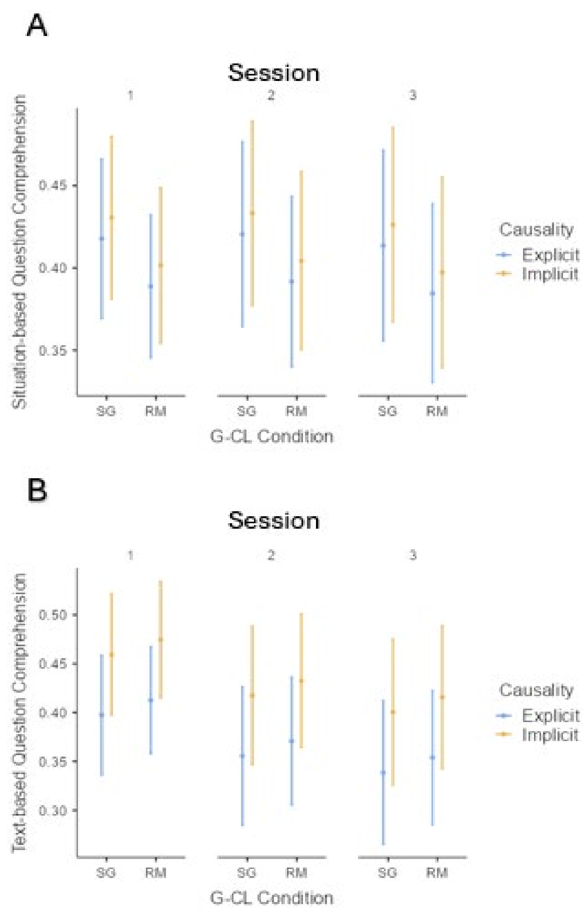
Note: The error bars here represent 95% Confidence Intervals.

Levels of Representation

The next set of analyses evaluated whether G-CL, causality and delay influence comprehension differently based on the level of representation of the information in the questions. Specifically, questions tapping into the text base and those tapping into the situation model were evaluated separately. Two 2 (Graphic Organizer: Readymade vs. Self-Generated) x 2 (Causality: Explicit vs. Implicit) x 3 (Delay: Session 1, Session 2, and Session 3) mixed ANOVAs were run separately for the situation-based and text-based questions. For text-based comprehension questions, there was a marginal effect of Causality, $F(1, 11) = 3.185$, $p = .076$ (see Figure 9), with those who received Explicit causal connections ($M = 0.38$, $SD = 0.24$) having lower comprehension compared to those with Implicit causal connections ($M = 0.44$, *SD*

= 0.2*).* Additionally, there was a significant G-CL x Causality interaction, $F(1, 11) = 4.837$, $p = .029$ (see appendix G). A Tukey post-hoc analysis showed a significant difference between the Readymade Explicit group ($M = 0.37$, $SD = 0.22$) performing worse than the Readymade Implicit group ($M = 0.48$, $SD = 0.20$; $t = -3.039$, $p = .016$). No significant effects were found in a parallel analysis on the situation-based questions.

Figure 5-2. Average [A] Situation-based Comprehension Performance and [B] Text-based Comprehension Performance, both broken up by G-CL Condition, Session, and Causality



Note: The error bars here represent 95% Confidence Intervals.

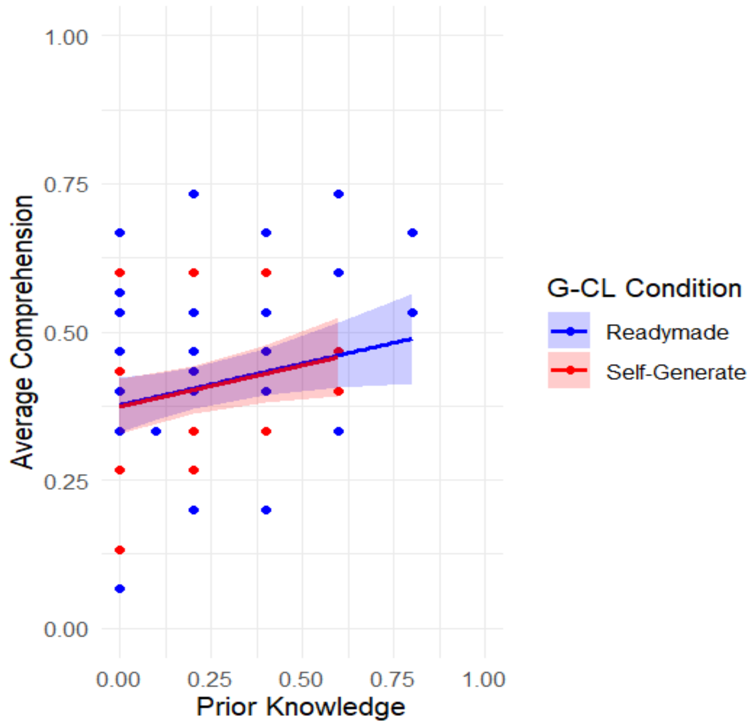
Do Individual Differences Explain the Effect of G-CL on Comprehension?

Three separate general linear models were run to identify whether the interaction of individual difference variables (i.e., prior knowledge, motivation, and self-reported G-CL) and G-CL had an impact on comprehension performance. All regressions included the G-CL condition manipulation (Readymade vs. Self-Generate), an individual difference variable, and the interaction between the two. Below, the output from each regression is discussed separately for each individual difference variable.

Prior Knowledge

Prior Knowledge is a critical variable to consider when looking at the impact of G-CL condition on comprehension performance. Specifically, low prior knowledge people typically need more scaffolding of information compared to high prior knowledge people who already have schemas for the content and do not need the same fine-grained scaffolding (Kalyuga et al., 2002; i.e., expertise reversal effects). This makes it necessary to ask if prior knowledge had an impact on comprehension in this study as well. First, there was a main effect of Prior Knowledge, $t = 0.310, p = .757$, on comprehension performance, such that higher levels of prior knowledge were associated with higher levels of comprehension performance (see Figure 10). However, the main effect of G-CL condition, $t = 2.115, p = .037$, and the Prior Knowledge x G-CL condition interaction, $t = -0.560, p = .577$, were not significant (see appendix H).

Figure 5-3. Effects of G-CL condition and Prior Knowledge on comprehension performance

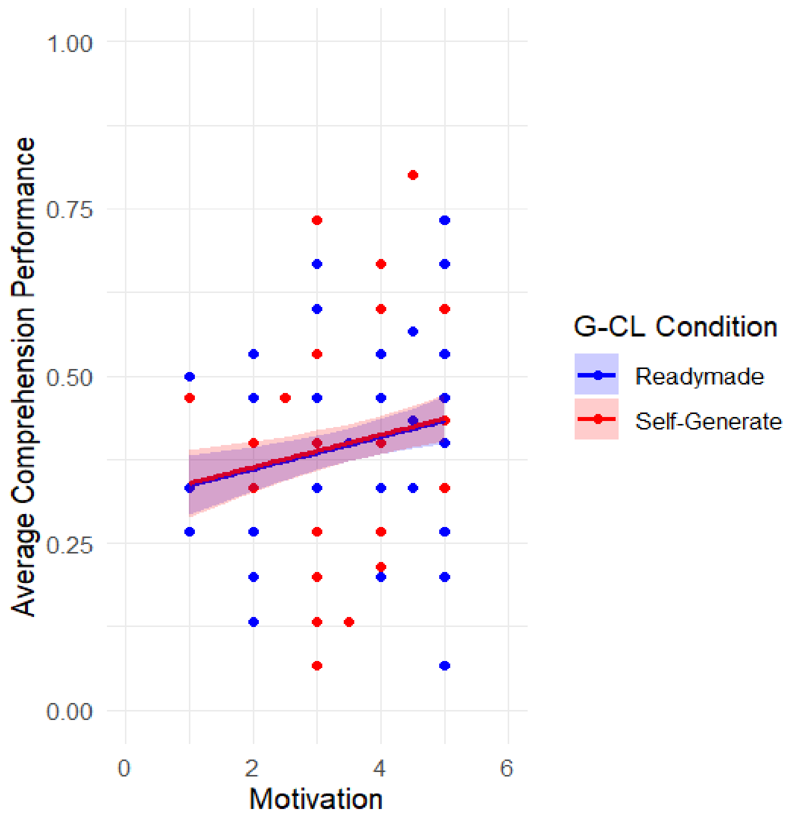


Note: The error bars in this figure are 95% Confidence Intervals.

Motivation Assessment

Next, I evaluated the influence of motivation, G-CL condition and their interaction on comprehension. The main effect of Motivation on comprehension was significant, $t = 2.298$, $p = .023$, such that higher levels of reported motivation were associated with better comprehension performance (see Figure 11). However, the main effect of G-CL, $t = -0.097$, $p = .923$, and the Motivation x G-CL condition interaction were not significant, $t = 0.158$, $p = .874$ (see appendix H).

Figure 5-4. Effects of G-CL condition and motivation on comprehension performance

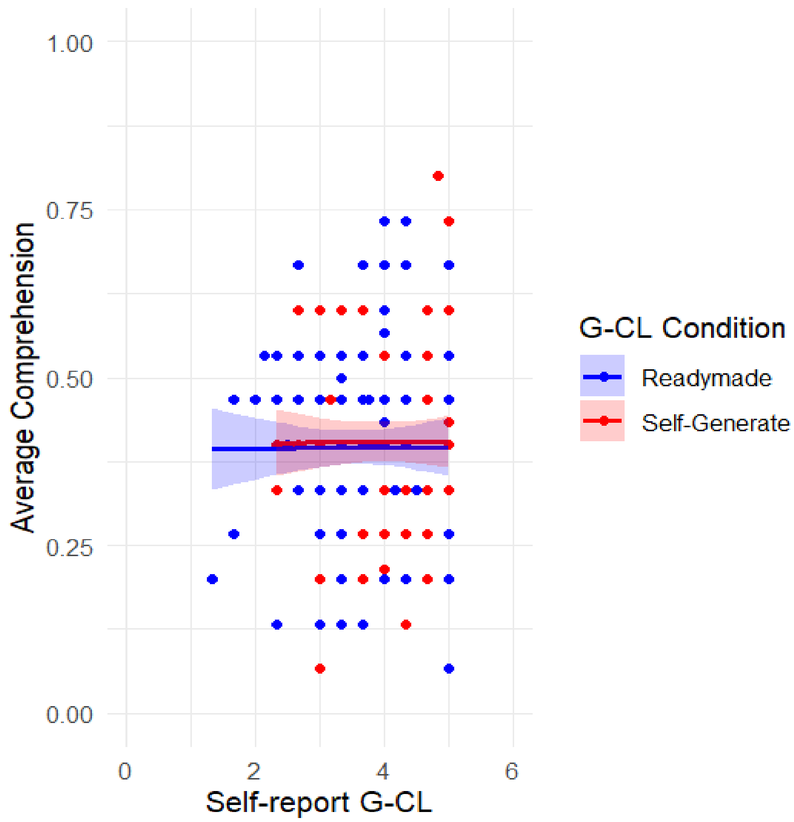


Note: The error bars in this figure are 95% Confidence Intervals.

Self-Reported G-CL Assessment

Next, I evaluated the influence of self-reported G-CL, the G-CL condition and their interaction on comprehension. Previous research shows that self-reported CL (i.e., indirect measures) and manipulations of CL (i.e., direct measures) are tapping into the same variable (e.g., Hoogerheide et al., 2019). This analysis would serve as a manipulation check to verify that I replicate these findings. However, there were no significant effects of self-reported G-CL, $t = 0.190, p = .85$, G-CL condition, $t = 0.306, p = .85$, or their interaction on comprehension, $t = -0.228, p = .82$ (see Figure 12; Appendix H).

Figure 5-5. Effects of Self-reported G-CL and G-CL condition on comprehension performance



Note: the error bars here are 95% confidence intervals.

Chapter 6 - Discussion

This study aimed to examine the impact of cognitive load on encoding and retrieval processes, with four main sub-goals to explore. I evaluated (1) the limits of G-CL, (2) the role of causality, (3) the longitudinal effects of both G-CL and causality and (4) the influence of various individual differences.

Effects of Germane Cognitive Load on Comprehension

Research on the benefits of G-CL on comprehension of expository materials is mixed (i.e., the redundancy effect; Mayer et al., 2005; Sweller 1994). In the current study, learners who self-generated information on the graphic organizers did not outperform those who received a readymade graphic organizer on a comprehension test. I manipulated G-CL by having participants either self-generate the information in a structured graphic organizer or having participants study a readymade graphic organizer (i.e., information was already filled in). The results provided support for the *Cognitive Load Hypothesis*, suggesting that the demands of learning new content and immediately applying it afterward may have imposed a cognitive load on learners. Thus, while self-generation can enhance learning in some contexts (Colliot & Jamet, 2020b), there are limits to its effectiveness.

Furthermore, there was a significant effect of G-CL self-generated accuracy performance on comprehension: higher G-CL performance was predictive of better comprehension scores. This suggests that participants' success on this component of the study positively influenced their overall comprehension, hinting that with higher G-CL scores, there might have been stronger support for the *Self-Generation Hypothesis*.

Notably, it seems that participants were challenged beyond what would be desirable levels of difficulty (Bjork & Bjork, 2020), as their performance was significantly lower than in previous studies with similar tasks. For instance, in the current study, the mean performance was 29% correctly completed text fields on the graphic organizer, whereas previous studies by Colliot and colleagues have thrown out participants who have lower than 80% accuracy (2018a, 2018b, 2019, 2020, 2021, 2022). Aspects of the methodology that likely increased cognitive load and potentially exceeded the optimal difficulty level were the participants' inability to 1) revisit earlier slides and/or 2) seek clarification on the presented material, both of which would typically be available in a face-to-face class.

The former point is a key difference between this study and prior work by Colliot and colleagues. Previously, participants completed a Graphic Organizer while they still had access to the learning materials (Colliot & Jamet, 2020a), whereas participants in the current study had to complete the Graphic Organizer after the lecture. This methodological shift may have impacted the results in several ways. One way in which the results were likely influenced was that without access to the lecture material, participants often filled in the Graphic Organizer with incorrect information (i.e., 49.5% of incorrect responses were commissions, whereas 50.5% were omissions of information). When learners retrieve the wrong information immediately after learning and do not receive corrective feedback, this leads to retrieval interference (i.e., a conflict between the original information and the wrong/commission information; Erdman, 2011; Bridge & Paller, 2012; Wallner et al., 2022) and has been shown to become exacerbated over time (Bridge & Paller, 2012; Wallner et al., 2022). In fact, a wealth of information has shown that corrective feedback is critical to long-term comprehension (Erdman, 2011; Erdman & Chan, 2012). Thus, future work should include a form of feedback for the graphic organizer information. This would support previous research that has found that the readymade condition consistently outperformed self-generation conditions on an immediate test and after a 2-day delay, replicating their earlier findings (Colliot & colleagues, 2018b, 2019, 2020a). Those individuals presented with fully accurate, instructor-generated information seemed to have a comprehension advantage over other individuals that were asked to generate that information.

Another way that the current results may have been impacted was that attention might have waned because there was no continuous engagement during the learning phase (i.e., instead learners passively watched the lecture), resulting in reduced performance. Much work has demonstrated the benefits of active learning in traditional classrooms (e.g., Bonney, 2024; Beimel et al., 2024) and online learning (e.g., Smith & Rixner, 2020; Irani et al., 2020; Bonney, 2024; Beimel et al., 2024). Thus, it appears that the version of the G-CL task used here was more challenging than in prior studies and that other changes to the lecture may have benefitted comprehension. These findings indicate that there is value in making slides or content accessible outside of lecture recordings in courses to support learning.

Effects of Causality on Comprehension

Research on the usefulness of causal connections in comprehension needs to be further investigated in an educational setting, because work with expository-based materials is not used as often (Zinar et al. 1990; van Silfhout et al., 2014; Ushiro et al., 2015), rather most work has been done on narrative text (Zwann et al., 1995; Pettijohn et al., 2016; Pettijohn & Radvansky, 2016; Radvansky, 2012; Radvansky & Zacks 2017). To evaluate the role of causality, participants either received explicit causal connectives tying information together throughout the lecture or not (i.e., implicit causal connectives). Surprisingly, though, there was no effect of causality on comprehension, providing support for the *Implicit Sufficient hypothesis* (i.e., implicit causal connections are sufficient for good comprehension). Further, causal connectives were not more or less effective for comprehension depending on learners' level of G-CL.

Effects of Delay Comprehension

The role of G-CL and causality needs to be evaluated after a meaningful delay to better understand whether there are lasting benefits, rather than only assessing comprehension immediately after learning. Results provide some evidence in favor of the *Delay Hypothesis* (i.e., over time memory decreased). Interestingly, these results were different depending on whether delay was treated as continuous (i.e., number of days since learning) or categorical (i.e., session number). Specifically, when delay was treated categorically, I found that the self-generation group (i.e., G-CL condition) had better comprehension than the readymade group when learners were given explicit causal connections, but only at Session 2. This 3-way interaction was not observed when delay was treated as a continuous variable. However, in the continuous delay analysis, I observed 2-way interactions with Causality and G-CL, but no significant pairwise effects were observed in the post-hoc analyses. Interestingly, I did not observe an effect of delay on either text-based or situation-based questions either. This result does not replicate prior work evaluating forgetting curves for the different levels of representation (Fisher & Radvansky, 2018). However, I should note that they used narrative texts in their study, whereas I used expository text in the current study. More work is needed to better understand the role that G-CL and causality play in the comprehension of different types of information as their representations are consolidated over time.

Effects of Individual Differences on Comprehension

Finally, I evaluated whether all these effects described above were further impacted by individual difference variables (i.e., prior knowledge, motivation/engagement, and perceived difficulty). Unfortunately, working memory will not be discussed in-depth since I was not able to reliably measure it in this study using the OSPAN. Since performance on the secondary, interleaving task (i.e., the math task) was well below the desired threshold for most participants, this measure was not included in any of the analyses and its impact on comprehension cannot be considered. Previous research would suggest this is an important variable to consider when learning new content and storing that information for long-term retention (e.g., Barlett, 1932). Thus, it will be important for future work to further evaluate the role that working memory plays in the relationships between G-CL, causality and comprehension.

I did, however, observe an effect of both motivation and prior knowledge on comprehension, such that higher motivation and more prior knowledge was associated with better comprehension. Importantly, these effects did not change depending on the G-CL or causality manipulations, which indicates that motivation and prior knowledge are important for learning regardless of how much germane cognitive load a learner is under or how explicit the causal relationship are. The motivation result supports previous work (Multu-Bayraktar et al., 2019; Mayer, 2014), suggesting that motivation should be considered when evaluating learning and comprehension over time.

The results related to prior knowledge were somewhat surprising given past work and relevant theories. Specifically, prior work has repeatedly demonstrated expertise reversal, which is a phenomenon that describes how instructional strategies that are effective for novice learners can become less effective or even counterproductive as learners gain expertise (Kalyuga et al., 2003). In other words, the optimal way to present information is different for low-knowledge individuals (i.e., smaller fragments) compared to high-knowledge individuals (i.e., broader points to build on existing schema). However, in the current study, prior knowledge did not interact with cognitive load or causality. One explanation for the lack of interactions was that the average level of prior knowledge was low (M = 0.212, SD = 0.216), indicating that the current sample was mostly made up of novices. Future work should seek out experts and novices in this area to evaluate how experts perform in this same task. Another explanation for why is that my manipulation of G-CL was too difficult for even high-prior knowledge learners to perform well. This idea is supported by an insignificant correlation between prior knowledge and GO accuracy ($r = 0.228$, $p = .132$). This suggests that even when individuals were familiar with the content, they were still not doing well on the task overall (i.e., it was too difficult).

Finally, there were no significant results involving difficulty as measured by the self-reported G-CL assessment. This variable was included to serve as a manipulation check (i.e., previous research has shown that completing a cognitive load task is more demanding for participants; Zwann & Radvansky, 1998; Sweller, 1994), but it failed to replicate previous research. While there was a weak, negative correlation between prior knowledge and difficulty ($r = -0.188$, $p = .055$), suggesting that when the content was new information (i.e., less prior knowledge) they were rating the task as most difficult, no other effects were observed. These null results suggest that the G-CL manipulation itself was too difficult (i.e., evidenced by the extremely low accuracy of the graphic organizers). If we consider the learned helplessness literature, it is possible that the G-CL manipulation was too difficult and that participants felt a lack of control over their performance, leading to a learned helplessness when completing the GO task (i.e., low scores on the GO task; Maier & Seligman, 1976).

Finally, the types of questions/level of assessments did have an impact on comprehension performance. Specifically, when looking at multiple choice questions, the implicit group who were also seeing the readymade Graphic Organizer had higher performance than other groups. When looking at the text-based questions, those who were in the Explicit group and completing the self-generated Graphic Organizer had higher performance.

Limitations of the Current Study

There were several limitations within this study, related to the participants, the tasks, and the available data or overall format for analyses.

Participant-related limitations

First, given the longitudinal design, attrition from session 1 to session 3 was an issue, resulting in a loss of data from ~40 participants. Since one of the main goals of this study was concerned with the retention of information over time, it does pose a problem for data that exists in the final session. Specifically, any analyses evaluating interactions with delay may have been underpowered, preventing me from observing my effects of interest.

Second, although about half of my participants were college students, the experiment did not take place in a real online class, which means they were not being incentivized with a course grade like they would in a real class. Further, nearly all participants had low prior knowledge (i.e., average prior knowledge was 21%), meaning that the PowerPoint itself may not have been optimally organized for this category of learners. Although not evaluated in the current dataset, the expertise reversal effect has shown that different learners need information structured in different ways (Kalyuga et al., 2003). This effect could have worked against the participants in the current study and hurt performance on the comprehension questions and/or the Graphic Organizer generation task.

Third, participants were recruited in different ways: online Prolific participants and Simpson College students in psychology courses. With that said, the location variable had no impact on analyses related to the main predicted results, showing no overall differences in comprehension ($p = .108$), Graphic Organizer accuracy ($p = .225$), or prior knowledge ($p = .194$). However, some exploratory analyses using the individual difference variables showed differences amongst the groups. Specifically, the groups differed on the self-reported motivation assessment ($p < .0001$), with Prolific participants reporting that they were more motivated (M = 3.59, SD = 1.19) than the Simpson participants (M = 2.96, SD = 0.914). Further, the groups differed on the G-CL self-reported assessment ($p < .0001$), such that Prolific participants reported being under higher self-reported G-CL (M = 4.19, SD = 1.20) than the Simpson participants (M = 3.39, SD = 0.72). Finally, if only some participants were more motivated (i.e., under higher G-CL like expected), it could mean that only the Prolific participants were completing the task as intended. With this, I asked the question "How much effort did you put into this task" and there was a difference between the samples ($p < .0001$), such that the Prolific participants reported putting in more effort (M = 4.10, SD = 0.98) compared to Simpson participants (M = 3.02, SD = 1.10). While this is still a self-report measure, it does suggest that the two groups did not put in equal effort, which in turn could lead to differences in the impact of the G-CL manipulations. Relatedly, all individuals approach reading in different ways, such as when they are reading for entertainment verses for a test (Narvaez et al., 1999). Mainly, research has found that when reading for tests, people made more repetitions and evaluations compared to those reading for entertainment. This same general effect likely occurs when learning new lecture material.

Fourth, since this study was conducted online, it is possible that there were higher levels of E-CL than what would be found in a face-to-face setting. I will argue that issues of this nature could be a factor in face-to-face learning as well, but are often addressed by instructor methods (e.g., repeating materials when something happens, closing a door to lessen distractions, etc.). Importantly, these same distractions may be present if the study was completed in a lab setting, as opposed to online, but they could be slightly more controlled (i.e., like a face-to-face situation). This factor of E-CL is hard to control in any environment and this variance would hopefully be captured with the motivation/effort questions. While E-CL issues can arise in any learning environment, this study highlights their potential impact within online learning, an increasingly prevalent format, particularly in college settings. It is not uncommon for online learning to be used as a supplement and/or replacement for face-to-face learning, which makes it important to understand how easily pedagogy translates into these online environments.

Task-related limitations

First, one of the primary manipulations in this study was the manipulation of germane cognitive load, and this task appears to be too difficult for all participants (mean performance was 29%). Previous research did manipulate when participants viewed the Graphic Organizer (Colliot & Jamet, 2020a), and they reported better comprehension when they presented it sequentially, instead of simultaneously after the learning material, specifically for the self-generation condition. This study served as the primary support for why I chose to set up my study in the sequential nature that I did, but there is other research that supports this idea for learner comprehension improvement (e.g., Robinson & Kiewra, 1998). Specifically, they argue that a simultaneous Graphic Organizer presentation can draw attention away from non-represented text and spend too much time focusing on the Graphic Organizer itself – decreasing retention for other material that was not in the Graphic Organizer. I did not replicate these ideas in my own research though, so there is potentially something methodologically different about my design. It is possible that this version was harder compared to previous versions that Colliot and colleagues completed.

Second, the causality manipulation used in the current study may not have been strong enough to produce benefits in comprehension. Previous research has demonstrated that causal connectives, such as “because” improve recall (Caron et al., 1988) but did not use expository materials. Other research has shown clear advantages in using causal connectives in expository texts (van Silfhout et al., 2014; Cevasco & van den Broek, 2019; Ushiro et al., 2015; Linderholm et al., 2000), but some research suggests that connectives may be beneficial to only low-knowledge readers (McNamara et al., 1996) and/or when metacognition is high (Welie et al., 2016). Interestingly, when texts heavily rely on causal connectives, high-knowledge readers experience increased cognitive load (i.e., like the expertise reversal effect; McNamara et al., 1996; O’reilly & McNamara et al., 2009). Additionally, these texts can overload working memory for low-knowledge readers who are trying to balance the integration of causal connectives and new information in the text itself (McNamara et al., 1996; Yan-ron, 2013). Importantly, this work suggests that low-knowledge readers may not maintain causal connectives overtime especially when texts are difficult in nature, making it hard to measure long-term impacts of causal manipulations without some sort of scaffolding in place (Ushiro et al., 2015; McNamara et al., 1996).

Additionally, this manipulation of causality assumes that participants in the implicit condition will not make the causal connections (that were provided explicitly to the other group) during their learning phase. This, however, is not something that can be assessed in the current study. Future work could evaluate the extent to which learners in the implicit condition generate the causal connectives on their own. Previous research in this area suggests that students recall information in a manner that matches how it was presented to them (i.e., descriptive text gets recalled in a descriptive way; Meyer et al., 2018), which would suggest that participants may not automatically generate causal connections. However, causality is incredibly important for us to successfully navigate and understand our world (Zwann & Radvansky, 1998; Zwann et al., 1995;

Trabasso & van den Broek, 1985; Bohn-Gettler et al., 2014), thus, generating causal connections may be a well-learned process that naturally happens during the learning of expository texts.

Third, the working memory measure used in this study, OSPAN, did not reliably assess actual working memory capacity. Scores were consistently too low for the measure to serve as a meaningful predictor or covariate in the analyses. Anecdotally, similar issues have been observed in our Memory and Aging Lab since studies have transitioned to online platforms using OSPAN. Thus, future online studies should consider using other span tasks to provide reliable results.

Finally, this study design did not leverage transfer-appropriate processing (Craik & Lockhart, 1972; Craik & Tulving, 1975; Morris et al., 1977). That is, the encoding and retrieval stages may not have aligned well with each other, and thus, did not optimize comprehension performance in this study. Specifically, at encoding, participants learned expository information through a lecture and then re-encoded information with a graphic organizer. This graphic organizer emphasized the hierarchical structure of the concepts (e.g., semantic and episodic memory being part of a larger category: declarative memory). However, at retrieval, participants completed a mixture of multiple choice and short answer questions. These questions did not require students to recreate the lecture material word-for-word (i.e., as in the original encoding opportunity) nor did they require participants to recreate the hierarchical structure of the material (i.e., as in the graphic organizer). Instead, these questions required the retrieval of text-based and situation model level information. Further, some of these questions require students to transfer the information learned in the lecture to an applied situation.

When encoding and retrieval conditions match, memory and comprehension should be higher than when these conditions do not match (Morris et al., 1977). Thus, future work should evaluate the effects of causality and graphic organizers under conditions in which transfer-appropriate processing can optimally benefit students' comprehension.

Future Directions & Implications

While this study has its limitations, it provides valuable insights and clear paths for future research. Several key recommendations emerge from the current findings. First, future studies should track the amount of time participants spend on the G-CL task to determine whether an objective measure of effort influences comprehension. Additionally, better control over delay periods, such as ensuring consistent intervals (e.g., exactly seven days), would improve the reliability of longitudinal studies. Further, an important next step in this research is to ensure that participants encode the correct information when they complete the G-CL task. As previously discussed, if participants do not encode information correctly from the start, then it is unsurprising that they struggle to recall it correctly later. To address this, future studies should integrate feedback into the self-generation G-CL condition to ensure participants encode the correct information before being assessed on their recall and comprehension.

Transitioning to in-person testing is another avenue worth exploring, as it could address concerns related to E-CL often associated with online learning. In-person settings would allow students to minimize distractions, ask questions, and receive face-to-face guidance, if needed. This format could also clarify whether the challenges observed in this study stem from the online context or the inherent difficulty of the tasks. Conducting controlled lab-based studies could help validate manipulations before moving to broader online experiments.

However, it is still important to explore the role of instructor presence in online learning. Recent research presents mixed findings on the inclusion of an instructor's face in lecture videos. Some studies suggest that seeing an instructor can improve content learning (Gu et al., 2024), motivation, engagement (Alemdag, 2022; Colliot & Jamet, 2018), and cognitive load (Alemdag, 2022), with the social-cue hypothesis linking these benefits to enhanced learning and comprehension (Mayer et al., 2003; Moreno & Mayer, 2007). However, other research found no measurable learning benefit and even suggests that including unnecessary images (such as the instructor's face) might increase cognitive load (Heidig et al., 2024; Mayer, 2021; Gu et al., 2024; Sweller, 1994). Since participants in the current study only heard the instructor's voice, future research should investigate how the presence or absence of an instructor's face influences comprehension and engagement.

To improve participant engagement and learning outcomes, future studies could also incorporate interactive elements. For instance, pop-up prompts could ask participants if they understand the material or not. If they say yes, they could provide a summary of the content in their own words for verification and, if not, the program could provide additional explanations to the participant. This approach could simulate classroom interactions and ensure participants remain attentive to the content, while also allowing researchers to analyze metacognitive abilities of the participants throughout the learning process. Self-paced learning is another promising direction, as research shows it can improve transfer ability, particularly for learners who need more time to process information (Sorden, 2005; Mayer et al., 2018). This is an aspect unique to online learning that would be particularly important at the individual student- and group-level when structuring courses for the future (i.e., online or in-person). Numerous studies have shown that when reading expository texts, reading time increases when people are in a study mode vs entertainment mode (Narvaez et al., 1999). Tracking viewing times for each slide could further illuminate comprehension and motivation levels during different types of content. Specifically, researchers could assess whether participants that take longer also have increased comprehension compared to those who spend less time on the content. This investigation could only be done if the learning phase was self-paced.

Additionally, encouraging participants to take notes may enhance immediate and delayed comprehension and recall (Rahmani & Sadeghi, 2011). However, effective note taking might require explicit instruction, such as teaching participants how to use graphic organizers. These adjustments and enhancements would not only address the study's limitations but also contribute to a more comprehensive understanding of effective learning strategies in both online and in-person settings.

A good next step in this line of research would be to dive into the impacts of testing effects with a G-CL task. Specifically, prior research has shown that long-term retention of information improves when a student is actively testing and practicing retrieval of the learned information (Rowland, 2014). Retrieval practice through this testing process has been found to enhance memory retention over time (Roediger & Karpicke, 2006a). This process supports the strengthening of retrieval mechanisms (Bjork, 1975; Karpicke & Blunt, 2011), facilitates error correction (Pashler et al., 2005), and encourages the deeper elaboration of material (Carpenter, 2009). Additionally, this type of retrieval practice aligns cognitive processes during learning with those required for assessment, which would promote transfer-appropriate processing (Craik & Lockhart, 1972; Craik & Tulving, 1975; Morris et al., 1977). However, several factors can influence the effectiveness of these effects, including the format of the test (e.g., multiple choice vs. recall-based questions). Specifically, research suggests that recall-based questions are more beneficial than multiple-choice or recognition-based formats for promoting retrieval practice (Butler, 2010). Further, ensuring accuracy during retrieval is critical, highlighting the importance of corrective feedback in learning (Kang et al., 2007). Thus, in future work using graphic organizers, researchers should provide participants with feedback as to what content was correctly and incorrectly retrieved. Furthermore, spacing effects play a crucial role in optimizing retrieval, as spaced study sessions allow learners to engage with the material multiple times across different intervals, leading to stronger retention (Cepeda et al., 2006). In the context of the current study, researchers could manipulate when and how often the participants complete the graphic organizers and the comprehension questions. Investigating these factors within a G-CL framework would provide deeper insight into how testing effects influence learning and memory retention over time.

Summary

This study highlights the nuanced relationship between cognitive load, causality, and individual differences in shaping comprehension of expository materials. The findings show that even though previous research found benefits to self-generation on comprehension, it can sometimes hinder learning due to the increased cognitive demands they impose. This suggests that access to accurate, instructor-provided information helps reduce E-CL and support retention. Causality showed mixed effectiveness on comprehension, with explicit causal connections facilitating better mental model construction, but the effectiveness varied across sessions, emphasizing the importance of timing and retrieval opportunities.

These results have important implications for online learning environments. Instructors should incorporate explicit causal structures and provide organized materials to support comprehension and manage cognitive load effectively. Additionally, individual differences, such as prior knowledge and motivation levels, play a critical role in learning outcomes and should be considered when designing instructional strategies. Future research should focus on the long-term effects of self-generation and causal mapping, particularly in diverse educational settings.

By refining our understanding of these factors, educators can develop more effective strategies to enhance student learning and retention over time.

References

Alemdag, E. (2022). Effects of instructor-presented videos on learning, cognitive load, motivation, and social presence: A meta-analysis, *Education & Information technologies,* 27, 12713-12742.

Bartlett, F.C. (1932). Remembering: A study in experimental and social psychology. Cambridge University Press.

Bartlett, B.J. (1978). *Top-level structure as an organizational strategy for recall of classroom text.* Unpublished doctoral dissertation, Arizona State University.

Beimel, D., Tsoury, A., & Barnett-Itzhaki, Z. (2024). The impact of extend and validity in active learning methods across online and face-to-face education on students' course evaluations, *Frontiers in Education,* 9:1432054. doi: 10.3389/feduc.2024.1432054

Bjork, R. A. (1994). Memory and metamemory considerations in the. *Metacognition: Knowing about knowing*, *185* (7.2).

Bjork, A., & Bjork,. E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum.

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition, 9 (4), 475-479.*

Bohn-Gettler, K. M. (2014). Does Monitoring Event Changes Improve Comprehension? *Discourse Processes*, 51(5-6), 398-425, DOI: 10.1080/0163853X.2014.913338

Booney, K. M. (2024). An Investigation of the Impact of online and in-person delivery of undergraduate biology instruction, *The American Biology Teacher, 86*(5), 289-297.

Bridge, D.J. & Paller, K.A. (2012). Neural Correlates of Reactivation and Retrieval-induced Distortion, *Journal of Neuroscience, 32*(35), 12144-12151.

Brunken, R., Plass, J.L., & Leutner, D. (2010). Direct Measurement of Cognitive Load in Multimedia Learning, *Educational Psychology, 38*(1), 53-61.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1118.

Cain, K., & Nash, H.M. (2011). The influence of connectives on young readers' processing and comprehension of text, *Journal of Educational Psychology, 103*, 429-441.

Caron, J., Micko, H. C., & Thüring, M. (1988). Conjunctions and the recall of composite sentences. *Journal of Memory and Language, 27*(3), 309–323. https://doi.org/10.1016/0749-596X(88)90057-5

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

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132, 354-380.

Cieniak, G., Shceiter, K,, & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load?, *Computers in Human Behavior,* 25, 315-324.

Colliot, T., & Jamet, É. (2018a). How does adding versus self-generating a hierarchical outline while learning from a multimedia document influence students' performances?

Computers in Human Behavior, 80, 354–361. http://dx.doi.org/10.1016/j.chb.2017.11.037.

Colliot, T., & Jamet, É. (2018b). Does self-generating a graphic organizer improve students' learning? *Computers & Educatio*n, 126, 13–22.

Colliot, T. & Jamet, E. (2019). Asking Students to b active learners: the effects of totally or partially self-generating a graphic organizer on students' learning performances, *Instructional Science,*

Colliot, T. & Jamet, E. (2020). Effects of self-generated graphic organizers on learning depend on in-task guidance, *Journal of Computer Assisted Learning,* 1-10.

Colliot, T. & Jamet, E. (2021). Improving students' learning by providing a graphic organizer after a multimedia document, *British Journal of Educational Technology, 1-14.*

Colliot, T., Kiewra, K. A., Luo, L., Flanigan, A. E., Lu, J., Kennedy, C., & Black, S. (2022). The effects of graphic organizer completeness and note-taking medium on computer-based learning, *Education and Information Technologies,27* (2), 2435-2456.

Cook, L.K. (1983). *Instructional effects of text structure-based reading strategies on the comprehension of prose.* Unpublished doctoral dissertation, University of California, Santa Barbara. Cited in Mayer, R. Aids to text comprehension. Paper presented at the 90th annual convention of the American Psychological Association, Washington, D.C., August 19

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294. https://d

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal behavior,* 11, 671-684.oi.org/10.1037/0096-3445.104.3.268

Crosson, A.C. & Lesaux, N.K. (2013). Does knowledge of connectives play a unique role in the reading comprehension of English learners and English-only students? *Journal of Research in Reading, 36*(3), 241–260. DOI:10.1111/j.1467-9817.2011.01501.x

Duffin, E. (2021). College Enrollment in the United States from 1965 to 2019 and Projections up to 2029 for Public and Private Colleges. *Statista.*

Ebbinghaus, H. (1885/1962). Memory: A contribution to experimental psychology. New York: Dover.

Erdman, M.R. & Chan, J.C.K. (2012). Providing corrective feedback during retrieval practice does not increase retrieval-induced forgetting, *Journal of Cognitive Psychology, 25*(6), 692-703.

Finkenbinder, A.M. (1913). The curve of forgetting. *The American Journal of Psychology, 24*(1), 8-32.

Fisher, J. S., & Radvansky, G. A. (2018). Patterns of forgetting. *Journal of Memory and Language, 102,* 130–141. https://doi.org/10.1016/j.jml.2018.05.008

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedro A, Sciaini, Marco, Scherer, Cédric (2024). *viridis(Lite) - Colorblind-Friendly Color Maps for R*. doi:10.5281/zenodo.4679423, viridis package version 0.6.5, https://sjmgarnier.github.io/viridis/.

Geva, E. (2007). Conjunction use in school children's oral language and reading. In R. Horowitz (Ed.), Talking texts: How speech and writing interact in school learning (pp. 271–294). Hillsdale, NJ: Erlbaum

Graesser, A.C., McNamara, D.S., & Louwerse, M.M. (2003). What do Readers Need to Learn in Order to Process Coherence Relations in Narrative and Expository Text?

Gu, C., Peng, Y., Nastase, S.A., Mayer, R.E., & Li, P. (2024). Onscreen presence of instructors in video lectures affects learners' neural synchrony and visual attention during multimedia learning, *PNAS, 12(*121), 1-12.

Heidig, S., Beege, M., Rey, G.D., & Schneider, S. (2024). Instructor presence in instructional videos in higher education: three field experiments in university courses, *Education Technology Research Development.* *https://doi.org/10.1007/s11423-024-10391-9*

Heller, O., Mack, W., Seitz, J. (1991) Replikation der Ebbinghaus'schen Vergessenskurve mit der Ersparnis-methode: "Das Behalten und Vergessen als Function der Zeit". Zeitschrift für Psychologie 199: 3–18.

Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & van Gog, T. (2019). Enhancing Example-Based Learning: Teaching on Video Increases Arousal and Improves Problem-Solving Performance, *Journal of Educational Psychology, 11*(1), 45-56.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory, 6*(1), 1–12. https://doi.org/10.1037/0278-7393.6.1.1

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance, 7*(3), 604–610. https://doi.org/10.1037/0096-1523.7.3.604

Irani, S., & Denaro, K. (2020). Incorporating active learning strategies and instructor presence into an online discrete mathematics class, *Proceedings of the 51st ACM Technical Symposium on Computer Science Education,* 1186-1192. https://doi.org/10.1145/3328778.3366904

Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness. Cambridge, England: Cambridge University Press.

Kalyuga, S., Aryes, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect, *Educational Psychologist, 38*(1), 23-31

Karpicke, J. D., & Blunt, J. R. (2011). *Retrieval practice produces more learning than elaborative studying with concept mapping*. Science, 331(6018), 772-775.

Kang, S. H. K., McDermott, K.B., & Roediger III, H.L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention, *European Journal of Cognitive Psychology, 19*(4/5), 528-558.

Katayama, A.D. (1997). Getting Students Involved in Note Taking: Why Partial Notes Benefit Learners More Than Complete Notes.

Katayama, A. D., & Robinson, D. H. (2000). Getting Students “Partially” Involved in Note-Taking Using Graphic Organizers. The Journal of Experimental Education, 68(2), 119–133. https://doi.org/10.1080/00220970009598498

McNamara, D. S. & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence, *Discourse Processes,* 22, 247-288.

Linderholm, T., Everson, M., van den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more- and less-skilled readers’ comprehension of easy and difficult texts. Cognition and Instruction, 18 (4), 525-556.

Magliano, J. P., *Miller, J.*, & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, *15*, 533-545.

Magliano, J. P., Radvansky, G. A., Forsythe, J. C., Copeland, D. E.. (2014). Event segmentation during first-person continuous events. *Journal of Cognitive Psychology,* 26*, 649-661*.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, *41*, 31-48.

Mayer, R.E. (2014). Incorporating Motivation into multimedia learning, *Learning and Instruction,* 29, 171-173.

Mayer, B. J. F., Wijekumar, K., & Lei, P. (2018). Comparative signaling generated for expository texts by 4th-8th graders: variations by text structure strategy instruction, comprehension skill, and signal word, *Reading Writing,* 31, 1937-1968.

Mayer, R.E., Wells, A., Parong, J, & Howarth, J.T. (2019). Learner control of pacing of an online slideshow lesson: Does segmenting help?,*Applied Cognitive Psychology,* 33, 930-935.

McCarthy, K. S. & McNamara, D. S. (2021). The multidimensional knowledge in text comprehension framework, *Educational Psychologist, 56*(3), 196-214.

McCrudden, M., Schraw, G., Lehman, S., & Poliquin, A. (2007). The effects of causal diagrams on text learning, *Contemporary Educational Psychology*, 32, 367-388.

Meyer, B. J. F., & McConkie, G. W. (1973). What is recalled after hearing a passage? *Journal of Educational Psychology, 65*, 109-117.

Meyer, B.J.F. (1975). The organization of prose and its effects of memory. Amsterdam: North-Holland

Meyer, B. J. F. (1982). Reading research and the composition teacher: the importance of plans, *College Composition and Communication*, *33,* 37-49.

Meyer, B. J. F., & Freedle, R. O. (1984). Effects of discourse type on recall, *American Educational Research Journal*, 21, 121–143.

Meyer, B. J. F., & Poon, L. W. (2001). Effects of structure strategy training and signaling on recall of text, *Journal of Educational Psychology, 93,* 141–159.

Meyer, B. J. F., Ray, M. N., & Middlemiss, W. (2012). Children's use of comparative text signals: the relationship between age and comprehension ability, *Discours. Revue de linguistique, psycholinguistique et informatique, 10*, 1–25

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning From Text. *Cognition and Instruction*, *14*(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing, *Journal of verbal learning and verbal behavior*, *16*(5), 519-533.

Muller, G.E. and A. Pilzecker. 1900. Experimentelle Beitrage zur Lehre vom Gedachtnis. Z. Psychol. Erganzungsband 1: 1–300.

Multu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review, *Computers and Education,* 141, 1-22.

Murre, J.M.J., & Dros, J. (2015) Replication and Analysis of Ebbinghaus' Forgetting Curve. *PLoS ONE,* 10(7): e0120644. https://doi.org/10.1371/journal.pone.0120644

Narvaez, D., Van Den Broek, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading, *Journal of Educational Psychology*, *91*(3), 488.

Nourbakhsh, N., Chen, F., Wang, Y., & Calvo, R. A. (2017). Detecting Users' Cognitive Load by Galvanic Skin Response with Affective Interference, *ACM Transactions on Interactive Intelligent Systems, 7*(3), 1-20.

O'reilly, T. & McNamara, D.S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers, *Discourse Process, 43*(2), 121-152.

Paas, F. G., & Van Merriënboer, J. J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. Journal of Educational Psycholog*y, 86*(1), 122–133. https://doi.org/10.1037/0022-0663.86.1.122

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist,* 38, 63–71. http://dx.doi.org/10.1207/ S15326985EP3801_8

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words?. *Journal of experimental psychology: Learning, Memory, and Cognition*, *31*(1), 3.

Pettijohn, K. A. & Radvansky, G. A. (2016). Narrative event boundaries, reading times, and expectation, *Memory and Cognition, 44,* 1064-1075.

Pettijohn, K. A., et al. (2016). Event boundaries and memory improvement, *Cognition, 1(*48). 136-144.

Plass, J.L. & Kalyuga, S. (2018). Cognitive Load as a Local Characteristic of Cognitive Processes: Implications for Measurement Approaches. In R.E. Zheng (Ed.), *Cognitive*

Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice (pp. 59-74)*,* Routledge/Taylor & Francis Group.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 509–522. https://doi.org/10.1037/0278-7393.2.5.509

Potter, M., Staub, A., Radó, J., & O'Connor, D.H. (2002). Recognition memory for briefly presented pictures: the time course of rapid forgetting. *Journal of experimental psychology. Human perception and performance, 28 5*, 1163-75 .

Radossawljewitsch, P.R. (1907) Das Behalten und Vergessen bei Kindern und Erwachsenen nach experimentellen Untersuchungen [An experimental investigation of the retention and forgetting among children and adults]. Leipzig: Nemnich.

Radvansky, G. A., & Copeland, D. E. (2001). Working memory and situation model updating. *Memory & Cognition, 29*(8), 1073-1080.

Radvansky, G. A., & Copeland, D. E. (2004). Working memory span and situation model processing. *The American journal of psychology,* 191-213.

Radvansky, G. A., Copeland, D. E., & Zwaan, R. A. (2005). A novel study: Investigating the structure of narrative and autobiographical memories. *Memory, 13*(8), 796–814.

Radvansky, G.A., Tamplin, A., & Krawietz, S. (2010) Walking through doorways causes forgetting: Environmental integration. *Psychonomic Bulletin & Review.* 17, 900-904.

Radvansky, G.A. (2012). Across the Event Horizon, *Current Directions in Psychological Science, 21*(4), 269-272.

Radvansky, G.A. & Zacks, J.M. (2017). Event Boundaries in Memory and Cognition, *Current Opinion in Behavioral Sciences,* 17, 133-140.

Rahmani, M. & Sadeghi, K. (2011). Effects of note-taking training on reading comprehension and recall, *The Reading Matrix, 11*(3), 116- 128.

Roediger, H.L., & Karpicke, J.D. (2006). The power of testing memory: Implications for educational practice, *Memory.* 10, 319-332.

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice, *Perspectives on Psychological Science, 1*(3), 181-210.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. Journal of Personality and Social Psychology, 35, 677-688.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect, *Psychological Bulletin, 140*(6), 1432-1463.

Sarkar D (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5, http://lmdvr.r-forge.r-project.org.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior, 6*(1), 156–163. https://doi.org/10.1016/S0022-5371(67)80067-7

Smith, R. & Rixner, S. (2020). Design and Evaluation of a collaborative online computational thinking course, *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education,* 342-348. https://doi.org/10.1145/3341525.3387388

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1*(3), 341–358. https://doi.org/10.1207/s1532690xci0103_4

Sorden, S. D. (2005). A Cognitive Approach to Instructional Design for Multimedia Learning, *Informing Science Journal,* 8, 263-279.

Stull, A. T., & Mayer, R. E. (2007). Learning by doing versus learning by viewing: Three experimental comparisons of learner-generated versus author-provided graphic organizers, *Journal of Educational Psychology, 99*(4), 808–820.

Sweller, J. (1994). Cognitive Load Theory, Learning Difficulty, and Instructional Design, *Learning and Instruction,* 4, 295-312.

Trabasso, T., & Van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24*(5), 612–630. https://doi.org/10.1016/0749-596X(85)90049-X

Unsworth, N., & Engle, R.W. (2007). The Nature of Individual Differences in Working Memory Capacity: Active Maintenance in Primary Memory and Controlled Search from Secondary Memory. *Psychological Review, 114*(1), 104-132.

Ushiro, Y., Hamada, A., Kimura, Y., 名畑目, 真., 雅也, 細., 大樹, 加., & 葉子, 渡. (2015). Building Causally Coherent Mental Representations and Learning From Expository Texts Among Japanese EFL Readers.

van Silfhout, G., Evers-Vermeul, J., Mak, W. M., & Sanders, T. J. M. (2014). Connectives and layout as processing signals: How textual features affect students' processing and text representation. *Journal of Educational Psychology, 106*(4), 1036–1048. https://doi.org/10.1037/a0036293

Wallner, L., Nickl, A. T., & Bäuml, K.-H. T. (2022). When study capacities are limited and deadline is fixed—How practice type and practice timing influence recall of practiced and unpracticed material. *Journal of Applied Research in Memory and Cognition, 11*(4), 545–553. https://doi.org/10.1037/mac0000024

Welie, C., Schoonen, R., Kuiken, F., & van den Bergh, H. (2017). Expository text comprehension in secondary school: for which readers does knowledge of connectives contribute the most?, *Journal of Research in Reading, 40*(1), 42-65.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. In *Journal of Statistical Software* (Vol. 40, Issue 1, pp. 1–29). http://www.jstatsoft.org/v40/i01/

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.4, https://github.com/tidyverse/dplyr, https://dplyr.tidyverse.org.

Wolfe, M. B. W. (2005). Memory for Narrative and Expository Text: Independent Influences of Semantic Associations and Text Organization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(2), 359–364. https://doi.org/10.1037/0278-7393.31.2.359

Yan-ron, Z. (2013). Consciousness of Causal Inference in Expository Reading. *Psychological development and education*.

Zinar, S. (1990). Fifth-Graders' Recall of Propositional Content and Causal Relationships from Expository Prose, *Journal of Reading Behavior, 22*(2), 181-199. https://doi.org/10.1080/10862969009547703

Zwaan, R.A., Magliano, J.P., & Graesser, A.C. (1995). Dimensions of Situation Model Construction in Narrative Comprehension, *Journal of Experimental Psychology, 12*(2), 386-397.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The Construction of Situation Models in Narrative Comprehension: An Event-Indexing Model. *Psychological Science*, *6*(5), 292–297. https://doi.org/10.1111/j.1467-9280.1995.tb00513.x

Zwaan, R.A. & Radvansky, G.A. (1998). Situation Models in Language Comprehension and Memory, *Psychological Bulletin, 123*(2), 162 -185.

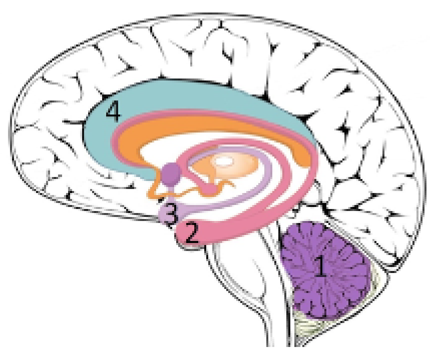
Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading, 2*(3), 199–220.

Zu, T., Hutson, J., Loschky, L. C., & Rebello, N. S. (2020). Using eye movements to measure intrinsic, extraneous, and germane load in a multimedia learning environment. *Journal of Educational Psychology, 112*(7), 1338-1352.

Appendix A - Learning Phase Questions

Question Set A

1. What is an engram?
    a. When information is processed, an engram (memory trace) is created
2. If a kindergartener is asked to remember who the first president was, what kind of memory would they rely on to remember that?
    a. Semantic, Declarative
3. Which non-declarative memory is least affected by aging?
    a. Procedural Memory



4. In the diagram above, which number corresponds to the amygdala?
    a. 1
    b. 2
    c. 3
    d. 4
5. In which situation would the amygdala be used?
    a. brushing your teeth
    b. organizing your desk
    c. Walking in a forest and you suddenly see a snake slithering across the path

d. observing the mesmerizing movements of colorful fish and coral reefs

6. Following damage to a cerebral structure, the HM patient is no longer able to access old memories and is no longer able to acquire new knowledge. However, his cognitive abilities as well as his general knowledge of the world are preserved. Which type of declarative memory is kept in the case of the HM patient?
    a. Procedural Memory
    b. Emotional Memory
    c. Semantic Memory
    d. Episodic Memory
7. What distinguishes episodic memory from semantic memory?
    a. semantic memory is general knowledge, and episodic memory is memory for experiences
8. Taylor and Trevor go to a cafe for coffee and scones and they decide to focus on different cues for their long-term retention of the event. Taylor takes in the smells and uses her olfactory sensory memory while Trevor takes in the sounds and uses his echoic sensory memory. Who will have a better chance of remembering information from that day?
    a. Taylor
9. You are participating in a quiz show, and the host asks you: What lobe is responsible for iconic memory and what type of information does it store?
    a. Occipital Lobe & Visual information
10. You go to the grocery store and realize that you forgot the list, so you are trying to remember the list of things that you needed; what sensory memory component would you rely on to do this?

a. **Iconic Memory**
b. Gustatory Memory
c. Somatosensory Memory
d. Echoic Memory

11. Short-term and Working memory are both sub-systems of which main type of memory?
    a. Sensory memory
    b. **Immediate memory**
    c. Long-term memory
    d. Declarative memory
12. This part of immediate memory is responsible for managing and allocating attentional resources.
    a. Phonological Loop
    b. **Central Executive**
    c. Visuospatial Sketchpad
    d. Echoic Memory
13. Which component of working memory is responsible for processing visual information?
    a. Phonological Loop
    b. Central Executive
    c. **Visuospatial Sketchpad**
    d. Echoic Memory
14. What is the average length of retention of information in short-term memory?
    a. **~20 seconds**

15. If you are waiting in the doctor's office and hear your name called, which component of working memory is processing that information?

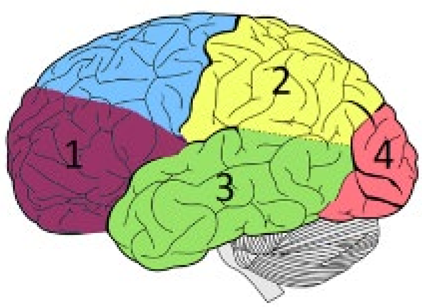
    a. Phonological Loop

    b. Central Executive

    c. Visuospatial Sketchpad

    d. Iconic memory

Question Set B



1. On the diagram above, which number corresponds to the parietal lobe

    a. 1

    b. 2

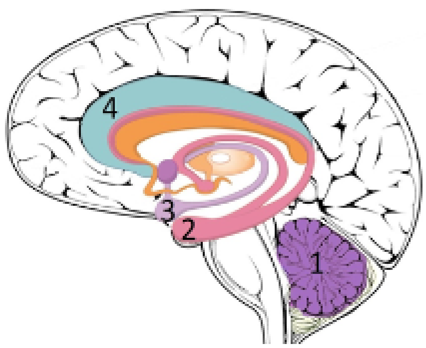
    c. 3

    d. 4

2. When a child is learning about the revolutionary war in 5th grade and they remember they learned that George Washington was president during this time, they are using which type of memory?

    a. Episodic Memory

3. Can you recall a time when a particular experience evoked strong emotions within you, such as a moment of immense joy, fear, or sadness? In which type of memory is this stored? (be precise)
    a. **LT, non-declarative, emotional**



4. In the diagram above, which number corresponds to the brain structure involved in procedural memory?
    a. **1**
    b. 2
    c. 3
    d. 4
5. There are two types of long-term memory, what are they?
    a. **Declarative, non-declarative**
6. Name something in the real world that you think is similar to semantic memory
    a. **Dictionary, encyclopedia, hard drive, etc.**
7. Following damage to a cerebral structure, the HM patient is no longer able to access old memories and is no longer able to acquire new knowledge. However, his cognitive

abilities as well as his general knowledge of the world are preserved. Which type of declarative memory is affected in the case of the HM patient?

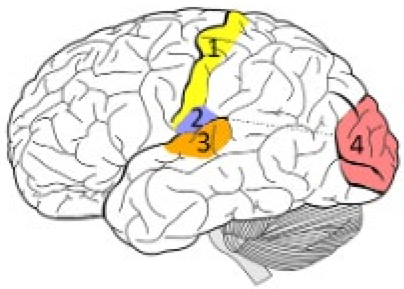
a. Procedural Memory
b. Emotional Memory
c. Semantic Memory
d. Episodic Memory

8. Entering High-school or having your first kiss are first-time experiences that are more often remembered by people over 40, this is referred to as what?

a. Recollection phenomenon
b. Adolescent Recollection
c. Reminiscence Bump
d. Remembrance Spike

9. Emmeline was telling her husband, Gerome, about what time he needed to leave for dinner and then asked him, "Did you hear what I just said?" He was able to tell her the correct time even though he wasn't paying attention. Gerome was able to do so because her words were briefly maintained by his ____.

a. Iconic Memory
b. Gustatory Memory
c. Somatosensory Memory
d. Echoic Memory



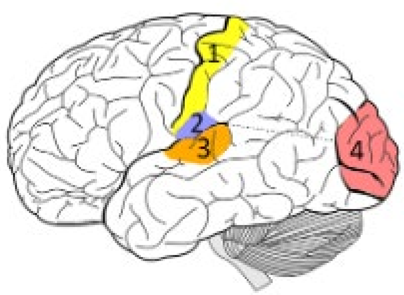
10. Which number corresponds to the visual sensory area?

    a. 1

    b. 2

    c. 3

    d. 4



11. Which number corresponds to the somatosensory area?

    a. 1

    b. 2

    c. 3

    d. 4

12. If you are presented with the letters 'C'. 'I'', and 'A' to remember for a later assessment, you could use this strategy to remember the letters easier.

    a. Chunking

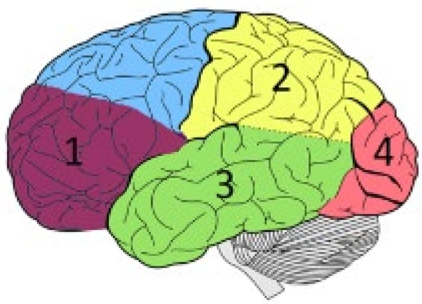
b. Visualization

c. Priming

d. Encoding

13. What distinguishes working memory from short-term memory?

a. Working memory also processes information, unlike short-term memory



14. In the diagram above, which number corresponds to where short-term memory is located?

a. 1

b. 2

c. 3

d. 4

15. Once information has made it through sensory memory, it is being processed in which memory system?

a. Working memory

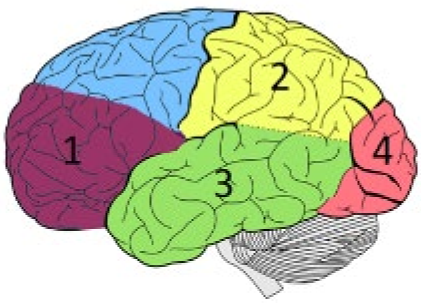
b. Short-term memory

c. Long-term memory

d. Non-declarative memory

<u>Question Set C</u>

2. Think about a specific skill or action that you have learned, such as riding a bicycle or playing a musical instrument. In which type of memory is this knowledge stored? (be precise)
    a. **LT, non-declarative, procedural**



3. On the diagram above, which number corresponds to the temporal lobe?
    a. 1
    b. 2
    c. **3**
    d. 4
4. Semantic and Episodic memory are both subsystems of what type of memory?
    a. Short-term Memory
    b. **Declarative Memory**
    c. Non-declarative Memory
    d. Working Memory
5. An amnesic patient forgets what is happening and is unable to acquire new information in memory. However, by drawing a star several times their performance improves. Given the performance of the star drawing task, what type of non-declarative memory is preserved here?

a. Procedural Memory

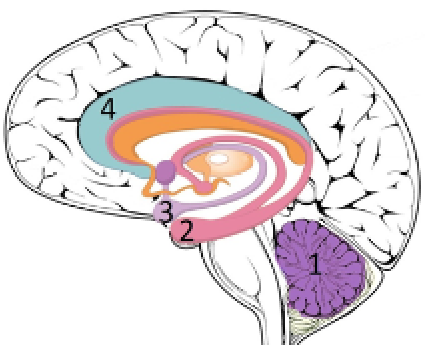
b. Emotional Memory

c. Semantic Memory

d. Episodic Memory

6. Before what age do we observe almost no memories?

    a. 3yo



7. In the diagram above, which number corresponds to the hippocampus?

    a. 1

    b. 2

    c. 3

    d. 4

8. What are the 2 types of non-declarative memories?

    a. Procedural and emotional

9. Following damage to a cerebral structure, the HM patient is no longer able to access old memories and is no longer able to acquire new knowledge. However, his cognitive

abilities as well as his general knowledge of the world are preserved. Which brain structure is affected in the case of the HM patient?

   a. **Seahorse**

10. Which type of sensory memory is involved when you get a paper cut?

    a. **Tactile Memory**

11. Olfactory memory is located inside of the ____ lobe.

    a. **Temporal**

12. What is the name of the most used sensory memory, accounting for nearly 80% of information transmitted to our brain?

    a. **Iconic Memory**

13. In 1957, Miller discovered the memory span was ______.

    a. **7 +/- 2**

14. If a child is repeating a list of items to help them remember they are likely around what age?

    a. **6-7 yo**

15. Which module of working memory do you think causes the effects seen in this graph?

    a. **Phonological loop**

    b. Central executive

    c. Visuouspatial sketchpad

    d. Iconic memory

16. Regina was at a work event and was introduced to a new colleague named "Ramone". In order to remember their name, Simone kept repeating "Ramone, Ramone, Ramone" in her head. This type of mental repetition is referred to as _.

a. Subvocalization

Appendix B - Pre-test Knowledge Questions

1. Name working memory modules (maximum 3 points):
    a. Phonological (articulatory) loop
    b. visual-spatial notebook
    c. central administrator (central executive system)
2. Name the main types of declarative long-term memory (maximum 2 points):
    a. Semantic memory
    b. episodic memory
3. What is the most important brain structure for memory? (maximum 1 point):
    a. Hippocampus/temporal lobe
4. What is the span of working memory? (maximum 1 point):
    a. About 7 items
5. Which author is at the origin of the memory span? (maximum 1 point):
    a. Miller in 1956

Appendix C - Motivation/G-CL Assessment Questions

Original scales (Moreno, 2007)

- Motivation Assessment
    a. How interesting was it to learn about essential teaching skills today?
    b. How entertaining was it to learn about essential teaching skills today?
    c. How eager would you be to learn about a different educational psychology topic in the same conditions you learned today?
    d. How motivating was it to learn about essential teaching skills today?
    e. How much did the computer program help you understand essential teaching skills?
    f. How helpful was the computer program for learning about essential teaching skills?
- Self-Reported G-CL Assessment
    g. How difficult was it to learn about essential teaching skills with the computer program?

Questions used in this study

- Motivation Assessment
    - How interesting was it to learn about memory structures?
    - How entertaining was it to learn about memory structures?
    - How eager would you be to learn about other related topics in the future?
- Self-Reported G-CL Assessment
    - How difficult was the task that you did about the lecture over memory structures?
    - How difficult were the questions that you completed about the lecture over memory structures?
    - How much effort did you invest to learn about memory structures during the lecture?

Appendix D - Result Comparisons and Output from model predicting comprehension performance, including Session (Categorical), Causality, Condition, and Motivation Self-Report

When looking for the best fit model, I compared 3 different model structures after determining that the best link-function was squareroot and the best random effects option was to only include the subject varying at the intercept. Model 2 included the full-factorial of Causality, categorical Session, G-CL condition, covariates of Self-reported G-CL and Motivation. Model 3 included the full-factorial of Causality, categorical Session, G-CL condition, covariates of Self-reported G-CL. Model 4 included the full-factorial of Causality, categorical Session, G-CL condition, covariates of Self-reported Motivation. The best fit model was Model 4, as determined by the lowest AIC and BIC values (see table 1 below). The output from this model is displayed in Table 2 below.

Appendix. Table 6-1. AIC and BIC values for each of the models compared.

Model	AIC	BIC
Model 2	-291.94	-237.79
Model 3	-289.53	-238.76
Model 4	-293.97	-243.18

Appendix. Table 6-2 . *Output from model predicting comprehension performance, including Session (Categorical), Causality, Condition, and Motivation Self-Report*

	Estimate	*SE*	*t*	*p*
Intercept	0.561	0.04	13.961	<.0001*
G-CL Condition [GO]	-0.031	0.038	-0.808	.419
Causality [Implicit]	0.011	0.038	0.288	.773
Session 2	-0.055	0.021	-2.569	.01*
Session 3	-0.034	0.023	-1.449	.147
Motivation Report	0.021	0.009	2.108	.035*
G-CL Condition X Causality	0.047	0.059	0.798	.4246
G-CL Condition X Session 2	0.140	0.037	3.742	.0001*
G-CL Condition X Session 3	0.019	0.037	0.511	.609
Causality X Session 2	0.094	0.034	2.729	.006*
Causality X Session 3	0.079	0.037	2.123	.034*
G-CL Condition X Causality X Session 2	-0.235	0.054	-4.332	<.0001*
G-CL Condition X Causality X Session 3	-0.130	0.057	-2.295	.022*

Note: Session 2 indicates a comparison between session 2 and 1; Session 3 indicates a comparison between session 3 and 1.

Appendix E - Full model output when predicting comprehension performance, including Session (Categorical), Causality and GO Proportion Correct.

Appendix. Table 6-3. *Output from model predicting comprehension performance, including Session (Categorical), Causality and GO Proportion Correct.*

	Estimate	*B*	*t*	*p*
Intercept	0.407	0	26.98	<.001
Session 2	-0.009	-0.067	-0.26	.794
Session 3	-0.014	-0.103	-0.39	.696
Causality	-0.013	-0.096	-0.42	.669
Proportion Correct	0.142	0.231	2.14	.035
Session 2 X Causality	-0.087	-0.634	-1.25	.214
Session 3 X Causality	-0.089	-0.653	-1.26	.213

Note: Session 2 indicates a comparison between session 2 and 1; Session 3 indicates a comparison between session 3 and 1.

Appendix F - Result Comparisons and output for model predicting comprehension performance, including Session (Continuous), Causality, and Self-Report G-CL

When looking for the best fit model, I compared 3 different model structures after determining that the best link-function was squareroot and the best random effects option was to only include the subject varying at the intercept. Model 2 included the full-factorial of Causality, categorical Session, G-CL condition, covariates of Self-reported G-CL and Motivation. Model 3 included the full-factorial of Causality, categorical Session, G-CL condition, covariates of Self-reported G-CL. Model 4 included the full-factorial of Causality, categorical Session, G-CL condition, covariates of Self-reported Motivation. The best fit model was Model 3, as determined by the lowest AIC and BIC values (see table 4 below). The output from this model is displayed in Table 5 below.

Appendix. Table 6-4. *AIC and BIC values for each of the models compared.*

Model	AIC	BIC
Model 2	-307.5	-267.69
Model 3	-308.3	-271.81
Model 4	-306.3	-269.81

Appendix. Table 6-5. *Output from model predicting comprehension performance, including Session (Continuous), Causality and Self-Report G-CL*

	Estimate	*SE*	*t*	*p*
Intercept	0.657	0.047	13.885	<.0001*
G-CL Condition [GO]	0.012	0.037	0.335	.738
Causality [Implicit]	0.029	0.038	0.753	.451
Delay	-0.003	0.002	-1.909	.056*
G-CL Report	-0.009	0.012	-0.819	.412
Condition X Causality	-0.02	0.057	-0.347	.728
G-CL Condition X Delay	-0.01	0.003	-2.948	.003*
Causality X Delay	0.006	0.003	2.266	.023*
G-CL Condition X Causality X Delay	0.002	0.005	0.463	.644

Appendix G - Model output for exploratory analyses looking at types of questions and the levels of representation as an outcome variable.

Appendix. Table 6-6 Model output for analysis looking at the impact of full factorial of Condition, Causality, and Session to predict Multiple Choice Average Performance

	F	*p*
Causality	0.949	.331
G-CL Condition	0.198	.657
Session	0.101	.904
Causality X G-CL Condition	5.556	.019*
Causality X Session	0.002	.998
G-CL Condition X Session	0.425	.654
Causality X G-CL Condition X Session	1.210	.300

Appendix. Table 6-7 Model output for analysis looking at the impact of full factorial Condition, Causality, and Session to predict Short-answer question performance

	F	*p*
Condition	0.0554	.814
Session	0.2741	.761
Causality X G-CL Condition	1.4927	.223
Causality X Session	0.9146	.402

G-CL Condition X Session	0.0895	.914
Causality X G-CL Condition X Session	0.0996	.905

Appendix. Table 6-8. Model output for analysis looking at the impact of full-factorial Condition, Causality, and Session to predict Situation-based question performance

	F	*p*
Causality	0.09261	.761
G-CL Condition	1.07751	.300
Session	0.01913	.981
Causality X G-CL Condition	0.00738	.932
Causality X Session	1.09738	.336
G-CL Condition X Session	0.40379	.668
Causality X G-CL Condition X Session	0.68627	.505

Appendix.Table 6-9. Model output for analysis looking at the impact of full-factorial of Condition, Causality, and Session to predict Text-based question performance

	F	*p*
Causality	3.185	.076
G-CL Condition	0.103	.749
Session	1.087	.339
Causality X G-CL Condition	4.837	.029
Causality X Session	0.159	.853
G-CL Condition X Session	1.169	.313
Causality X G-CL Condition X Session	0.302	.739

Appendix H - Model output for exploratory analyses looking at the influence of individual difference variables as a predictor of comprehension performance

Appendix. Table 6-10. Model output for analysis using Condition and Prior Knowledge to predict comprehension performance

Predictor	Estimate	*SE*	*t*	*p*
Intercept	0.37	0.026	14.293	<.0001
G-CL Condition	0.012	0.038	0.310	.757
Prior Knowledge	0.167	0.079	2.115	.037*
Prior Knowledge X G-CL Condition	-0.072	0.128	-0.560	.577

Appendix. Table 6-11. Model output for analysis using Condition and Self-Reported G-CL to predict comprehension performance

Predictor	Estimate	*SE*	*t*	*p*
Intercept	0.385	0.058	6.629	<.0001

G-CL Condition	0.031	0.101	0.306	.76
Self-Reported G-CL	0.003	0.016	0.190	.85
Self-Reported G-CL X G-CL Condition	-0.006	0.026	-0.228	.82

Appendix. Table 6-12. Model output for analysis using Condition and Motivation to predict comprehension performance

Predictor	Estimate	*SE*	*t*	*p*
Intercept	0.315	0.036	8.703	<.0001
G-CL Condition	-0.006	0.062	-0.097	.923
Motivation	0.023	0.010	2.298	.023*
Motivation X G-CL Condition	0.003	0.016	0.158	.874