12-2021

Effects of Static and Dynamic Visuals on the Learning of Science Concepts in the Secondary-School Classroom

Theodore Johnson
University of San Francisco, tjohnson352@gmail.com

The University of San Francisco

EFFECTS OF STATIC AND DYNAMIC VISUALS ON THE LEARNING OF SCIENCE CONCEPTS IN THE SECONDARY-SCHOOL CLASSROOM

A Dissertation Presented
to
The Faculty of the School of Education
Learning & Instruction Department

In Partial Fulfillment of the
Requirements for the Degree
Doctor of Education

by
Theodore L. Johnson
San Francisco
December 2021

ABSTRACT

Effects of Static and Dynamic Visuals on the Learning of Science Concepts in the Secondary-School Classroom

The current study is grounded in the cognitive theory of multimedia learning. The investigator considered how embedding science text with visuals could affect secondary-school students' ability to retain the information they learn (rote learning) and transfer the new knowledge to an unfamiliar problem (meaningful learning). Furthermore, this study explored how the type of visuals (static versus vs. dynamic visual) and text (audio vs. print) affect science learning. The data generated was sourced from student participants in a secondary-school biology classroom.

The purpose of this study was to investigate how prior knowledge and the integration of information modalities (i.e., text, audio, static visual, dynamic visuals) promotes rote learning (information retention) and meaningful learning (knowledge transfer) in science. The study was used also to investigate how the interaction of prior knowledge, which was coded as expertise level, with information modality effects learning and cognitive load.

The study was based on a quasi-experiment that included a pretest, intervention, and posttest phase. The pretest assessed prior knowledge of the subject matter and established a baseline knowledge score. 117 participants were assigned to one of four treatment groups. For Group 1, the learning material was exclusively text. Group 2 had textual information with embedded pictures that corresponded with the concepts in the text. Group 3 had animation with the text subscripted in a video. And, Group 4 was

provided with a fully animated version of the video that included audio narration instead of subscripted text.

Three sets of response variables were generated from the collected data: (a) rote learning scores, (b) meaningful learning scores, and (c) cognitive load scores. The between-group differences in the response variables were evaluated via analysis of variance (ANOVA) using the SPSS Statistics software package. The ANOVA results revealed statistically significant effects only for rote learning and the cognitive load associated with rote learning. No statistically significant effect was detected for meaningful learning, the learning intervention, and their associated cognitive loads. Furthermore, the interaction of prior knowledge (i.e., learner expertise) with information modality did not have statistically significant effects on any of the responding variables.

SIGNATURE PAGE

This dissertation, written under the direction of the candidate's dissertation committee and approved by the members of the committee, has been presented to and accepted by the Faculty of the School of Education in partial fulfillment of the requirements of the degree of Doctor of Education. The content and research methodologies presented in this work represent the work of the candidate alone.

Theodore L. Johnson	December 14, 2021
Author	Date

Dissertation Committee

Xornam Apedoe, PhD	December 14, 2021
Chairperson	Date
Patricia Busk, PhD	December 14, 2021
Chairperson	Date
Mathew Mitchell, PhD	December 14, 2021
Chairperson	Date
Sarah Capitelli, PhD	December 14, 2021
Chairperson	Date

ACKNOWLEDGEMENTS

During my time at the University of San Francisco, there are several individuals who supported me throughout this transformative process and, without them, achieving success would not have been possible. I want to specially recognize these special individuals:

- Thanks to **Professor Dr. Xornam Apedoe** for her relentless patience and dedication to academic excellence. Professor Apedoe, who served as the chair of my dissertation committee and as my advisor, provided clear and specific guidance throughout the dissertation writing process. She painstakingly read through the many edits and offered constructive feedback that made me think more critically about my writing. Without her support, I simply would have been at a loss of where to even begin putting together this manuscript. Professor Apedoe, you always made me feel like I was making steady progress, even when I felt otherwise.
- Thanks to **Professor Dr. Patricia Busk** for her meticulous commitment to excellence and attention to detail. Professor Busk will be remembered for alway pushing me, and all of her students, to continuously strive to evolve into a better version of ourselves. You amazed me the amount of effort you put into providing feedback. During the writing process, you provided invaluable guidance on statistical analyses that made it possible for me to make sense of my data. Thanks for always believing in me!
- Thanks to **Professor Dr. Mathew Mitchell** for always inspiring me to become an even better educator. There is no other teacher who could make the most complex ideas and concepts seem like common sense. You are one of the most impressive teachers I have met in my many years of working in the profession. It was in my first course under your guidiane that I was introduced to cognitive load theory and became inspired to do this research. Thank you for always being amazing, yet appearing to be absolutely unaware of it.
- Thanks to my cohort of doctoral students (i.e., **Dr. David Sul, Dr. Natalie Chiang, Patricia Garcia, Letta Mayer, Taiye Obada, Juhee Mo**) who were my partners in the struggle. Thanks for all the support and encouragement. Special thanks to **Dr. David Sul** and **Dr. Natalie Chiang** for providing additional support during the writing process.

Of course, none of this would have been possible without the support of my family, so I want to specially recognize some key figures in my life.

- To my dearest and loving wife, **Professor Dr. Annick Wibben**, for being the rock that keeps me grounded, for pushing me to grow and evolve into a better version of myself, and for always believing in me.
- My daughters, **Klara Malaika Johnson** and **Luise Amani Johnson**, for constantly reminding me of why it all matters.
- My parents, **William Alexander Johnson**, and **Theonie Enid Rita Johnson,** for instilling an appreciation and love from learning from an early ag

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION TO THE STUDY

The current study extended from three research areas in cognition and instruction: multimedia learning, expertise reversal effect, and generative drawing effect. This study added to each of these bodies of research with a focus on secondary-school science education. The study focuses on how to best tailor multimedia instructional design to meet the needs of novice and advanced students of secondary science. This first chapter, which includes seven sections, is an overview of the study. The first section presents the statement of the problem, emphasizing why this work is relevant to teaching and learning practices. The second section provides the purpose of the study, including an overview of the study design and the relevant independent and dependent variables. The third section provides a detailed review of the theoretical framework for the study. It begins with a summary of the key features of cognitive load theory (CLT) and the relationship of CLT to Baddeley's model of working memory. The section ends with a description of the cognitive theory of multimedia learning and the multimedia design principles. CLT forms the foundation of the cognitive theory of multimedia learning, the theory in which the current study is grounded. The fourth section is the Background and Need, which provides a detailed justification for the investigation by reflecting on relevant prior and ongoing research and teaching and learning practice. Finally, the last three sections provide the research questions, definition of terms, and the chapter summary.

Statement of the Problem

Understanding and appreciating science at even the most basic level requires a certain amount of imagination. Science deals with abstract ideas and concepts that are extensions of inferences derived from empirical data rather than actual observations.

When using imagination to construct mental images, objects that are typically invisible to the naked eye can be visualized in the mind's eye. To appreciate this point, consider a student who explores the concept of photosynthesis at the secondary-school level. Although the student might readily recognize that the leaves of plants are green, they might not understand why this is the case or how it relates to photosynthesis. To understand these concepts, one must construct mental images based on various principles about the plants' microscopic and molecular structures and physiology. Attempting to facilitate the process, the teacher might convey that chlorophyll pigments have a structure that uniquely supports their function, including light absorption and reflection within specific color ranges of the light spectrum. The teacher might further describe the light spectrum and how the energy captured by chlorophyll propagates through relevant biochemical pathways for eventual use in glucose synthesis. This information, however, is far too abstract and meaningless for the novice learner who has no way of visually and kinesthetically interacting with chlorophyll and other accessory molecules. If these molecules were observable at the macro level, the student could better appreciate the unique relationship between their molecular structure and function. For example, they would see the chlorophyll molecule specifically absorbing violet, blue, yellow, orange, and red lights while reflecting a green light ray. Moreover, they also would notice how the light absorption initiates a cascade of logically derived chemical events that culminates with glucose synthesis.

Conveying what happens in plants' leaves during photosynthesis requires many interacting information elements. The conceptual complexity needed to understand photosynthesis is common in other aspects of science education. Presenting the relevant information exclusively with words (orally or textually) is insufficient for novice learners who have yet to connect various crucial underlying information elements or (sub)concepts. With the use of models such as illustrations (visuals) and 3-D manipulatives (kinesthetics), however, the students can address confusion and misconceptions and fill in the missing gaps in their knowledge (McTigue & Slough, 2010, Meneses, Escobar, & Vélez, 2018). Besides conceptual abstraction in science, new matriculants also may find it encumbering that science texts use academic language that references unfamiliar concepts and complex causal relations (Meneses et al., 2018). Furthermore, they also must overcome the added challenge of deciphering domain-specific vocabulary and syntax that differ from conventional language (Uccelli et al., 2015). For example, although in chemistry, the chemical equations that represent chemical reactions share some similarities to the algebraic equations used in mathematics, the two types of equations differ in their syntax and symbols. In chemistry, the chemical equation for photosynthesis is,

$$6\mathbf{CO}_2 + 6\mathbf{H}_2\mathbf{O} \rightarrow \mathbf{C}_6\mathbf{H}_{12}\mathbf{O}_6 + 6\mathbf{H}_2 \qquad (1)$$

This equation can be represented algebraically in mathematics using three separate equations for each of the three chemical elements involved in the reaction.

$$\text{Carbon } (\mathbf{C}): \quad 6\mathbf{C} = 6\mathbf{C} \qquad (2)$$

$$\text{Oxygen } (\mathbf{O}): \quad 6(2\mathbf{O} + \mathbf{O}) = 6\mathbf{O} + 6(2\mathbf{O}) \qquad (3)$$

$$12\mathbf{O} + 6\mathbf{O} = 6\mathbf{O} + 12\mathbf{O}$$

$$18\mathbf{O} = 18\mathbf{O}$$

$$\text{Hydrogen } (\mathbf{H}): \quad 6(2\mathbf{H}) = 12\mathbf{H} \qquad (4)$$

$$12\mathbf{H} = 12\mathbf{H}$$

In chemistry, the → takes the place of the = to symbolize equivalency. The advanced learner will understand this syntactic difference and appreciate that the chemical equation conveys greater meaning than all three algebraic equations combined. In addition to representing quantitative equivalency, the chemical equation also represents a reaction process in which compounds change forms over time. Considering the relative complexity of the syntax and semantics conveyed in science writing, it is not surprising that the cognitive demand for processing science text can overwhelm novice learners and cause them to lose the motivation to learn.

The situation is far from hopeless. There is ample evidence that prior knowledge and literacy skills are key factors affecting science text comprehension (Kloser, 2016; McNamara, Ozuru, & Floyd, 2011). Teachers can help students overcome these challenges by considering their prior knowledge while planning and scaffolding instruction. For example, in a study by Kloser (2016), the investigators reported that when textual information included epistemic evidence (i.e., relevant empirical evidence) supporting the scientific claims, high-school students displayed better comprehension than traditional science text. This finding, which is similar to that of other investigators (e.g., Cartiff et al., 2020; Greene & Yu, 2016; Lin & Chan, 2018; Rosman et al., 2019; Vieira et al., 2017), suggests that science text embedded with information that logically justifies claims can facilitate meaningful learning. In the cognitive theory of multimedia learning, verbal and visual information are assumed to integrate to convey meaning (Mayer, 2002). Thus, it may be that such evidence-enriched text engages the learner in imaginative thought processes that promote the construction of self-generated visuals of abstract concepts. Epistemically supported claims might help learners recreate the cascade of molecular events from which the relevant scientific inferences originate. If this assumption is correct, engaging the learner's imaginative visualization about key concepts may be critical for science learning.

Cognitive load theory (Sweller, 2020), which explains human cognitive processing, may reveal why self-generated and borrowed visuals may improve science comprehension. According to this theory, the human mind can handle only between 5 and 9 distinct chunks of information (i.e., distinct information elements or concepts) at any given time (Miller, 1956). Learning cannot proceed if the number of interacting information elements exceeds this cognitive-load limit. In the photosynthesis example, the novice learner must negotiate a host of new science-specific terms that, taken separately, will exceed their cognitive capacity and prevent learning progression. Even for the advanced learner with some prior knowledge, many (sub)concepts (i.e., visible light propagation, light absorption, hydrolysis) must be understood separately to comprehend photosynthesis fully. Learning (sub)concepts compartmentally using words alone could entail far too many segmented chunks of information than the student can process simultaneously. The collapse of cognitive processing in science occurs when the learner cannot evaluate each information chunk separately before combining them to construct a visual representation of the overarching idea or concept (Kloser, 2016; Norris & Phillips, 2003). If the learner could construct an accurate visual of each (sub)concept, the breadth of information would consolidate into fewer chunks to reduce the cognitive-load. The generation of self-constructed visuals requires encoding skills to convey meaning from each bit of relevant information (McNamara et al., 1996). Once encoded, cognitive processes dynamically integrate all information while generating and refining the student's mental representations (McNamara et al., 1996).

The novice learner likely is incapable of independently generating visuals that involve multiple complex element interactions, but they could benefit from teacher-provided (borrowed) visuals. According to McTigue and Slough (2010) and others (e.g., Schnotz 2014; Uttal & O'Doherty, 2008), borrowed visuals help illustrate phenomena and processes that cannot be observed directly or show relations that words alone cannot articulate fully. Several studies on multimodal learning and cognition have confirmed that visuals positively affect cognitive processing (Kalyuga & Singh 2015; Mayer & Gallini, 1990; Mayer & Sims, 1994). There is a need, however, for more empirical evidence of the benefits of multimodal reading comprehension in high-school science literacy and the effects of embedded visuals on science academic achievement at the secondary-school level. Several prior studies have examined verbal texts and images separately with minimal consideration to their combined effect on meaning-making in science (Firat, 2017; Höffler & Leutner, 2007; Kühl et al., 2011; Meneses et al., 2018). For those studies (e.g., Kühl et al., 2011; Lin & Dwyer, 2010) that examined the interaction of text and visuals, the focus was on college- and university-age students rather than school-age students. The current study focused on secondary-science education to investigate how visuals affect students' ability to retain what they learn (rote learning) and transfer the new knowledge to an unfamiliar problem (meaningful learning). Furthermore, the study will explore whether the multimedia effect varies depending on the mode of visuals (i.e., static versus dynamic visual) and text (i.e., audio versus print) on learning.

Purpose of the Study

The purpose of this study was to investigate how prior knowledge and the integration of information modalities (i.e., text, audio, static visual, dynamic visuals) can promote rote learning (information retention) and meaningful learning (knowledge transfer) in science. The finding presented in this body of work extended from data previously collected in an experimental pretest-posttest study.

The independent variable for this study was the mode of information presentation, which had four levels: (a) Text only (control), (b) Text with embedded pictures, (c) Subscripted animation (i.e., text and animated video), and (c) Full animation (i.e., audio and animated video). The four dependent variables were cognitive-load, rote learning, and meaningful learning. Cognitive load was measured using an instrument that evaluated mental effort, extraneous load, and germane-load. Rote learning was measured using an instrument that included a conceptual recall and a nonconceptual recall questionnaire. Finally, meaningful learning was measured using an instrument that included a knowledge transfer questionnaire. This study adds to the growing research on multimedia learning literature that visual representation of scientific concepts can enhance literacy and conceptual comprehension in science and how this effect interacts with the learner's prior knowledge to effect learning.

In the learning-styles literature, the term visual learner describes an approach to learning based on the user's preference for a visual presentation of information such as graphs, concept maps, illustrations, or other images (Dunn, 2003; Fleming & Mills, 1992). The term visual learner is now so commonly used by educators and the public that it often loses meaning and relevance regarding teaching and learning practices. What is more, several investigators provide empirical evidence that the various learning-styles lack validity and reliability as tools for predicting learning outcomes (Coffield et al., 2004; Kirschner, 2017; Veenman et al., 2003). A growing body of research in multimedia learning, however, supports the notion that integrating visuals with words into learning material can improve information retention and knowledge transfer (Mayer & Pilegard, 2014), irrespective of the learner's self-identified learning-style. Although many teachers intuitively might appreciate that integrating information from multiple modalities (i.e., visual, sound, kinesthetic) can support learning, they may not always point to empirical evidence that supports this claim. The current study considered multimedia learning effectiveness by evaluating how embedding static or dynamic visuals with text or auditory-based information can affect student learning in science.

The work presented in this study will help address misconceptions concerning the learning-style hypothesis when emphasizing the importance of cognitive load theory for elucidating how learning happens. Even though the evidence against the learning-style hypothesis (see Coffield et al., 2004; Kirschner, 2017; Veenman et al., 2003), the concept remains popular among teachers. For a student who self-identifies as a visual learner (i.e., visual learning-style), there also can be lasting benefits to developing the skills needed to work with verbal information. By assigning this student to a particular learning-style, they might lose out on developing these other skills. Students who gain proficiency at using various information modalities may have more opportunities to broaden their knowledge and understanding because they may be better equipped to access information from a broader range of sources. Based on the existing empirical support for the dual-channel concept (Baddeley's model; Baddeley & Hitch, 1974), there also could be immediate benefits to the learner when the instructional material incorporates information that uses both words and images. According to Baddeley's model, when the learner uses multimedia content that includes words and pictures, information flows through the visual and auditory channels, rather than a single channel, to improve information processing efficiency (Baddeley, 2000, 2013; Baddeley & Hitch, 1974). Thus, the current study evaluated whether the modality effect (i.e., integration of information modalities) can improve science learning outcomes.

In addition to the modality effect mentioned above, the study also explored how prior knowledge interacts with visual representations on learning outcomes. According to Schwambornet al. (2010), when learners create drawings while reading, they activate their generative processing, resulting in improved learning. Van Meter and Garner (2005) described how generative processing happens during learning. Initially, the learner purposely selects relevant information from the provided text and organizes the selected information into a verbal model. Then, through a dynamic process, the learner uses both the original text and the verbal model to guide them in constructing a coherent visual representation.

Although generative learning may work for advanced learners, a novice learner may find the provided text far too complicated for effective visualization (Carney & Levin, 2002). In this case, the novice learner might benefit from teacher-provided (i.e., borrowed) visuals that complement the text. The current study evaluated this assumption by examining how prior knowledge interacts with teacher-provided visuals to affect learning. Taken together, the findings from this study highlight the need for additional research on engaging students in generative drawing as practical learning tools. Additionally, by contrasting the performance of advanced and novice learners under varying instructional support (presence or absence of borrowed visuals), the current study could add to the literature on the expertise reversal effect (see Kalyuga, 2014)

Theoretical Framework

The current study had its theoretical basis in the cognitive theory of multimedia learning (Mayer, 2002), which incorporates assumptions from cognitive load theory and Baddeley's model of the working-memory (Baddeley 2000, 2013; Baddeley & Hitch, 1974). This section, which provides an overview of the theoretical framework for the study, comprises four parts. Cognitive load theory is presented in the first part as the foundational theory for explaining human cognitive processing. In the second part, Baddeley's model is described, detailing the importance of the dual-channel subsystems (reviewed in Baddley, 2000) to cognitive load theory and an understanding of human cognitive architecture. In the third part, the cognitive theory of multimedia learning is introduced, with a brief overview of its significance to the study is provided. Finally, a summary of how cognitive load theory, Baddeley's model, and cognitive theory of multimedia learning is relevant to teaching and learning practices are found in the last.

Cognitive load theory

Cognitive load theory (Sweller, 2015; Sweller et al., 2011; Sweller et al., 1998) is a novel theory for explaining how the human mind processes information during learning, thinking, and problem-solving. It follows one's understanding of human cognitive architecture, including the structure and function of sensory memory, the working-memory, and long-term memory. Learning, which enhances one's ability to engage intellectually with the environment, relies on the information from prior knowledge stored in long-term memory. Intellectual activity stalls without this requisite information and its encoded knowledge (Sweller, 2020). Cognitive load theory provides a plausible explanation for how the flow and processing of information happen during learning. As such, it is the basis for a variety of experimental research focused on identifying factors that hinder learning and for developing strategies to alleviate the effects of adverse factors (for a summary of relevant instructional strategies, see Kalyuga, 2015; Sweller et al., 1998; Sweller et al., 2011).

Knowledge acquisition

According to cognitive load theory, knowledge is a collection of all information stored in long-term memory (Sweller et al., 2011). When one receives new information from the environment, they engage the mental processes that regulate learning and promote knowledge expansion. Learning happens deliberately or innately, depending on whether it builds primary or secondary knowledge. Primary knowledge is adaptive biologically and encodes the skills human ancestors evolved an innate predisposition to learn (Geary & Birch, 2016). Because of their importance to survivorship, these ancestors needed to learn such skills quickly and effortlessly. Although the primary knowledge for encoding them requires learning, in most cases, we acquire them innately without the need for explicit instruction and study (i.e., deliberate learning; cf., Geary, 2008; Sweller, 2020). In instances of learning difficulties associated with their acquisition, however, instructional intervention needs consideration (Prasada, 2000).

The end goal of teaching and learning is to expand secondary knowledge. Whereas primary knowledge encodes biologically adaptive skills applicable to multiple domains (Sweller, 2015), secondary knowledge is domain-specific and encodes culturally relevant behavior and skills (for a review, cf., Prasada, 2000; Sweller, 2015). In other words, secondary knowledge supports the learner with orienting and navigating various aspects of the sociocultural environment that is unique to their home or community. Therefore, secondary knowledge must be learned deliberately by studying relevant domain-specific concepts (Kirschner et al., 2006).

Schema. The knowledge generated and stored in long-term memory arises from unique experiences and differential study of secondary information. Initially, one's (prior) knowledge comes almost exclusively from primary biological information, and the process of learning is the ongoing alteration and expansion of this information (Sweller, 2020; Sweller & Sweller, 2006). Owing to the uniqueness of each person's environment and experiences that cause the changes, this alteration can vary among individuals. Consequently, each person develops an ever-changing and uniquely complex information web of prior knowledge. Furthermore, according to Sweller (2020), as the information evolves, it dynamically influences how one acquires and constructs future knowledge. As such, any future learning requires this constant state of information alteration.

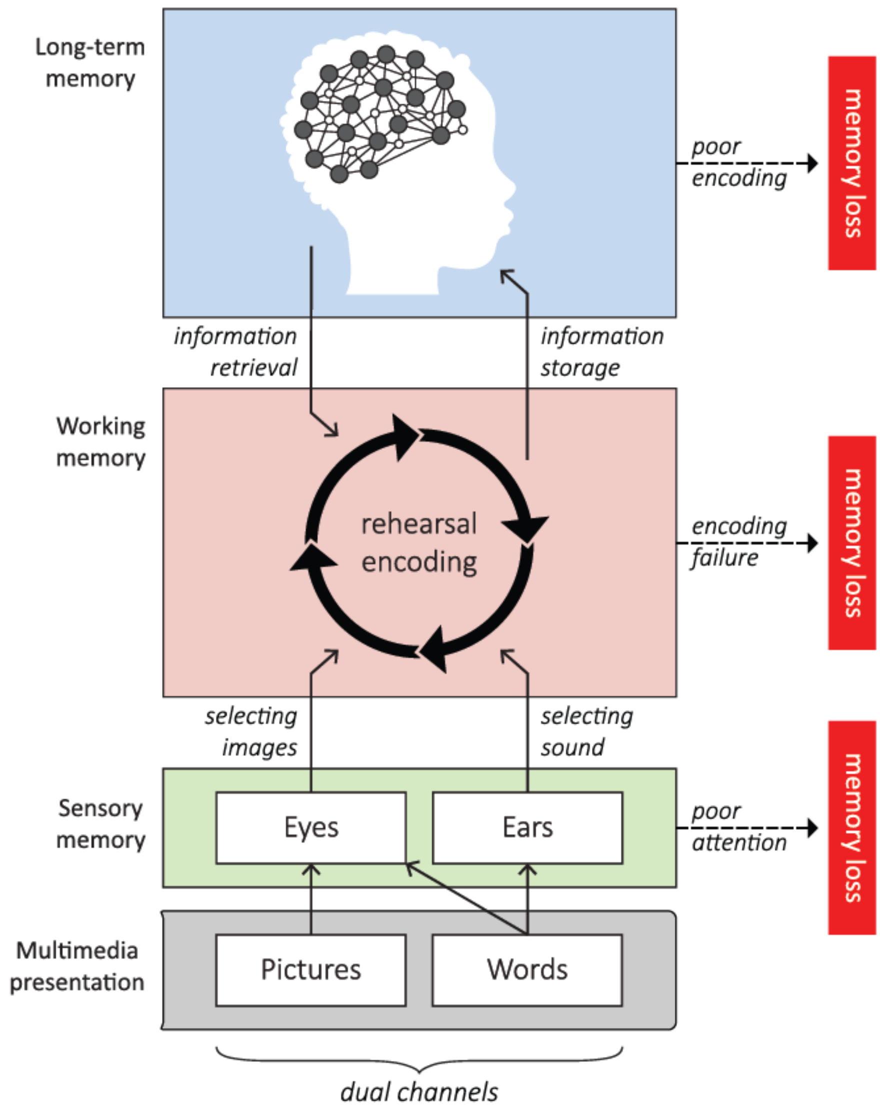
The information web stored in long-term memory is the full embodiment of the learner's primary and secondary knowledge store (Geary, 2008). Although learners can generate secondary knowledge, they borrow most from other people (e.g., teachers, book authors, film producers; Geary, 2008). Borrowed information can alter the learner's prior knowledge to convey new meaning or improve or diminish existing meaning. The encoded meanings, which are imperfect schematic representations of reality, helps to make sense of the world (Sweller et al., 2011).

In cognitive load theory, knowledge is a collection of interacting information elements (Sweller et al., 2011). The interactions between these elements produce knowledge structures called schemas that organize the information in long-term memory. For example, a particular schema could be a simple two-element interaction with minimal context for assigning meaning or a chunk of information made of multiple interconnected subschemas that together can convey complex meaning. Thus, according to the borrowing and reorganizing principle (Sweller et al., 2011), the goal of learning may be to construct increasingly chunkier schemas by fusing independent subschemas.

Once acquired, a schema is stored in long-term memory until retrieved to help make sense of new information (Sweller et al., 2011). When learners study a particular concept, they retrieve and refine the relevant schemas, improving understanding and concept mastery. Mastery is an effect of schema automation, which happens after repeated schema retrieval and rehearsal (Sweller et al., 1998). Once automated, the learner needs not to exert conscious effort to engage with the schema. Therefore, automation frees up space in the working-memory for nonautomated schemas because the learner needs to only momentarily mount the automated schema during moments when cognitive processing requires it. Ultimately, automation reduces the response time for retrieving the schema from long-term memory, thus lowering cognitive demand and enhancing the processing power and interpretation (Guida et al., 2012; Martin & Evans, 2020).

Information reservoirs. According to cognitive load theory, the processing, integration, and storage of schema involve three memory reservoirs: long-term memory, the working-memory, and sensory memory (Sweller et al., 2011). Each of these reservoirs is distinguishable by its function, limits on the amount of information they hold (information capacity), and the length of time to hold the information (temporal span).

How information flows between the three memory reservoirs is illustrated in Figure 1 (for a review, e.g., Atkinson & Shiffrin, 1968; Mayer, 2014a; 2014b). Long-term memory stores all knowledge until a specific schema is needed to process new information (Sweller et al., 2011). Long-term memory has neither a limit on information capacity nor a limit on temporal span (Bahrick et al., 1975; Sperling, 1960). Nevertheless, it lacks information-processing capability (Moreno & Park, 2010; Sweller et al., 2011). Information processing happens in the working-memory and sensory memory (Atkinson & Shiffrin, 1968; Baddeley 1992). Sensory memory, which has an unlimited information capacity and fleetingly short temporal span, receives all new information from the environment. The learner's selective attention determines which incoming information elements will transfer from the sensory memory to the working-memory. Those information elements that are ignored rapidly decay and fade away from memory.



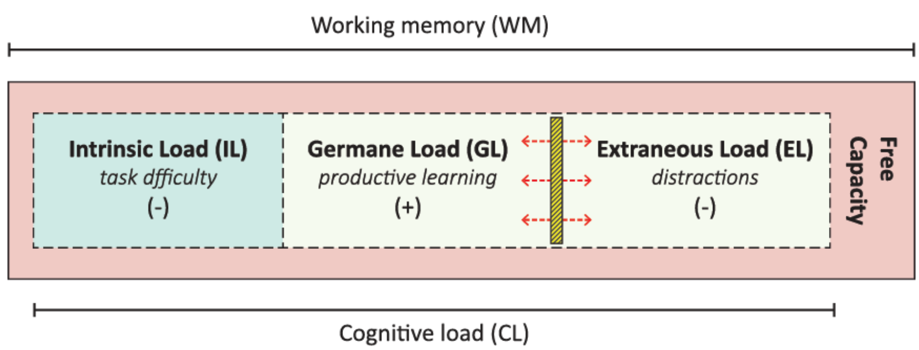
Note: Information flows through the dual visual-audio channel from sensory memory to the working-memory where it is processed prior before transferring to long-term memory. The figure illustrates the Atkinson and Shiffrin model of human cognitive architecture. Based on the model by Atkinson & Shiffrin, 1968, p. 93 and Mayer, 2014b, p. 66.

Figure 1. The human cognitive processing model

Working-memory has a finite information capacity of 7±2 chunks of information (Miller, 1956) and a temporal span that extends for the duration of conscious interaction with the information. In the working-memory, new information is reconciled with prior knowledge to construct a new schema or refine preexisting schema (Sweller et al., 2011). Understanding the architecture of the working-memory and the mechanism through which it operates remains an active area of research. The most promising model for elucidating how information is filtered and passed on for the working-memory processing is Baddeley's model (Baddeley 2000, 2013; Baddeley & Hitch, 1974).

The architecture of working memory

Cognitive load theory investigators have sought to decipher the architecture and function of the working-memory to understand the factors that regulate learning. The demand (cognitive-load) that learning puts on cognition restricts the information processing capacity of the working-memory (Hanham, Leahy, & Sweller, 2017; Sweller et al., 2011). A visual representation of cognitive-load is found in Figure 2., which shows that cognitive-load comprises three distinct parts: intrinsic (cognitive) load, extraneous (cognitive) load, and germane (cognitive) load (Sweller, 2010). Furthermore, germane and extraneous loads have a conflicting effect on learning because they occupy a shared space separated by a dynamic boundary (Hanham et al., 2017; Sweller et al., 2011).



Note: According to cognitive load theory, cognitive-load (CL) operates within the working-memory. Total CL is derived from the task difficulty (Intrinsic load; IL), the extent of productive learning (Germane load; GL), and distractions (Extraneous load; EL). GL and EL dynamically share a space within CL such that when either one increases, the other decreases. (+) and (-) indicate a positive or negative effect that the types of loads have on learning, respectively. Modified from Moreno & Park, 2010, p. 18.

Figure 2. Working-memory and cognitive load theory

Intrinsic load. Intrinsic load represents the complexity (or difficulty) of the learning information or learning task (Pollock, Chandler, & Sweller, 2002), which varies directly with the number of interacting elements involved (Hanham et al., 2017; Sweller et al., 2011). This element-interactivity (i.e., information or task difficulty and complexity) is intrinsic to the learning information or task and, thus, unaffected by differential instruction (Mayer & Moreno, 2010; Paas et al., 2003). When cognitive demand exceeds the learner's cognitive load, however, the teacher can reduce interactivity by scaffolding instruction (Wong et al., 2020). For example, the teacher could divide the (grand) schema into smaller, more manageable subschemas, allowing the learner to acquire each relevant subschema separately before combining them into the (grand) schema later. Until this (grand) schema is acquired, the learner cannot fully appreciate the overarching concept, and learning stalls or remains incomplete. Thus, teachers must understand how elements within a particular domain convey meaning and consider element interactivity when developing learning plans and lessons.

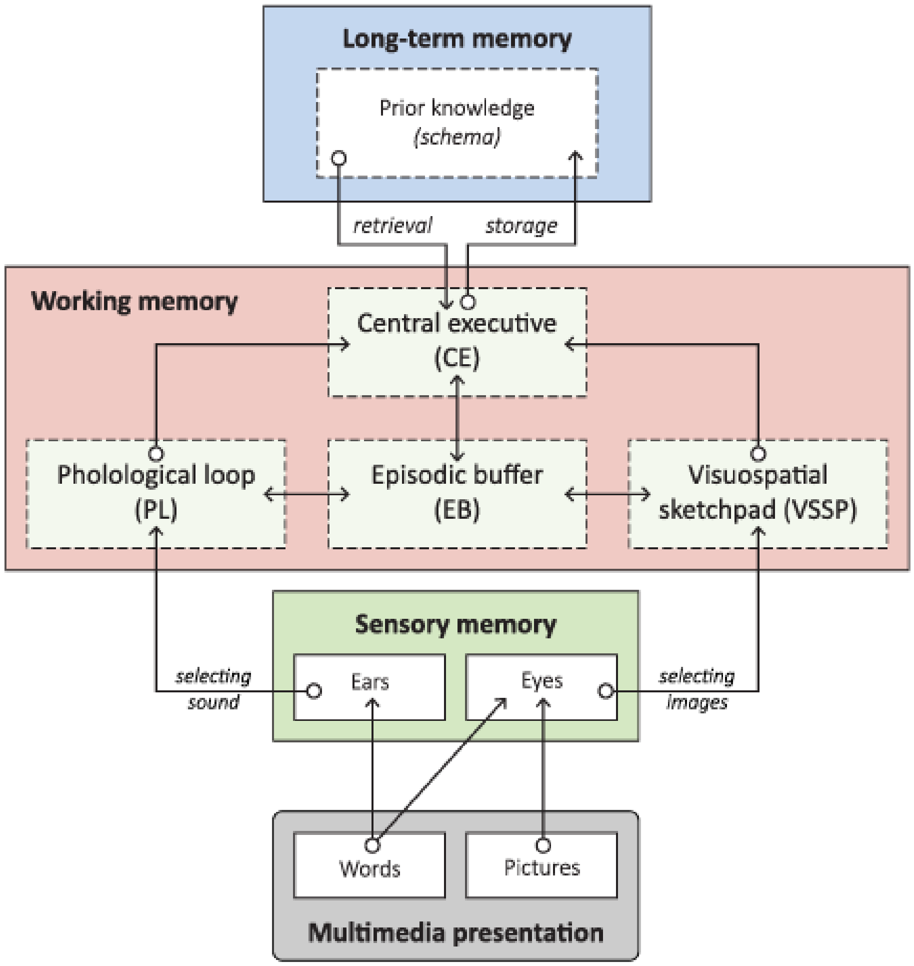
Extraneous and germane-loads. The learner must negotiate the relevance of various information elements to the learning task during the learning process. Not all of the information received in sensory memory will be relevant. Irrelevant information includes negative factors that constitute the extraneous load; they require additional and unnecessary processing that lowers the extent of meaningful learning (Sweller, 2010). Meaningful learning happens because of the germane-load capacity: the cognitive space for productive information processing of the relevant conceptual information (Sweller et al., 2011). Germane-load measures the amount of mental work involved during learning, including the effort needed to process relevant information and construct schemas. As illustrated by Figure 2 (p. 15), extraneous load and germane-load are linked dynamically, thus, the extraneous load is counterproductive to learning. During instructional design, the teacher can minimize the extraneous load by avoiding negative contributing factors, such as those associated with classroom structure (e.g., student grouping, seating arrangement), classroom culture, instructional format, and the mode of information presentation and instructional delivery (Eitel et al., 2020). Thus, although instructional design cannot alter germane-load directly, it can facilitate germane-load by reducing extraneous load and using appropriate instructional techniques that guide the learner in practices that facilitate germane-based cognition. The current study applies two different scales that are based on self-assessment surveys to measure extraneous load and germane load.

Baddeley's Model of Working Memory

The multimedia approach to teaching and learning is common in modern classrooms. Teachers use instructional and learning material constructed with various sound, text, and visual modalities in multimedia learning. For example, in a biology class, students might explore the concept of natural selection using computer-simulated laboratory activities that have embedded animations, graphs, diagrams, or audible narratives. These embedded features support students by managing the high number of interacting elements, allowing them to construct or upgrade relevant schemas more easily. Many computer-aided learning resources also have interfaces with embedded exercises that scaffold students' discovery of relevant concepts and the relationships between various conceptual elements.

Multimedia learning fits well with the concept of schema formation as described in cognitive load theory (Sweller, 1998), mainly when the teaching resources include visuals that help the learner organize and make sense of the conceptual information. In this regard, visuals are physical expressions of schemas that the teacher already has stored long-term memory and wants to transfer to the learner. Although cognitive load theory incorporates the idea of selective attention in the sequestration of information from sensory registers, it has yet to explain how the working-memory deals with different information modalities. Instead, proponents of cognitive load theory often rely on Baddeley's model of the working-memory (Baddeley, 2000, 2013; Baddeley & Hitch, 1974) as a theoretical framework for assumptions about multimedia learning.

In Figure 3, Baddeley's model depicts the working-memory as an information processing system comprising four subsystems: (a) central executive (CE), (b) phonological loop (PL), (c) the episodic buffer (EB), and (d) visuospatial sketchpad (VSSP). In agreement with cognitive load theory, Baddeley (2013) proposed that new information from the surroundings first mounts onto sensory memory. Then, the learner's selective attention determines which newly arriving information elements will transfer to the working-memory. In the working-memory, the arriving information mounts at either the PL or the VSSP. The PL receives and processes verbal information, and the VSSP deals exclusively with visual and spatial information. Although the VSSP initially receives text as visual elements, all textual information converts to verbal code and then transfers to the PL for further processing and storage. In other words, when one reads a text, the working-memory immediately converts the text to a virtual sound and passes it on from VSSP to the PL. The EB is the intermediary between the two channels, allowing information to switch between verbal and visual modes (Baddeley, 2000). The CE is where learning (i.e., schema construction and alteration) occurs (Baddeley, 2013); it has four critical responsibilities: (a) monitoring and coordinating the activities of the other
four critical responsibilities: (a) monitoring and coordinating the activities of the other



Note: Baddeley's model depicts the working-memory as an information processing system comprising four subsystems: (a) central executive (CE), (b) phonological loop (PL), (c) the episodic buffer (EB), and (d) visuospatial sketchpad (VSSP). The model proposes that information flows through a dual visual and audio channel that the VSSP and PL moderate. Modified from Baddeley, 2000, p.418.

Figure 3. Baddeley's model of the working-memory

three subsystems and connecting them to long-term memory, (b) regulating attention, (c) transferring information through and between the three memory reservoirs (i.e., sensory memory, the working-memory, and long-term memory), and (d) encoding information.

Cognitive load theory incorporates two assumptions from Baddeley's model: (a) the working-memory has a dual subsystem for processing verbal and visuospatial information (Mayer, 2014b) and (b) each subsystem has a limited capacity for processing information (Baddeley, 1999). Although cognitive load theory considers both, it emphasizes assumptions about the limited capacity over assumptions about the dual-channel (Shuler et al., 2011). These borrowed assumptions fit nicely into the broader claim that the working-memory has a 7±2 capacity limit (Miller, 1956). Baddeley's (2000, 2013) assumptions explain why multimedia modes of information presentation could benefit learners. If the dual subsystem exists, then cognitive load theory should accommodate Baddeley's model in the working-memory construct. The two channels should fit within the space dynamically shared by germane and extraneous loads because the subsystems process and encode information. Together the four subsystems constitute the 7±2 capacity limit of the working-memory (Miller, 1956), and each has a limited capacity for handling information (Baddeley, 1999). Understanding how the role of the dual-channel subsystems is relevant to teaching and learning because if the learning material is presented exclusively via a single information modality (e.g., teacher's oral presentation), only one of the learner's sensory registers (e.g., PL) engages. The other register (e.g., VSSP) remains dormant, reducing the working-memory capacity. Based on this assumption, it could be argued that teachers should design lessons and select learning material that maximizes students' opportunities to engage the two sensory registers. Embracing a multimedia approach to teaching and learning grounded in the cognitive theory of multimedia learning may facilitate the proper use of cognitive resources while learning. The current study evaluated this assumption by comparing learning outcomes under mono- and multimedia modes of information presentation.

Cognitive Theory of Multimedia Learning

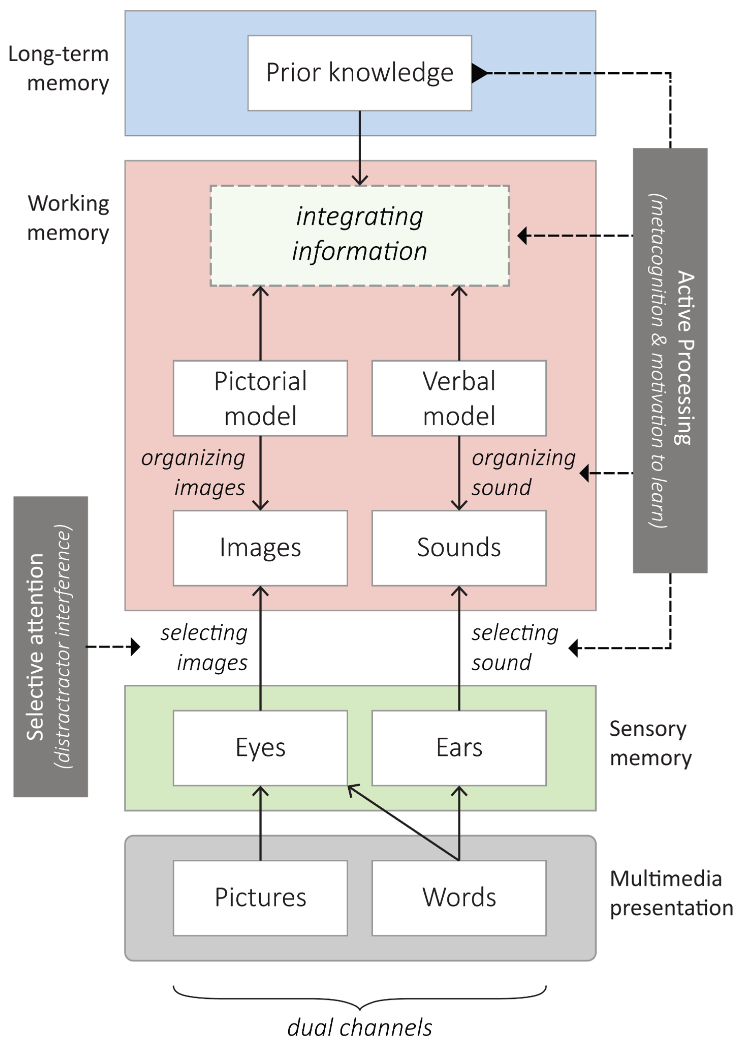
The use of multimedia resources in teaching and learning practices is now an everyday occurrence in classrooms. The multimedia hypothesis from the cognitive theory of multimedia learning states that learning occurs more readily when the information integrates words and visuals (Mayer, 2002a). In multimedia learning, the learner constructs mental representations from words (e.g., printed and aurally narrated text) and visuals (e.g., illustrations, animations, photos, videos, diagrams, graphs, charts), then integrates them with relevant prior knowledge. Although the benefit of learning from multiple information modalities might seem obvious, teachers and resource developers do not always apply empirically grounded multimedia design principles (Van Merrienboer & Kester, 2014). Instead, they seem to assume that merely embedding words with pictures in text-based learning material is sufficient for appropriately scaffolding instruction (Mayer & Massa, 2003).

The dual-channel system

The cognitive theory of multimedia learning is emerging as one of the most promising theories for applying evidence from cognitive load theory to educational practices. Although grounded in cognitive load theory, when addressing how different information modalities propagate through the working-memory, the cognitive theory of multimedia learning borrows from two prevailing working-memory models. The first (representational model) is Pavio's dual-coding theory (Pavio, 1986), which assumes that the working-memory operates via a dual-channel system with a verbal channel that processes language and a non-verbal channel that processes non-articulate sound and images. The second (sensory-modality model) is Baddeley's model (Baddeley, 2000, 2013; Baddeley & Hitch, 1974), which also includes a dual-channel system; however, the channels begin with information received as sensory signals (i.e., words and pictures) at receptors in the ears and eyes, respectively, before assignment to either a verbal or pictorial category. The two working-memory models are similar in their assumptions of how information handling occurs. Thus, further assumptions about the working-memory will follow Baddeley's model for simplicity and consistency.

Information processing and propagation

The cognitive theory of the working-memory (Figure 4) assumes that information processing occurs through a dual audio-visual channel in the working-memory (Mayer, 2014a; 2014b). Before transfer and storage in long-term memory, the channeled



Note: Mayer's model of human cognitive architecture incorporates the dual-channel assumption from Baddeley's model. The learner's selective attention filters incoming information in favor of germane elements to the learning task. The learner's metacognition and motivation to learn and the active processing of the selected information are critical during the information integration process. Modified from Mayer, 2014b, p.66.

Figure 4. Mayer's model of human cognitive architecture

information integrates with prior knowledge to form new or expand preexisting schemas. For example, during instruction or study, the learner receives information from a multimedia presentation (e.g., textbook, video, lecture) as either word (i.e., aural narration, printed text) or pictures: printed text initially enters the sensory memory via photoreceptors in the eyes. In contrast, spoken words and other sound elements enter through receptors in the ears. Ideally, multimedia presentations should only include information that promotes the intended learning objectives. Unfortunately, even under ideal circumstances, elements that distract from learning are unavoidable; yet, the sensory memory information decays so rapidly that it permanently disappears if unattended for even a brief period (Bahrick et al., 1975; Sperling, 1960). Thus, the learner could benefit from developing strategies for sifting through the sea of incoming information to identify those few relevant elements for transfer to the working-memory.

Cross-channel representations

Once transferred to the working-memory, sound and visual elements could undergo cross-channel representation to convey additional meaning (Mayer, 2014b). For example, when the learner reads a printed passage describing an insect, they initially receive all of the information as a series of separate visual elements (shapes of the letters) through the visual channel (i.e., photoreceptors in the eye). Then, in the working-memory, the learner metacognitively converts the pictures (letters and words) to sound elements for redirection from the visual channel to the audio channel. Finally, active cognitive processes (Mayer, 2009; Wittrock, 1989) integrate the converted auditory elements and relevant prior knowledge to convey the passage's intended meaning.

Further processing of the meaningful information could return the information to the visual channel when the learner virtually "hears" the description and imagines an image of the insect. According to Baddeley's model, the visuospatial sketchpad (VSSP) is the visual channel subsystem, the episodic buffer (EB) is the cross-channel exchanger, the phonological loop (PL) is the auditory subsystem, and the central executive (CE) is the information integrator and meaning encoder (Mayer, 2014a; 2014b). Information from the descriptive passage propagates from sensory memory (SM) through the working-memory: SM → VSSP → CE → EB → PL → CE → EB → VSSP. The information initially flows from sensory memory in the optic neurons to the visual channel, the visuospatial sketchpad (VSSP). Continued propagation through this visual channel moves the information to the central executive (CE). Each visual element (i.e., the letters in the text) decodes to convey phonetic meaning associated with the sounds they represent. The episodic buffer moderates dynamic interaction between the central executive and the two channels to decode and integrate the information elements into a coherent meaning. For example, the decoded information elements convert to auditory information from the central executive and proceed to the episodic buffer (EB1) for crossover to the auditory channel, that is, the phonological loop. From the phonological loop, they can return to CE for imaginative processing to form an insect image based on the meaning encoded by the passage's description. The information then switches back over to the visual channel (VSSP). Similarly, information initially received as sound at the phonological loop could reversibly switch between the two channels.

Multimedia Design Principles

Teachers who know about cognitive load theory and cognitive theory of multimedia learning may be better equipped to select and develop learning material that effectively supports student learning. An instructional design that adheres to multimedia design principles, however, may improve learning outcomes without a complete theoretical understanding. Twelve basic principles of multimedia design discussed in the literature are provided in Table 1. These basic principles form the foundation of several other advanced principles, including the generative drawing principle and the expertise reversal principles. The generative drawing principle is "when learners create drawings while reading text, they initiate generative processing that leads to better learning outcomes" (Leutner & Schmeck, 2014, p. 434). This principle relates to the idea that constructing self-generated images requires imagination to facilitate the cognitive integration of relevant information elements into schemas. Thus, when students create drawings while learning, they display some aspect of their cognitive processing that hints at learning progression. The other advanced multimedia design principle is the expertise reversal principle, which states that "high-information instruction is beneficial for novices when compared with the performance of novices who receive a low-guidance format but disadvantage for expert learners, when compared with the performance of experts who receive a low-guidance format" (Kalyuga, 2014, p. 579). During (grand) schema construction, the learner builds and modifies multiple subschemas that are later assembled to convey the big idea or concept of the (grand) schema (Wong et al., 2020). As a result, some subschemas will remain incomplete even for the advanced learner who is not yet an expert. The gaps in knowledge will fill organically as new relevant information arrives and future cognitive processing occurs during information retrieval and rehearsal.

Because science concepts often are abstract, as advanced learners construct subschemas, they are almost always doing so based on incomplete information perceived or interpreted using imagination (i.e., generative visualization). These subschemas are not necessarily incorrect; however, they only partially convey the full meaning of the (sub)ideas or (sub)concepts that they represent. The incomplete jigsaw of (sub)concepts

Table 1

12 Basic Multimedia Design Principles

Principles	Description
Multimedia principle	"Learning with words and pictures is more effective than learning with words alone." (Butcher, 2014, p. 175)
Modality principle	"Under split-attention conditions, presenting some information in visual mode and other information in auditory mode can expand effective the working-memory capacity, reducing excessive cognitive-load" (Low & Sweller, 2014, p. 227).
Redundancy principle	"Presenting the same information concurrently in multiple forms (or unnecessarily elaborating on previously presented information) can interfere with rather than facilitate learning" (Kalyuga & Sweller, 2014, p. 247).
Signaling principle	"Multimedia learning materials are more effective when cues are added that guide learners' attention to the relevant elements of the material or highlight the organization of the material" (Van Cog, 2014, p. 263).
Coherence principle	"People learn more deeply from a multimedia message when extraneous material is excluded rather than included" (Mayer & Fiorella, 2014, p. 280).
Spatial contiguity principle	"People learn more deeply from a multimedia message when corresponding words and pictures are presented near rather than far from each other on the page or screen" (Mayer & Fiorella, 2014, p. 280).
Temporal Contiguity Principle	"People learn more deeply from a multimedia message when corresponding animation and narration are presented simultaneously rather than successively" (Mayer & Fiorella, 2014, p. 280).
Segmenting principle	"People learn more deeply when a multimedia message is presented in learner-paced segments rather than as a continuous unit" (Mayer & Pilegard, 2014, p. 317).
Pretraining principle	"People learn more deeply from multimedia messages when [given] names and characteristics of the main concept" (Mayer & Pilegard, 2014, p. 317).
Personalization Principle	"People learn more deeply from a multimedia message when words are in conversational style rather than formal style" (Mayer, 2014c, p. 345).
Voice Principle	"People learn more deeply from multimedia messages when words are spoken in a human rather than in a machine voice" (Mayer, 2014c, p. 345).
Image Principle	"People do not necessarily learn more deeply from a multimedia message when the speaker's image is on screen rather than not on the screen" (Mayer, 2014c, p. 345).

interact dynamically as the learner integrates them into a (grand) schema. Confusion can arise when the teacher provides redundant information differently or includes additional information that the learner cannot resolve. Any number of the established subschemas previously encoded and integrated with prior knowledge might be affected, risking that the (grand) schema's framework becomes unstable and unraveling. Thus, according to the expertise reversal effect (Kalyuga, 2014), although it is tempting for teachers to provide students with a full basket of information, they should avoid redundancy when working with advanced learners. Depending on the complexity of the instructional concepts, there are times when less is indeed more with these advanced learners. The current study investigated this claim by evaluating the effects of learning level of knowledge expertise on learning outcomes.

Background and Need

Using anecdotal and empirical evidence, teachers and researchers have long pondered the idea that each person has a unique set of traits that determine how they interact with the environment and learn (Knoll et al., 2017; Mayer & Massa, 2003; Pashleret al., 2009). The variation in traits extends to a range of intrinsic and experiential factors associated with age, prior knowledge, values and beliefs, interests, motivation, attitude, culture, and intelligence. These differences present unique challenges for teachers who plan and design lessons for heterogeneous populations of students with varying learning needs. Presumably, with knowledge of the entire catalog of influencing factors, teachers might tailor instruction perfectly to the students' needs. There could be as many unique learning-styles as unique individuals (Coffield et al., 2004), however, complicating any effort at individual accommodation. Instead, to tackle the issue of individual differences, educational psychologists have identified three mental functions—cognition, conation, and affection—that collectively encapsulate the range of influencing factors on learning (Snow et al., 1996). Because secondary knowledge expansion is the goal of learning, these mental functions must affect learning and may require instructional design consideration. The current study focused specifically on cognition, which encapsulates several other (sub)functions (e.g., thinking, knowledge recall, knowledge transfer, evaluation, reasoning, problem-solving) involved with learning (Sweller et al., 2011). Although an indepth discussion is unwarranted here, it is worth noting that the conation factors (i.e., motivation and volition to learn) and affective factors (i.e., temperament and emotion) can alter the learner's selective attention and active processing of task-relevant information (Lavie et al., 2004; Mayer, 2011; Snow et al., 1996; also, see Figure 4, p. 22). As such, conation and affection are peripherally relevant in the current study.

Literacy is a crucial factor in science education influencing cognition and conceptual understanding of concepts. In the secondary-school classroom, multimedia information presentation is commonly used to promote science literacy. The remainder of this chapter offers an overview of literacy challenges in secondary-school science education rooted in domain-specific conceptual complexity and the human cognition limits described by cognitive load theory. With a focus on the cognitive theory of multimedia learning as a theoretical framework, this section also includes information about why students typically struggle with understanding science concepts and how the Next Generation Science Standards (NGSS; National Research Council, 2013) is tackling the issue. It culminates with a description and justification for the current study by associating promising empirical evidence for generative drawing and the expertise reversal principles as potential literacy intervention considerations in secondary-school science education.

The Challenge of Literacy in Science

In 2012, the National Research Council published A Framework for K-12 Science Education (National Research Council, 2012), a guide for developing and implementing science education programs with holistic consideration of science knowledge and practices. The NGSS (National Research Council, 2013), an extension of this framework, outlines the grade-level competencies based on three principal dimensions: (a) practices of science and engineering, (b) crosscutting concepts that unify science and engineering, and (c) core ideas from the physical sciences, life sciences, earth and space sciences, and engineering, technology and applications of science. The goal of these new science standards was to establish research-based benchmarks that science educators could integrate into their curriculum and lesson design to stimulate students' interests in science and expand their science knowledge and skills (National Research Council, 2012, 2013).

The work leading up to the NGSS is an ongoing and enduring effort extending back since the 1990s (National Research Council, 2012; Sadler & Brown, 2018). During this period, academics, policymakers, and teachers sought to counter the growing illiteracy level in science (for a historical review, see Sadler & Brown, 2018). Therefore, there is a need to articulate a consensus description of science to tackle the problem effectively. Although lacking a concrete and straightforward definition, one can conceptually perceive science as a system that builds and organizes humanity's collective knowledge of the natural world (Wilson, 1990). Nevertheless, because of the vastness of the universe, there is unlimited information yet to be discovered. Therefore, the National Research Council committee recognized a need to update instructional programs with new information and understandings and correct previous misconceptions and inaccuracies with ongoing discoveries.

There were two issues that the K-12 framework committee had to overcome in their revamping the science curriculum: (a) how to structure science-education programs to accommodate the growing breadth and depth of scientific knowledge and (b) how to accommodate the need for continuous and dynamic updates of scientific knowledge and skills based on current and future scientific discoveries. NGSS, which reflects the National Research Council K-12 science-curriculum framework, promotes a trend toward learning scientific practices and concepts centered on a few core ideas (depth of knowledge) over the rote learning of many scientific facts (i.e., breadth of knowledge; National Research Council, 2012). Accordingly, the curriculum framework promotes the idea that science education should provide opportunities for students to ask questions, solve problems, construct and test models, investigate and explain phenomena, collect and analyze data, make inferences, and communicate ideas and understanding (AAAS, 1993; NGSS, 2013). By engaging in these practices, students will better understand how scientists work as they become skilled at critically evaluating scientific works to discern meaning. As such, rote learning of facts is secondary, whereas the overarching core ideas in science and unifying concepts are the anchor points of an NGSS-based science curriculum (Chessnutt et al., 2018; Krajcik et al., 2014). Factual information is still essential, but only when it supports students in developing their understanding of the core ideas and unifying concepts.

The National Research Council claims that the NGSS standards and the National Research Council K-12 curriculum framework can foster more indepth learning of science content and promote foundational science knowledge and skills. The following quote by the National Research Council committee that developed the NGSS reflects this sentiment:

The framework is designed to help realize a vision for education in the sciences and engineering in which students, over multiple years of school, actively engage in scientific and engineering practices and apply crosscutting concepts to deepen their understanding of the core ideas in these fields." (National Research Council, 2012, p.8)

Results from several studies (Engels et al., 2019; Gale et al., 2019, Rachmawati et al., 2019; Wen et al., 2019) have provided data supporting the claim that the NGSS curriculum framework is leading to improvements in science academic achievement. Engels et al. (2019) examined the effectiveness of NGSS aligned year-long educational programs at improving science literacy. In this study, Engels et al. (2019) incorporated both project-based learning and place-based education to measure students' attitudes (i.e., affection) toward science and how these attitudes correlated with the students' ability to apply science skills (i.e., cognition). Results showed that as science literacy skills improved, the student's confidence level in science practices likewise increased. The improvement in attitude suggests that the NGSS-based educational program can positively affect science knowledge and literacy skills (i.e., science academic achievement). In other words, the confidence that a student gains as a result of improvements in science academic achievement could translate to more enthusiasm and motivation to engage actively in science practices (i.e., improved conation). By increasing students' participation in science practices, they experience a positive recursive effect on science knowledge and literacy skill. Subsequently, due to this reciprocal recursive effect, the student's attitude (i.e., affection) toward science again improves. Similar findings were reported by Rachmawati et al. (2019) and Wen et al. (2019), who use NGSS-based learning material and instructional approaches to measure student engagement in science practices and science achievement. They also reported that NGSS-guided practices increased willingness to engage in science practices and improvements in academic achievement.

The literature suggests that the NGSS curriculum framework can promote improved literacy and academic achievement in science. Additionally, the literature also indicates that progress in science literacy may correlate directly with positive attitudes toward science (Engels et al., 2019; Rachmawati et al., 2019; Wen et al., 2019) and that this improvement is most noticeable when students engage in the epistemic exploration of science concepts (Miller et al., 2018). Recall that, according to Mayer's model of human cognitive architecture (see Figure 4, p. 22), the flow of information between the memory reservoirs depends on active cognitive processing (Mayer 2014a, 2014b), which involves both metacognition and the motivation to learn. Perhaps the reciprocal recursive effect observed in various studies (e.g., Engels et al., 2019; Rachmawati et al., 2019; Wen et al., 2019) is predicted by the cognitive theory of multimedia learning's assumption that active processing and selective attention promote information flow and integration (see Figure 4, p. 22; Mayer, 2014a). The following section presents research on how concept modeling and mental visualization may help moderate student attitudes toward science and the reciprocal recursive effect on science knowledge and literacy skills.

Multimedia Learning in Science Education

The current study was situated in a secondary-school biology classroom. Science education presents unique opportunities for exploring the efficacy of cognitive load theory, Baddeley's model, and cognitive theory of multimedia learning due to the abstract nature of many science concepts requires that students integrate various information elements. Students who struggle with science may do so because they have difficulty moving from the abstract to the concrete. For example, in molecular biology, which deals with particles and molecules at the microscopic level or smaller, there is no tangible way for learners to visualize and manipulate such molecules in their native states. Describing them aurally or in print typically is not sufficient for students who struggle to grasp the relevant concepts. Using models such as illustrations and 3-D manipulatives, the teacher can help students tackle their underlying confusion and misconceptions (Rau, 2017). Besides conceptual abstraction in science, students also struggle because science texts tend to use academic language that references unfamiliar concepts and complex causal relations, including cross-disciplinary vocabulary and compact and embedded syntax that differ from everyday language (Meneses et al., 2018). It is not unusual that the cognitive demand for processing science text overwhelms students, leading to a loss of interest in the subject matter or the motivation to persevere.

The situation is far from hopeless. There is ample evidence that prior knowledge and literacy skills are key factors effecting science text comprehension (Kloser, 2016; McNamara et al., 2011). For example, in a study by Kloser (2016), the investigators discovered that text that includes epistemic evidence for scientific claims improved comprehension and promoted meaningful learning. It may be that such evidence-enriched text engages the learner in imaginative thought processes that promote the construction of self-generated visuals of the abstract concepts (Kloser, 2016). Thus, instruction that engages imaginative visualization about relevant scientific concepts may enhance science knowledge. The current study explored this idea by testing the generative drawing principle that producing personal visuals based on the information provided during learning tasks can lead to better outcomes (Leutner & Schmeck, 2014).

Visuals help illustrate phenomena and processes that are difficult to observe directly or explain relations that are difficult to describe with words alone (McTigue & Slough, 2010). Several studies on multimodal learning and cognition confirm that visuals benefit cognitive processing (Levin & Mayer, 1993; Mayer & Gallini, 1990; Mayer & Sims, 1994). There is little direct empirical evidence, however, that multimodal reading comprehension and embedding science text with visuals improves science literacy and academic achievement at the secondary-school level. Instead, prior studies examined verbal texts or images separately with minimal consideration to their combined effect on meaning-making in science (Firat, 2017; Höffler & Leutner, 2007; Kühl et al., 2011; Meneses et al., 2018). Furthermore, those studies that examined the interaction of text and visuals (e.g., Kühl et al., 2011; Lin & Dwyer, 2010) focused on college and university-age students. The current study considered how embedding science text with visuals can affect secondary-school students' ability to retain the information they learn (rote learning) and transfer the new knowledge to an unfamiliar problem (meaningful learning). Furthermore, this study explored how the type of visuals (static versus vs. dynamic visual) and text (audio vs. print) affect science learning.

The Need for the Study

The two multimedia design principles of interest in this study are the modality and expertise reversal principles. Although there is empirical evidence to support these principles, two features of the existing research leave room for further study. First, many existing studies were conducted in a controlled laboratory setting (Butcher, 2014). For such a controlled research design, various affective factors could influence students' emotions and temperament (Snow et al.,1996). These factors could interact subsequently with conative factors that effect the learner's motivation to learn and cognitive control

(Mayer, 2011). Although such studies are essential for establishing the design principles' validity, they may not predict students' learning outcomes in an actual classroom setting. The current study used an actual biology classroom of secondary-school students to evaluate the design of two principles' utilitarian reliability. Furthermore, the science concepts selected for the study align with the established curriculum framework for the science program at the school. Although the study design was controlled, the learning content reflects the learning resources typical of a regular instructional unit.

The second feature of the existing research that leaves room for further investigation is that the empirical database needed to establish the design principles' validity remains incomplete. There is some evidence, however, that when students use paper-and-pencil to produce and display visual representations of concepts, learning improves (Leopold & Leutner, 2013; Schmeck et al., 2012). More evidence is needed to establish this finding, particularly for computer-based learning environments. The current study added to the literature by examining how the learner's use of static and dynamic visuals correlates with rote and meaningful learning measures. As such, not only did the current investigation reflect the reality of the learning environment within a typical classroom setting, but it also provided additional insight into how a learner negotiates meaning from the provided information based on the mode of information presentation. In addition, the findings provided insight into dynamic cognitive processing involved in schema construction and information transfer.

Research Questions

The study attempts to answer three research questions about the multimedia approach to teaching and learning. All questions are quantitative in nature; however, the last question relies on qualitative data that are coded into quantitative data as described in instrumentation.

1. *The modality effect*. To what extent is there an effect of information modality (i.e., text, pictures, video, sound) on rote learning and meaningful learning of science concepts, as measured by participants' responses to recall and transfer questions, respectively?
2. *The expertise reversal effect.* What effect do prior knowledge (i.e., advanced vs. novice learners) and the use of borrowed visuals have on rote and meaningful learning, as measured by participants' responses to factual recall and transfer questions, respectively?
3. *Cognitive load.* What effect do prior knowledge (i.e., advanced vs. novice learners) and the use of borrowed visuals have on cognitive-load, as measured by participants' responses to the cognitive-load questionnaire?

Definition of Terms

The section provides an overview of relevant terms that are used throughout this body of work. It should serve as a quick reference for understanding how the various terms are used here.

Advanced learner: A term used to refer to those learners who have progressed along the continuum from a novice learner to an expert such that they have considerable foundation domain-specific knowledge, but still lack the expert level competency. The advanced learner is most similar to the advanced beginner, the second or five steps from novice to expert as described by Dreyfus & Dreyfus (1986). In this study, the term advanced learner is contrasted with novice learner in order to conveniently drive discussion about knowledge acquisition. Although the prior knowledge of participants is measured using the pretest instrument, participants are not explicitly categorized as advanced or novice learners. Instead, a correlational analysis is used to evaluate whether there is a potential effect of prior knowledge on rote learning and meaningful learning and cognitive-load.

Baddeley's model: A hypothetical model that describes the working-memory as an information processing system with four subsystems: (a) central executive, (b) phonological loop, (c) the episodic buffer, and (d) visuospatial sketchpad (Baddeley & Hitch, 1974).

Borrowed visual: Any information presented visually in the learning material in order to help the learner make sense of the information. The idea of borrowed visuals extends from the borrowing and reorganization principle, which states that all of the secondary knowledge stored in long-term memory is borrowed from other people. Borrowed visuals are specifically borrowed information that is presented as a visual representation rather than as text or sound (Sweller, 2011).

Central executive: The Baddeley's subsystem that is responsible for (a) monitoring and coordinating the activities of the other three subsystems and connecting them to long-term memory, (b) regulating attention, (c) transferring information through and between the three memory reservoirs (sensory, working, and long-term memories), and (d) encoding information (Baddeley, 2000).

Cognitive load capacity: The cognitive processing limit of the working-memory, estimated at 7±2 chunks of information (Miller, 1956).

Cognitive load theory: A novel theory for explaining how the human mind processes information during learning, thinking, and problem-solving (Sweller, 2015). In this study, cognitive-load is measured using three scales: (a) the mental effort scale, (b) the germane load scale, and (c) the extraneous load scale.

Cognitive theory of multimedia learning: A psychological theory that is based on the current understanding of human cognitive architecture as an information processing system operating within the limits of the cognitive load capacity (Mayer, 2014b; Moreno & Park, 2010; Paas & Sweller, 2014).

Coherence principle: A multimedia design principle that states that people learn more deeply from a multimedia message when extraneous material is excluded rather than included (Mayer & Fiorella, 2014, p. 280).

Conceptual recall: The recall of abstract ideas (i.e., concepts) from long-term memory that was explicitly presented in the learning material. Conceptual recall measures the extent to which the learner recalled explanative information or succeeded at formulating schema or knowledge structures (Mayer & Gallini, 1990). In the current study, conceptual recall is categorized as one evidence of rote learning.

Conceptual recall questionnaire: The instrument used in this study to measure conceptual recall. The CRQ questionnaire contained one free-response question item that required participants to explain how new species arise from preexisting species through natural selection. Responses were scored using the rubrics provided in Table 2 (p. 86)

Dynamic visual: A visual representation that has moving elements such as real-time videos and illustrated animations; visuals presented digitally and that change due to automated animation (e.g., movies) or virtual manipulation by the learner (i.e., computer simulation; Butcher, 2014).

Encode: A term that refers to the process of cognitively assigning meaning to information (Sweller, 2015).

Episodic buffer: The Baddeley model's subsystem that temporarily holds information from the other subsystems and long-term memory until required by the central executive. The episodic buffer can store multimodal information (Baddeley, 2000).

Epistemic evidence: A piece of supporting empirical evidence or logical argument that is used to justify a scientific claim (Lin & Chan, 2018).

Experts: Individuals who are highly proficient, skilled, and knowledgeable in a particular domain. They can effectively think about and solve domain-specific problems by identifying patterns in relevant information (National Research Council, 1999).

Extraneous load: The cognitive-load dedicated to processing irrelevant or unrelated information that distracts the learner during learning tasks (Sweller, 2015). In this study, the extraneous load is measured using the extraneous load scale.

Extraneous load scale: A subjective unidimensional instrument used to measure the cognitive resource devoted to internal and external elements that distract from learning. The version used in the current study is a 7-point rating scale that is a derivative of one developed by Bratfisch et al. (1972). The Bratfisch et al. (1972) version was modified according to an approach similar to that used by Cheng and Beal (2020).

Germane load: The cognitive-load that accounts for the amount of mental work involved during learning, including the effort needed to process relevant information and construct the conceptual schemas (Sweller, 2015).

Germane load scale: A subjective unidimensional instrument used to measure the cognitive resource devoted to learning. The version used in the current study is a 7-point rating scale that is a derivative of one developed by Bratfisch et al. (1972). The Bratfisch et al. (1972) version was modified according to an approach used by Cheng and Beal (2020).

Image Principle: A multimedia design principle that states that people do not necessarily learn more deeply from a multimedia message when the speaker's image (i.e., a talking head) is on screen rather than not on the screen (Mayer, 2014c, p. 345).

Information retention: The ability of the learner to remember and recall information stored in long-term memory. The extent of information retention is a key measure of rote learning; but, although information retention is critical to meaningful learning, it is the ability of the learner to transfer and apply the retained information to a novel problem or situation that constitutes meaningful learning (Sweller, 2015).

Information transfer: A term that is used (a) to refer to the flow of information between memory reservoirs, and (b) to refer to the transfer of retained information or knowledge from a familiar to a novel problem or situation. In the latter instance, information transfer is instead referred to as knowledge transfer. (Sweller, 2015).

Instructional material: The teaching or learning resource (e.g., text, audio, visuals, kinesthetic supplies) used to guide the learning process. For this study, the term refers to any of the science learning materials in print or animation that participants use to access the instructional content. The generalization of the term is limited to learning material relevant to the science classroom. The extent to which participants acquired the relevant knowledge from the learning material is measured based on their performance on three questionnaires: (a) the conceptual recall, (b) the nonconceptual recall, and (c) knowledge transfer questionnaires.

Intrinsic load: A measure of the complexity or difficulty of the learning information or learning task (Pollock, Chandler, & Sweller, 2002).

Knowledge transfer: The learner's ability to apply acquired knowledge to a new and unfamiliar situation or problem. Knowledge transfer is measured in this study using the knowledge transfer questionnaire (Sweller, 2015).

Knowledge: A collection of all information stored in long-term memory (Sweller, 2015).

Learning: is the process through which a learner acquires new knowledge.

Long-term memory: The memory reservoir that stores all of the knowledge until needed for cognitive processing (Baddeley, 1992).

Meaningful learning: The process of learning in which new information is understood and contextualized with prior knowledge. To achieve meaningful learning, the learner should demonstrate an ability to apply what was learned to a novel situation or problem (Mayer, 2014a).

Mental effort scale: A subjective unidimensional instrument used to measure mental workload. The version used in the current study is a 7-point rating scale that is a derivative of one developed by Bratfisch et al. (1972). The scale was modified to construct scales that measure extraneous and germane loads according to an approach used by Cheng and Beal (2020).

Modality principle: A multimedia design principle that states that under split-attention conditions, presenting some information in visual mode and other information in auditory mode can expand effective the working-memory capacity, reducing the excessive cognitive-load (Low & Sweller, 2014, p. 227).

Multimedia learning: Learning from words and pictures (Mayer, 2014b).

Multimedia principle: A multimedia design principle that states that learning with words and pictures is more germane to knowledge acquisition and comprehension than learning with words alone (Butcher, 2014, p. 175).

Nonconceptual recall: The recall of concrete information from long-term memory that was explicitly presented in the learning material. Nonconceptual recall measures the extent to which the learner succeeded at recalling non-explanative information (Mayer & Gallini, 1990). In the current study, conceptual recall is categorized as one form of rote learning.

Nonconceptual recall questionnaire: The instrument used in this study to measure nonconceptual recall. The NRQ questionnaire contained 18 equally weighted items (e.g., 3 checkboxes, 2 T/F, 1 numerical response, and 12 multiple choices).

Novice learner: A term used to refer to individuals who, unlike advanced learners and experts, are minimally proficient, skilled, and knowledgeable in a particular domain. Novices lack sufficient foundational prior knowledge or experience with the learning content, but they can become experts by acquiring extensive domain-specific knowledge, skills, and strategies that improve their ability to identify problems, organize and interpret data and other information, and formulate solutions to relevant problems. In this study, the term novice learner is contrasted with the advanced learner in order to conveniently drive discussion about knowledge acquisition. Although the prior knowledge of participants is measured using the pretest instrument, participants are not explicitly categorized as advanced or novice learners. Instead, a correlational analysis is used to evaluate whether there is a potential effect of prior knowledge on rote learning and meaningful learning and cognitive-load (National Research Council, 1999).

Personalization Principle: A multimedia design principle that states that people learn more deeply from a multimedia message when words are conversational rather than formal (Mayer, 2014c).

Phonological loop: The Baddeley's subsystem that is responsible for receiving and processing verbal information (Baddeley, 2000).

Pretraining principle: A multimedia design principle that states that people learn more deeply from a multimedia message when they know the main concept's names and characteristics (Mayer & Pilegard, 2014, p. 317).

Primary knowledge: Any is biologically adaptive knowledge that encodes the skills for which humans evolved an innate predisposition to learn, e.g., the skills associated with walking, first language acquisition, and suckling. Primary knowledge is learned innately and therefore requires little or no deliberate effort to acquire (Geary & Birch, 2016).

Redundancy principle: A multimedia design principle that states that presenting the same information concurrently in multiple forms (or unnecessarily elaborating on previously presented information) can interfere with rather than facilitate learning (Kalyuga & Sweller, 2014, p. 247).

Rote learning: The memorization of information based on repetition. Rote learning may happen without the learner fully understanding how the information connects to their prior knowledge. In this study, rote learning is measured using two questionnaires: (a) the conceptual recall, and (b) the nonconceptual recall questionnaire (Mayer, 2014a).

Schema: A knowledge structure that the learner uses for organizing information in long-term memory (Sweller, 2015).

Secondary knowledge: Any domain-specific knowledge that encodes culturally relevant behavior and skills, and that is acquired through the borrowing and reorganizing principle (Paas, 2014; Prasada, 2000).

Self-generated visual: Any visual representation such as pictures or diagrams that are produced by the learner based on the learner's interpretation of textual or aural information presented in the learning material. The idea of self-generated visuals extends from the generative drawing principle when students create drawings while reading text, generative processing initiates, which leads to better learning outcomes (Schwamborn et al., 2010).

Segmenting principle: A multimedia design principle that states that people learn more deeply when a multimedia message is presented in learner-paced segments rather than as a continuous unit (Mayer & Pilegard, 2014, p. 317).

Sensory memory: The memory reservoir that initially receives new information from the environment (Baddeley, 1992).

Static visual: The traditional form of visuals that include only non-moving pictures, iconic symbols, diagrams, or graphics. Unlike dynamic visuals such as animations and videos, static visuals are typically presented in print or digitally on screen and have a static or fixed form (Butcher, 2014; Hegarty, 2014; Lowe & Schnotz, 2014)

Temporal Contiguity Principle: A multimedia design principle that states that people learn more deeply from a multimedia message when corresponding animation and narration are presented simultaneously rather than in temporal succession (Mayer & Fiorella, 2014, p. 280).

Visuals: Any information that is presented as pictures, iconic symbols, diagrams, or graphics, rather than textually or aurally (Butcher, 2014; Hegarty, 2014; Lowe & Schnotz, 2014).

Visuospatial sketchpad: The Baddeley's subsystem responsible for receiving and processing visual and spatial information (Baddeley, 2000).

Voice Principle: A multimedia design principle that states that people learn more deeply from a multimedia message when the words are spoken in a human voice rather than in a machine voice (Mayer, 2014c, p. 345).

Working-memory: The information processing memory reservoir. It is in the working-memory that all information processing occurs, including the integration and encoding of new information with prior knowledge (Sweller, 2015).

Summary

The cognitive theory of multimedia learning is the theoretical basis for designing and analyzing the effects of the modality and expertise reversal principles. Concerning the modality principle, the study investigated the effect of different modes of information presentation on learning and the extent to which prior knowledge moderates the learning process. The focus on prior knowledge also presents an opportunity to add to the growing body of empirical research on the expertise reversal effect. Considering how the integration of textual, visual, and sound elements facilitates learning could provide insight into how information processing differs between advanced and novice learners and how borrowed visuals can alter this difference.

The cognitive theory of multimedia learning extends from cognitive load theory, which explains the architecture and function of human cognition. Growing empirical evidence for cognitive load theory leads to a better understanding of how the human mind selects, processes, and stores information and how the new information combines with prior knowledge to construct new knowledge or refine existing knowledge. How the human mind is structured remains to be fully delineated; however, Baddeley's working-memory model provided a reasonably well-supported explanation that fits well with cognitive load theory and cognitive theory of multimedia learning. Baddeley's model described human cognitive architecture as a dual-channel system that processes visual and auditory information. This dual-channel system may explain why learning from multimodal information may benefit students, particularly when learning conceptually complex science material. If Baddeley's model holds, teachers could better appreciate the structure of the human mind and understand how it promotes cognitive processing and learning. By doing so, teachers can better structure their instructional design by appropriately integrating different information modalities to support the learning needs of students.

CHAPTER II

REVIEW OF THE LITERATURE

The current study extended from three research areas in cognition and instruction: multimedia learning, expertise reversal effect, and generative drawing effect. This study added to each of these bodies of research with a focus on secondary-school science education. As such, the study focuses on how to best tailor multimedia instructional design to meet the needs of novice and advanced students of secondary science. The first section of the review presents novel teaching and learning challenges associated with science literacy, emphasizing scientific literacy and its relevance to the current study. The second section presents a review of multimedia learning and dual-channel processing literature, focusing on works that align with literacy in secondary-science education. This second section also provides a description of the standard research methods for evaluating learning outcomes under mono- and multimodal conditions. The third section contains a review of ongoing research on the generative drawing effect within the context of secondary-science literacy. The review culminates with a discussion of how educators could use the expertise reversal effect, self-generated visuals (i.e., generative drawing effect), and the dual-channel concept to guide differentiated multimedia instructional design.

The Next Generation Science Standards and Literacy in Science

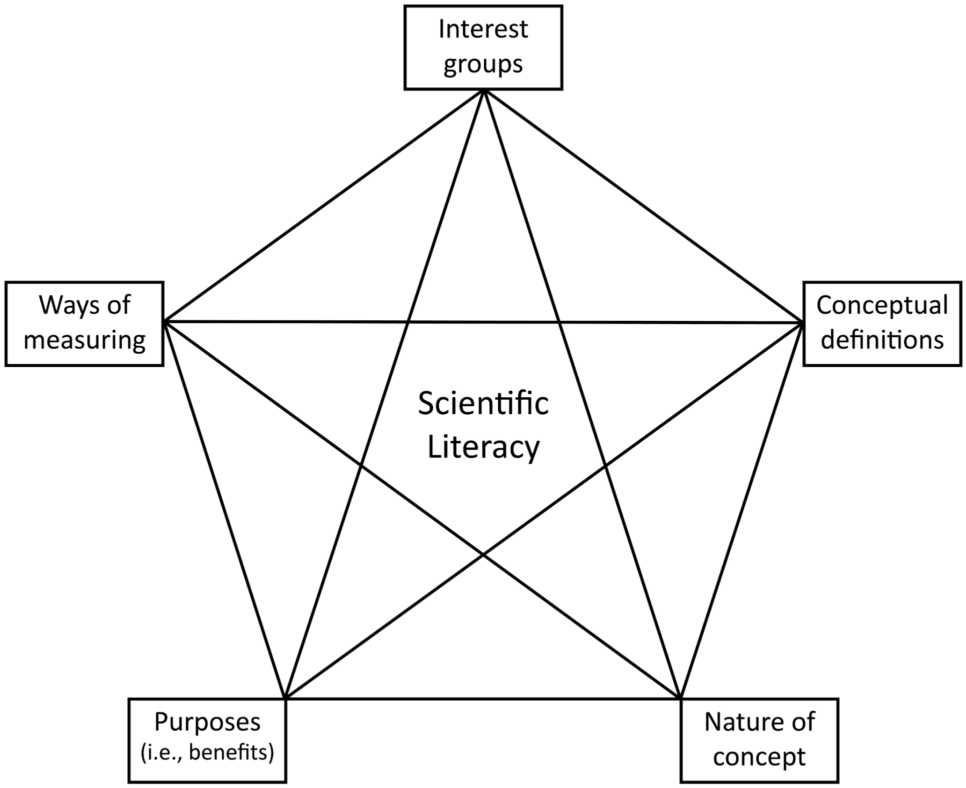
The development of science as a modern system for expanding and organizing humanity's knowledge of the natural world (Wilson, 1990) has been an arduous journey extending through many centuries. In ancient Greece, philosophers like Aristotle, Hippocrates, and Pythagoras used deductive reasoning to explain reality. Today, it is the scientific method that reigns supreme. Deductive reasoning is still relevant, but only as an exercise within the scientific method (Dunbar & Klahr, 2012). As such, an understanding of nature now relies on an internationally accepted set of common standards used for analyzing and interpreting experimental and field data. Whereas in ancient Greece, this knowledge was accessible only to a few elite learned individuals of the privileged class (Lopez, 2019), today, the expectation is that every member of society must possess a minimal level of scientific literacy (Laugksch, 2000). What does this mean? Of course, indepth mastery of scientific concepts remains out of reach for all but a few domain experts. For others, Pella (1976) described seven characteristics of the scientifically literate person (cited in Laugksch, 2000):

1. The scientifically literate person understands the nature of scientific knowledge;
2. The scientifically literate person accurately applies appropriate science concepts, principles, laws, and theories in interacting with his universe.
3. The scientifically literate person uses processes of science in solving problems, making decisions, and furthering his understanding of the universe;
4. The scientifically literate person interacts with the various aspects of [their] universe in a way that is consistent with the values that underlie science;
5. The scientifically literate person understands and appreciates the joint enterprises of science and technology and the interrelationship of these with each and with other aspects of society;
6. The scientifically literate person has developed a richer, more satisfying, more exciting view of the universe as a result of his science education and continues to extend this education throughout his life;
7. The scientifically literate person has developed numerous manipulative skills associated with science and technology. (pp. 76-77)

Pella's (1976) definition of scientific literacy was the first of its kind to articulate clearly what science education should aim toward (Laugksch, 2000). The definition, however, lacked utility as a set of guiding principles for science teachers and program developers in this original form. Several other scholars (e.g., Arons, 1983; Branscomb, 1981; Gabel, 1976; Miller, 1983; Shen, 1975) build on Pella's (1976) work to refine the term and give it definitive specificity and relevance for teaching and learning practices. For example, Shen (1975) delineated three subcategories: practical, cultural, and civic scientific literacy. Practical literacy deals with the application of scientific knowledge to solve practical problems. Cultural literacy informs about science's influence and effect on the human condition and human cultural achievements. Moreover, civic scientific literacy informs about the effect of science on public policies. Branscomb (1981) highlighted the importance of scientific literacy in reading, writing, and comprehension of scientific works. Furthermore, Arons (1983) emphasized the importance of applying scientific knowledge and reasoning skills to solve problems and make decisions.

The initial work in the 1970s and 1980s was about conceptualizing scientific literacy as a prerequisite to building purposeful science education programs for addressing current societal needs. The term science literacy, however, remains controversial (Jenkins, 1994) and lacks a universally accepted definition. Following an extensive review of relevant literature, Laugksch (2000) developed a conceptual overview (Figure 5) that reflects why it is difficult to reach a consensus. He discovered five categories of factors (i.e., interest groups, conceptual definitions, ways of measuring, purpose, and nature of concept) that influence scientific literacy interpretation (for an overview, see Laugksch, 2000). Accordingly, "these different interpretations give scientific literacy the appearance of being an ill-defined and diffused —and controversial —concept" (p. 74). Based on a diverse set of interests and stakeholders, Laugksch (2000) suggested that, instead, one should consider the term in context to its immediate intended purpose.

In U. S.-based education, the purpose of science literacy is well-articulated by the American Association for the Advancement of Science (AAAS) in Project 2061 (American Association for the Advancement of Science, 1989). Project 2061 is an



Note: Laughsch's model includes five categories of factors (i.e., interest groups, conceptual definitions, ways of measuring, purpose, and nature of concept) that influence scientific literacy interpretation. Reproduced from Laugksch, 2000, p.74.

Figure 5. Laugksch's conceptual overview of scientific literacy

ongoing collaboration between educators, academics, policymakers, and industry that was initially conceived in 1989 as a multiphase effort to reform U. S. science, technology, engineering, and mathematics education with a particular focus on science literacy. The following statement reflects the AAAS' definition of science literacy:

> The science-literate person is one who is aware that science, mathematics, and technology are interdependent human enterprises with strengths and limitations; understands key concepts and principles of science; is familiar with the natural world and recognizes both its diversity and unity; and uses scientific knowledge and scientific ways of thinking for individual and social purposes. (AAAS, 1989, p. 1)

This definition of scientific literacy broadly expresses what students should know and be able to do concerning science knowledge and practices and is central to science education reform in the U. S. As such, Project 2061 set the foundation for the current transformative science curriculum in the U. S. and beyond. The National Research Council subsequently published A Framework for K-12 Science Education (National Research Council, 2012), which builds on Project 2061 by providing educators with a guide for developing and implementing instructional programs with holistic consideration of science knowledge and practices. The Next Generation Science Standards (NGSS) (National Research Council, 2013), an extension of this framework, outlines grade-level competencies based on three principal dimensions: (a) practices of science and engineering, (b) crosscutting concepts that unify science and engineering, and (c) core ideas from the physical sciences, life sciences, earth and space sciences, and engineering, technology and applications of science. The goal of these new science standards is to establish research-based benchmarks that science educators could incorporate into curriculum and lesson designs to stimulate students' interests in science and expand their knowledge and skills (National Research Council, 2012, 2013).

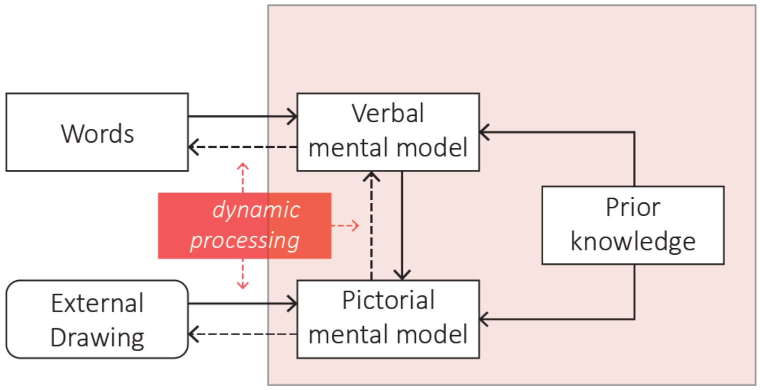
The rapid expansion of information in the current technology age presents two challenges that the NGSS curriculum framework had to address: (a) how to structure science education programs to accommodate the growing breadth and depth of scientific knowledge and (b) how to accommodate the need for continuous and dynamic updates of scientific knowledge and skills with ongoing scientific discoveries. NGSS promotes a trend toward building scientific knowledge and skills centered on a few core ideas (i.e., depth of knowledge) over the rote learning of a wide range of scientific facts (i.e., breadth of knowledge) (National Research Council, 2012). Accordingly, NGSS stressed the importance of providing students opportunities to ask questions, solve problems, construct and test models, investigate and explain phenomena, collect and analyze data, make inferences, and communicate ideas and understanding (AAAS, 1993; NGSS, 2013).

These practices naturally expose students to epistemic evidence for scientific claims, thus promoting conceptual understanding through the "discovery" process. The idea is that, by engaging in these practices, students will better understand how scientists work as they become skilled at critically evaluating scientific works to discern meaning. As such, rote learning of facts is secondary, whereas the overarching core ideas in science and unifying concepts form the anchor points of an NGSS-based science curriculum (Chessnutt et al., 2018; Krajcik et al., 2014). Factual information is still important but only when it supports students in developing their understanding of the core ideas and unifying concepts.

The National Research Council claimed that the NGSS standards and the National Research Council K-12 curriculum framework could foster more indepth learning of science content and promote foundational science knowledge and skills. Indeed, results from several studies (Engels et al., 2019; Gale et al., 2019; Rachmawati et al., 2019; Wen et al., 2019) validate this claim. For example, Engels et al. (2019) examined the effectiveness of NGSS aligned year-long educational programs at improving science literacy via a study of students in grades 10 to 12 that incorporated both project-based learning and place-based education. The researchers measured three factors: (a) students' attitudes toward science (i.e., the relevance of science), (b) students self-perceived ability to apply science skills (i.e., scientific method, data collection, data analysis), and (c) collaborate and communicate scientific understanding (i.e., present research findings) with peers and adults. The study applied a pretest-posttest assessment protocol (n = 230 and n = 207) that incorporated 15 coded items grouped in three categories to measure the three factors (Cronbach coefficient alpha $\geqq$ 0.80, 0.85, and 0.77, respectively). On all coded items, the investigators reported a pre- to posttest code frequency change ranging from 0% to 17% with Exploratory Factor Analysis Eigenvalues of 4.90, 1.77, and .96, respectively. These results suggest that the confidence associated with science academic achievement improvements could translate to greater enthusiasm and motivation to actively engage in science practices (i.e., changes in attitude toward science). Perhaps, increasing students' participation in science practices initiates a positive recursive effect on science knowledge and literacy skill. Subsequently, due to a reciprocal recursive effect, the student's attitude toward science (e.g., motivation to learn) again improves. In studies involving NGSS-based learning material and pedagogical approaches, Rachmawati et al. (2019) and Wen et al. (2020) also reported similar benefits of the NGSS curriculum for student engagement in science practices and science academic achievement. Rachmawati et al. (2019) reported a 133% increase in task performance scores of students who used NGSS-oriented learning tools relative to the control-group. This improvement was directly attributed to improved engagement with the learning material and motivation to learn. Wen et al. (2020) also reported that NGSS-based inquiry learning activities especially benefited low-science-achieving students, who “conducted more data analysis than other students and demonstrated adequate inquiry engagement" (p. 1).

When taken together, the literature suggests that the NGSS curriculum framework is improving science academic achievement. Additionally, there is some evidence that science literacy improvements may correlate directly with positive attitudes toward science (Engels et al., 2017; Rachmawati et al., 2019; Wen et al., 2019). What is more, a review of the literature by Miller et al. (2018) reported that this effect was most obvious when students explored scientific concepts by examining and justifying claims using supporting empirical evidence or logic (Miller et al., 2018). Such empirical evidence and logical explanation that supports scientific claims are referred to as epistemic evidence, and their use in teaching and learning practices is central to the NGSS curriculum (Miller et al., 2018). Such supporting evidence encourages students to engage with the learning material in more meaningful ways. Indeed, Kloster (2016) found that students who read epistemically considered text were more likely to provide specific justification for claims and less likely to accept the text as the authority. Simply embedding learning material with epistemic resources, however, is insufficient for promoting sophisticated epistemological commitment from students. Evidence from an earlier think-aloud study (Kloster, 2013) suggests that the benefits associated with epistemic evidence can only occur when students actively engage and interact with the epistemic evidence. Kloster argued that the learner must think about the relevant information and consider how it empirically supports and reinforces the presented claims (Kloster, 2013).

A possible explanation for why active engagement with epistemic evidence promotes science literacy and learning is that, when learning from epistemically supported text, the learner's imagination engages in constructing mental models based on the presented claims while simultaneously evaluating the model against the supporting epistemic evidence. Such a dynamic construction-evaluation process helps clarify misconceptions to promote learning. This hypothesis fits well with what is already known about generative processing (Leutner & Schmeck, 2014; Mayer 2014a). Accordingly to Leutner and Schmeck (2014), when learning from text, the learner dynamically selects conceptually relevant words and organizes and integrates them into a verbal mental model (see Figure 6). Simultaneously, the verbal model translates into a visual (pictorial) mental model through the dynamic processing of information from the text, the evolving visual (pictorial) mental model, and relevant prior knowledge. The ultimate goal is to



Notes: A generative drawing activity engages the learner in generative information processing to promote the construction of schematic verbal and pictorial mental models. The process depends on dynamic information processing (represented by the reversible solid and dashed arrows) and culminates with self-produced drawings that are external representations of the mental models (modified from a translated version by Leutner & Schmeck, 2014, p. 434; original German version by Schmeck, 2010, p. 29).

Figure 6. Generative information processing schematic

refine and equate the visual and verbal mental models to be as conceptually accurate as possible within the limits of the available information. Examining the self-generated drawings that students produce in the absence of drawing cues could offer information about the dynamic cognitive processing involved in constructing conceptual representations that transfer to long-term memory as evolving schema. These drawings can inform about the thought processes happening as the student organizes and integrates information. Because these drawings are physical representations of schema, the complexity and evolution of these generative drawings can hint at the extent of generative learning and conceptual mastery. The current study investigated the interaction of learner-produced generative drawings against prior knowledge and type of learning material (i.e., text only, text + visual, text + animation, and sound + animation) with task performance. Results for this analysis may provide additional support for the generative drawing effect while also informing about how it operates under multimedia conditions and in conjunction with varying levels of learners' prior knowledge.

Information Processing and Learning

It is well established that, under certain conditions, visuals can promote understanding of complex scientific concepts when the requisite knowledge is lacking or incomplete (Kloser, 2016; McNamara et al., 2011). The learner's prior knowledge of the subject matter is one factor that effects how much the borrowed visuals matter. An advanced learner who has sufficient prior knowledge and the relevant schema can generate their own visual (i.e., mental models and drawings), which is not the case with the novice learner who lacks the relevant prior knowledge and preexisting schema. Instead, the novice learner can benefit from additional support in the form of a teacher-provided (borrowed) visual. What is the theoretical basis for this assumption? According to Mayer's (2014a, 2014b) model of human cognitive architecture (see Figure 4 in Chapter 1), learning is an active process during which information flows through the dual (auditory and visual) channel between the three memory reservoirs (i.e., sensory memory, the working-memory, and long-term memory; Mayer 2014a). By presenting the information as text and visual, the novice learner can maximize both channels' utility. Each channel has a limited information capacity, which makes it advantageous to present information bimodally rather than unimodally. Support for this notion comes from a study by Saults and Cowen (2007) that evaluated the central capacity limit of the working-memory when participants had to recall information under unimodal (i.e., text or auditory) or bimodal (i.e., text and auditory) memory conditions. Saults and Cowen reported that information retention for bimodal memory load improved over unimodal memory load. They expected that information retention would compound under the bimodal conditions; however, although retention increased compared with the unimodal conditions, the number of items retained was less than double. Nevertheless, the evidence indicates that modality-specific memory can improve information retention, highlighting the multimodal effect's benefit. More work is needed to investigate the apparent muting of the compounding effect of dual-channel information processing. One possibility could be associated with the redundancy effect, where similar information presented in both an auditory and a visual mode could cause cross-modal interference during memory retrieval. Saults and Cowen also suggested that selective attention may function as a memory storage device. If the learner's attention fails to disperse information held within each channel equally, the compounding effect reduces, especially for the channel that receives less attention.

Selective attention

When performing a learning task, most of the new information arriving at the sensory memory is distractors that are irrelevant to the learning task (Pinto et al., 2013; Sasin, 2021). Irrespective of the information modality (i.e., auditory or visual), if these distractors capture the learner's attention, extraneous load increases, and germane load decreases, hindering learning (Sweller, 2010). Learning also suffers if the learner lacks sufficient discriminatory knowledge and skills to distinguish between task-relevant elements and distractors: consequently, one's ability to attend selectively to task-relevant information elements while ignoring distractors affects learning. Compared with novice learners, advanced learners possess better discriminatory knowledge and skills for selecting task-relevant information elements (Lavie, 2004). There are two mechanisms of selective attention described in the literature: top-down (or goal-oriented) and bottom-up (or stimulus-driven) attention (for a review, see Kastner & Ungerleider, 2000). The top-down attentional system is voluntary and goal-driven, which operates under deliberate executive control that actively maintains internal representations of processing priorities in favor of task-relevant information (Corbetta & Shulman, 2002; Petrucci & Pecchinenda, 2017;). With top-down attention, the learner must have enough discriminatory knowledge and skills to identify and select task-relevant elements while ignoring the irrelevant distractors (Lavie, 2004).

Bottom-up attention is involuntary and attention-grabbing and caused by stimuli (e.g., flashing light, a loud bang, highlighted text, visual cues within learning material) that differ considerably from the background (Corbetta & Shulman, 2002). Learning material such as textbooks typically is written and organized with elements that engage goal-driven and stimulus-driven attention. For example, a worked example of a mathematics problem initially might describe the purpose (i.e., the goal) of the worked example (i.e., applying the Pythagorean theorem to calculate the space diameter of a cuboidal structure). By establishing the goal, the learner is primed to maintain top-down attention while reading and following the solution's steps. The same worked example might interject with bolded subheadings or a diagram of the cuboidal space with space diagonal drawn in an attention-grabbing red color. These alterations in information presentation facilitate bottom-up attention by calling the reader's attention to key instructional elements in the text. In this regard, when used appropriately, both types of attention mechanisms can enhance germane load (learning) in the working-memory by promoting the preferential deployment of cognitive resources to the attended elements (Pinto et al., 2013). What happens, however, when the learner lacks the relevant prior knowledge and skills to make sense of the information? Under the goal-driven condition, demand on executive control processes could exceed the learner's cognitive capacity to discriminate between task-relevant and task-irrelevant information (Petrucci &

Pecchinenda, 2017). Competition from distractors will increase, leading to a rise in extraneous load and a drop in germane load (Sweller, 2011).

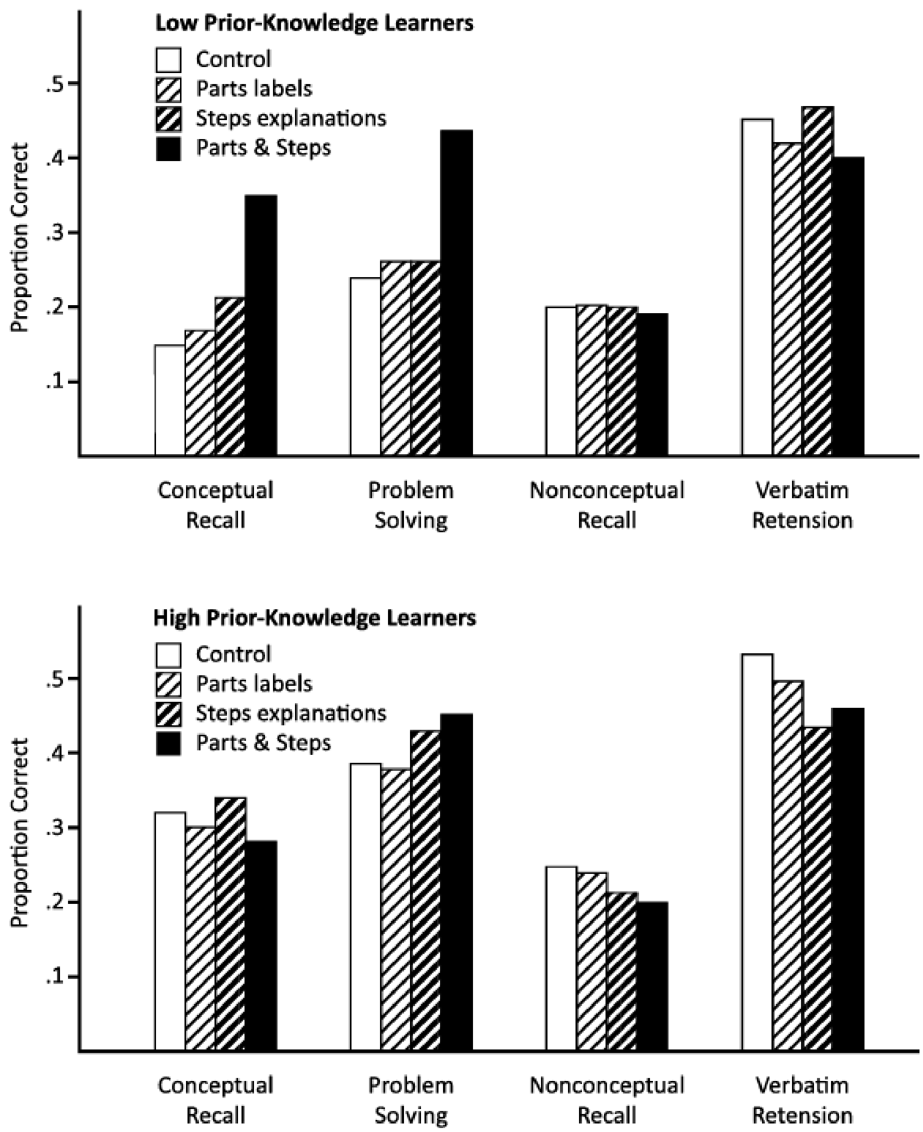
Although teachers should consider strategies that promote selective attention to task-relevant information, the abstract nature and complexity of many science concepts can mean that students will not always possess sufficient discriminatory knowledge and skills for identifying and selecting task-relevant information in a goal-driven manner. The likelihood of such a lack of discriminatory knowledge is especially so for novices relative to advanced learners (Kalyuga, 2014). Advanced learners may have some task-relevant prior knowledge and schema to help them identify, select, and interpret task-relevant information. The novice learner who lacks the task-relevant prior knowledge or schema may be less competent at targeting their selective attention on task-relevant elements. Under such conditions, the novice learner will require additional teacher support with identifying, selecting and interpreting information.

Borrowed visuals as learning support tools

Without appropriate instructional guidance and support, distractors can outcompete task-relevant elements for the novice learner's attention. Indeed, investigations on the expertise reversal effect show that novice learners and advanced learners respond differently to varying degrees of instructional support (Jiang et al., 2018; Kalyuga et al., 2003; Kalyuga & Sweller, 2014). The advanced learner who already has preexisting schemas to facilitate information processing may become confused when additional teacher support presents redundant information that contradicts elements already stored in their preexisting schema. For novice learners, however, additional instructional support can lead to improvements in learning outcomes. For example, a study by Mayer and Gallini (1990) demonstrated that novice learners benefit from teacher-provided (borrowed) visuals that help convey meaning coded in the accompanying text. The investigators conducted experiments in which they asked university students to read expository passages on a hydraulic braking system's function and respond to a series of questions. Students were assigned to four different treatment groups as follows: (a) no illustration (text only control-group), (b) embedded parts illustrations (model of the braking system with labeled parts), (c) embedded steps illustrations (model of the braking system with explanations of major actions occurring at each step along the braking process), and (d) embedded parts and steps illustrations (combination of treatment b and c).

The results (see Figure 7) show that the explanative illustrations (treatment d) resulted in statistically significant improvement in the performance of novice learners (i.e., low prior knowledge) on measures of explanative recall ($\eta^2 = .48$), creative problem solving ($\eta^2 = .44$), and conceptual recall ($\eta^2 = .40$). The investigators also evaluated the extent to which explanative illustrations improved problem-solving performance compared to non-explanative illustrations. Analysis of variance comparing the four treatment groups revealed statistically significant differences in problem-solving performances ($\eta^2 = .28$). Furthermore, a Dunnett's Test (at $p \leq .05$) revealed that the explanative illustration was the only treatment that resulted in statistically significant outperformance of the control-group. Consistent with predictions from the expertise reversal effect, the advanced (i.e., high prior knowledge) learners did not benefit from the explanative illustrations.

The study by Mayer and Gallini (1990) provided evidence validating the expertise reversal effect. In addition to highlighting the need for differentiating instructional



Note: Mayer & Gallini, 1990, p. 719

Figure 7. Mayer and Gallini posttests by treatment results

support to meet the students' needs based on domain-specific prior knowledge, they showed that embedding science text with illustrations can benefit novice learners who might struggle to connect multiple interacting elements required to understand and conceptualize the information. Because schema construction is the target of learning, when the number of interacting elements exceeds the learner's working-memory capacity, learning stalls (Baddeley, 1999; Miller, 1956). In Mayer and Gallini's (1990) study, the provided visuals are schematic representations of the target knowledge and concept. Not all treatment groups that had access to visuals, however, benefited. The groups who were provided illustrations embedded with either the parts labels or the explanations of the steps (rather than both) showed no statistically significant difference in performance relative to the control-group, suggesting that illustrations alone are not intrinsically beneficial. The visual must have sufficient explanatory information to interpret the intended meaning that the visual conveys. Mayer and Gallini (1990) outlined specific guidelines for effective illustrations that include the requirement that the accompanying text is appropriate for the intended learning outcome. For example, if the instructional goal is to improve conceptual understanding, explanatory text is more advantageous than descriptive or narrative text (Mayer & Gallini, 1990). Explanatory science texts use logic and empirical evidence to explain why a particular phenomenon occurs, and they do so by providing students opportunities to examine, evaluate, and justify claims. Explanatory texts are good sources of epistemic evidence that can enhance science literacy and learning outcomes (Kloster, 2013; Miller et al., 2018). Another condition is that effective illustrations should complement the instructional goal by including explanations missing from the learner's prior knowledge. This condition touches on two essential understandings. Based on the expertise reversal principle, providing additional supporting information that the advanced learner already knows can be counterproductive and should be avoided. Second, for the novice learner who lacks sufficient domain-specific prior knowledge, embedding visuals with complementary explanatory text could improve comprehension. Thus, an illustration should only be used for learners who lack the prior knowledge that the illustration encapsulates.

Meaningful and rote learning

In the study by Mayer and Gallini (1990), the improvements reported for novice learners were observed for conceptual recall and problem solving but not for nonconceptual recall or verbatim retention. Conceptual recall measures the extent to which the learner succeeded at building a "runnable mental model [or schema about] how one state change affects another" (Mayer & Gallini, 1990, p. 717). These four measures (i.e., conceptual and nonconceptual recall, problem-solving, and verbatim retention) assess the two types of learning: rote learning and meaningful learning. Mayer (2002) defined rote learning and meaningful learning concerning knowledge retention and knowledge transfer as follows:

> Retention is the ability to remember material in much the same way it was presented during instruction. Knowledge transfer is the ability to use what was learned to solve new problems, answer new questions, or facilitate learning a new subject matter. (p. 227)

With rote learning, the learner can recall information but neither transfer their knowledge to solve problems nor apply the knowledge to new situations. Rote learning is a prerequisite to meaningful learning, however, because it builds foundational knowledge. With meaningful learning, the learner demonstrates both knowledge retention (rote learning) and knowledge transfer. The application of newly acquired knowledge to solve unfamiliar problems is the best evidence of knowledge transfer. Meaningful learning, therefore, relies on knowledge (schema) construction, involving the learner's selective attention to task-relevant information and integration of incoming information with prior knowledge (Mayer, 1999). Thus, by measuring the learner's ability to recall conceptual information, one can gauge the extent to which they have effectively acquired or constructed the relevant schema for guiding task performance during problem-solving.

Furthermore, earlier work by White and Frederiksen (1987) supports the notion that learners who possess relevant schemas can better execute qualitative reasoning, a key component of problem-solving.

Expertise reversal principle

The expertise reversal effect initially was conceived in cognitive load theory as a redundancy principle product (Kalyuga, 2014). Accordingly, information that is beneficial for the novice learner can be redundant for the advanced learner. The redundant information can confuse or distract the advanced learner. In either instance (i.e., distraction or confusion), precious cognitive resources will be sequestered (Sweller, 2020). If the cognitive resource is used for processing distracting elements, the extraneous load will occupy a larger portion of the cognitive space it dynamically shares with the germane load (Sweller, 2020). When redundancy confuses the learner, it is likely due to differences in information presentation modes (Kalyuga & Sweller, 2014). The advanced learner interprets incoming information based on an established schema.

According to the generative drawing principle of cognitive theory of multimedia learning, when a learner uses drawings to represent textual information, they must construct pictorial and verbal models of the information (see Figure 6, p. 56) in the working-memory (Leutner & Schmeck, 2014). The pictorial and verbal models dynamically integrate to form a schema, which guides the learner in drawing a physical representation of the schema. This entire process, including the processing of new task-relevant information, the construction of mental models, schema generation, and drawing, is engaged when the learner participates in a generative drawing activity. When the same information appears in two different forms (i.e., auditory and visual) or is redundant with the learner's prior knowledge, the extraneous load will increase in the working-memory. Consequently, the germane load otherwise dedicated to processing task-relevant information will decrease. Task performance for advanced learners will decline when the learning material contains information already coded in prior knowledge. Evidence supporting the expertise reversal effect comes from studies in worked examples (Renkl & Atkinson, 2003), generative visualization (Cooper et al., 2001), segmenting principle (Spanjers et al., 2011), and schema generation (Armougum et al., 2020), to mention a few. For example, Armougum et al. (2020) measured cognitive-load and task performance of expert and novice train travelers under normal versus disturbed virtual environments. The assumption was that experts, but not novices, have preexisting schemas for navigating the virtual environment. Experts outperformed novices under normal but not under the disturbed conditions as predicted by the expertise reversal effect. Measures of cognitive-load provided insight into the mechanism of the expertise reversal effect. Accordingly, the disturbed environment presented unexpected events that interfered with the routines and automated actions coded for by expert participants' established schema. Thus, experts, but not novices, needed to dispense more cognitive resources for modifying the relevant stored schema in order to accommodate the unexpected events and complete the virtual task.

Findings similar to that of Armougum et al. (2020) come from studies investigating task performance within an academic setting. Kühl (2021) investigated the effect of prior knowledge on learners' ability to infer dynamic features from static visuals. The assumption here was that advanced and novice learners would benefit differently from static versus dynamic (animations) visuals. In building the rationale for the investigation, Kühl (2021) argued the following:

> Animations can have objective informational advantages over static pictures if crucial dynamic features cannot be depicted in static pictures. While learning through animations only requires learners to read off the dynamic features, learning with static pictures requires learners to infer these dynamic features. (p. 1)

The benefit of presenting information sequentially to convey the temporally functional relationships can give animation an advantage over static images. Static images, however, allow the learner to interact with and study visual elements in a self-paced manner (Ploetzner et al., 2020). On the contrary, animations' dynamic nature might overwhelm the learner's cognitive-load due to greater element interactivity and transitional changes (Castro-Alonso et al., 2014; Wong & Paas, 2018). Given these contradictions, Kühl (2021) wanted to determine whether the benefit of using static visuals was a product of the learner's level of expertise in the domain.

The results from Kühl's (2021) study were consistent with predictions of the expertise reversal effect. Participants with low-prior knowledge showed statistically significant improvement in performance with animations versus static visuals on measures of dynamic factual knowledge, nondynamic factual knowledge, and information transfer. Higher prior knowledge participants experience marginal benefits from static visuals and lower performance from animation. Consistent with the expertise reversal effect is that the novice learners consistently benefited from animations on all measures, and advanced learners performed worse relative to novices when using animations. Underperformance by advanced learners under the animated condition could be due to the added processing required to inhibit the animation's contradictory schema. If added processing plays a role, however, there should be a statistically significant benefit from static visuals for advanced learners, but the effect was not statistically significant.

That the advanced learner did not benefit from static visuals on any of the three measures is not surprising. Measures of dynamic and nondynamic factual information are rote learning measures because they are about recalling factual information. Nevertheless, rote knowledge forms the foundation on which all meaningful learning happens. The fact that the advanced learners performed better than novices on both of these measures (i.e., dynamic and nondynamic factual knowledge test) under the static condition indicates that their rote learning improved. Why did they not perform better on knowledge transfer (meaningful learning) under static conditions? Could it be that the advanced learners failed to imagine deliberately the inferred dynamic events visually conveyed in the animation under static conditions? The current study provided findings that informed about this possibility.

Assessments and Measurements

The current study used seven instruments (Appendix A) grouped into three categories, that is, rote learning, meaningful learning, and cognitive-load, to measure the dependent variables. The literature regarding the use of these instruments as valid and reliable measures of the dependent variable is reviewed here.

Instruments for rote and meaningful learning

The idea of rote learning being different from meaningful learning originates from early work by Katona (1940, 1942), who instead used the terms rote and meaningful. In the original publication (Katona, 1940), participants learned card tricks by memorization (rote learning) or by understanding (meaningful learning), and their task performances were evaluated. Based on the task performance results, Katona made three critical generalizations: (a) meaningful learning is more time consuming than rote learning, (b) knowledge retention after meaningful learning is greater than knowledge retention after rote learning, and (c) knowledge transfer after meaningful learning is greater than knowledge transfer after rote learning.

Hilgard et al. (1953) tested Katona's generalizations to better understand the underlying mechanics of learning responsible for the differences observed between the two learning groups. The investigators used the same card trick design from Katona's study, maintaining the essential experimental features but making procedural modifications to enhance statistical treatment and analyses. Based on their research results, the Hilgard team reported the following: (a) Katano's finding that learning for understanding resulted in significantly higher knowledge transfer was verified, however, Hilgard et al. (1953) was concerned that, although the understanding group performed better on transfer tasks, their success rate remained very low, (b) Katano's finding that the learning activity for the understanding group required more time than the rote learning group on learning tasks was verified, however, the difference was attributed to the extra instruction time for explaining the card tricks, and (c) Katano's finding that knowledge retention after learning for understanding is greater than knowledge retention after rote learning was rejected because the difference in retention was not statistically significant.

Hilgard et al.'s (1953) study has been influential to the current understanding of rote and meaningful learning. In explaining the poor performance of the understanding group on transfer tasks, they noted that some participants in this group became impatient with the understanding method and adopted the rote memorization method once an answer was achieved. As such, these participants failed to acquire an understanding of the card tricks fully. They, therefore, applied the rote method to the unfamiliar transfer tasks because they had not learned and, thus, could not apply the understanding method. These participants from the understanding group made the same kinds of errors observed with the rote learning group. This insightful observation suggests that not only does meaningful learning require more learning time, but it also requires the learner to engage in the learning activity actively. Inclusion in the understanding group alone did not lead to improved learning; the learner also needed to apply themselves to acquiring the relevant knowledge and skills. In other words, active engagement with the learning process may be a prerequisite for meaningful learning when instructional conditions are germane to meaningful learning. This assumption is widely supported by empirical evidence from several studies (Lavie et al., 2004; Lin & Chan, 2018; Mayer & Gallini, 1990; Paas, 1992; Schwamborn et al., 2010).

According to Mayer (2002b), knowledge retention and knowledge transfer are the two most important educational goals. Retention is the recall of conceptual and nonconceptual information explicitly presented during instruction or learning activity (Mayer, 200b; Mayer & Gallini, 1990). Knowledge transfer involves applying the retained knowledge in answering or resolving unfamiliar questions or problems, respectively, or using the retained knowledge to facilitate learning something new (Mayer, 2002; Mayer & Wittrock, 1996).

Measures of rote learning. Mayer and Gallini (1990) used a novel approach to measure rote learning and meaningful learning that included verbatim and constructed response questions. For rote learning, they used two different posttest questionnaires. The first posttest, which was used to measure conceptual and nonconceptual recall, required the participants to reconstruct the learning material by writing down as much as they could remember from the passage they read. The investigators used rubrics to score participants' constructed responses based on the number of conceptual elements (out of 35 possible), and nonconceptual elements (out of 60 possible) recalled from the learning material. They categorized explanatory and non-explanatory ideas as conceptual and nonconceptual recall, respectively. The second posttest was a questionnaire that included 15 items that required verbatim responses.

The investigators strictly evaluated participants based on the learning material's information with both the constructed and verbatim responses. The current study applied a similar approach to developing a rote learning instrument for measuring the recall of conceptual and nonconceptual information explicitly presented in the learning material. For conceptual recall, a single questionnaire item required participants to explain how new species arise from preexisting species through the process of natural selection. Consistent with the Mayer and Gallini (1990) approach, participants' responses will be scored for explanatory elements. The constructed response was not evaluated for nonconceptual recall because the low maximum limits on the number of nonconceptual elements included in responses would leave a narrow reliability window for statistical analysis. Instead, a 19-item nonconceptual recall questionnaire that covers the full range of the learning content was developed. This approach, which others have used (Kühl, 2021; Spanjers et al., 2011), combines two features of Mayer and Gallini's (1990) method (i.e., from the constructed response and verbatim questionnaire).

Measures of meaningful learning. Mayer and Gallino (1990) measured meaningful learning by evaluating the extent to which participants could transfer the knowledge acquired from the learning material to an unfamiliar problem. The investigators developed a problem-solving instrument that required participants to reflect on concepts recalled from the learning material and consider how they could be transferred and applied to address the problems presented. For example, the learning material included an expository passage that explained how several types of mechanical pumps (e.g., centrifugal, sliding vane, lift, and bicycle tire pumps) operate. Although the text included all of the relevant conceptual and nonconceptual information, however, it did not explain explicitly how to resolve the problems presented in each of the following instrument question items:

1. What could be done to make a pump more reliable, that is, to make sure it would not fail?
2. What could be done to make a pump more effective, that is, to move more liquid or gas more rapidly?
3. Suppose you push down and pull up the handle of a lift pump several times, but no water comes out. What could have gone wrong?
4. Why does water enter a lift pump? Why does water exit from a lift pump?
5. The text you read mentioned a "screw pump that consisted of a screw rotating in a cylinder," but the text did not really explain how it works. Based on your understanding of how pumps work, please write your own idea of how you think a screw pump could be used to move water. (Mayer & Gallini, 1990, p. 721)

Participants had to recall the relevant conceptual and nonconceptual information and apply the recalled knowledge while actively considering the problem. Recall that in the study by Hilgard et al. (1953), the investigators expressed concerns that the understanding group's overall performance, although greater than the rote learning group, was far below what was expected. This finding aligns with the idea supported by others (Lavie et al., 2004; Lin & Chan, 2018; Mayer & Gallini, 1990; Paas, 1992; Schwamborn et al., 2010) that active engagement with the learning process may be a prerequisite for meaningful learning when instructional conditions are germane to meaningful learning.

The current study builds on Mayer and Gallini's (1990) approach in measuring meaningful learning by using a problem-solving item that requires participants to recall information presented in the learning material and apply that information to resolve a problem that was not explicitly discussed or explained.

Instruments for cognitive load measure

Mental effort. A variation of the mental effort rating scale developed initially by Bratfisch, and his colleagues (Bratfisch et al., 1972) was used in the current study. The efficacy of using the Bratfisch, Borg, and Dornic scale (1972) to gauge learners' cognitive-load is well established and supported in the literature (Ayres, 2006; Cheng & Beal, 2020; Paas, 1992; Paas & Van Merriënboer, 1994). Empirical and theoretical studies (Borg, 1978; Borg et al, 1971) also suggested that some modifications to the scale (e.g., scale type, scale category, magnitude estimation, verbal labels) do not have a statistically significant effect on efficacy. For example, in the study by Cheng and Beal, the scale type was modified from a 9-point to a 5-point rating scale. Furthermore, in addition to the mental effort category described in the original version, Cheng & Beal (2020) added two additional types: lesson difficulty and strategy difficulty. In the current study, the mental effort scale was similarly modified as follows: (a) the scale was adjusted to a 7-point rating scale, (b) the wording of the scale item was adjusted to give greater specificity to the study's learning and performance task, and (c) two additional categorical items were added to measure extraneous and germane loads.

Extraneous and germane loads. Existing empirical and theoretical studies (Borg, 1978; Borg et al., 1971) suggested that some modifications to the scale (e.g., scale type, scale category, magnitude estimation, verbal labels) do not have a statistically significant effect on efficacy. Adjusting how items are phrased could, however, provide additional insight into the working-memory function. Cheng & Beal (2020) modified the Botgand Dornic scale (1972) by adding two other categories—i.e., lesson difficulty and strategy difficulty—to their cognitive-load instrument. In the current study, additional data were collected to evaluate extraneous and germane loads and gain additional insight into participants' working-memory during task performance. A similar approach was used to generate two other cognitive-load scales for measuring extraneous and germane loads. According to Sweller (2010), the extraneous load increases when the learner becomes distracted by irrelevant elements to the learning task. For extraneous load, an item was developed to measure the extent of external distractions by asking participants to rate how much they had to review the learning material due to distractions. A second item measured internal distractions by asking participants to rate how much they worried about not understanding the learning material. Although worrying may lead to processing some new information, it is internally generated. What is more, worrying is distracting because it is driven by emotions and causes the learner to focus on elements that are irrelevant to the learning task. The germane load was measured by evaluating meaning-making during the learning activity using items that asked participants to rate the amount of thinking they had to do during the learning activity to make sense of unfamiliar and familiar vocabulary in the passage, respectively. The use of such subjective rating scales is well established in cognitive-load research and is "the most influential and widely used instrument in cognitive-load measure" (Zheng & Greenberg, 2018, pp. 48).

Summary of the Literature

The NGSS curriculum framework (National Research Council, 2012) redefines science education's goal to focus on science literacy rather than just the rote learning of a breath of factual information. Students should understand the critical underlying concepts in science that are generalizable across the various domains, and they should appreciate the epistemic evidence that supports the relevant scientific claims. Due to the complexity of many scientific concepts, however, novice learners who lack sufficient prior knowledge can experience cognitive overload when trying to make sense of complex ideas that include many interacting information elements (Carney & Levin, 2002). As such, they might be unable to dynamically integrate new information with prior knowledge and construct verbal and pictorial mental models that eventually lead to schema generation. Teachers might support their students in meaningful learning by scaffolding their instruction with the learner's level of expertise in mind. Although advanced learners might independently conceptualize information and construct mental models, novice learners cannot. Visual elements in learning materials that stimulate goal- and stimulus-driven attention and teacher-provided (borrowed) visuals may help the novice learner progress when learning would otherwise stall (for a review, see Kastner & Ungerleider, 2000).

The benefit of tailoring the instructional intervention to the learner's level of prior knowledge is well-established; however, more research is needed to understand fully how the mode of information presentation in science is related to the expertise reversal effect. This need for additional research is growing as advances in technology make it increasingly easier for teachers to integrate various information modalities into daily classroom instruction. It is not well-understood how multimedia, including animations and computer simulations, effect learners in different ways. Future research will help provide a better insight into how multimedia learning effects learners. Through the work presented in this study, the investigator hopes to provide some insight into how multimedia learning intervention interacts with prior knowledge to promote learning.

CHAPTER III

METHODOLOGY

The purpose of this study was to investigate how prior knowledge and the integration of information modalities (i.e., text, audio, static visual, dynamic visuals) promotes rote learning (information retention) and meaningful learning (knowledge transfer) in science. The finding presented in this body of work extended from data previously collected in an experimental pretest-posttest study. This pretest-posttest study used secondary-science instructional content to investigate the effect of borrowed and self-generated visuals on rote learning and meaningful learning. The study explores whether the multimedia effect on learning varies depending on the mode of visuals (i.e., static versus dynamic visual) and text (i.e., audio versus print). Furthermore, it evaluated the effectiveness of borrowed and self-generated visuals in science on novice and advanced learners. The findings from this investigation added to the growing body of research in the multimedia learning literature that visual representation of scientific concepts can enhance literacy and conceptual comprehension in science. This chapter provides (a) an overview of the research questions, (b) the details of the research design, (c) a description of the sample, (d) the procedure for protecting human subjects, (e) treatments, and (f) information on the instruments, procedure, data analysis, and limitations of the study design.

Research Questions

The study answers three research questions about the multimedia approach to teaching and learning. All questions are quantitative in nature.

1. *The modality effect*. To what extent is there an effect of information modality (i.e., text, pictures, video, sound) on rote learning and meaningful learning of science

concepts, as measured by participants' responses to recall and transfer questions, respectively?

2. *The expertise reversal effect.* What effect do prior knowledge (i.e., advanced vs. novice learners) and the use of borrowed visuals have on rote and meaningful learning, as measured by participants' responses to factual recall and transfer questions, respectively?
3. *Cognitive load.* What effect do prior knowledge (i.e., advanced vs. novice learners) and the use of borrowed visuals have on cognitive-load, as measured by participants' responses to the cognitive-load questionnaire?

Research Design

The study was based on data collected in a pretest-posttest quasi-experimental study. A schematic diagram of the research design is provided in Figure 8. This section provides a general overview of the structure of the research design described and discussed in greater detail throughout this chapter.

Independent variables

The independent variable was the mode of information modality that varied by treatment group as follows: (a) Group 1, which was the control-group, had access to learning material that was exclusively presented as text; (b) Group 2 had access to the same textual information as Group 1, but with embedded static visuals (i.e., pictures) that corresponded with the concepts in the text; (c) Group 3 was similar to Group 2, except that the visuals were animated and the text was subscripted in a video; and (d) Group 4 was similar to Group 3, the video was a full animation with the text instead of being subscripted was vocalized.

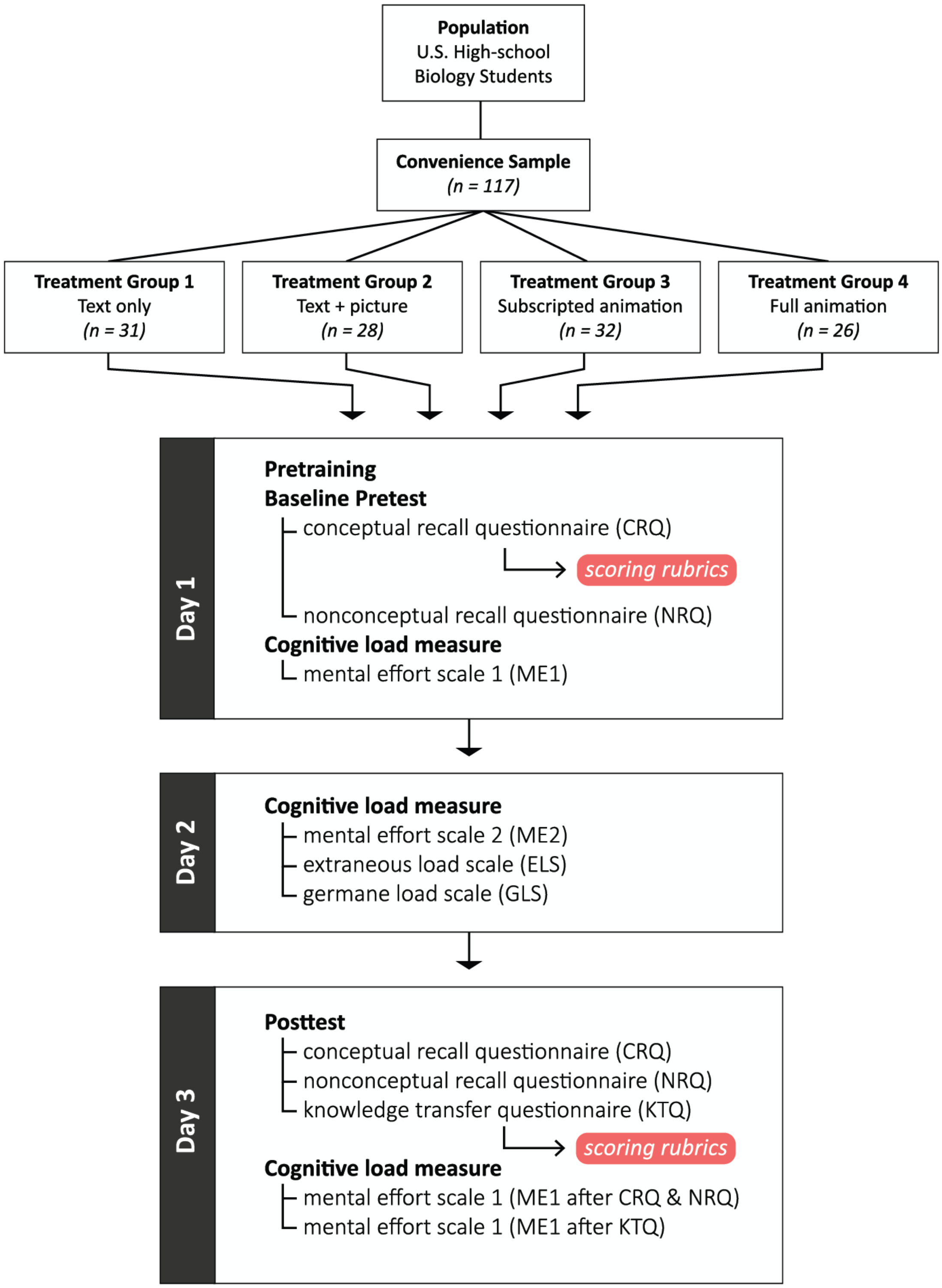


Figure 8. Schematic overview of the research design

Dependent variables

The dependent variables could be assigned to two categories: (a) measures of relevant knowledge and (b) measures of cognitive load.

Measures of knowledge. For the relevant knowledge, there were three independent variables, prior knowledge, rote learning, and meaningful learning. Prior knowledge (baseline score) was measured *before* the treatment using the nonconceptual recall questionnaire (NRQ) and the conceptual recall questionnaire (CRQ). Rote learning (posttest score 1; PTS1), which was compared to prior knowledge, was measured *after* the treatment using a repeat administration of the same nonconceptual recall questionnaire (NRQ) and the conceptual recall questionnaire (CRQ). Meaningful learning (posttest score 2; PTS2) was measured *after* the treatment using the knowledge transfer questionnaire (KTQ).

Measures of cognitive load. For cognitive load, there were three instruments used: (a) the mental effort scale, which was administered immediately after the pretest recall questionnaires (pretest NRQ and CRQ), after the treatment, after the posttest recall questionnaires (posttest NRQ and CRQ), and after the posttest knowledge transfer questionnaire; and (b) the extraneous load and (c) germane load scales, which were administered along with the mental effort scale immediately after the treatment only. The mental effort scale alone was used to calculate the cognitive load for the pretest (CLS1), and posttest recall (CLS2), and posttest knowledge transfer (CLS3). The three scales together were used to calculate the cognitive load for the treatment (i.e., the learning activity).

Overview of the study

The investigation was conducted in three phases (Phases 1, 2, and 3) over three days and took place in a biology classroom where the participants usually had their lessons. Phase 1 consisted of a pretraining exercise and a pretest. The pretest included two types of knowledge recall questionnaires: a conceptual recall questionnaire and a nonconceptual recall questionnaire.

Each phase ended with a cognitive-load measure that included the mental effort (all phases), extraneous load (Phase 2 only), and germane load (Phase 2 only) scales. Two variations of the mental effort scale were used: version 1 (phase 1 and 3) and version 2 (phase 2). Version 1 and version 2 differed only in a slight wording adjustment to give greater specificity to the pretest-posttest items and learning activity, respectively.

Phase 2 consisted of the learning activity based on instructional content on the theory of evolution by natural selection. During the learning activity, participants had access to loose-leaf and lined sheets of paper and pencils for drawing and note-taking based during the learning activity. The notes and drawings were collected and stored for later analysis. After the learning activity, participants completed a cognitive-load measure that included the mental effort, extraneous load, and germane load scales.

Phase 3 consisted of a posttest that included the same conceptual and nonconceptual recall questionnaires used in Phase 1, plus an additional knowledge transfer questionnaire. In Phase 3, the cognitive-load measure using the mental effort scale (version 1) was repeated twice, first immediately after completing the two recall questionnaires and again after completing the knowledge transfer questionnaire.

Sample

The sample came from a college preparatory high-school in California's San Francisco Bay Area. The school is recognized as one of the highest performing public high-schools in California and nationally, as evident by having been recognized four times as a National Blue Ribbon School, eight times as a California Distinguished School, and one time as a Gold Ribbon School. The school also has one of the most extensive Advanced Placement programs globally, offering 30 advanced placement courses distributed over 130 classes. Students generally complete two or more AP courses before graduating and have an average weighted grade point average of 3.87.

Admission to this high-school traditionally has been highly competitive and merit-based, attracting some of the most academically high-achieving students from the local municipality. The school's total enrollment is approximately 3,000 students with the following ethnic breakdown: 60% Asian or Pacific Islander, 15% White, 10% Hispanic or Latino, and 2% African American. Also, 40% of students are classified as socioeconomically disadvantaged and 2% as English learners.

For the current study, participants came from a convenience sample of 117 students, 76 girls (66%) and 41 boys (34%). The ethnic distribution of the sample was as follows: 78 Asian/Pacific Islander (67%), 22 White (19%), 15 Hispanic or Latino (13%), and 1 African American (<1%). Ninety-two of the participants were enrolled in AP biology, and 25 were enrolled in Introductory biology. All of the students in the introductory course were 9th graders. The AP biology students were a mixture of grades 10 through 12 with a distribution of 18, 50, and 25 students in grades 10, 11, and 12, respectively. A prerequisite for enrolling in AP biology as an 11th or 12th grader was evidence of a C-level grade or better in the ninth-grade introductory biology course and concurrent or previous (with C-level grade) enrollment in introductory chemistry. The more stringent prerequisite for 10th graders included the following: A-level grade in their ninth-grade introductory biology course; a recommendation from their introductory biology teacher; and concurrent enrollment in introductory chemistry. This stringency meant that, in AP biology, 10th graders were typically much higher academic achievers in science than their upper-class counterparts. Participants from the sample were assigned to treatment groups randomly, as described in the procedure section.

Protection of Human Subjects

The investigator adhered to all of the ethical standards and policies of the University of San Francisco and the school district's institutional review board (IRB). Also, they followed all of the human research protection regulations of the U. S. Department of Health, Education, and Welfare (1979).

Institutional Review Boards

Application for IRB approval was made to the IRB of the University of San Francisco and the respective school district with a letter of support from Matthew Mitchell, Ph.D. (see letter of support in Appendix B). The USF's IRB approved the study's IRB application on the grounds that it “did [not] require further IRB review or oversight as it is a standard educational improvement project" (see the USF IRB approval letter in Appendix C). The school district's IRB also approved the IRB application (see school district's IRB approval letter in Appendix D), which cleared the way for the study to be conducted. Both of the IRB approval notices were received via email messages. The investigator adhered to all ethical standards and policies of both the University of San Francisco and the school district's IRB for the protection of Human Subjects and all human research protection regulations of the U. S. Department of Health, Education, and

Welfare (1979). In compliance with the established rules and regulations, parents, guardians, and students were provided a letter explaining the purpose of the research study, along with a Parent/Guardian Consent Form and a Student Assent form for signing (Appendix E). Students took home both sets of forms and returned them before the start of the investigation.

Benefits and protections

There were no known risks to participants. There were clear benefits, however, in that participants had the opportunity to learn new concepts in biology. Participants' data initially were associated with the school-district-administered student email address. The email address was a sure way to ensure that data collected during the three phases remained linked. Once all of the data were linked, personal identification information, including email addresses, was removed. The unlinking of personal information occurred before data analysis to protect the participants' privacy.

The study was conducted with minimal interruption to normal teaching and learning because the learning content fit within the defined instructional timeframe and curriculum framework. As such, the content knowledge that was explored aligns with the district's graduation requirement. Furthermore, the pedagogical strategies used in the study are not atypical for U. S. classrooms. Specifically, students used laptop computers to interact independently with the learning material and assess the extent of their learning.

Treatment Description

All treatment groups engaged in the same learning activity; however, although the learning material in all groups consisted of the same factual and conceptual content, the presentation mode differed by treatment group. There was one control-group (Group 1) and three treatment groups (Groups 2, 3, and 4). Two of the groups (Group 1 and Group

2) were presented with static information (see Appendix F for printed copies), and the other two groups (Group 2 and 3) were given the same information as videos (see Appendix F for the links to the videos). Participants were assigned to a treatment group according to the following procedure:

1. The four sets of learning material were stored digitally on the school's Google Drive server.
2. The HTTP links to each set of learning material were generated using the link Shortening website tinyurl.com.
3. The four shortened hyperlinks were used to produce the *Links to the learning material* sheet (Appendix F), which included eight copies of each link (32 copies total).
4. For each of the four classes of student participants, one of the *Links to the learning material* sheet was printed.
5. The sheet was cut into strips, each with one of the four links printed on it.
6. The strips of paper were placed on the classroom desk one at a time without consideration to the link or student.
7. Students were assigned to desks in the order that they entered the room, rather than according to the seating arrangement of their regular lessons.

The participants in Group 1 had access to the learning content exclusively in a text format (text only). Group 2 (text + picture) was given the same textual information as Group 1 with the following modification: the main concepts were provided visually as embedded static pictures that were arranged with spatial contiguity to relevant text. The pictures used in the learning material for Group 2 were sourced from the animated videos used with treatment Groups 3 and 4. Group 3 (text + video) was given the same textual information as Group 1 with the following modification: (a) the main concepts were presented visually as part of an animated video and (b) the textual information was embedded in the video as subscripted text that was spatially and temporally synchronized with the relevant concepts. Group 4 (audio + video) had access to the same textual information from Group 1, but with the following modification: (a) the main concepts were presented visually as part of an animated video, (b) the textual information was instead presented aurally by animated characters in the animation, and (c) all participants of Group 4 had access to earphones to listen to the audio without causing a distraction to other participants. The earphones were either the participants' personal earphones or one supplied.

There were a few controlled variables across the groups. The first was that each participant received paper and pencil to use during the learning activity for note-taking, including drawing visual representations of concepts. The second was that each participant had access to the same type of laptop computer (i.e., Macbook Air) to access the relevant learning material. The third is that all of the data collection happened in the same learning environment, which was the classroom where the participants regularly took their lessons.

Instructional Unit

The learning activity involved instructional content based on a unit on evolution by natural selection, which is a central idea of the NGSS curriculum framework (National Research Council, 2013). The learning material provided information on the mechanism of evolution and covered a variety of relevant concepts such as speciation, ecological performance and selection mechanisms, sources of individual variation, genetic recombination and mutation, and the process of natural selection. Throughout the presented narrative, logical arguments were epistemologically explored to address the question of why and how relevant events that facilitate evolution occur. A good example of the use of such epistemic evidence is the inclusion of the story of the Galapagos finches as supporting evidence and as an example of the evolution in action.

Instrumentation

The study used seven instruments to measure the dependent variables (Appendix A). The instruments were grouped into three categories: (a) rote learning, (b) meaningful learning, and (c) cognitive-load.

The rote learning instrument

The study used two instruments for measuring rote learning: the nonconceptual and the conceptual recall questionnaires. Participants had 15 minutes to complete the two questionnaires in sequential order.

Nonconceptual recall. The nonconceptual recall questionnaire is a measure of rote learning that is used and described by Mayer and Gallini (1990). Drawing on Mayer and Gallini's (1990) work, 18 items were developed that required participants to recall factual information about evolution by natural selection. Twelve of the items were multiple-choice with four answer options each. Three of the items were checkboxes with four answer options. For the checkbox items, participants had to select all possible correct answer choices for each. Two of the items were true or false items. And, one of the items required participants to type in a numerical response. Each of the 18 items was weighted at one point, giving a total score for the nonconceptual recall questionnaire of 18 points.

Conceptual recall. According to Mayer and Gallini (1990), “learners who have built a runnable mental model are more likely to [recall concepts], as compared to students who have not built mental models" (p. 718). Thus, conceptual recall measures the extent to which the learner has a preexisting schema that encodes the recalled concept. Conceptual recall requires learners to explain or describe a concept that was explicitly presented or taught. The one item included in the questionnaire was scored using the rubric (Table 2). The item required participants to explain how new species arise from Table 2

Conceptual Recall Questionnaire Rubric

Concept	3 points	2 points	1 point	0 points	Points
Reproduction	Mentions or infers reproduction and describes the involvement of **2 features** of reproduction to the process of speciation: excess reproduction and inheritance.	...1 feature...	...no feature...	No mention of or inference to reproduction.	
Individual difference (uniqueness)	References individual difference (uniqueness) and mentions **2 sources** of uniqueness: mutations and recombination.	...1 source...	...no source...	No mention of individual difference	
Conditions for natural selection	Mentions or infers **3 conditions** for natural selection: diversity, selection factors, and competition.	...2 conditions...	...1 condition ...	No mention of any condition for natural selection	
Effects of natural selection	**Mentions 2 effects** of natural selection: survival of the fittest and inbreeding.	...infer or mentions… 2 effects...	...infers or mentions ...1 effect...	No mention of the effects of natural selection.	
Speciation	Mentions that speciation eventually occurs because of **3 factors**: inbreeding, divergence, reproductive isolation	...2 factors..	...1 factor...	No mention of speciation	
				Total	

preexisting species through natural selection. Speciation and natural selection were presented explicitly in the learning material. The rubric was developed based on only the learning material's information. It was used for scoring the participants' responses and were not shared with the participants. The responses were scored for five key concepts from the learning material: reproduction, individual differences (uniqueness), natural selection conditions, effects of natural selection, and speciation. The possible range for each key concept was 0 to 3 points according to the rubrics. Thus, for the five key concepts, the item had a total possible range of 0 to 15 points.

The meaningful learning instrument

The study used an instrument, the knowledge transfer questionnaire, for measuring meaningful learning. The knowledge transfer questionnaire (Appendix A) evaluated how well participants could apply the knowledge acquired during the learning activity to a problem that was not addressed explicitly by the learning material. The questionnaire included two items. The first item was a multiple-choice question that required participants to use what they learned to infer about how populations of organisms change over time. There are four possible answer options provided as a multiple-choice selection: (a) populations tend to decrease over time, (b) populations tend to increase over time, (c) populations tend to remain steady, and (d) populations tend to fluctuate over time. The ideal answer was option (c), followed by option (d). Options (a) and (b) are illogical because either instance would result in the extinction of all populations; however, a detailed explanation of this rationale is not warranted here. The scoring rubric for Item 1, which had a point range from 0 to 2, is provided in Table 3.

The second item of the knowledge transfer questionnaire was an open-ended constructed response question that required participants to provide a rationale for their answer choice on item 1. Answers to this question were scored on a 2-point scale for the extent of a clear causal relationship and rationale. The scoring rubric for Item 2, which had a point range from 0 to 2, is provided in Table 4. The scores from Item 1 and 2 were combined to form a raw knowledge transfer score that ranged from 0 to 4 points.

Table 3

Knowledge Transfer Questionnaire Item 1 Rubric

Selected response	**Score**	**Total score**
Steady populations	2 points	
Fluctuating populations	1 point	
Decreasing populations	0 points	
Increasing populations	0 points	

Table 4

Knowledge Transfer Questionnaire Item 2 Rubric

Selected Response from Item 1	2 points Full causal relationship	1 point Some casual relationship	0 points No causal relationship	Points
Steady populations	Over-reproduction causes intense competition for limited resources, leading to natural selection favoring the fittest individuals.			
Fluctuating populations	Changes in the environment cause fluctuation in resources, which causes populations to fluctuate.			
Increasing populations	Migration can present access to excess resources, or new adaptations can enhance competitive advantage. Either can lead to a population increase.			
Decreasing populations	Populations that fail to adapt to changing environments progressively decrease and eventually go extinct.			

The cognitive load instrument

The participant's overall cognitive-load was measured subjectively using three scales: the mental effort, extraneous load, and germane load scales. An overview of these scales, including the phases used, questions or items, and the rating scale ranges are provided in Table 5.

Table 5

Cognitive Load Instrument for Mental Effort, Extraneous Load and Germane Load

Phase	Questions/Items	7-point rating	Scales
Phase 1 & 3	1. How difficult did you find the questions on this page?	Not difficult (1) – Very difficult (7)	Mental Effort 1 (ME1)
Phase 2	2. How difficult was it for you to understand the content of the learning material?	Not difficult (1) – Very difficult (7)	Mental Effort 2 (ME2)
Phase 2	3. How often did you find yourself going back over the content of the learning material because you got distracted?	Not often (1) – Very often (7)	Extraneous load (ELS1)
Phase 2	4. How much did you find yourself worrying about not understanding the learning material?	Not often (1) – Very often (7)	Extraneous load (ELS2)
Phase 2	5. Once you were engaged with the learning material, how much thinking did you have to do to make sense of <u>un</u>familiar vocabulary in the passage?	Not much (1) – Very much (7)	Germane load (GLS1)
Phase 2	6. Once you were engaged with the learning material, how much thinking did you have to do to make sense of familiar vocabulary in the passage?	Not much (1) – Very much (7)	Germane load (GLS2)

Mental effort scale. A variation of the mental effort rating scale developed initially by Bratfisch and his colleagues (Bratfisch et al., 1972) was used in the current study. The efficacy of using the Bratfisch et al. (1972) scale to gauge the cognitive-load of learners is well established and supported in the literature (Ayres, 2006; Cheng & Beal, 2020; Paas, 1992; Paas & Van Merriënboer, 1994). Furthermore, empirical and theoretical studies (Borg, 1978; Borg et al., 1971) suggest some modifications to the scale (e.g., scale type, scale category, magnitude estimation, verbal labels) do not affect efficacy in a statically significant way. For example, in the study by Cheng and Beal (2020), the scale type was modified from a 9-point to a 5-point rating scale. Furthermore, in addition to the mental-effort category described in the original version, Cheng and Beal (2020) added two additional types: lesson difficulty and strategy difficulty.

In the current study, the Bratfisch et al. (1972) scale was modified to a 7-point rating scale using only the mental effort category. The investigator created two versions of this scale: mental effort scale 1 (ME1) and mental effort scale 2 (ME2). The ME1 was written with specificity to the task performance on the nonconceptual recall, conceptual recall, and knowledge transfer questionnaire of Phases 1 and 3. The ME2 was written with specificity to the task performance during the learning activity of Phase 2. After participants completed the 18-items from the two pretest recall questionnaires, they immediately responded to ME1. The scale was repeated in Phase 3 after participants completed the three posttest questionnaires.

Extraneous and germane load scales. As was previously mentioned, empirical and theoretical studies (Borg, 1978; Borg et al., 1971) suggest that some modifications to the scale (e.g., scale type, scale category, magnitude estimation, verbal labels) do not have a statistically significant effect on the efficacy. Adjusting how items are phrased, however, could provide additional insight into the working-memory function. Evidence of this comes from Cheng and Beal (2020), who added two other categories—that is, lesson difficulty and strategy difficulty—to their cognitive-load instrument. To measure extraneous and germane loads during the learning activity of the current study, a similar approach was used for generating two additional cognitive-load scales. According to Sweller (2010), the extraneous load increases when the learner becomes distracted by irrelevant elements to the learning task. Thus, based on the work of Borg (1978) and Borg et al. (1971), which is supported by Cheng and Beal (2020) and Sweller (2010), two items (Appendix A) were developed to evaluate external and internal distractions that participants may have experienced during the learning activity. External distractions were anything emanating from the classroom environment (i.e., noise, odor, temperature, light intensity) that was not germane to the learning activity and, thus, caused the participant to be distracted from the learning process. Internal distractions, however, were self-generated thought processes due to affective factors that were not germane to the learning task and, thus, also caused distraction from the learning task.

The first item measured the extent of external distractions by asking participants to rate how much they had to review the learning material due to distractions that they experienced in the classroom. The second item measured internal distractions by asking participants to rate how much they worried about not understanding the learning material. Although worrying may lead to processing some new information, it is generated internally, meaning that it leads to retrieval of information already stored in long-term memory. What is more, worrying can be distracting because it is driven by emotions, thus causing the learner to focus on elements that are irrelevant to the learning task.

To measure germane load, two items (Appendix A) were constructed that evaluated the extent of meaning-making during the learning activity. The two items asked participants to rate the amount of thinking they had to do during the learning activity to make sense of unfamiliar and familiar vocabulary in the passage, respectively.

Procedure

The investigation was conducted in three phases (Phases 1, 2, and 3) over 3 days in a biology classroom where the participants usually took their lessons. A schematic diagram of the research design is provided in Figure 8 (p. 77).

Prior to Phase 1, the investigator, who was also a teacher at the study site, received preliminary support from the school's principal and the chair of the science department pending IRB approval. Two IRB applications were made. The first was made to the University of San Francisco Institutional Review Board for the Protection of Human Subjects and the second to the school district's Research, Planning, and Assessment Department. Both were approved. Following IRB approval, letters and forms for consent and assent were sent out to parents and students, respectively. Once the consent and assent forms were returned, Phase 1 was cleared to proceed. The remainder of this section of the chapter provides a detailed overview of the research design.

Phase 1: Pretraining and baseline score

Phase 1 took place on Day 1 and consisted of a pretraining exercise, a pretest, and a cognitive-load measure. During the pretraining exercise, the investigator provided participants an overview of the investigative process, including a description of the three phases and the treatment groups (text only, text + visual, text + video, and audio + video). Group assignments were not made, however, until Phase 2. The participants were told that the pretest and posttest questionnaire's purpose was to evaluate their prior knowledge of the subject matter and the extent of their knowledge gained after the learning activity. They were not informed about the purpose of the cognitive-load measure. But, they received training on using Google Sheets to respond to questions about the learning content and the cognitive-load measure. They also were informed that they would have access to two sheets of paper that they could use to take notes, including to make conceptual diagrams.

Immediately following the pretraining exercise, participants were given 15 minutes to respond to the pretest. The pretest was administered via the Google Forms platform through an account provided by the local public-school system. As such, the participants' responses and their district-provided email addresses were collected and stored automatically. The email addresses, which are unique to each participant, allowed the investigator to use them as IDs for linking together data collected at different times. The purpose of the pretest was to gather information about the participants' prior knowledge about evolution by natural selection relevant to the content covered during the learning activity in Phase 2. The pretest included three questionnaires: Nonconceptual recall, conceptual recall, and cognitive-load questionnaires.

After completing the pretest, the cognitive-load questionnaire (CLQ) was presented within the same Google Forms. The cognitive-load measure included the mental effort scale (version 1; ME1) based on the scale initially developed by Bratfisch and his colleagues (Bratfisch et al., 1972).

Phase 2: Learning activity and treatment

Phase 2 took place on Day 2 and consisted of a learning activity and a cognitive-load measure. Before the learning activity, each participant was assigned to one of the four treatment groups as previously described (p. 82). They were also provided a Macbook Air laptop and paper for notetaking and drawings. The laptops were turned on and loaded up with Chromebook web browsers before the participants entered the classroom. Next to the laptop was a small strip of paper with one of four TinyURL links to the respective learning material (see Table 6). Participants assigned to Group 4 (full animation) also had access to headphones for listening to the audio part of the video. They were previously informed that they would have the option of using personal earphones or one provided by the researcher. Prior to starting the study, all participants were familiar with the use of Macbook Air laptops and the Google Apps (i.e., Forms, Docs, and Drive) used during the learning activity (Phase 2) and the questionnaires because of ongoing and regular use of these tools during previous classroom instruction.

Table 6

Links to Learning Materials

Group	Treatment	Links
1	Text only	http://tinyurl.com/txabc
2	Text + picture (text + static visual)	http://tinyurl.com/y59fsc66
3	Subscripted animation (text + dynamic visual)	http://tinyurl.com/yy2fqlks
4	Full animation (audio + dynamic visual)	http://tinyurl.com/vdabc

Once all participants understood what was expected of them, the learning activity could proceed; they were allowed to type the provided TinyURL link into the browser address bar to access the respective learning material. Participants then had 30 minutes to study the presented concepts. At the end of the 30 minutes, sharing permissions to access the learning material was rescinded via the investigator's Google account.

Immediately after completing the learning activity, the cognitive-load questionnaire was presented using Google Forms. This questionnaire included three scales: mental effort (version 2; ME2), extraneous load, and germane load. In addition, participants were asked to select one of the four types of learning material that they used during the learning activity, which was used to associate the treatment group (independent variable) with the participants' responses. The form also automatically recorded timestamps and the participants' unique school email addresses. The email addresses were used as identifiers to associate all of the participants' responses from Phase 1, 2, and 3. ME2 was based on the scale initially developed by Bratfisch and his colleagues (Bratfisch et al., 1972) but differed slightly from ME1 in that the wording of the mental effort question was modified to provide further specificity to the learning activity rather than the two recall questionnaires. The extraneous and germane load scales measured the participants' extraneous (distractions) and germane (learning) loads, respectively.

Phase 3: Posttest

Phase 3 took place on Day 3 and consisted of the posttest and two cognitive-load measures. The posttest was presented using the Google Forms platform and included three sets of question items: a conceptual recall questionnaire (CRQ), a nonconceptual recall questionnaire (NRQ), and a knowledge transfer questionnaire (KTQ). The CRQ and NRQ, which evaluated the extent of rote learning, were repeated from the pretest. The KTQ assessed the extent of meaningful learning. Participants had 15 minutes to complete both parts of Phase 3.

The posttest was embedded with two repeats of ME1 from the cognitive-load instrument. The first repeat of ME1 came after participants completed the CRQ and NRQ, and the second after the KTQ. Thus, the cognitive-load assessment occurred separately for tasks that evaluated rote learning and meaningful learning.

Data analysis

The data for this study were generated using the instruments described in the Instrumentation section of this chapter and provided in Appendix A. The SPSS Statistics software package was used for all statistical calculations, which are summarized and discussed in Chapter 4. The significance level for the family-wise type I error rate is set to .10. Partial η^2 and Cohen's d were used to calculate the effect size, with criteria specification provided by Cohen (1988; see Table 7).

Table 7

Cohen's Effect Size Criteria

Method	small	medium	high
Cohen's d	.20	.50	.80
Eta-squared (η^2)	.01	.06	.14

The remainder of this section provides an overview of the data-analysis procedure for pretest, posttest, and cognitive-load scores. It contains information regarding the statistical analyses used to address each of the research questions.

The Raw Data

The original sample size for this study included a total of 122 participants. Four participants were eliminated because they were absent from Phase 2 and 3 of the study and, therefore, did not have the crucial set of 12 data points needed to generate the rote learning, meaningful learning, and cognitive load scores. For the remaining 118 participants, 8 were missing the three data points from the Phase 1 variables (Phase 1 NRQ; Phase 1 CRQ; Phase 1 ME1) because they were absent on the first day of the study. All of these 8 participants were from the same class period, which was scheduled as the first lesson of the day. It is not unusual for the first lesson of the day to have a high number of students arriving late. Because they were all late to the lesson, they could not be admitted into the classroom. Instead of eliminating them from the study, the missing data points were imputed with the class average by treatment level for each missing variable. Thus, for this class, the average values of the three missing variables were calculated for each treatment group and those averages were imputed for the participants by treatment assignment. The descriptive statistics for the complete raw data set, including the imputed values, are provided in Table 8.

Table 8

Descriptive Statistics for the Raw Data

Variable	*N*	Mean	SD	% Max
Phase 1 NRQ	118	8.59	2.11	47%
Phase 1 CRQ	118	2.28	1.13	15%
Phase 1 ME1	118	5.07	1.08	73%
Phase 2 ME2	118	2.48	1.37	35%
Phase 2 ELSa	118	2.75	1.43	39%
Phase 2 ELSb	118	2.83	1.63	40%
Phase 2 GLSa	118	2.37	1.29	34%
Phase 2 GLSb	118	2.23	1.30	32%
Phase 3 NRQ	118	15.06	2.11	84%
Phase 3 CRQ	118	3.89	1.41	26%
Phase 3 KTQ	118	2.31	1.34	58%
Phase 3 ME1a	118	2.87	1.42	41%
Phase 3 ME1b	118	4.27	1.49	61%

As expected, the results from the descriptive statistics showed that the participants performed on average better on measures of nonconceptual recall than conceptual recall. The mean raw scores for Phase 1 NRQ were 3.8 times greater than that of Phase 1 CRQ. Likewise, the mean raw scores from Phase 3 NRQ was 3.9 times greater than that of Phase 3 CRQ. Furthermore, on both measures of knowledge recall, there was a noticeable improvement in mean scores following the Phase two treatment. The mean raw NRQ score increased from 8.54 (Phase 1) to 15.06 (Phase 3), a 1.8 times increase. Likewise, the mean raw CRQ score increased from 2.24 to 3.89, a 1.7 times increase.

The mean knowledge transfer scores (Phase 3 KTQ) were noticeably lower than both measures of knowledge recall; however, this is deceiving because the maximum raw score for KTQ was 4-points as compared to 18-points for the NRQ and 15-points for the CRQ. A more meaningful comparison is the %Max value, which is a measure of the percent of the maximum possible point that the mean score represents. In fact, the mean KTQ score was 58% of max, which was better than three of the four other knowledge measures. Phase 3 NRQ was better at 84% of max.

Measures of cognitive-load also yielded predictable results. After the treatment, the mental effort associated with knowledge recall (i.e., ME1 of Phase 1 and ME1a of Phase 3) dropped from 5.08 to 2.87 on the 7-point rating scale, a 44% decrease. This suggests that, on average, participants exerted less cognitive-load after having engaged in relevant learning during the Phase 2 treatment. The mental effort associated with the learning activity was 51% lower than for the pretest, and 14% lower for the posttest. This suggests that, on average, participants exerted more mental effort on the pretest and posttest than during the learning activity. This might indicate a disconnect between perceived information complexity and reality.

Pretest: Prior knowledge: Baseline Score

Results from the pretest provided a baseline score and information on the extent of the participants' prior knowledge associated with the conceptual and nonconceptual recall of information explicitly presented in the learning material. The baseline score only measured the extent of knowledge related to rote learning and not meaningful learning and was generated from the conceptual and nonconceptual recall questionnaires (CRQ and NRQ, respectively). The CRQ and NRQ were each scored separately to generate a raw score. The raw scores from these two measures were modified to a 10-point scale according to the procedure described below so that they were equally weighted.

Nonconceptual recall. To measure prior nonconceptual knowledge, the nonconceptual questionnaire (NRQ), which included 18 items, was administered during Phase 1 of the study. Each of the 18 items was weighted at one point to give a maximum raw score of 18 points. The raw NRQ score was then divided by 1.8 to generate a modified score on a 10-point scale.

Conceptual recall. To measure prior conceptual knowledge, the conceptual recall questionnaire (CRQ), which included one free-response question, was administered. Responses from the CRQ were scored based on five items addressed in the learning material during Phase 2. Each of these items received an integer score of 0, 1, 2, or 3 points based on the criteria outlined in the rubrics in Table 2. The maximum raw score on the CRQ was 15 points. This raw score was divided by 1.5 to generate a modified score on a 10-point scale.

Baseline score. The modified conceptual recall score was averaged with the modified nonconceptual recall score to generate the baseline score, which has a maximum possible value of 10 points.

The descriptive statistics for the modified scores of the nonconceptual recall questionnaire (NRQ) and conceptual recall questionnaire (CRQ), which were averaged together to generate the baseline score, are provided in Table 9. Results of a one-way analysis of variance (ANOVA) supported the null hypothesis that there were no statistically significant between-group differences on measures of Pretest NRQ, Pretest CRQ, and Baseline prior knowledge.

Table 9

Descriptive Statistics for the Modified Baseline Scores

Instrument	*N*	Mean	SD
Pretest NRQ	118	4.77	1.17
Pretest CRQ	118	1.52	0.75
Baseline (averaged)	118	3.14	0.80

Note: These scores are derivatives of the raw NRQ and CRQ scores from Phase 1 (refer to the data analysis section, pp. 98-100, for a complete explanation of how the baseline scores were calculated)

Posttest scores: Rote and meaningful learning

Two posttest scores were generated: (a) posttest score 1 (PTS1) and (b) posttest score 2 (PTS2). The PTS1 is the rote learning score based on the same nonconceptual recall and conceptual recall questionnaires used in the pretest. The PTS2 is the meaningful learning score generated from the knowledge transfer questionnaire of the posttest.

Rote learning: Posttest score 1. PTS1 is the rote learning score that was generated from participants' responses to the same nonconceptual recall and conceptual recall questionnaires used in the pretest. PTS1 is comparable to the baseline score and was derived from the raw NRQ and CRQ scores via the same method used to calculate the baseline scores (pp. 98-100). The only difference is that the data came from the posttest rather than the pretest responses. The descriptive statistics for the modified scores of the nonconceptual recall questionnaire (NRQ) and conceptual recall questionnaire (CRQ), which were averaged together to generate the PTS1 score, are provided in Table 10.

Table 10

Descriptive Statistics for the Rote Learning Score (PTS1)

Instrument	*N*	Mean	SD
Posttest NRQ	118	8.37	1.17
Posttest CRQ	118	2.59	0.94
PTS1 (averaged)	118	5.48	0.77

Note: These scores are derivatives of the raw NRQ and CRQ scores from Phase 3 (refer to Instrumentation, pp. 101-102, for a complete explanation)

Meaningful learning: Posttest score 2. The PTS2 is the meaningful learning score generated from participants' responses to the posttest knowledge transfer questionnaire (KTQ). The knowledge transfer questionnaire included two items administered during Phase 3. The first item was a four-option multiple-choice question that asked participants to predict the impact of environmental changes on populations of organisms. Although the learning material did not address this problem explicitly, participants could use their newly acquired knowledge to make relevant inferences. Item 1 received either 2, 1, or 0 points based on the scoring rubrics in Table 3 (p. 88).

Item 2 of the knowledge transfer questionnaire was an open-ended constructed response question that required participants to provide a rationale for their answer choice on item 1. Answers to this question were scored on a 2-point scale based on the extent of a clear causal relationship and rationale for the selected response to Item 1. The scoring rubric for Item 2 had a point range from 0 to 2 and is provided in Table 4 (p. 88). When combined, the raw value for PTS2 was 2 + 2 = 4 points. This raw score was multiplied by 2.5 to adjust it to a 10-point scale. Three sample responses and their scores on the rubrics are presented in Appendix G.

The PTS2, the meaningful learning score, was generated from the knowledge transfer data through a two-step process: (a) the knowledge transfer score was determined using the knowledge transfer rubrics and (b) the score was modified to a 10-point scale. Table 11 provides the descriptive statistics for the PTS2.

Table 11

Descriptive Statistics for the Meaningful Learning Scores

Instrument	*N*	Mean	SD
PTS2 (KTQ)	118	5.78	3.36

Note: The PTS2 scores are derivatives of the raw KTQ scores from Phase 3 (refer to Instrumentation, pp. 101-102, for a complete explanation)

Cognitive load scores

The participants' overall cognitive-load was measured subjectively using three scales: (a) the mental effort scale, (b) extraneous load scale, and (c) germane load scale. An overview of these scales (Table 12), including the phases used, question items, and the rating scale, is provided in Table 5 (p. 89). Four cognitive-load scores were generated: (a) CLS1 for the prior knowledge instrument of the pretest, (b) CLS2 for the rote learning

Table 12

Cognitive load scores

Cognitive load score	Phase	Task	Scale
CLS1	Phase 1	Pretest	ME1
CLS2	Phase 3	Posttest	ME1
CLS3	Phase 3	Posttest	ME1
CLS4	Phase 2	Learning	ME2/ELS/GLS

Note: CLS1-4 are the four cognitive-load scores. ELS and GLS are the extraneous and germane load scales, respectively. ME1 and ME2 are the two versions of the mental effort scales. The items used to generate data for ME1 and ME2 were semantically similar except that ME1 was specific to the pretest and posttest (Phases 1 and 3), and ME2 was specific to the learning activity instrument of the posttest, (c) CLS3 for the meaningful learning instrument of the posttest, and (d) CLS4 for the learning activity. CLS1, 2, and 3 were all measured using just the mental effort scale. CLS4 was an average score of all three scales using the data collected in Phase 2. The scores for each scale were based on results from one (e.g., mental effort scale) or two (extraneous and germane load scales) items scored on a 7-point rating scale. For the extraneous and germane load scales, scores from their respective two items were averaged. All of the raw scores were modified to a 10-point scale by dividing the raw score by .7. The descriptive statistics for the cognitive-load scores, including the germane and extraneous load scores (GLS and ELS), are provided in Table 13.

Table 13

Descriptive Statistics for the Cognitive Load Scores

Variable	Scale	*N*	Mean	SD
CLS1 (Pretest)	ME1	118	7.24	1.54
CLS2 (Posttest)	ME1a	118	4.10	2.03
CLS3 (Posttest)	ME1b	118	6.10	2.14
ME2 (Learning)	ME2	118	3.55	1.96
ELS (Learning)	ELS	118	3.99	1.81
GLS (Learning)	GLS	118	3.28	1.75
CLS4 (Learning)	ME2, ELS, GLS	118	3.61	1.60

Because the metal effort, extraneous load, and germane load all measure cognitive-load, these scales' reliability, and internal consistency was statistically evaluated using Cronbach coefficient alpha, with $\boldsymbol{\alpha} \geq .65$ set as the minimum threshold. Next, three correlation analyses were performed between the (a) two items of the extraneous load scale, (b) two items of the germane load scale, and (c) between all of the items that together account for the cognitive-load score. The last analysis included the two items each of the ELS and GLS and the mental effort scale (ME2). The results of these analyses are reported in Table 14.

Table 14

Internal Consistency for the Cognitive Load Scales

Analysis	N	α value
ELS Items 1 & 2	118	0.54
GLS items 1 &2	118	0.89
ELS 1 & 2, GLS 1 & 2, and ME2	118	0.82

The coefficient alpha reliability estimate for the extraneous load items fell below the minimum threshold, **α** = .54. The two items (i.e., ELS1 and ELS2) were intended to measure the level of distraction that participants experienced during the learning activity. The question items, however, were framed in slightly different ways to measure the extent of external and internal distractions, respectively. Internal distractors include any self-generated thoughts that were unrelated to the learning task. External distractors included all other types of distractions that were externally generated and irrelevant to the learning task. As elements that reduced the participant's attention to the learning activity, the internal and external distractors could both contribute to extraneous-load by occupying more of the working memory space dynamically shared with germane-load. That the coefficient alpha reliability estimate fell below the minimum threshold could be due to the fact that the two items were designed to detect distractors originating from different sources. While it is beneficial to explore the extent of these two sources of distraction, it might have been better to measure internal and external distraction separately through the use of different scales that each had multiple question items. The scale used in this analysis, thus, has no truly comparative set of items for reliably evaluating internal consistency. Instead, the result of the efficient alpha reliability estimate suggests that the two items may indeed have measured factors that were not exactly identical.

The coefficient alpha reliability estimate for the germane load items was acceptable, $\boldsymbol{\alpha} = .89$, suggesting that the two items were reliably consistent at measuring productive learning of the relevant information. The coefficient alpha reliability estimate for all of the cognitive load items together (i.e., ELS1, ELS2, GLS1, GLS1, and ME2) was acceptable, $\boldsymbol{\alpha} = .82$, suggesting the items together were reliably consistent at measuring total cognitive load as a function of extraneous and germane load.

Research questions

The current study is based on three research questions that focus on the effects of information modality (Question 1), prior knowledge on learning (Question 2), and the interaction of prior knowledge and borrowed visuals on cognitive-load (Question 3). This section provides an overview of how these questions were investigated

Question 1

Research question 1 examines the effects of the mode of information presentation (independent variables) on three dependent variables: (a) rote learning, (b) meaningful learning, and (c) cognitive-load. To address the research question, a one-way analysis of variance (ANOVA) was used to determine whether there was a statistically significant between-group difference on the mean scores for each responding variable. Information modality (i.e., treatment level) was entered as the independent variable and the meaningful learning, rote learning, and cognitive load scores were entered separately as the dependent variables (Figure 9). As a control, the analysis was also performed on the pretest results of prior knowledge scores, that is, the pretest NRQ, pretest CRQ, and baseline scores (see Table 9, p. 100).

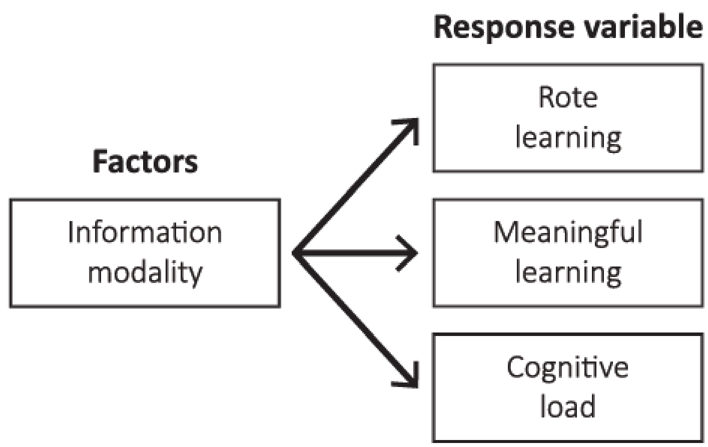


Figure 9. Schematic of the one-way ANOVA analysis

Question 2

Research question 2 examined the expertise reversal effect as a function of information modality (i.e., treatment; independent variable) and prior knowledge (i.e., expertise level) on rote learning score (PTS1; response variable 1) and meaningful learning (PTS2; response variable 2). This question evaluated the assumption that increased visual support through the use of borrowed visuals interacts inversely with prior knowledge on learning. Accordingly, participants with high prior knowledge were predicted to perform better under the text-only treatment (Treatment 1) and worse on the animated treatment (Treatment 3). The prior knowledge scores were first categorized by expertise as low, medium, or high PrKn. To accomplish this, the descriptive statistics of the prior knowledge scores were obtained (Table 15) to determine the mean and range of the scores. The upper third of the range was assigned expertise of *high* PrKn, the middle third to *medium* PrKn, and the lower third to *low* PrKn. Next, the knowledge gained was calculated by taking the difference between the rote learning (PTS1) and the prior knowledge (PrKn) scores. The mean values of these knowledge gain scores were then compared by treatment and expertise, and the significance of observed differences between the groups was tested via ANOVA.

Table 15

Descriptive Statistics of PrKn and Expertise Levels

N	PrKn mean	PrKn min	PrKn max	Low (*n* = 31)	Medium (*n* = 28)	High (*n* = 33)
92	3.09	1.67	4.67	< 2.67	2.67 - 3.67	> 3.67

A two-way analysis of variance (ANOVA) was used to determine whether or not there were statistically significant between-group differences in the mean learning scores. Mode of information presentation (i.e., treatment level) and expertise level were entered as fixed factors (Figure 10).

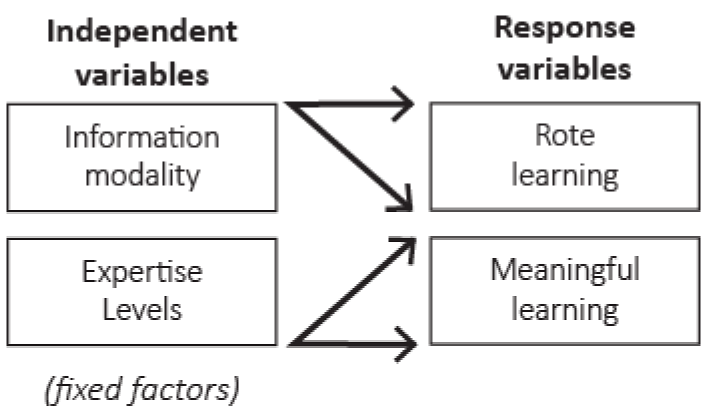


Figure 10. Schematic of the ANOVA analysis on learning scores

The response variables were the rote and meaningful learning scores. This analysis allowed for testing whether information modality had an effect on rote learning score (PTS1) and meaningful learning scores (PTS2) after accounting for prior knowledge expertise. Important to note that that the ANOVA required the segmenting of the sample into 12 groups (i.e., 3 expertise x 4 treatment levels). The largest of these groups (expertise level 2 under treatment level 3) included 17 participants. These sample sizes, with n < 30, meet the minimum requirements to fulfill the central limit theorem.

Question 3

Research question 3 examined the expertise reversal effect as a function of information modality (i.e., treatment; independent variable) and prior knowledge (i.e., expertise level) on the cognitive-load associated with the rote learning task (CLS2) and the meaningful learning task (CLS3). This question evaluates the assumption that increased support through the use of borrowed visuals interacts directly with the cognitive-load with sensitivity to prior knowledge. Accordingly, participants with higher prior knowledge should experience lower cognitive-load under text-only treatment (Treatment 1) and high cognitive-load under the animated treatment (Treatment 3).

A two-way analysis of variance (ANOVA) was used to determine whether or not there were statistically significant between-group differences in the mean cognitive load scores. Mode of information presentation (i.e., treatment level) and expertise level were entered as fixed factors (Figure 11).

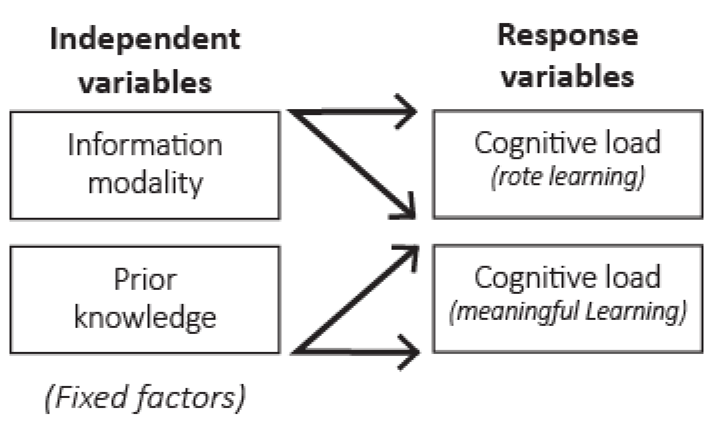


Figure 11. Schematic of the ANOVA analysis on cognitive load scores

The responding variables were the cognitive load scores associated with rote learning (CLS2) and meaningful learning (CLS3). This analysis allowed for testing whether the information modality had an effect on cognitive load after accounting for prior knowledge.

CHAPTER IV

RESULTS

The purpose of this study was to investigate how prior knowledge and the integration of information modalities (i.e., text, audio, static visual, dynamic visuals) promotes rote learning (information retention) and meaningful learning (knowledge transfer) in science. This chapter includes the results of the data analyses that focused on addressing the three research questions. It concludes with a summary of these results.

Research Questions

The research study addressed three research questions that focused on the effects of information modality (Question 1), prior knowledge on learning (Question 2), and the interaction of prior knowledge and borrowed visuals on cognitive-load (Question 3). This section provides an overview of how these questions were investigated.

Question 1

Research question 1 examined the effects of the mode of information presentation (independent variables) on three sets of dependent variables: (a) knowledge recall (PTS1), (b) knowledge transfer (PTS2), and (c) cognitive load (CLS2, 3, and 4). Prior knowledge (PrKn) and its associated cognitive load (CLS1) were also included for baseline comparison. The descriptive statistics for all of the dependent variables by treatment are provided in Table 16, and visual comparisons of the means from the four treatment groups for each independent variable are provided in the graphs presented in Appendix H.

A one-way analysis of variance (ANOVA) was conducted to evaluate the statistical significance of the between-group differences in the means. A schematic

Table 16

Descriptive Statistics for Dependent Variables

Variables	Statistics	Text only	Text + picture	Subscripted animation	Full animation	Total
Pretest NRQ	*M*	4.84	4.45	4.68	5.15	4.77
	n	31	28	33	26	118
	SD	0.93	1.43	0.85	1.40	1.17
Pretest CRQ	*M*	1.55	1.41	1.52	1.60	1.52
	n	31	28	33	26	118
	SD	0.78	0.73	0.92	0.51	0.75
PrKn	*M*	3.19	2.93	3.10	3.38	3.14
	n	31	28	33	26	118
	SD	0.72	0.93	0.71	0.82	0.80
Posttest NRQ	*M*	8.42	8.15	8.13	8.82	8.37
	n	31	28	33	26	118
	SD	0.93	1.14	1.45	0.97	1.17
Posttest CRQ	*M*	2.32	2.55	2.69	2.85	2.59
	n	31	28	33	26	118
	SD	1.10	0.70	0.92	0.93	0.94
PTS1	*M*	5.37	5.35	5.41	5.84	5.48
	n	31	28	33	26	118
	SD	0.74	0.65	0.94	0.60	0.77
PTS2	*M*	5.97	6.25	5.61	5.29	5.78
	n	31	28	33	26	118
	SD	2.86	3.94	3.54	3.11	3.36
CLS1	*M*	7.40	7.13	7.72	6.58	7.24
	n	31	28	33	26	118
	SD	1.59	1.61	1.42	1.38	1.54
CLS2	*M*	3.92	4.23	4.81	3.30	4.10
	n	31	28	33	26	118
	SD	1.77	2.00	2.55	1.20	2.03
CLS3	*M*	6.13	6.12	6.54	5.49	6.10
	n	31	28	33	26	118
	SD	1.89	2.36	2.29	1.92	2.14
ME2	*M*	3.64	3.47	3.85	3.13	3.55
	n	31	28	33	26	118
	SD	1.95	1.84	2.33	1.57	1.96
ELS	*M*	4.08	3.65	4.33	3.82	3.99
	n	31	28	33	26	118
	SD	1.77	1.98	1.86	1.62	1.81
GLS	*M*	3.29	3.60	3.48	2.69	3.29
	n	31	28	33	26	118
	SD	1.66	1.77	2.12	1.22	1.75
CLS4	*M*	3.29	3.57	3.89	3.21	3.61
	n	31	28	33	26	118
	SD	1.66	1.67	1.87	1.25	1.60

representation of the ANOVA used for investigating this research question is provided in Figure 9 (p. 106). Levene's test of equality of variances for the dependent variables (i.e., PrKn, PTS1, PTS2, CLS1, CLS2, CLS3, and CLS4) was statistically significant only for CLS2. Thus, the assumption of homogeneity of variance was met on all of the dependent variables except for the cognitive load associated with rote learning (CLS2). The ANOVA results (Table 17) were statistically significant only for PTS1, CLS1, and CLS2 (significance level set at $p = .10$). ANOVA results on the Posttest NRQ and CRQ scores were statistically significant only for Posttest NRQ, which suggests nonconceptual recall had a greater effect on the PTS1. The effect size for Posttest NRQ, however, was small ($\eta^2 = .05$), which suggests a small level of practical importance. The effect size was medium for PTS1 ($\eta^2 = .06$), CLS1 ($\eta^2 = .07$), and CLS2 ($\eta^2 = .07$), suggesting a medium level of practical importance for CLS1 and CLS2.

Table 17

One-way ANOVA for the Dependent Variables

Variables	F	p	η^2	Power
Pretest NRQ	1.74	.16	0.04	.58
Pretest CRQ	0.31	.82	0.01	.19
PrKn	1.49	.22	0.04	.52
Posttest NRQ	2.16*	.06	0.05	.66
Posttest CRQ	1.65	.18	0.04	.55
PTS1	2.49*	.06	0.06	.72
PTS2	0.42	.74	0.01	.22
CLS1	3.00*	.03	0.07	.80
CLS2	2.97*	.04	0.07	.79
CLS3	1.17	.33	0.03	.43
ME2	0.69	.56	0.02	.30
ELS	0.82	.49	0.02	.33
GLS	1.44	.23	0.04	.50
CLS4	0.88	.45	0.02	.35

* Statistically significant at the .10 level

The results of Post hoc analyses (Table 18) using the Tukey HSD criterion with CLS1 was significantly lower under the full animation condition (Treatment 4; M = 6.58, SD = 1.38) as compared to the subscripted animation condition (Treatment 3; M = 7.72, SD = 1.42). The results also indicated that the average cognitive load associated with CLS2 was significantly lower under the full animation condition (Treatment 4; M = 3.30, SD = 1.20) as compared to the subscripted animation condition (Treatment 3; M = 4.81, SD = 2.55).

Table 18

Post Hoc Results for CLS1 and CLS2

Variables	1	2	3	4
1. text only		0.28	-0.32	0.82
2. text + picture	0.32		-0.60	0.55
3. subscripted animation	0.89	0.57		1.15*
4. full animation	-0.62	-0.94	-1.51*	

* Statistically significant at the .10 level
Note: CLS1 mean differences are above the diagonal and CSL2 mean differences are below the diagonal.

Question 2

Research question 2 examined the expertise reversal effect as a function of information modality (i.e., treatment; independent variable) and prior knowledge on rote learning score (PTS1; response variable 1) and meaningful learning (PTS2; response variable 2). Participants' prior knowledge scores were used for assignment of expertise level (i.e., low, medium, or high expertise) according to the procedure described in the methodology chapter (pp. 106-107). This question evaluated the assumption that increased visual support through the use of borrowed visuals would interact with prior knowledge inversely on learning. A two-way analysis of variance (ANOVA) was used to determine whether or not there were statistically significant between-group differences in the mean learning scores. Mode of information presentation (i.e., treatment level) and prior knowledge (i.e., expertise level) were entered as fixed factors (Figure 10, p. 108). The response variables were the mean rote (PTS1) and meaningful learning scores (PTS2), which are presented in Table 19. This analysis allowed for testing whether

Table 19

Descriptive Statistics for PTS1 and PST2 (Treatment by Expertise)

Variable	Treatment	Expertise	*n*	*M*	*SD*
PTS1	1: Text only	1 (low)	8	5.35	0.46
		2 (medium)	14	5.19	0.84
		3 (high)	9	5.69	0.76
	2: Text + picture	1 (low)	14	5.12	0.75
		2 (medium)	6	5.39	0.29
		3 (high)	8	5.72	0.52
	3: Subscripted animation	1 (low)	9	5.04	1.46
		2 (medium)	17	5.48	0.65
		3 (high)	7	5.72	0.59
	4: Full animation	1 (low)	6	5.47	0.37
		2 (medium)	8	5.85	0.67
		3 (high)	12	6.01	0.60
PTS2	1: Text only	1 (low)	8	5.63	3.72
		2 (medium)	14	5.18	2.68
		3 (high)	9	7.50	1.77
	2: Text + picture	1 (low)	14	5.89	4.11
		2 (medium)	6	6.67	4.65
		3 (high)	8	6.56	3.52
	3: Subscripted animation	1 (low)	9	4.72	3.84
		2 (medium)	17	5.59	3.80
		3 (high)	7	6.79	2.38
	4: Full animation	1 (low)	6	5.00	2.74
		2 (medium)	8	5.63	3.72
		3 (high)	12	5.21	3.10

information modality had an effect on rote learning score (PTS1) and meaningful learning scores (PTS2) after accounting for prior knowledge (i.e., expertise level).

Levene's test of equality of variances for the response variables (i.e., PTS1, PTS2) was statistically significant only for PTS1. The assumption of homogeneity of variance, therefore, was met only for meaningful learning (PTS2). To further evaluate the extent to which nonconceptual and conceptual recall contributed to the overall rote learning score, Levene's tests were also performed on the posttest NRQ and CRQ scores. The results of these analyses were statistically significant only for nonconceptual recall (posttest NRQ), reflecting the results of the previous analysis (question 1, p. 112) that the nonconceptual recall had a greater effect than did the conceptual recall on the PTS1. The ANOVA results (Table 20) for the interaction of treatment and expertise levels were statistically

Table 20

Two-way ANOVA for Learning (Treatment by Expertise)

Source	Variables	SS	df	MS	*F*	η^2
Treatment	Posttest NRQ	5.48	3	1.83	1.45	.04
	Posttest CRQ	3.47	3	1.16	1.43	.04
	PTS1	2.50	3	0.83	1.50	.04
	PTS2	17.12	3	5.71	0.48	.01
Expertise	Posttest NRQ	4.78	2	2.39	1.90	.04
	Posttest CRQ	6.31	2	3.15	3.88*	.07
	PTS1	5.03	2	2.51	4.53*	.08
	PTS2	25.36	2	12.68	1.07	.02
Treatment* Expertise	Posttest NRQ	12.56	6	2.09	1.66	.09
	Posttest CRQ	7.66	6	1.28	1.58	.08
	PTS1	1.19	6	0.20	0.36	.02
	PTS2	27.23	6	4.54	0.38	.02
Error	Posttest NRQ	133.50	106	1.26		
	Posttest CRQ	85.93	106	0.81		
	PTS1	58.73	106	0.55		
	PTS2	1254.03	106	11.83		
Total	Posttest NRQ	156.32	117			
	Posttest CRQ	103.37	117			
	PTS1	73.92	117			
	PTS2	1323.74	117			

* Statistically significant at the .10 level

significant for neither PTS1 and nor posttest NRQ. These results, therefore, lack sufficient evidence to reject the null hypothesis that there were no statistically significant between-group differences for the interaction of prior knowledge (i.e., expertise) with the mode of information presentation on learning.

Question 3

Research question 3 examined the expertise reversal effect as a function of information modality (i.e., treatment; independent variable) and prior knowledge (i.e., expertise level) on the cognitive-load associated with the rote learning task (CLS2) and the meaningful learning task (CLS3). Participants' prior knowledge scores were used for assignment of expertise level (i.e., low, medium, or high expertise) according to the procedure described in the Methodology chapter (pp. 106-107). This question evaluates the assumption that increased support through the use of borrowed visuals interacts directly with the cognitive-load with sensitivity to prior knowledge (i.e., expertise level).

A two-way analysis of variance (ANOVA) was used to determine whether or not there were statistically significant between-group differences in the mean cognitive load scores (Table 21). Mode of information presentation (i.e., treatment level) and prior knowledge (i.e., expertise level) were entered as fixed factors (Figure 11, p. 109). The response variables were the cognitive load scores associated with rote and meaningful learning. This analysis allowed for testing whether information modality had an effect on the cognitive load scores associated with rote learning score (CLS2) and meaningful learning scores (CLS3) after accounting for prior knowledge (i.e., expertise).

Levene’s test of equality of variances for the response variables (i.e., CLS2, CLS3) was statistically significant only for CLS2. The assumption of homogeneity of variance, therefore, was rejected only for the cognitive load associated with rote learning

Table 21

Descriptive Statistics for CLS2 and CLS3 (Treatment by Expertise)

Variable	Treatment	Expertise	*n*	*M*	*SD*
CLS2	1: Text only	1 (low)	8	4.11	1.78
		2 (medium)	14	3.88	2.20
		3 (high)	9	3.81	1.01
	2: Text + picture	1 (low)	14	4.59	2.11
		2 (medium)	6	4.05	2.10
		3 (high)	8	3.75	1.86
	3: Subscripted animation	1 (low)	9	4.29	2.67
		2 (medium)	17	5.71	2.58
		3 (high)	7	3.27	1.36
	4: Full animation	1 (low)	6	3.33	1.95
		2 (medium)	8	3.39	1.06
		3 (high)	12	3.21	0.89
CLS3	1: Text only	1 (low)	8	6.61	2.01
		2 (medium)	14	5.41	1.87
		3 (high)	9	6.83	1.56
	2: Text + picture	1 (low)	14	5.82	2.47
		2 (medium)	6	6.19	3.34
		3 (high)	8	6.61	1.31
	3: Subscripted animation	1 (low)	9	6.03	1.86
		2 (medium)	17	7.56	2.19
		3 (high)	7	4.69	1.79
	4: Full animation	1 (low)	6	4.76	1.73
		2 (medium)	8	5.89	1.42
		3 (high)	12	5.60	2.32

(CLS2). The ANOVA results (Table 22) for the interaction of treatment and expertise levels, however, was statistically significant for the cognitive load associated with meaningful learning (CLS3) but not rote learning (CLS2). The ANOVA results for CLS3, however, did not have a high level of confidence due to the results of Levene's test of homogeneity of variance. These results, therefore, lack sufficient evidence to reject the null hypothesis that there were no statistically significant between-group differences in cognitive load due to the interaction of prior knowledge with the mode of information presentation on cognitive load.

Table 22

Two-way ANOVA for Cognitive Load (Treatment by Expertise)

Source	Variables	SS	df	MS	*F*	η^2
Treatment	CLS2	16.89	3	5.63	1.47	0.04
	CLS3	11.73	3	3.91	0.92	0.03
Expertise	CLS2	10.87	2	5.43	1.41	0.03
	CLS3	4.09	2	2.04	0.48	0.01
Treatment* Expertise	CLS2	24.47	6	4.08	1.06	0.06
	CLS3	59.82	6	9.97	2.33*	0.12
Error	CLS2	407.52	106	3.85		
	CLS3	452.17	106	4.27		
Total	CLS2	459.75	117			
	CLS3	527.81	117			

* Statistically significant at the .10 level

Summary of Results

The purpose of this study was to evaluate the utilitarian reliability of the modality effect and the expertise reversal principles within the context of an actual biology classroom of secondary-school students. Specifically, this study considered the effects of the mode of information presentation (i.e., text, audio, static visual, dynamic visuals) on knowledge acquisition and cognitive-load and the interaction of prior domain-specific knowledge with information modality on rote and meaningful learning.

Research question 1 guided the investigation about whether there is an effect of the mode of information presentation on learning and cognitive-load. The modality effect was detected for rote learning and its associated cognitive load, and also for the cognitive load associated with meaningful learning. The results for the meaningful learning cognitive load, however, failed to meet the assumption of homogeneity of variance, and therefore, the results cannot be reliably interpreted. Although the ANOVA analysis did not yield a statistically significant result for all of the variables, it is worth considering the overall trend (Appendix H) for the mean values of the groups. Concerning measures of rote learning (PTS1), the ANOVA results of which were found to be statistically significant, (Table 16) the full animation group (Group 4) outperformed the other groups. Interestingly, the ANOVA results from comparing the cognitive-load scores that coincide with the PTS1 from the four groups also yielded a statistically significant effect of mode of information presentation (Table 16). PTS1 and CLS2 were derived from the posttest and measured rote learning and the associated cognitive-load, respectively. CLS2 and PST1 were the two most statistically significant results (Table 16), suggesting that the modality effect may be more relevant for rote learning than meaningful learning. Such a trend is predicted by the expertise reversal principle (Kalyuga, 2014).

Concerning meaningful learning (PTS2), Group 2 (text+visual) presented the highest mean score of 6.25, followed by Group 1 (text-only) with a mean score of 5.97 and Group 3 (text+video) with a mean of 5.61. Full animation (Group 4) had the lowest mean score of 5.29. Although the observed differences between the four groups on measures of meaningful learning were not statistically significant, the trend is somewhat indicative of the predictions based on the expertise reversal principle. While for rote learning, the text conditions (Groups 1 and 2) outperformed the animation conditions (Groups 3 and 4), the reverse was observed for meaningful learning.

CHAPTER V

DISCUSSION OF RESULTS

The purpose of this study was to investigate how prior knowledge and the integration of information modalities (i.e., text, audio, static visual, dynamic visuals) promotes rote learning (information retention) and meaningful learning (knowledge transfer) in science. This chapter, which focuses on digesting and interpreting the results, is organized into seven sections: (a) summary of the study, (b) summary of the findings, (c) limitations of the study, (d) discussions of the findings, (e) conclusion, (f) implications of the study for future research, and (g) implications of the study for teaching and learning practices.

Summary of the Study

The two multimedia design principles of interest in this study are the modality and expertise reversal principles. Although there is empirical evidence to support these principles, two features of the existing research leave room for further investigation. First, many current studies were conducted in a controlled laboratory setting (Butcher, 2014). For such controlled research designs, there is a range of affective factors that could influence students' emotions and temperament (Snow et al., 1996). These factors could interact subsequently with conative factors that effect the learner's motivation to learn and cognitive control (Mayer, 2011). Although such studies are essential for establishing the design principles' validity, they may not always predict students' learning outcomes in an actual classroom setting. The current study used a typical biology classroom of secondary-school students to evaluate the design of the modality principle and the expertise reversal principle utilitarian reliability. Furthermore, the science concepts selected for the study align with the established curriculum framework for the science program at the school. Although the study design is controlled, the learning content reflects students' learning resources during a regular instructional unit.

The second feature of existing research that leaves room for further investigation is that the empirical database needed to establish the design principles' validity remains incomplete. The current study will add to the literature by examining how the learner's use of static and dynamic visuals correlates with rote and meaningful learning measures. As such, not only does the current investigation reflect the reality of the learning environment within a typical classroom setting but also provides additional insight into how a learner negotiates meaning from the provided information based on the mode of information presentation. In addition, the collected data provided insight into dynamic cognitive processing involved in schema construction and the transfer of such evolving schema to long-term memory.

The current study attempted to answer three research questions about the multimedia approach to teaching and learning.

1. *The modality effect*. To what extent is there an effect of information modality (i.e., text, pictures, video, sound) on rote learning and meaningful learning of science concepts, as measured by participants' responses to recall and transfer questions, respectively?
2. *The expertise reversal effect.* What effect do prior knowledge (i.e., advanced vs. novice learners) and the use of borrowed visuals have on rote and meaningful learning, as measured by participants' responses to factual recall and transfer questions, respectively?

3. *Cognitive load.* What effect do prior knowledge (i.e., advanced vs. novice learners) and the use of borrowed visuals have on cognitive-load, as measured by participants' responses to the cognitive-load questionnaire?

Summary of the Findings

Research question 1 examined the effects of the mode of information presentation (independent variables) on three sets of dependent variables: (a) knowledge recall (PTS1), (b) knowledge transfer (PTS2), and (c) cognitive-load (CLS2, 3, and 4). Knowledge recall was scored using a rote learning instrument that included a nonconceptual recall questionnaire (NRQ) and a conceptual recall questionnaire (CRQ). Prior knowledge (PrKn) and its associated cognitive load (PTS1) were also included for reference purposes. Prior knowledge was scored using the rote learning instrument. Meaningful learning was scored using the knowledge transfer questionnaire that required participants to apply the knowledge acquired during the learning activity to a new problem that was not addressed explicitly in the learning activity. Cognitive load was scored using an instrument that included the mental effort scale (MES), the extraneous load scale (ELS), and the germane load scale (GLS).

Results of the one-way analyses of variance (ANOVA) on each dependent variable were statistically significant PTS1 and its associated cognitive load (CLS2), and for the cognitive load associated with prior knowledge (CLS1). The result for CLS2, however, was questionable because the data failed Levene's test of equality of variances. There was no statistically significant effect for PTS2 and its associated cognitive-load (CLS3), or for the cognitive load associated with the learning activity (CLS4). The effect size was medium for all statistically significant variables. To further explore the source of between-group difference in knowledge recall, ANOVA was performed on the results of the posttest nonconceptual recall questionnaire (NRQ) and conceptual recall questionnaire (CRQ). Because the knowledge recall score (PTS1) is derived from the results of NRQ and CRQ, performing ANOVA on these two variables informed about how they each contributed to rote learning. The ANOVA results on these variables were statistically significant for the posttest NRQ with a small effect size, but not for the posttest CRQ, which had a small effect size.

Research question 2 examined the expertise reversal effect as a function of information modality (i.e., treatment; independent variable) and PTS1 and PTS2. Participants' prior knowledge scores were used for assignment of expertise level (i.e., low, medium, or high expertise) according to the procedure described in the methodology chapter (pp. 110-111). This question evaluated the assumption that increased visual support through the use of borrowed visuals would interact with prior knowledge inversely on learning. A two-way ANOVA was used to determine whether or not there were statistically significant between-group differences in the mean learning scores. Mode of information presentation (i.e., treatment level) and prior knowledge (i.e., expertise level) were entered as fixed factors (Figure 10, p. 108). The response variables were the rote and meaningful learning scores. This analysis allowed for testing whether information modality had an effect on PTS1 and PTS2 after accounting for prior knowledge (i.e., expertise level). The ANOVA results for the interaction of treatment and expertise levels were statistically significant for neither PTS1 nor posttest NRQ. These results, therefore, lack sufficient evidence to reject the null hypothesis that there were no statistically significant between-group differences of the interaction of prior knowledge with the mode of information presentation on learning.

Research question 3 examined the expertise reversal effect as a function of information modality (i.e., treatment; independent variable) and prior knowledge (i.e., expertise level) on the cognitive-load associated with the rote learning task (CLS2) and the meaningful learning task (CLS3). Participants' prior knowledge scores were used for assignment of expertise level (i.e., low, medium, or high expertise) according to the procedure described in the Methodology chapter (pp. 106-107). This question evaluates the assumption that increased support through the use of borrowed visuals interacts directly with the cognitive-load with sensitivity to prior knowledge (i.e., expertise level).

A two-way ANOVA was used to determine whether or not there were statistically significant between-group differences in the mean cognitive load scores. Mode of information presentation (i.e., treatment level) and prior knowledge (i.e., expertise level) were entered as fixed factors (Figure 11, p. 109). The response variables were the cognitive load scores associated with rote and meaningful learning, that is, CLS2 and CLS3, respectively. This analysis allowed for testing whether information modality had an effect on CLS2 and CLS3 after accounting for prior knowledge (i.e., expertise). The assumption of homogeneity of variance based on Levene’s test was met only for CLS2. The ANOVA results for the interaction of treatment and expertise levels, however, were not statistically significant for any of the responding variables, including CLS2. These results, therefore, lack sufficient evidence to reject the null hypothesis that there were no statistically significant between-group differences of the interaction of prior knowledge with the mode of information presentation on cognitive load.

Limitations

The current study takes place in a traditional classroom setting, a more naturalistic instructional and learning environment. However, several limitations are worth mentioning. This section provides an overview of these limitations related to sampling, reliability, and validity of the instruments used to measure learning and cognitive-load and the complexity of the learning material.

The participants in the study came from a convenience sample of students who are part of a high-achieving and majority Asian population that is not reflective of the broader population of high-school students in the United States, which calls to question whether the findings of the study are generalizable to the general population. Also, for research questions 2 and 3, the data from each treatment level were further segmented by expertise levels, resulting in group samples as small as six participants, a sample size far too small to yield results that can be reliably generalized to a broader population.

Initially, the current study was intended to serve as the first phase of a broader and more refined study. The intent was to evaluate the validity and reliability of the various instruments used to measure the dependent variables, particularly the cognitive-load instruments used for detecting extraneous and germane loads. With to closure of schools due to the Covid pandemic access to students was limited and the planned subsequent study is delayed. As a result, opportunities to refine the instruments and the overall study were minimal.

One reason for conducting a pilot study was to ensure that the learning material was sufficiently complex for the Central Limit Theorem to apply regarding the performance distribution of participants, mainly when correlating performance with prior knowledge. The lack of a pilot study, however, means that there was no opportunity to modify the complexity of the learning material accordingly.

In Mayer and Gallini's (1990) study, nonconceptual recall and verbatim retention, both being measures of rote learning, were segmented into two separate dependent variables. In the current study, however, the two were combined into a single instrument comprising 18 items, which could be a point of concern because a proper pilot study was not conducted to evaluate the validity of this instrument as a measure of rote learning. Also, because the same set of questions was used on both the pretest and posttest, there is concern that participants may have been influenced by the pretest questions to focus their attention more on elements from the learning material linked to these questions. If this situation occurred, it could have affected their performance on the posttest.

The study relied on participants to self-report their mental effort and their extraneous and germane load. In its original construct, the Paas (1992) scale may not be understood by students and could lead to misinterpretation and incorrect responses. Therefore, the mental effort scale was modified slightly using more appropriate terms for the age group. Additionally, the extraneous and germane load scales were developed as added measures of cognitive-load. Because the metal effort, extraneous load, and germane load all measure cognitive-load, these scales' reliability and internal consistency were statistically evaluated using Cronbach coefficient alpha, with $\boldsymbol{\alpha} \geq .65$ set as the minimum threshold.

The two extraneous load items, ELS1 and ELS2 measured the extent of learning distractions caused by external factors (e.g., distracting sounds, presence of others, etc.) and internal factors (e.g., self-generated stress caused by worrying about not understanding the learning material) and were analyzed for internal consistency and intercorrelation. The Cronbach coefficient alpha derived from the correlation of these two items was .54, which suggests that the two items fail to meet the minimum threshold for internal consistency. This result was somewhat expected. Although both items measure extraneous load, they pool data from different sources of extraneous loads and should not always correlate as indicated by the interitem correlation matrix, which yielded a coefficient of .37. Although this value does not indicate a strong correlation, it is above the reliability limit of .15. Additional items that included some reversely worded statements for measuring internally and externally derived sources of external loads would have strengthened this instrument.

The two germane load instruments, GLS1 and GLS2, measured the amount of thinking that was required to understand unfamiliar and familiar vocabulary respectively. Similar to the extraneous load items, these two items were designed to pool data from two different sources. Yet, the Cronbach coefficient alpha derived from the correlation of these two items was .89, which suggests that the two had a high level of internal consistency. This might indicate that the two items measured understanding in general without regard to vocabulary. The use of additional items that included some reversely worded statements for measuring understanding would have strengthened this instrument and provided clarification.

When all of the cognitive-load items were correlated together for internal consistency, a high Cronbach coefficient alpha of .82 was returned. Additionally, the interitem correlation matrix returned statistically significant correlation coefficients for all paired items with significance at the .01 level. As indicated above, however, the use of additional reversely worded items would improve the reliability of the instrument.

For those participants with low writing skills, their response on the conceptual recall and knowledge transfer items may suffer from an inability to articulate what they know clearly and can do. Because the instrument is intended to evaluate conceptual understanding and problem-solving instead of writing skills, any failure to communicate knowledge could artificially reduce the participants' scores.

Discussion of Findings

One focus of this investigation was whether there is an effect of information modality on learning and cognitive-load. It is reasonable to expect that the cognitive processes reflected by the cognitive-load and meaningful and rote learning scores should interact. A further reflection on the ANOVA analysis suggests that relevant cognitive processes may indeed interact. In both instances, the full animation group (Group 4) performed better than the other groups, which is consistent with what would be expected for novice learners who benefit from additional learning support in form of borrowed visuals (Mayer & Gallini, 1990). In Mayer and Gallini (1990), novice learners were found to benefit from borrowed visuals that help convey meaning coded in the accompanying text. Although the Mayer study used static visuals and text, the key point was that explanative illustrations provided additional learning support and, thus, resulted in statistically significant improvement in the performance of novice learners on measures of information retention. The full animation group in the current study experienced a higher level of learning support than was used in the Mayer and Gallini (1990) study. This group would therefore be expected to yield even better learning outcomes when compared to any of the other treatment groups. What is more, the CLS2 results indicate that the full animation group experienced the lowest cognitive-load while completing the recall questionnaire, which measured rote learning. This may explain why this group also exhibited the highest mean rote learning score. Consistent with what is already known about the architecture of the working-memory (see Modified from Moreno & Park, 2010), the lower cognitive-load suggests that the working-memory capacity could still accommodate additional information processing that was germane to the learning process. An additional investigation would be necessary to fully explore whether full animation could better facilitate rote learning.

With respect to the measure of meaningful learning (PTS2), the text+picture group presented the highest mean score, followed by the text-only and the text+video groups, and the full animation group had the lowest mean score. The observed between-group differences, however, were not statistically significant. Consequently, any attempt to interpret the meaningful learning results is highly speculative. It is worth noting that, while full animation seems to facilitate rote learning, it may be less effective than the other information modalities at promoting meaningful learning. In fact, it seems that providing less visual and audio support to learners may promote higher-level learning outcomes. This is evident by the fact that the two groups that received the most support (text+video and full animation), performed the worst. These results align well with what would be predicted by the expertise reversal effect (Kalyuga, 2014). It is possible that the low expertise participants did not contribute to the differences in meaningful learning (PTS2) that was observed across treatment groups since they may have lacked the foundational knowledge required to effectively respond to the PTS2 questionnaire. Whether or not this is the case can only be resolved with further investigation using a large enough sample size for segmenting by treatment level and expertise. However, the results appear to align with the predictions of the expertise reversal effect if the observed differences in PTS2 are attributed primarily to the high expertise participants. As such, the high expertise participants seem to be sensitive to information modality such that the higher the learning support (i.e., integration of pictures, animation, and sound), the worst they performed on meaningful learning tasks. Kühl's (2021) study similarly reported that prior knowledge interacted inversely with learning support (i.e., static visual → full animation) on learning. However, the current study suggests that information modality has a greater effect on meaningful learning than rote learning. This finding aligned well with previous research. According to Kalyuga (2014), more advanced learners require less instructional support in the form of visuals for which they already have pre-existing schemas. The animation groups provided additional visual support, while the text treatments provided none (Group 1) or reduced (Group 2) visual support. The participants' performance on meaningful learning tasks (PTS2) is sensitive to the mode of information presentation that aligns with the expertise reversal principle, suggesting that the main between-group differences are due to the more advanced learner. If this is the case, it could be that the instrument used in this study was not sufficiently sensitive at detecting the participants' expertise level. The results from the analyses from research questions 2 and 3 support this assumption. Both analyses yielded no statistically significant effect of the interaction of prior knowledge with the mode of information presentation.

CLS3 is the measure of the cognitive-load associated with meaningful learning as measured by the knowledge transfer questionnaire. The mean CLS3 scores from treatment groups 1 (text only), 2 (text+picture) and 3 (text+video) are in line with what would be expected if the expertise reversal effect (Kalyuga, 2014) assumption is correct. It may be that individuals who achieve more meaningful learning are also better at generating their own visuals and may become confused when presented with borrowed visuals. Accordingly, information that is beneficial for the novice learner can be redundant for the advanced learner. Sweller (2020) suggests that redundant information can confuse or distract the advanced learner, thus sequestering away precious cognitive resources. While this may be the case, the results of the ANOVA regression analysis on posttest scores (dependent variable) with prior knowledge (independent variable) did not yield a statistically significant effect. Additional investigation with a larger sample size may be needed to explore this possibility.

CLS2 and PST1 were the two most statistically significant results (Table 16, p. 111), suggesting that the modality effect may be more relevant for rote learning than meaningful learning. Such a trend is predicted by the expertise reversal principle (Kalyuga, 2014). The rote learner is still building foundational knowledge and may not have sufficient pre-existing schema to generate visuals independently from the text-only learning material. Consequently, the rote learners may lag in working memory ability to process relevant information and modify pre-existing schema appropriately under the text-only condition. It may be that the modality effect on rote learning is even more pronounced than was detected in this study. Perhaps, the learning material was not sufficiently complex to moderate a broad enough span in cognitive load between the various groups even under fixed cognitive-load conditions. The results of the analyses on the cognitive-load may have hinted at this possibility. It was noted that after the treatment, the mental effort associated with knowledge recall (i.e., ME1 of Phase 1 and ME1a of Phase 3) dropped by 44%. This suggests that, on average, participants exerted less cognitive-load after having engaged in relevant learning during the Phase 2 treatment. The mental effort associated with the learning activity was 51% lower than for the pretest, and 14% lower for the posttest. In other words, participants reported cognitive load that was highest for the pretest and lowest for the learning activity, suggesting that more mental effort was exerted on the pretest and posttest than during the learning activity. Assuming that some participants would have been unfamiliar with the concepts presented in the learning material, it was assumed that cognitive load would progressively decrease as the student progressed from phase 1 (pretest) to phase 2 (learning activity) and finally to phase 3 (posttest). The participants, however, reported greater cognitive demand on the posttest than on the learning activity. This might indicate that the complexity of the learning material was not sufficiently high for discriminating between participants based on their prior knowledge. While the participants who experienced the full animation reported lower cognitive load, their rote learning scores did not reflect a similar magnitude of improvement compared to the text-only group. Overall, this suggests that the participants perceived the learning material as less difficult than their results. But, if the relevant instrument used in this study failed to detect the participants' level of expertise, then the performances of the more advanced participants may have masked those of the others.

Conclusion

The results were most meaningful concerning rote learning and the associated cognitive-load. Rote learning is fundamental to meaningful learning in that it endows the learner with the basic set of relevant information needed to build complex schemas. Thus, rote learning precedes any meaningful learning that might occur (Mayer, 2002; Mayer, 2014a). As such, more advanced learners might already have progressed further along the continuum of rote knowledge and are better grounded in their understanding of basic concepts. The novice learner, however, may require additional support to acquire this basic knowledge. The results from the current study indicate that full animation resulted in the best overall performance on the rote learning posttest. This result supports the dual-channel subsystems (Baddley, 2000). The use of full animation with sound and visuals allows the learner to process relevant incoming information through two rather than one sensory channel (Mayer, 2014b). By doing so, the learner is not restricted to the capacity limit of either one of the channels and should therefore experience a lower cognitive-load. This assumption is supported by the fact that there was a statistically significant effect of information modality on cognitive-load, with the full animation group reporting the lowest cognitive-load on the rote learning posttest. Conversely, the subscripted animation group reported the highest cognitive-load. These conflicting results concerning the modality of the visual aid are somewhat perplexing and might be a potential area for further investigation. This observation might support Baddeley's (2013) proposal that textual information propagates through the working-memory. Accordingly, all textual information is initially received in the working-memory via the visual channel and is later converted to a verbal code and transferred to the audio channel for further processing and storage. It could be that when the learning material includes animation and associated subscripts, the visual channel becomes overtaxed, resulting in the higher cognitive-load that was observed. Future research could explore the combined interaction of animation and embedded subscripts on cognitive-load and potential effects on rote learning.

Implications for Research

It is important to emphasize that, even with observed trends, the small sample size of the current study presents significant limitations regarding the generalization of some of the findings. In addition, because sample size effects the observed power of statistical analyses, much of the statistical analyses that involved the interaction of multiple independent variables required that the sample size be further segmented. Nevertheless, with this in mind, the findings of this study may still present several implications for current and future research that are worth mentioning.

The results from the rote learning data also indicated that the text-only group reported the second-lowest cognitive-load. Perhaps, this effect was due to the performance of participants with more advanced prior knowledge. By splitting the groups according to expertise, it would be possible to determine whether this was the case. However, an attempt to conduct this analysis resulted in sample sizes that were too small to detect a statistically significant effect. Future investigations could explore this possibility with the use of larger sample sizes.

The study was inconclusive about how the expertise reversal effect interacted with the mode of information presentation on learning and cognitive-load. There may be two possible reasons for this issue, sample size, and the complexity of the concepts presented in the learning material. Concerning information complexity, to detect a difference in performance between the novice and advanced learner, there ought to be sufficiently complex concepts and information that challenge the individual differentially based on their prior knowledge. Without this spread comprehension ability, differences in performance may not be realized. For this reason, the original plan for this study was to involve a pilot study that would guide the selection of learning material that appropriately targeted different levels of expertise. Future investigations should involve methods to appropriately select learning material that differentially challenges learners based on their prior knowledge of the relevant concepts.

Implications for Practice

The findings from this study present several important implications for educational practice. First, this study adds to the growing body of research (e.g., Mayer & Pilegard, 2014), supporting the notion that integrating visuals with words in learning material can improve information retention and knowledge transfer. Second, although the findings in this study were inconclusive for meaningful learning, it supports this assumption regarding rote learning. The data analyses on rote learning suggest that the learner's expertise may influence the efficacy of the different combinations of information modalities in the learning material. The full animation proved most effective, and with reliably statistical significance for conceptual and nonconceptual recall, the findings were inconclusive for the other treatments. However, the analyses of the associated cognitive-loads indicate that the unimodal treatment, i.e., Text-only, was more effective than the bimodal treatment, Text+picture. Based on these findings, it appears that prior knowledge may influence the pattern of performance for each treatment. It may be that for rote learning, full animation works well for all learners, particularly those with low prior knowledge. Full animation can guide learners who lack sufficient conceptual understanding for generating relevant visuals independently. Conversely, the advanced learner might not need instructional support. Indeed, investigations on the expertise reversal effect show that novice and advanced learners respond differently to varying degrees of instructional support (Jiang et al., 2018; Kalyuga et al., 2003; Kalyuga & Sweller, 2014). The advanced learner who already has preexisting schemas to facilitate information processing may become confused when additional teacher support presents redundant information that contradicts elements already stored in their preexisting schema. For novice learners, however, additional instructional support can lead to improvements in learning outcomes. For example, Mayer and Gallini's (1990) study demonstrated that novice learners benefit from teacher-provided (borrowed) visuals that help convey meaning coded in the accompanying text.

The results of the rote learning data analyses suggest that the advanced learner was not adversely affected by additional instructional support, as would be assumed from the expertise reversal principle (Kalyuga & Chandler, 2003). Although an additional investigation is required to explore this possibility, it could mean that the power of the expertise reversal effect fades with the conceptual complexity of the learning content. In other words, when the complexity of the instructional material is low, there may be less need to consider the expertise reversal effect in the combination of information modalities. Instead, the teacher should pay special attention to the students with a low level of prior knowledge.

REFERENCES

Alesandrini, K. L. (1981). Pictorial–verbal and analytic– holistic learning strategies in science learning. *Journal of Educational Psychology*, *73*, 358–368.

American Association for the Advancement of Science. (1989). *Project 2061—Science for all Americans.* Washington, DC: AAAS.

Armougum, A., Gaston-Bellegarde, A., Joie-La Marle, C., & Piolino, P. (2020). Expertise reversal effect: Cost of generating new schemas. *Computers in Human Behavior*, *111, 106406*. https://doi.org/10.1016/j.chb.2020.106406

Arons, A. B. (1983). Achieving wider scientific literacy. *Daedalus*, *112(2)*, 91–122.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (pp. 89-195) Oxford, England: Academic.

Ayres, P. (2006). Impact of reducing intrinsic cognitive-load on learning in a mathematical domain. *Applied Cognitive Psychology*, *20*, 287-298.

Baddeley, A. (1992). Working-memory. *Science*, *255*, 556-559

Baddeley, A. D. (1999). Cognitive psychology: A modular course.Essentials of human memory. Psychology Press/Taylor & Francis (UK).

Baddeley, A. (2000). The episodic buffer: A new component of the working-memory? *Trends in Cognitive Sciences*, *4*, 417-423.

Baddeley, A. (2013). Essentials of Human Memory. Abingdon, Oxon: Psychology Press.

Baddeley, A.D., & Hitch, G.J. (1974). Working-memory. In G. A. Bower (Ed.), *The Psychology of learning and motivation*, (pp. 47–89). Academic Press.

Bahrick, H. P., Bahrick, P. O., & Wittinger, R. P. (1975). Fifty years of memory for names and faces: cross-sectional approach. *Journal of Experimental Psychology: General*, *104*, 54-75.

Borg, G. (1978). Subjective aspects of physical work. *Ergonomics*, *21*, 215-220.

Borg, G., Bratfisch, O., & Dornic, S. (1971). On the problem of perceived difficulty. *Scandinavian Journal of Psychology, 12*, 249-260.

Branscomb, A. W. (1981). Knowing how to know. Science, *Technology, & Human Values*, *6(36)*, 5–9.

Bratfisch, O., Borg, G., & Dornic, S. (1972). *Perceived item-difficulty in three tests of intellectual performance capacity (Rep. No. 29)*. Stockholm, Sweden: Institute of Applied Psychology.

Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 174-205). Cambridge University Press.

Carney, R. N., Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, *14*, 5–26.

Cartiff, B. M., Duke, R. F., & Greene, J. A. (2020). The effect of epistemic cognition interventions on academic achievement: A meta-analysis. *Journal of Educational Psychology.* Journal of Educational Psychology, *113(3)*, 477–498. https://doi.org/10.1037/edu0000490

Castro-Alonso, J. C., Ayres, P., Wong, M., & Paas, F. (2018). Learning symbols from permanent and transient visual presentations: Don't overplay the hand. *Computers & Education*, *116*, 1–13. https://doi.org/10.1016/j.compedu.2017.08.011

Cheng, L., & Beal, C. R. (2020). Effects of student-generated drawing and imagination on science text reading in a computer-based learning environment. *Educational*

Technology Research & Development, *68(1)*, 225–247. https://doi.org/10.1007/s11423-019-09684-1

Chestnutt, K., Jones, M. G., Hite, R., & Cayton, E. (2018). Next generation crosscutting themes: Factors that contribute to students' understandings of size and scale. *Journal of Research in Science Teaching*, *55*, 876-900.

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). Learning-styles and pedagogy in post-16 learning: A systematic and critical review. LSRC reference, Learning & Skills Research Centre, London, UK.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Erlbaum.

Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining procedures and concepts. *Journal of Experimental Psychology: Applied*, *7*, 68-82.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 215–229. https://doi.org/10.1038/nrn755

Dreyfus, H. L., & Dreyfus, S. E. (1986). Mind over machine: The power of human intuition and expertise in the era of the computer. New York, NY: The Free Press.

Dunbar, K. N., & Klahr, D. (2012). *Scientific thinking and reasoning*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0035

Dunn, R. (2003). The dunn and dunn learning-style model and its theoretical cornerstone. In: R. Dunn and S. Griggs (Eds.), Synthesis of the Dunn and Dunn learning-styles model research, (pp. 1-6). New York: St. John's University's Center for the Study of Learning and Teaching Styles.

Eitel, A., Bender, L., & Renkl, A. (2020). Effects of informed use: proposed extension of the self-management effect. In A. Tindall-Ford, S. Agostinho & J. Sweller (Eds.), *Advances in cognitive load theory* (pp. 16-29). New York, NY: Routledge.

Engels, M., Miller, B., Jennewein, J., & Eitel, K. (2019). The confluence approach: Developing scientific literacy through project-based learning and place-based education in the context of NGSS. *Electronic Journal of Science Education, 23*, 33–57.

Firat, M. (2017). How real and model visuals affect the test performance of elementary students. *Computers in Human Behavior*, *71*, 258-265.

Fleming, N.D., & Mills, C. (1992). Not another inventory, rather a catalyst for reflection. *To Improve the Academy*, *11*, 137-155.

Gabel, L. L. (1976). The development of a model to determine perceptions of scientific literacy. Unpublished doctoral thesis, The Ohio State University, Columbus, OH.

Gale, J., Koval, J., Ryan, M., Usselman, M., & Wind, S. (2018). Implementing NGSS Engineering disciplinary core ideas in middle school science classrooms: Results from the field. *Journal of Pre-College Engineering Education Research (J-PEER)*, *9(1)*, Article 2.

Geary, D. (2008). An evolutionarily informed education science. *Educational Psychologist*, *43*, 179-195.

Geary, D., & Berch, D. (2016). Evolution and children's cognitive and academic development. In D. Geary, & D. Berch (Eds.), *Evolutionary perspectives on child development and education* (pp.217-249). Switzerland: Springer.

Greene, J. A., & Yu, S. B. (2015). Educating Critical Thinkers: The Role of Epistemic Cognition. *Policy Insights from the Behavioral and Brain Sciences*, *3*, 45–53.

Guida, A., Gobet, F., Tardieu, H., & Nicolas, S. (2012). How chunks, long-term the working-memory, and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: a two-stage framework. *Brain and cognition, 79 (3)*, 221-244.

Hall, V. C., Bailey, J., & Tillman, C. (1997). Can student-generated illustrations be worth ten thousand words? *Journal of Educational Psychology*, *89*, 677–681.

Hanham, J., Leahy, W., & Sweller, J. (2017) Cognitive load theory, element interactivity, and the testing and reverse testing effects. *Applied Cognitive Psychology*, 31, 265-280.

Hegarty, M. (2014). Multimedia Learning and the development of mental models. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 576–597). Cambridge University Press.

Hilgard, E. R., Irvine, R. P., & Whipple, J. E. (1954). Rote memorization, understanding, and transfer: An extension of Katona's card-trick experiments. *Journal of Experimental Psychology*, *46(4)*, 288-292. https://doi.org/10.1037/h0062072

Huberty, C. J., & Petoskey, M. D. (2000). Multivariate analysis of variance and covariance. In H. Tinsley and S. Brown (Eds.) *Handbook of applied multivariate statistics and mathematical modeling*. New York: Academic Press.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, *17(6)*, 722–738.

Jenkins, E. W. (1994). Scientific literacy. In T. Husen & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (Volume 9, 2nd ed., pp. 5345–5350). Oxford, UK: Pergamon Press.

Jiang, D., Kalyuga, S., & Sweller, J. (2018). The curious case of improving foreign language listening skills by reading rather than listening: expertise reversal effect. *Educational Psychology Review*, *30*, 1-27.

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning (pp.* 576–597). Cambridge University Press.

Kalyuga, S. (2015). *Instructional guidance: A cognitive-load perspective*. IAP Information Age Publishing.

Kalyuga, S., Ayres, P., & Chandler, P. (2003). The expertise reversal effect. *Educational Psychology*, *38*, 23–31.

Kalyuga, S., & Singh, A. M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*(4), 831–852. https://doi.org/10.1007/s10648-015-9352-0

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 247-262).

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*, 315–341.

Katona, G. (1940). *Organizing and memorizing: Studies in the psychology of learning and teaching*. New York: Columbia University Press.

Katona, G. (1942). Organizing and memorizing: a reply to Dr. Melton. *American Journal of Psychology*, *55*, 273-275.

Kirschner, P. A. (2017). Stop propagating the learning-styles myth. *Computers & Education*, *106*, 166-171. https://doi.org/10.1016/j.compedu.2016.12.006

Kirschner, P., Sweller, J., & Clark, R. (2006) While minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential and inquiry-based teaching. *Educational Psychologist*, *41*, 75-86.

Kloser, M. (2013). Exploring high-school biology students; engagement with more and less epistemologically considerate texts. *Journal of Research in Science Teaching, 50(10)*, 1232–1257. *Computers in Human Behavior*, *27*, 29-35.

Kloser, M. (2016). Alternate text types and student outcomes: An experiment comparing traditional textbooks and more epistemologically considerate texts. *International Journal of Science Education*, *38*, 2477–2499.

Knoll, A. R., Otani, H., Skeel, R. L., & Van Horn, K. R. (2017). Learning style, judgements of learning, and learning of verbal and visual information. *British Journal of Psychology*, *108*, 544-563

Krajcik, J., Codere, S., Dahsah, C. et al. (2014). Planning Instruction to Meet the Intent of the Next Generation Science Standards. *Journal of Science Teacher Education, 25*, 157–175. https://doi.org/10.1007/s10972-014-9383-2

Kühl, T. (2021). Prerequisite knowledge and time of testing in learning with animations and static pictures: Evidence for the expertise reversal effect. *Learning and Instruction*, *73*, 101457. https://doi.org/10.1016/j.learninstruc.2021.101457

Kühl, T., Scheiter, K., Gerjets, P., & Edelmann, J. (2011). The influence of text modality on learning with static and dynamic visualizations. *Computers in Human Behavior*, *27(1)*, 29–35. https://doi.org/10.1016/j.chb.2010.05.008

Laugksch, R. (2000). Scientific literacy: a conceptual overview. *Science Education*, *84*, 71-94.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133(3)*, 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Leopold, C., & Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learning and Instruction*, *22*, 1, 16-26.

Leutner, D., Leopold, C., & Sumfleth, E. (2009). Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. *Computers in Human Behavior*, *25*, 284–289.

Leutner, D., & Schmeck, A. (2014). The generative drawing principle in multimedia learning In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 433-448).

Levin, J. R., & Mayer, R. E. (1993). Understanding illustrations in text. In B. K. Britton, A. Woodward, & M. R. Binkley (Eds.), *Learning from textbooks: Theory and practice* (pp. 95–113). Lawrence Erlbaum Associates, Inc.

Lin, F., & Chan, C. K. K. (2018) Promoting elementary students' epistemology of science through computer-supported knowledge-building discourse and epistemic reflection, *International Journal of Science Education, 40:6*, 668-687.

Lin, H, Dwyer, F. M. (2010). The effect of static and animated visualization: A perspective of instructional effectiveness and efficiency. *Educational Technology Research and Development*, *58*, 155-174.

Lopez, R. (2019, August 28). Did sons and daughters get the same education in ancient Greece? [Online] *National Geographic*.

https://www.nationalgeographic.com/history/magazine/2019/07-08/education-in-ancient-greece/

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 227-246).

Lowe, R. K., & Schnotz, W. (2014). Animation principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 513–546). New York: Cambridge University Press.

Martin, A. J., & Evans, P. (2020). Load reduction instruction. In A. Tindall-Ford, S. Agostinho & J. Sweller (Eds.), *Advances in cognitive load theory* (pp. 16-29). New York, NY: Routledge.

Mayer, R. E. (1999). *The promise of educational psychology*. Upper Saddle River, NJ: Prentice-Hall.

Mayer, R. E. (2002a). Multimedia learning. *Psychology of Learning and Motivation*, *41*, 85-139.

Mayer, R. E. (2002b). Rote versus meaningful learning. *Theory into Practice*, *41(4)*, 226–232

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge, England: Cambridge University Press.

Mayer, R. E. (2011). *Applying the science of learning*. Upper Saddle River, NJ: Pearson.

Mayer, R. E. (2014a). Introduction to multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1-24). Cambridge University Press.

Mayer, R. E. (2014b). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43-71). Cambridge University Press.

Mayer, R. E. (2014c). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 345–368). Cambridge University Press.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 279-315).

Mayer, R., & Gallini, J. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, *82*, 715–726.

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology*, *95*, 833-846. https://doi.org/10.1037/0022-0663.95.4.833

Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive-load and manage intrinsic cognitive-load during multimedia learning. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 131–152). New York, NY: Cambridge University Press.

Mayer, R. E., & Pilegard, C. (2014). Principles for managing essential processing in multimedia learning: segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 316-344).

Mayer, R. E., & Sims, V. K. (1994). For Whom Is a Picture Worth a Thousand Words?: Extensions of a Dual-Coding Theory of Multimedia Learning. *Journal of Educational Psychology*. *86*, 389-40.

Mayer, R.E., & Wittrock, M.C. (1996). Problem-solving transfer. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Macmillan.

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.

McNamara, D., Ozuru, Y., & Floyd, R. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education*, *4*, 229–257.

McTigue, E., & Slough, S. (2010). Student-accessible science texts: Elements of design. *Reading Psychology*, *31*, 213–227

Meneses, A, Escobar, J., & Véliz, S. (2018). The effects of multimodal texts on science reading comprehension in Chilean fifth-graders: text scaffolding and comprehension skills, *International Journal of Science Education*, *40*, 2226-2244

Miller, E., Russ, Manz, E., R., Stroupe, & D., Berland, L. (2018). Addressing the epistemic elephant in the room: Epistemic agency and the next generation science standards. *Journal for Research in Science Teaching*. *55(7)*, 1053-1075. https://doi.org/10.1002/tea.21459

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63(2)*, 81–97

Miller, J. D. (1983). Scientific literacy: A conceptual and empirical review. *Daedalus*, *112(2)*, 29–48

Moreno, R., & Park, B.. (2010). Cognitive load theory: Historical development and relation to other theories. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 9–28). Cambridge University Press.

National Research Council. (1999). *How people learn: bridging research and practice.* Washington, DC: The National Academies Press. https://doi.org/10.17226/9457.

National Research Council. (2012). *A framework for k-12 science education: practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press. https://doi.org/10.17226/1

National Research Council. (2013). *Next generation science standards: for states, by states*. Washington, DC: The National Academies Press. https://doi.org/10.17226/18290.3165.

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, *87*, 224–240.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skills in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*, 429–434.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*, 1–4.

Paas, F., & Sweller, J. (2014) Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1-24). Cambridge University Press.

Paas, F., & van Merriënboer, J.J.G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*, 122-133.

Pashler, P., McDaniel, M., Rohrer, D., & Bjork, R. (2009). Learning styles. *Psychological Science in the Public Interest*, *9*, 105-119.

Pavio, A. (1986). Mental Representations: *A dual coding approach*. New York, NY: Clarendon Press.

Pella, M. O. (1976). The place or function of science for a literate citizenry. *Science Education*, *60,* 97–101.

Petrucci, M., & Pecchinenda, A. (2017). The role of cognitive control mechanisms in selective attention towards emotional stimuli. *Cognition & Emotion*, *31(7)*, 1480–1492. https://doi.org/10.1080/02699931.2016.1233861

Ploetzner, R., Berney, S., & B´etrancourt, M. (2020). A review of learning demands in instructional animations: The educational effectiveness of animations unfolds if the features of change need to be learned. *Journal of Computer Assisted Learning*, *36*, 838–860. https://doi.org/10.1111/jcal.12476

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, *12(1)*, 61–86.

Prasada, S. (2000). Acquiring generic knowledge. *Trends in Cognitive Sciences*, 4, 66-71.

Rachmawati, E., Prodjosantoso, A. K., Wilujeng, I. (2019). Next generation science standard in science learning to improve student's practice skill. *International Journal of Instruction*, *12*, 299–310.

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, *29(4)*, 717–761. https://doi.org/10.1007/s10648-016-9365-3

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive-load perspective. *Educational Psychologist*, *38*, 15-22.

Rosman, T., Mayer, A.-K., Merk, S., & Kerwer, M. (2019). On the benefits of 'doing science': Does integrative writing about scientific controversies foster epistemic beliefs? *Contemporary Educational Psychology*, *58*, 85–101.

Sadler, T. D., & Brown, D. E. (2018). Introduction to the special issue: A critical examination of the Next Generation Science Standards. *Journal of Research in Science Teaching*, *55*, 903–906.

Sasin, E., & Fougnie, D. (2021). The road to long-term memory: Top-down attention is more effective than bottom-up attention for forming long-term memories. *Psychonomic Bulletin & Review, 28(3)*, 937–945. https://doi.org/10.3758/s13423-020-01856-y

Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in the working-memory. *Journal of Experimental Psychology: General*, *136(4)*, 663–684. https://doi.org/10.1037/0096-3445.136.4.663

Schmeck, A., Mayer, R. E, Opfermann, M., Pfeiffer, V., Leutner, D. (2014). Drawing pictures during learning from scientific text: testing the generative drawing effect and the prognostic drawing effect. *Contemporary Educational Psychology*, *39(4)*, 275-286.

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning (pp.* 72-103). Cambridge University Press.

Schüler, A., Scheiter, K., & van Genuchten, E. (2011). The role of working memory in multimedia instruction: Is working memory working during learning from text and pictures? *Educational Psychology Review*, *23*, 389-411.

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology*, *102(4)*, 872–879. https://doi.org/10.1037/a0019640

Shen, B. S. P. (1975). Scientific literacy and the public understanding of science. In S. B. Day (Eds.), *Communication of scientific information* (pp. 44–52). Basel: Karger.

Snow, R. E, Corno, L., & Jackson, D. (1996). Individual-differences in affective and conative functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243-310). New York, NY: Simon & Schuster Macmillan.

Spanjers, I. A. E., Wouters, P., van Gog, T., & van Merrienboer, J. J. G. (2011). An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behaviour*, *27*, 46-52.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74(11)*, 1–29.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257-285.

Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review*, *22*:123–138.

Sweller, J. (2015). In academe, what is learned, and how is it learned? *Current Directions in Psychological Science*, *24*, 190-194.

Sweller, J. (2020). Cognitive load theory. In A. Tindall-Ford, S. Agostinho & J. Sweller (Eds.), *Advances in cognitive load theory* (pp. 1-11). New York, NY: Routledge.

Sweller, J., & Sweller, S. (2006) Natural information processing systems. *Evolutionary Psychology*, *4*, 434-458.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive load theory, *Explorations in the Learning Sciences*. New York, NY: Springer.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–28.

Uccelli, P., Barr, C., Dobbs, C., Phillips Galloway, E., Meneses, A., & Sánchez, E. (2015). Core academic language skills: An expanded operational construct and a novel instrument to chart school-relevant language proficiency in preadolescent and adolescent learners. *Applied Psycholinguistics*, *36(5),* 1077–1109.

US Department of Health, Education, and Welfare. (1979). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Retrieved from https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont- report/index.html

Uttal, D. H., & O'Doherty, K. (2008). Comprehending and learning from “visualizations”: a developmental perspective. In J. Gilbert (Ed.), *Visualization: Theory and practice in science education* (pp. 53–72). Netherlands: Springer.

Van Cog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 263-278). Cambridge University Press.

Van Merriënboer, J. J. G., & Kester, L. (2014). The four-component instructional design model: Multimedia principles in environments for complex learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning (pp.* 104–148). Cambridge University Press.

van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology*, *69*, 129–140.

van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology*, *31*, 142–166.

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawings: Literature review and synthesis. *Educational Psychology Review*, *12*, 261–312.

Veenman, M. V. J., Prins, F. J., & Verheij, J. (2003). Learning-styles: Self-reports versus thinking-aloud measures. *British Journal of Educational Psychology*, *73*, 357-372.

Vieira. R. D., Florentino de Melo, V., Avraamidou, L., & Lobato, J. A. (2017). Reconceptualizing Scientific Literacy: The Role of Students' Epistemological Profiles. *Educational Science*, *7*, 47.

Wen, C.-T., Liu, C.-C., Chang, H.-Y., Chang, C.-J., Chang, M.-H., Chiang, S.-H., Yang, C.-W., & Hwang, F.-K. (2020). Students' guided inquiry with simulation and its relation to school science achievement and scientific literacy. *Computers & Education*, *149*, 103830. https://doi.org/10.1016/j.compedu.2020.103830

White, B. Y., & Frederiksen, J. R. (1987). Qualitative models and intelligent learning environments. In R. W. Lawler & M. Yazdani (Eds.), *Artificial intelligence and education, Vol. 1*. Learning environments and tutoring systems (pp. 281–305). Ablex Publishing.

Wittrock, M. C. (1989). Educational psychology and the future of research in learning, instruction, and teaching. In M. C. Wittrock & F. Farley (Eds.), *The future of educational psychology* (pp. 75–89). Lawrence Erlbaum Associates, Inc.

Wong, M., Castro-Alonso, J. C., Ayres, P., & Paas, F. (2020) The effects of transient information and element interactivity on learning from instructional animations. In A. Tindall-Ford, S. Agostinho & J. Sweller (Eds.), *Advances in cognitive load theory* (pp. 16–29). New York, NY: Routledge.

Zheng, R. Z., & Greenberg, K. (2018). The boundary of different approaches in cognitive-load measurement: strength and limitations. In R. Z. Zheng (Ed.), *Cognitive load measurements and applications* (pp. 45–56). New York, NY: Routledge.

APPENDIX A
INSTRUMENTS QUESTION ITEMS

Rote learning Instrument

Nonconceptual Recall Questionnaire

1. How long has evolution been occurring?
    a. Hundreds of years
    b. Thousands of years
    c. Millions of years
    d. Billions of years
2. What is a species?
    a. Individuals that live in the same ecosystem
    b. Individuals that can interbreed with one another and produce offspring that can also reproduce.
    c. A group of individuals that share a common ancestry.
    d. A population of individuals that compete for resources with another population of individuals.
3. Which of the following are the two key points needed to understand how one animal can develop into a whole new species or animal? *Check all that apply*
    uniqueness of living creatures
    catastrophic events in the environment
    selection processes
    climate change
4. Which structure in the cell is made of DNA?
    a. Proteins
    b. Chromosomes
    c. Nucleus
    d. mitochondria
5. Which factors ensure that each individual in the population is unique?
    a. excess reproduction by parents
    b. environmental factors such as food, climate, and predators
    c. Heredity
    d. interbreeding with other species
6. What is the source of uniqueness among individuals?

a. the nucleus
b. Ribosomes
c. resources availability
d. DNA

7. Which of the following factors always contribute to heredity?
   a. recombination and mutation
   b. copulation and maturation
   c. recombination and copulation
   d. all of the above
8. Which of the following statements is true?
   a. Mutation causes random changeS in the DNA.
   b. Recombination happens after the sperm combines with the egg.
   c. Copulation is necessary before genetic recombination can occur
   d. Gametes have double the DNA of a normal cell.
9. T/F Only the less fit creatures are subjected to the process of natural selection
   a. True
   b. False
10. Which of the following are good examples of selection factors? *Check all that apply*
    Predators
    Parasites
    Climate
    changes in the environment
11. T/F Individuals of the same species will share the same set of traits and characteristics.
    a. True
    b. False
12. T/F Individuals of the same species will share the same set of traits and characteristics.
    a. True
    b. False
13. Excluding humans, organisms tend to
    a. only produce as many offspring as the environment can support.
    b. make considerable effort to produce offspring that are as different

from each other as possible.
   c. each have sufficient resources to survive and reproduce because they live in equilibrium with their environment.
   d. All of the above statements are correct.

14. How did Darwin's finches get to the Galapagos Islands?
   a. They evolved from non-avian species that once lived on the island.
   b. They were likely blown there from South America by a storm.
   c. There is no good explanation as to how they got there.
   d. All of the above statements are viable answers.
15. What were the initial conditions on the Galapagos Island like for finches?
   a. They struggled as the local predators preyed on them.
   b. There was a shortage of food.
   c. It was a finches paradise.
   d. Their population grew very slowly because there were not enough female finches.
16. What was the main factor that led to the early speciation of finches on Galapagos Islands?
   a. Predation
   b. Drought
   c. Disease
   d. Competition
17. Speciation of the finches occurred on the basis of which of the following finches traits?
   a. the sound of their chirps.
   b. the sizes of their beaks.
   c. the color of their feathers.
   d. their flight pattern.
18. Which of the following statements about finch species is correct?
   a. Individual finches mate primarily with other finches that use the same niche.
   b. Worm digging finches preferred mating with seed cracking finches in order to diversify the resources of their offspring.
   c. Over the course of many generations behavioral characteristics were reduced, enabling the finches to exploit a variety of ecological niches successfully.
   d. The various species of finches differ in their behavior, not their physical appearance.
19. How many species of Darwin finches are there?

Conceptual Recall Questionnaire

1. Using no more than 50 words, explain how new species evolve from

preexisting

Meaningful learning Instrument

Knowledge Transfer Questionnaire

1. Based on what you learned, when Darwin traveled around the world, what would he have noticed about the sizes of populations of organisms?
    a. They tended to decrease over time.
    b. They tended to increase over time.
    c. They tended to remain steady.
    d. They tended to fluctuate.
2. Using what you learned, briefly explain your answer to the above question.

Cognitive load Instrument

Mental effort scale 1

1. How difficult did you find the questions on this page?
   Not difficult ---1-2-3-4-5-6-7 --- very difficult

Mental effort scale 2

1. How difficult was it for you to understand the content of the learning material?
   not difficult ---1-2-3-4-5-6-7 --- very difficult

Extraneous load scale

1. How much did you find yourself going back over the content of the learning material because you got distracted?
   very little ---1-2-3-4-5-6-7 --- very much

2. How much did you find yourself worrying about not understanding the learning material?
   very little ---1-2-3-4-5-6-7 --- very much

Germane load scale

1. Once you were engaged with the learning material, how much thinking did you have to do to make sense of **un**familiar vocabulary in the passage?
   very little ---1-2-3-4-5-6-7 --- very much
2. Once you were engaged with the learning material, how much thinking did you have to do to make sense of the familiar vocabulary in the passage?
   very little ---1-2-3-4-5-6-7 --- very much

APPENDIX B

LETTER OF SUPPORT (MATHEW MITCHELL, PH.D.)

UNIVERSITY OF SAN FRANCISCO
CHANGE THE WORLD FROM HERE

School of Education
Department of Learning and Instruction
2130 Fulton Street
San Francisco, CA 94117-1071
Tel 415.422.6289

August 23, 2018

I am writing a very strong letter of support for Theodore Johnson's proposed study with AP and 9th grade Biology students. Mr. Johnson is currently an active doctoral student in the Learning & Instruction Department at the University of San Francisco. He is an outstanding doctoral student and I have every confidence that he will complete an exemplary doctoral dissertation study.

I have had Ted in several of my courses so I'm already aware of the care and thoroughness with which he approaches projects. He has a specific interest in the potential of multimedia instruction as a way to enhance student learning. In one course last year he did a lot of background reading in this area: so he's well aware of the research needs for additional research to undress currently unanswered questions in the field.

If all goes well this study will serve as a pilot study for Mr. Johnson's final dissertation study. I think he's presented a very reasonable and well-planned agenda for designing, collecting, and analyzing his data.

Ted has my full support for this proposed project and I will help him in any way I can moving forwards in terms of analysis and interpretation of the results.

If you have additional questions please feel free to contact me.

Yours sincerely,

Mathew Mitchell, Ph.D.
Professor
Learning & Instruction Department
School of Education
University of San Francisco
(415) 422–2794
mitchellm@usfca.edu

APPENDIX C

USF IRB CLEARANCE/APPROVAL LETTER

From: [redacted] <noreply@axiommentor.com>

To: [redacted]

Subject: IRB Review Not Required - IRB ID: 1112

Date: Fri, 21 Sep 2018 19:24:56 +0000

UNIVERSITY OF SAN FRANCISCO

To:	Theodore Johnson
From:	[redacted], IRB Chair
Subject:	Protocol #1112
Date:	09/21/2018

The protocol **1112. Impact of multimedia education on learning** has been reviewed by the IRB chair and found not to require further IRB review or oversight as it is a standard educational improvement project.

Please note that changes to your protocol may affect its exempt status. Please contact our office to discuss any changes you may contemplate.

Sincerely,

[redacted]

Professor & Chair, Institutional Review Board for the Protection of Human Subjects
University of San Francisco
irbphs@usfca.edu
USF IRBPHS Website

APPENDIX D
SCHOOL DISTRICT IRB APPROVAL LETTER

487

[redacted]

[redacted]

December 12, 2018

Theodore Johnson
University of San Francisco
2350 Turk Street
San Francisco, CA 94118

Dear Ted:

Thank you for your request for [redacted] School District's permission to conduct your research on the ***Impact of Multimedia Instruction on Learning.***

Our office has reviewed your request and approved it for one year. You must submit an annual renewal request using the most current district research application template, at least 3 months prior to your intended start date. This approval is at a central District office level and requires the approval of the principal and other collaborating individuals at the school site(s) where you intend to conduct your research. District approval in no way obligates any school site, teacher, principal, student, or other individual to participate in your study. ***Please present a copy of this approval letter when you request data or invite sites or individuals to participate in the research. Be sure to communicate to them clearly that participation is always optional.***

You are also required to submit to the Research, Planning and Assessment Department a report of the results of your study when it is completed. In addition, in keeping with the District's commitment to professional development and to ensure that all research is actionable and useful, it is critical that you share your work with the district and school community that assisted you in the course of your study. Please submit interim updates, according to the schedule detailed in your application, so we know your study is on track.

Please note that this approval grants permission only to conduct the research, and not to share or publish the results beyond the [redacted] community. Any approval to share/publish results must be obtained after the results of the study have been reviewed by our office.

Good luck with your research. Feel free to contact us if you have any questions.

Sincerely,

[redacted]

Supervisor
Research Planning and Assessment Department

APPENDIX E

INFORMED CONSENT LETTER AND FORM

UNIVERSITY OF SAN FRANCISCO

CHANGE THE WORLD FROM HERE

March 27, 2019

Dear Parents and Guardians:

In addition to my work at [redacted] as your child's biology teacher, I am also a doctoral candidate in the School of Education at the University of San Francisco. The purpose of this letter is to explain why I am requesting that your child participate in my research study. During this semester, I will be conducting a study on multimedia learning and instruction. Specifically, my research focuses on developing instructional material that effectively integrates information from various sources, such as digital and printed text, voiced audio and other sound elements, and static visuals and animations. While multimedia learning and instruction have been long practiced in our classrooms, many teachers, myself included, typically rely on “trial and error" to determine best instructional design practices. With the expansion of technology integration in lesson design, various new multimedia materials have become easily accessible to teachers. However, these resources are often embraced without a reflection on learning efficacy. I want to step back and study what works in terms of visual designs, learning platform architecture, embedded elements, and student learning needs.

The purpose of my study is to understand better how prior knowledge and the structure of and types of information modalities used for instruction impact both rote learning (i.e., information retention) and meaningful learning (i.e., information application and transfer). The study will take place during our unit on evolution and will fit seamlessly with the curriculum to minimize its impact on instructional time. Students will be randomly assigned to one of four groups based on the type of learning material used: text only, text + static visual, text + dynamic visual, and audio + dynamic visual. Following the data collection, all students will have access to all of the learning materials.

There are 3 phases to this study. Phase 1 is a pre-assessment similar to a typical quiz intended to gauge your child's prior knowledge of the topic. Phase 2 is the learning phase, when students review and study the provided learning material. Finally, phase 3 is the post-assessment phase that will happen after the study phase. Following Phase 3, I may wish to interview your child to get additional feedback.

Your child is not required to participate in this study, and it will not impact their grade in the course. It is worth noting that the study's activities are typical of a biology classroom and involve concepts specified in the curriculum. However, because I will be examining the generated data through a researcher's lens rather than a teacher, I am legally required to request permission from both you and your child. There are no known risks involved in this study, and your child will not receive any compensation for his or her participation. Your child's name will not be linked to any record documents. If you have any questions, please contact me at [redacted] or via email at [redacted].

This letter serves as a consent form for your child's participation. It will be kept by both Mr. [redacted], principal at [redacted] ([redacted]), and by [redacted], faculty advisor at the University of San Francisco School of Education ([redacted]). If you have any questions about your child's rights as a participant, you may contact the University of San Francisco Institutional Review Board (IRB) for human subjects tests at IRBPHS@usfca.edu. Please have your child return the signed form (next page) to me via [redacted] by Monday, September 10th.

Thank you for your support,

[redacted]

Theodore Johnson

Form of Parent/Guardian Consent for Child to Participate in Research Study

I have read the attached consent letter that describes the project entitled *Impact of Multimedia Instruction on Learning*, conducted by Theodore Johnson, a doctoral student at the University of San Francisco and my child's biology teacher. The study's purpose, data collection method, type of data to be collected, and how the data will be used were explained. I am also aware that I can ask questions about this research and was provided the necessary contact information. I have also been informed that my child's participation in this study is not compulsory and does not impact their grade in the course.

Parent/guardian, please check one of the boxes below.

☐ I give my consent for my child to participate in this study.

☐ I <u>do not</u> give my consent for my child to participate in this study.

____

Name of child (Please print clearly)

______________________________________ ______________________________ __________

Name of parent/guardian (Please print clearly) **Signature of parent/guardian** *Date*

****

Form of Student Assent to Participate in Research Study

- I agree to participate in the study entitled *Impact of Multimedia Instruction on Learning*.
- I understand that my participation in the study will involve three 15 minutes assessments of my knowledge and skills.
- I understand that my participation in the study is strictly voluntary. Agreeing or not agreeing to participate in the study will not affect my school status, grades, or opportunities in any way.
- I understand that I may withdraw from the study at any time, even after I begin participating.
- I understand that my privacy will be protected in that my name will not be linked to any collected data.
- I understand that if I have any questions about this study or my participation in it, I can communicate directly with my teacher (Theodore Johnson), who is also the researcher.

Student, please check one of the boxes below.

☐ I assent (agree) to participate in this study.

☐ I <u>do not</u> assent (agree) to participate in this study.

______________________________ ______________________________ __________

Student's name (Please print clearly) **Signature of student** *Date*

APPENDIX F

LEARNING MATERIAL

Links to the learning material

Link:
http://tinyurl.com/vdabc

Link:
http://tinyurl.com/yy2fqlks

Link:
http://tinyurl.com/txabc

Link:
http://tinyurl.com/y59fsc66

Link:
http://tinyurl.com/vdabc

Link:
http://tinyurl.com/yy2fqlks

Link:
http://tinyurl.com/txabc

Link:
http://tinyurl.com/y59fsc66

Link:
http://tinyurl.com/vdabc

Link:
http://tinyurl.com/yy2fqlks

Link:
http://tinyurl.com/txabc

Link:
http://tinyurl.com/y59fsc66

Link:
http://tinyurl.com/vdabc

Link:
http://tinyurl.com/yy2fqlks

Link:
http://tinyurl.com/txabc

Link:
http://tinyurl.com/y59fsc66

Link:
http://tinyurl.com/vdabc

Link:
http://tinyurl.com/yy2fqlks

Link:
http://tinyurl.com/txabc

Link:
http://tinyurl.com/y59fsc66

Link:
http://tinyurl.com/vdabc

Link:
http://tinyurl.com/yy2fqlks

Link:
http://tinyurl.com/txabc

Link:
http://tinyurl.com/y59fsc66

Link:
http://tinyurl.com/vdabc

Link:
http://tinyurl.com/yy2fqlks

Link:
http://tinyurl.com/txabc

Link:
http://tinyurl.com/y59fsc66

Link:
http://tinyurl.com/vdabc

Link:
http://tinyurl.com/yy2fqlks

Link:
http://tinyurl.com/txabc

Link:
http://tinyurl.com/y59fsc66

Text only

MECHANISM OF EVOLUTION

What is evolution?

Evolution is the development of life on Earth. This is a process that began billions of years ago and is still continuing to this day. Evolution tells U. S. how it was possible for the enormous diversity of life to develop. It shows U. S. how primitive protozoa could become the millions of different species that we see today.

Evolution is the answer to the question that we have all asked on seeing a dachshund and a great dane together: how is it possible for ancestors to have descendants that look so very different from them. In answering this question, we want to focus on animals, excluding other forms of life such as fungi and plants. The first question to ask is, how can one animal develop into a whole new species of animal.

Ah! But just a quick question - what exactly is the **species**? A species is a community of animals that is capable of producing offspring with one another, with those offspring also being capable of reproducing.

In turn to understand this answer better we need to take a closer look at the following points: the **uniqueness** of living creatures guaranteed through the excess production of offspring and heredity, and as a second key point, **selection**.

Uniqueness

Let's begin with uniqueness. Every creature that exists is unique and this is essential
for evolution. The members of a species may strongly resemble each other in appearance. However, they all have slightly different traits and characteristics. They may be a bit bigger, fatter, stronger, or bolder than their fellow animals.

So, what is the reason for these differences? Let's take a closer look at a creature. Every creature is made up of **cells**. These cells have a **nucleus**. The nucleus contains the **chromosomes**, and the chromosomes hold the **DNA**. DNA consists of different **genes**, and it's these genes that are life's **information carriers**. They contain instructions and orders for the cells, and determine the characteristics and traits that living creatures have. It's precisely this DNA that is unique to every creature. The DNA is slightly different from individual to individual, which is why each has slightly different characteristics.

How is the enormous range of DNA created? One key factor is the **excess production** of offspring in nature. We can observe that creatures generally produce far more offspring than is necessary for the survival of their species, with many offspring dying an early death as a result. Often there are even more offspring than the environment in which they live is able to support. This is one factor that helps increase diversity within a species. The more offspring that are produced, the more little differences occur, and this is what nature wants: as many little differences as possible.

The second major cause of the uniqueness of individuals occurs in **heredity** itself. By the way, heredity means the passing on of DNA to offspring. Two very interesting factors come into play in this process: **genetic recombination** and **mutation**.

Recombination is the random mixing of the DNA of two creatures when two creatures fall in love and mate they recombine their genes twice. The first time they do this separately when they generate the **gametes**, that is sperm and egg cells. The gamete takes half of the genes and shuffles them. The second recombination occurs when a male inseminates a female. The parents each provide 50% of their DNA, in other words 50% of their unique traits and characteristics. These are then recombined or mixed and the result is new offspring. These offspring have a random mix of the DNA and therefore, the traits and characteristics of their parents. This increases the diversity and differences within a species even further.

But, **mutations** are also important for evolution. Mutations are random changes in DNA. These can also be described as copying errors within the DNA triggered by toxins, or other chemical substances, or by radiation. A mutation exists when a part of the DNA
is altered. These changes are often negative and may result in illnesses such as cancer. However they may also have neutral or positive effects such as the blue eye color in humans, which is one such random mutation. In all cases, a mutation has to affect a gamete, that is, a sperm or egg cell, because only the DNA in the gametes is passed on to the offspring. This is also the reason why we protect our sexual organs during x-rays, while other parts of the body are not at risk.

In summary then, in the heredity process, creatures pass on their characteristics to their offspring in the form of DNA. Recombination and mutation changes the

DNA so that each child looks different from its siblings and receives a random mix of the characteristics of its parents. There is a key word here: **random**. All of these processes are based on chance. Random recombination and mutations result in individuals with random mixes of traits and characteristics, which in turn mix these randomly and pass them on.

But how can so much be down to chance when all living creatures are so perfectly adapted to their environment. For example, the stick insect, the hummingbird, and the frog fish. The answer is provided by the second key point: selection. Each individual is subjected to a process of **natural selection**. As we have learned each individual is somewhat different from its fellows, and there is extensive variation within a species. Environmental influences have an effect on living creatures. These so-called selection factors include predators, parasites, animals of the same species, toxins, changes in habitat, or the climate. Selection is a process that each individual is subjected to. Every creature has a unique mix of traits and characteristics. This mix helps them to survive in their environment or not. Anyone with an unsuitable mix will be selected from the environment, while those with the right mix survive and can pass on their enhanced traits and characteristics. This is why diversity is so important, and why creatures make so much effort to produce offspring that are as different as possible. They increase the likelihood that at least one of their offspring passes nature's selection process. They maximize their chances of survival.

A good example of natural selection can be seen in a group of finches living on a remote island. They are some of the most famous animals in the world of science and are known as **Darwin finches** after their discoverer **Charles Darwin**. And this is the story of those finches:

A few hundred years ago, a small group of finches was blown onto the Galapagos Islands in the middle of the Pacific, probably by a big storm. The finches found themselves in an environment that was completely new to them - a real finch paradise with an abundance of food and no predators. They reproduced rapidly and numerously and the islands were soon heaving with finches.

This meant that food supplies became increasingly scarce and the finch paradise was soon threatened with famine, and finch friends became competitors. This is when selection intervened. Their individuality and small differences - in this case their slightly different beaks - meant that some of the birds were able to avoid competing with their fellow finches. The beaks of some of the finches were more suitable for digging for worms, while other finches were better able to use their beaks for cracking seeds. The finches consequently sought out ecological niches and these niches were safe from excessive competition. They soon began to mate primarily with other finches that used the same niche. Over the course of many generations these characteristics were enhanced, enabling the finches to exploit their niches successfully. The differences between the worm diggers and the seed crackers became so large that they were no longer able to mate with one another. Different species emerged as a result.

Today there are 14 different species of finches living on the Galapagos Islands, all of which are descended from the same group of stranded finches. This is how new species are created by evolution: through the interaction of unique individuals, the excess production of offspring, recombination and mutation in heredity, and finally through selection.

Why is this so important? It tells U. S. where the variety of life comes from and why living creatures are so perfectly adapted to their habitats. But it also affects U. S. personally. Every person is the result of three and a half billion years of evolution, and that includes you. Your ancestors fought and adapted in order to survive. This survival was an extremely uncertain thing. If we consider the fact that 99% of all the species that have ever lived are extinct, then you can consider yourself part of a success story. The dinosaurs have disappeared, but you are alive watching this video because you're incredibly special, just like all the other creatures that exist today that are irreproducible and unique in the universe.

The information used in this document were sourced from:

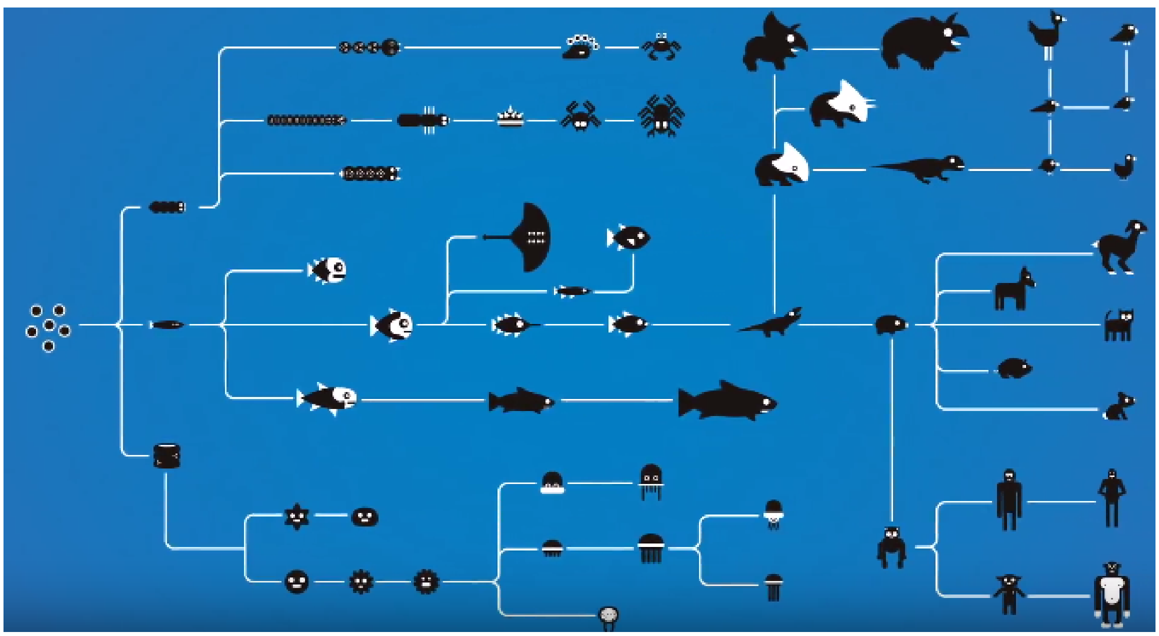

UNIVERSAL
UNIVERSAL PUBLISHING
PRODUCTION MUSIC
Thank you very much
for your support
Mechthild Reddanz, Markus Fenner,
Xuyen Dam and Cathrin Ziegler

Text + Static visual (picture)

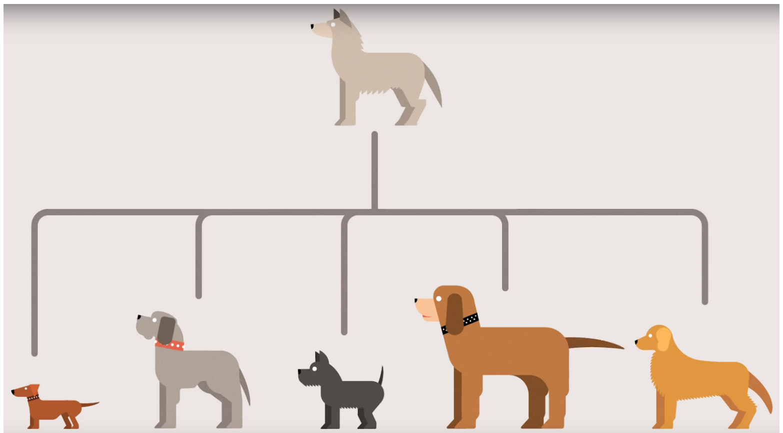
MECHANISM OF EVOLUTION

What is evolution?

Evolution is the development of life on Earth. This is a process that began billions of years ago and is still continuing to this day. Evolution tells U. S. how it was possible for the enormous diversity of life to develop. It shows U. S. how primitive protozoa could become the millions of different species that we see today.



Evolution is the answer to the question that we have all asked on seeing a dachshund and a great dane together: how is it possible for ancestors to have descendants that look so very different from them.

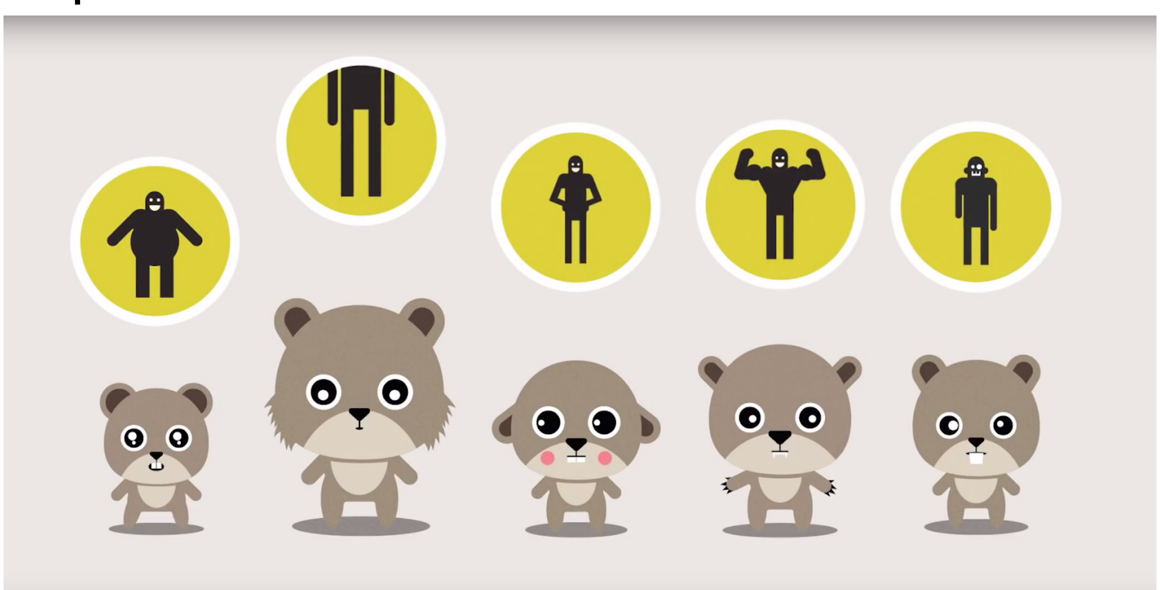


In answering this question, we want to focus on animals, excluding other forms of life such as fungi and plants. The first question to ask is, how can one animal develop into a whole new species of animal.

Ah! But just a quick question - what exactly is the **species**? A species is a community of animals that is capable of producing offspring with one another, with those offspring also being capable of reproducing.

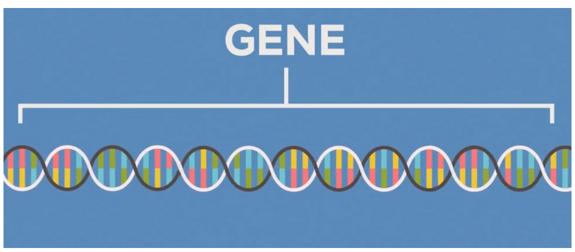
In turn to understand this answer better we need to take a closer look at the following points: the **uniqueness** of living creatures guaranteed through the excess production of offspring and heredity, and as a second key point, **selection**.

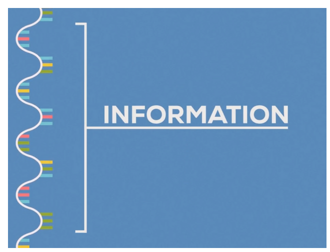
Uniqueness

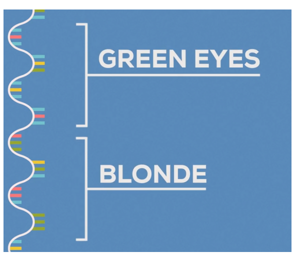


Let's begin with uniqueness. Every creature that exists is unique and this is essential
for evolution. The members of a species may strongly resemble each other in appearance. However, they all have slightly different traits and characteristics. They may be a bit bigger, fatter, stronger, or bolder than their fellow animals.

So, what is the reason for these differences? Let's take a closer look at a creature. Every creature is made up of **cells**. These cells have a **nucleus**. The nucleus contains the **chromosomes**, and the chromosomes hold the **DNA**. DNA consists of different **genes**, and it's these genes that are life's **information carriers**. They contain instructions and orders for the cells, and determine the characteristics and traits that living creatures have. It's precisely this DNA that is unique to every creature. The DNA is slightly different from individual to individual, which is why each has slightly different characteristics.



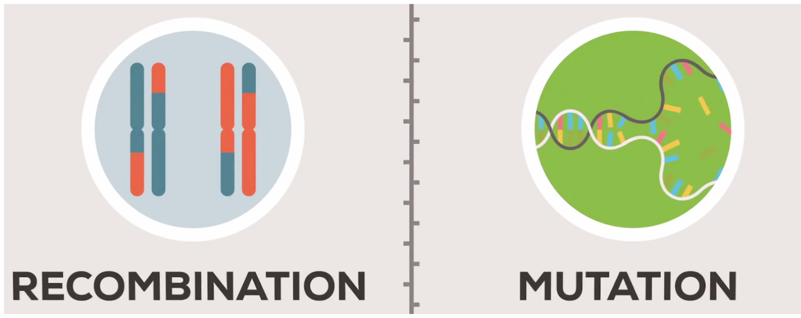




How is the enormous range of DNA created? One key factor is the **excess production** of offspring in nature. We can observe that creatures generally produce far more offspring than is necessary for the survival of their species, with many offspring dying an early death as a result.

Often there are even more offspring than the environment in which they live is able to support. This is one factor that helps increase diversity within a species. The more offspring that are produced, the more little differences occur, and this is what nature wants: as many little differences as possible.

The second major cause of the uniqueness of individuals occurs in **heredity** itself. By the way, heredity means the passing on of DNA to offspring. Two very interesting factors come into play in this process: **genetic recombination** and **mutation**.

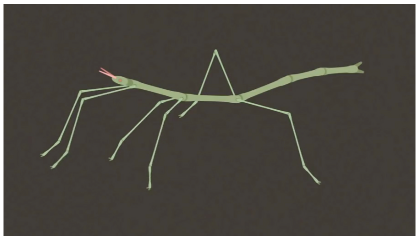
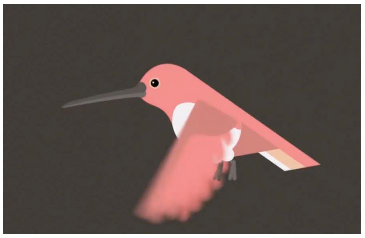
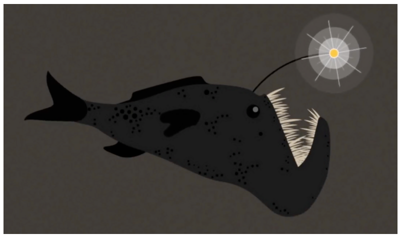


Recombination is the random mixing of the DNA of two creatures when two creatures fall in love and mate they recombine their genes twice. The first time they do this separately when they generate the **gametes**, that is sperm and egg cells. The gamete takes half of the genes and shuffles them. The second recombination occurs when a male inseminates a female. The parents each provide 50% of their DNA, in other words 50% of their unique traits and characteristics. These are then recombined or mixed and the result is new offspring. These offspring have a random mix of the DNA and therefore, the traits and characteristics of their parents. This increases the diversity and differences within a species even further.

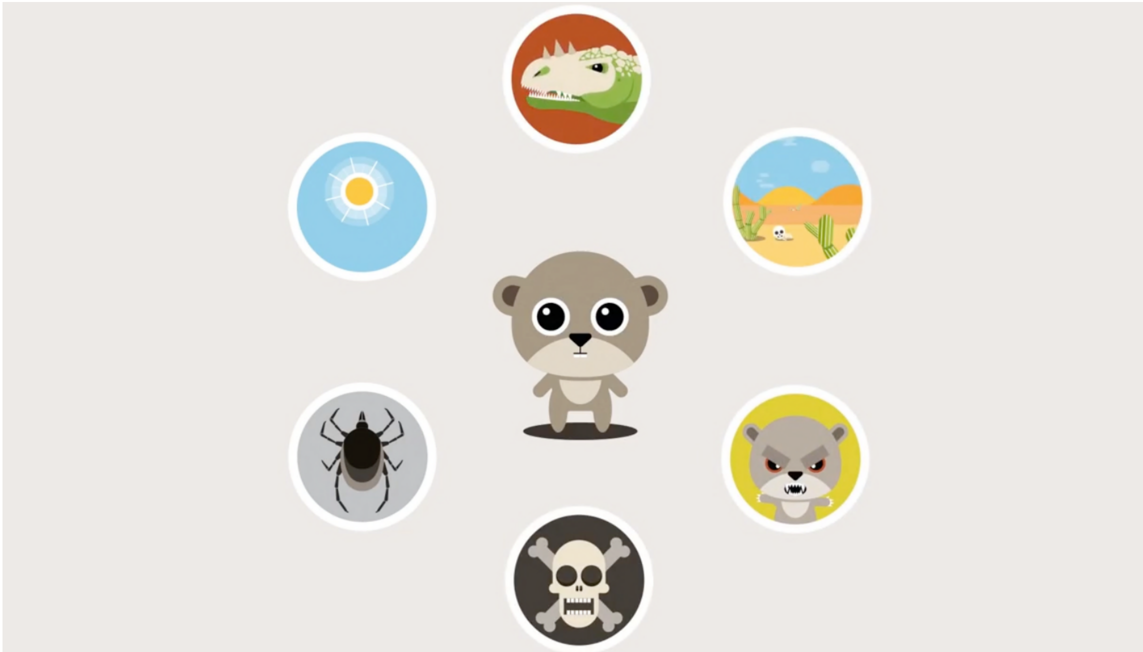
But, **mutations** are also important for evolution. Mutations are random changes in DNA. These can also be described as copying errors within the DNA triggered by toxins, or other chemical substances, or by radiation. A mutation exists when a part of the DNA
is altered. These changes are often negative and may result in illnesses such as cancer. However they may also have neutral or positive effects such as the blue eye color in humans, which is one such random mutation. In all cases, a mutation has to affect a gamete, that is, a sperm or egg cell, because only the DNA in the gametes is passed on to the offspring. This is also the reason why we protect our sexual organs during x-rays, while other parts of the body are not at risk.

In summary then, in the heredity process, creatures pass on their characteristics to their offspring in the form of DNA. Recombination and mutation changes the DNA so that each child looks different from its siblings and receives a random mix of the characteristics of its parents. There is a key word here: **random**. All of these processes are based on chance. Random recombination and mutations result in individuals with random mixes of traits and characteristics, which in turn mix these randomly and pass them on.

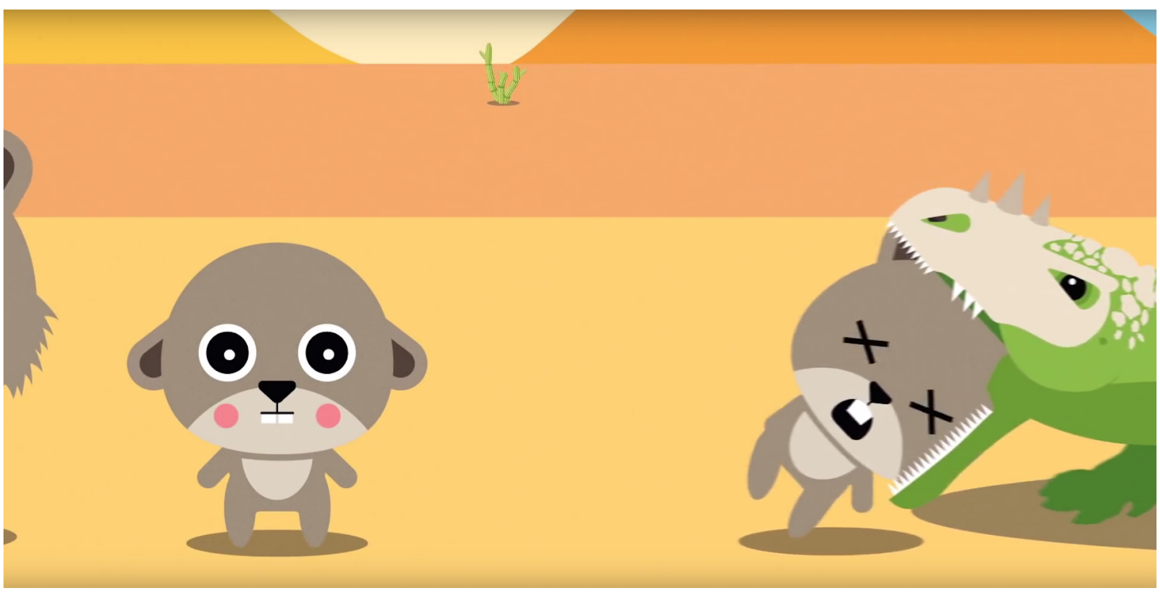
But how can so much be down to chance when all living creatures are so perfectly adapted to their environment. For example, the stick insect, the hummingbird, and the frog fish.

The answer is provided by the second key point: **selection**. Each individual is subjected to a process of **natural selection**. As we have learned each individual is somewhat different from its fellows, and there is extensive variation within a species. Environmental influences have an effect on living creatures. These so-called selection factors include predators, parasites, animals of the same species, toxins, changes in habitat, or the climate.



Selection is a process that each individual is subjected to. Every creature has a unique mix of traits and characteristics. This mix helps them to survive in their environment or not. Anyone with an unsuitable mix will be selected from the environment, while those with the right mix survive and can pass on their enhanced traits and characteristics.

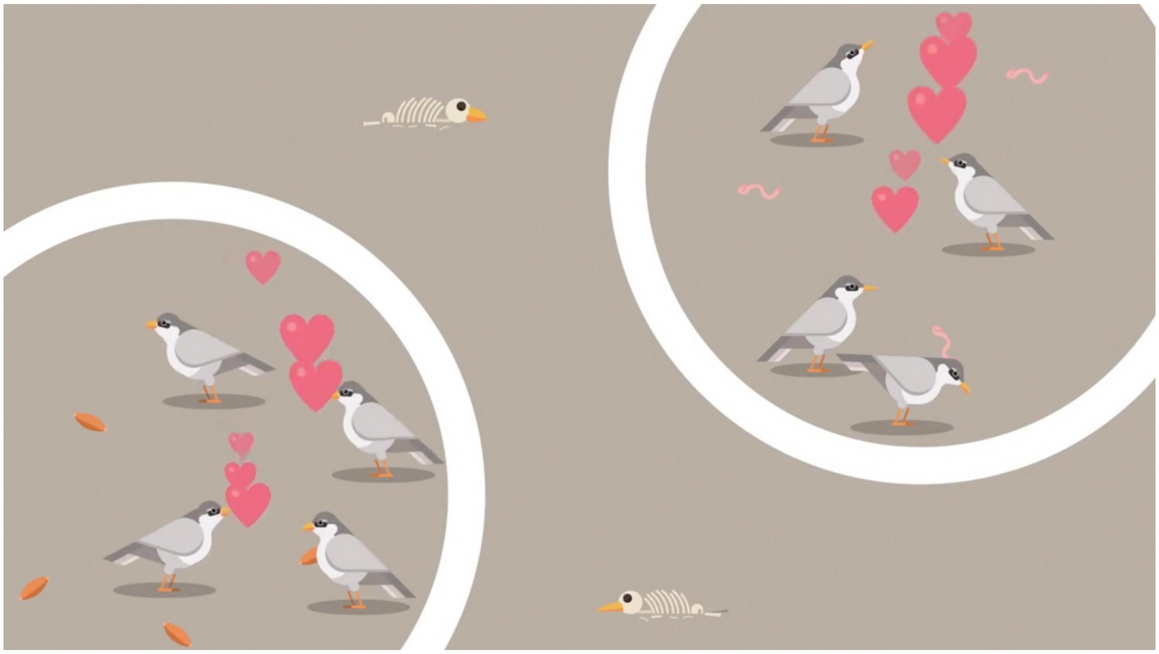


This is why diversity is so important, and why creatures make so much effort to produce offspring that are as different as possible. They increase the likelihood that at least one of their offspring passes nature's selection process. They maximize their chances of survival.

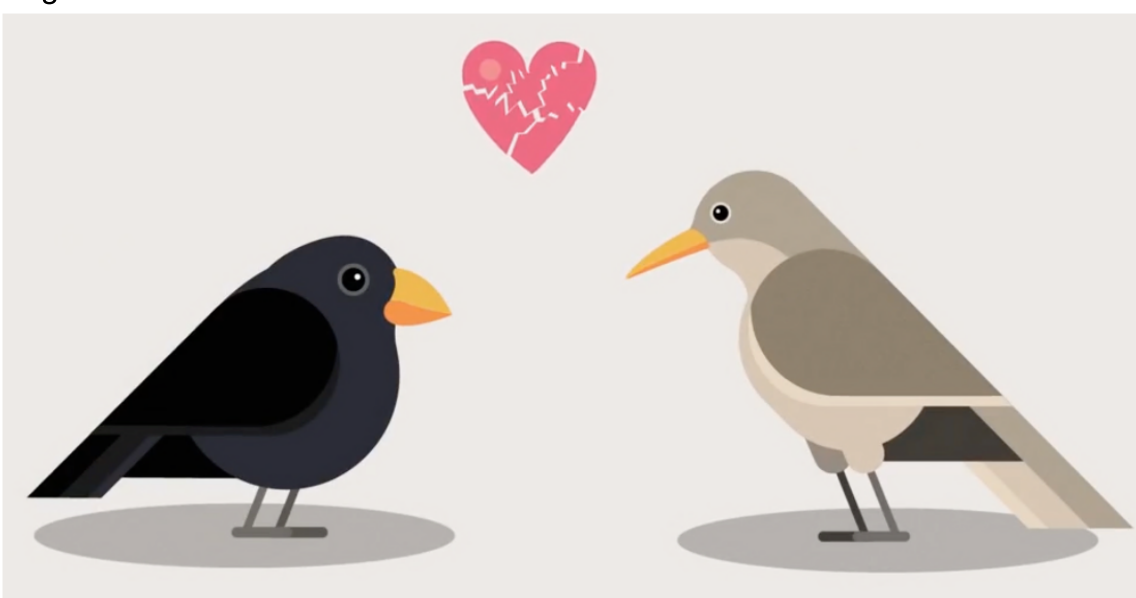
A good example of natural selection can be seen in a group of finches living on a remote island. They are some of the most famous animals in the world of science and are known as **Darwin finches** after their discoverer **Charles Darwin**. And this is the story of those finches:

A few hundred years ago, a small group of finches was blown onto the Galapagos Islands in the middle of the Pacific, probably by a big storm. The finches found themselves in an environment that was completely new to them - a real finch paradise with an abundance of food and no predators.

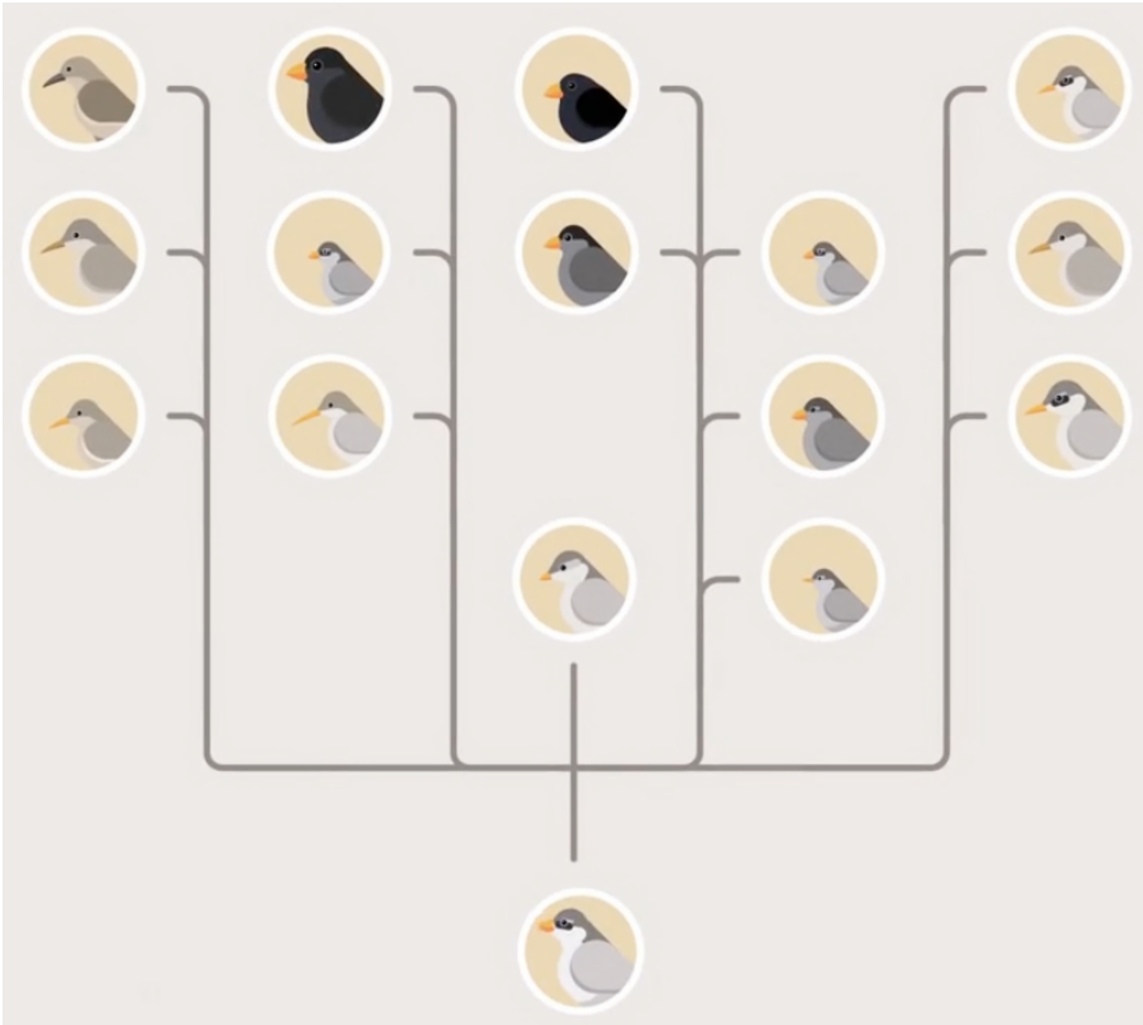
They reproduced rapidly and numerously and the islands were soon heaving with finches. This meant that food supplies became increasingly scarce and the finch paradise was soon threatened with famine, and finch friends became competitors. This is when selection intervened. Their individuality and small differences - in this case their slightly different beaks - meant that some of the birds were able to avoid competing with their fellow finches. The beaks of some of the finches were more suitable for digging for worms, while other finches were better able to use their beaks for cracking seeds. The finches consequently sought out ecological niches and these niches were safe from excessive competition. They soon began to mate primarily with other finches that used the same niche.



Over the course of many generations these characteristics were enhanced, enabling the finches to exploit their niches successfully. The differences between the worm diggers and the seed crackers became so large that they were no longer able to mate with one another.



Different species emerged as a result. Today there are 14 different species of finches living on the Galapagos Islands, all of which are descended from the same group of stranded finches.



This is how new species are created by evolution: through the interaction of unique individuals, the excess production of offspring, recombination and mutation in heredity, and finally through selection.

Why is this so important? It tells U. S. where the variety of life comes from and why living creatures are so perfectly adapted to their habitats. But it also affects U. S. personally. Every person is the result of three and a half billion years of evolution, and that includes you. Your ancestors fought and adapted in order to survive. This survival was an extremely uncertain thing. If we consider the fact that 99% of all the species that have ever lived are extinct, then you can consider yourself part of a success story. The dinosaurs have disappeared, but you are alive watching this video because you're incredibly special, just like all the other creatures that exist today irreproducible and unique in the universe.

The information used in this document were sourced from:

Text (subscript) + animation (Text + Video)

The subscripted video contains the same text as the Text Only material as a subscript that is synchronized with the animation. The static images (pictures) used in the Text + picture material were screenshot from the video and embedded in the text for that treatment. The video is accessible using the following link: http://tinyurl.com/yy2fqlks

Full animation (Audio + Video)

The video contains the information as the Text Only material, but instead, presented aurally. The static images (pictures) used in the Text + picture material were screenshot from the video and embedded in the text for that treatment. The video is accessible using the following link: http://tinyurl.com/vdabc

APPENDIX G

SAMPLE SCORING OF KTQ

The following table provides three sample responses to the knowledge transfer questionnaire. The responses were scored using the rubrics provided in Table 3.

Based on what you learned when Darwin traveled around the world, what would he have noticed about the sizes of populations of organisms?					
Sample	Item 1 responses	Item 1 score	Item 2 responses	Item 2 score	Total Score raw → adjusted
1	They tended to remain steady.	2	Darwin would have noticed that the population sizes of organisms would tend to remain steady over time. Looking at the finches as an example, when they first came to the Galapagos Islands they had a large population size due to excess amounts of resources. However, when the high population led to a famine, the population dropped. After the population dropped, only those with suitable traits for the environment survived, and then the population of the finches was in equilibrium with the resources in the environment and the environment itself. So while the population may fluctuate, or increase and decrease initially, Darwin would have noticed that the general trend of population sizes, particularly those in balance with their environment, would tend to remain steady.	2	4 → 10
2	They tended to fluctuate.	1	The sizes of populations depend on environmental factors and how much competition there is. If Darwin traveled around the world, different places would have different conditions, causing the sizes of populations to be different too.	1	2 → 5
3	They tended to increase over time.	0	They tend to increase over time because organisms tend to repopulate over time which leads to an increased population if all the organisms repopulate. This also leads to diversity and eventually different species that look very different but came from a common ancestor.	1	1 → 2.5

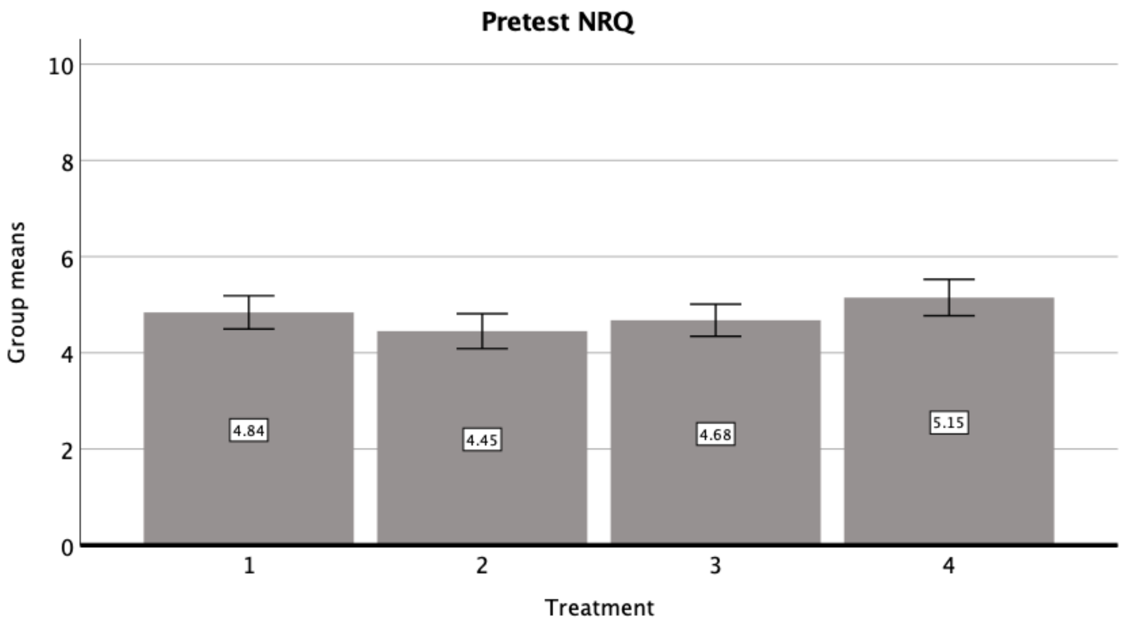
Sample 1 received the maximum score of 2 points each for the two responses, resulting in the highest possible total raw score of 4. On Item 1, by inferring that Darwin would have observed that populations tended to remain steady, the response earned 2 points. To earn an additional 2 points, the response to Item 2 should articulate a causal relationship that reflects the effects of (a) over-reproduction, (b) competition for limited resources, and (c) natural selection on the maintenance of a steady population. Sample 1 touched on each of these casual relationships. Although this sample did not explicitly include the term over-reproduction, it did include the effect of over-reproduction, i.e., large population size. Based on the explanation provided, it is reasonable to conclude that the participant made relevant inferences. Additionally, Sample 1 reflected on the other two causal relationships, i.e., competition for limited resources and natural selection, when describing survivorship under conditions of reduced resources.

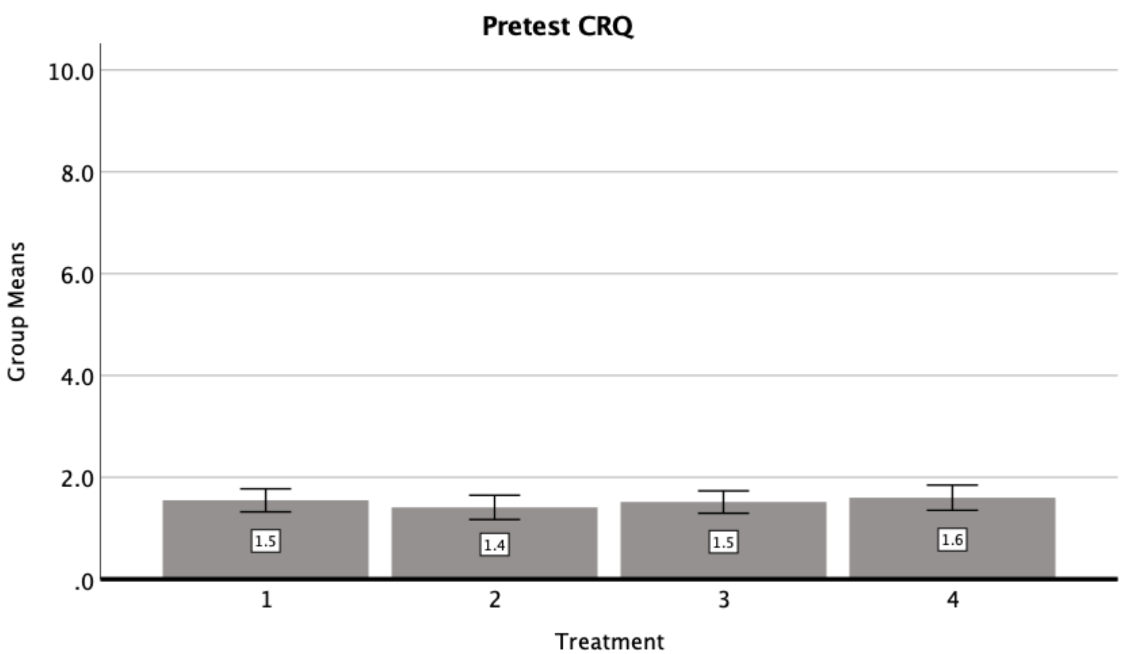
Sample 2 received a score of 1 point each for the two responses, resulting in a total raw score of 2 points. On Item 1, by inferring that Darwin would have observed that populations tended to fluctuate, the response earned 1 point. While populations can fluctuate, most populations at any given time would remain steady. The response in Sample 1 emphasized this by stating, “while the population may fluctuate, or increase and decrease initially, Darwin would have noticed that the general trend of population sizes, particularly those in balance with their environment, would tend to remain steady." However, the response in Sample 2 does not mention the general steady nature of population sizes. To earn an additional 2 points, the response to Item 2 needed to articulate a causal relationship that explains that (a) environmental changes can impact (b) resources availability, which (c) result in population fluctuating. The response to Item 2 received 1 point because it included a reference to competition and to the environment as factors that affect population sizes, causing fluctuations. However, the response simply mentioned these factors as contributing to population fluctuation without fully explaining the connection.

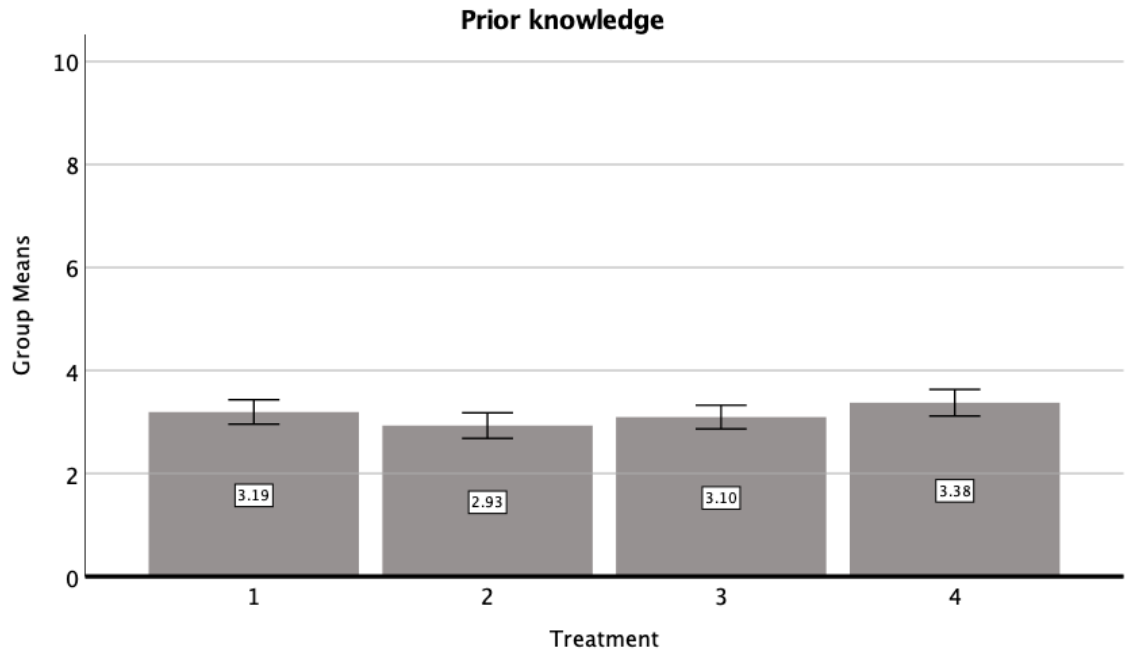
Sample 3 received a score of 0 and 1 points for Items 1 and 2, respectively, resulting in a total raw score of 1 point. On Item 1, by inferring that Darwin would have observed that populations tended to increase, the response earned 0 points. Assuming that one would observe populations tending to increase is irrational since such an event would require unlimited resources. Likewise, suggesting a trend of population decrease would result in all species immediately going extinct. Despite earning no points for the Item 1 response, the learning material did mention how the Darwinian finches population rapidly increased once they arrived on the Galapagos Islands. If the response provided a causal relationship that reflected relevant concepts from the learning material, the response could earn two points. To earn the additional two points, the Item 2 response must mention that (a) migration can present access to new resources, or (b) new adaptations can enhance competitive advantage, and (c) that either can lead to a population increase. The response to Item 2 does not reflect (a) nor (b); however, it does mention that “repopulate over time, which leads to an increased population." Therefore, the response was assigned 1 point because it provides a marginal causal relationship for population increase.

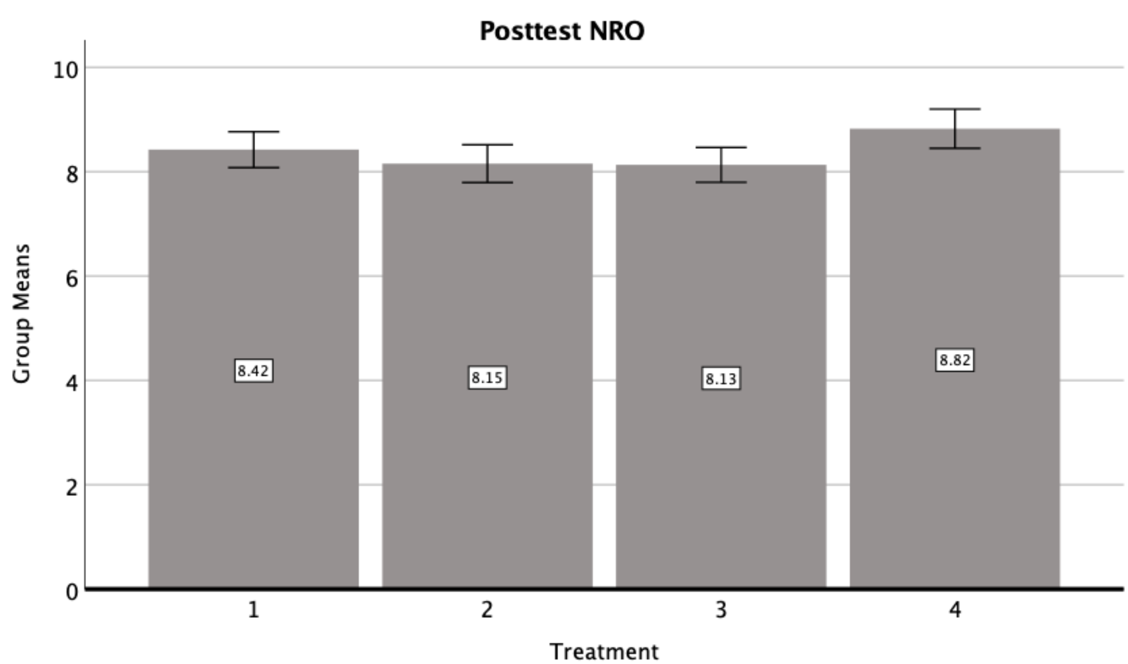
APPENDIX H

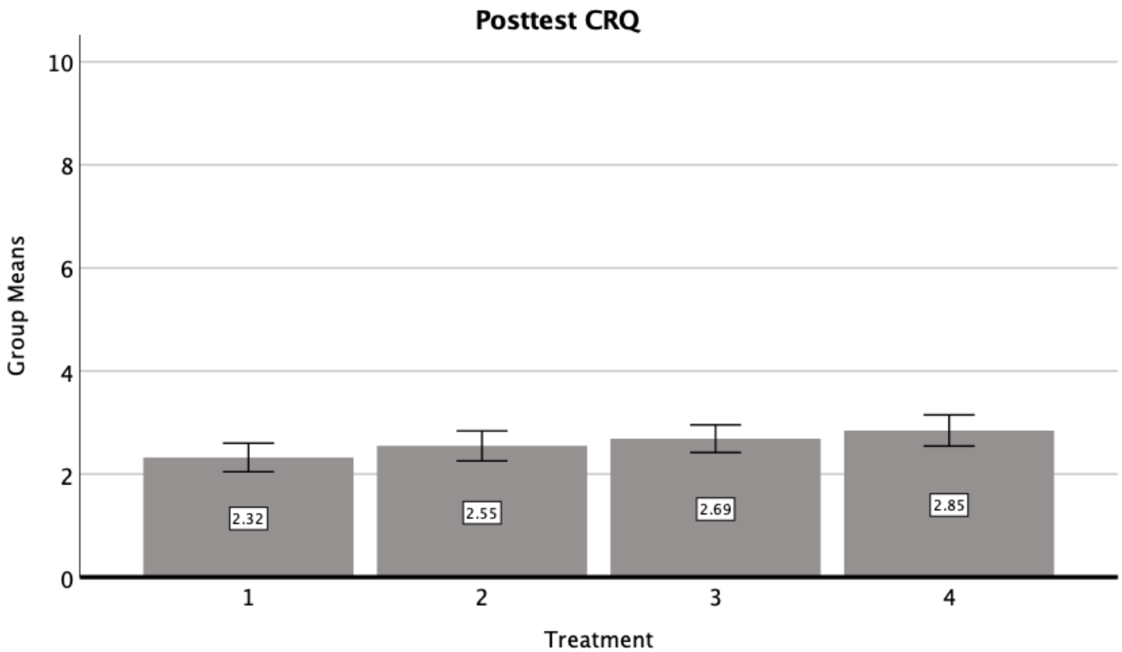
GRAPHS OF ANOVA RESULTS (QUESTION 1)

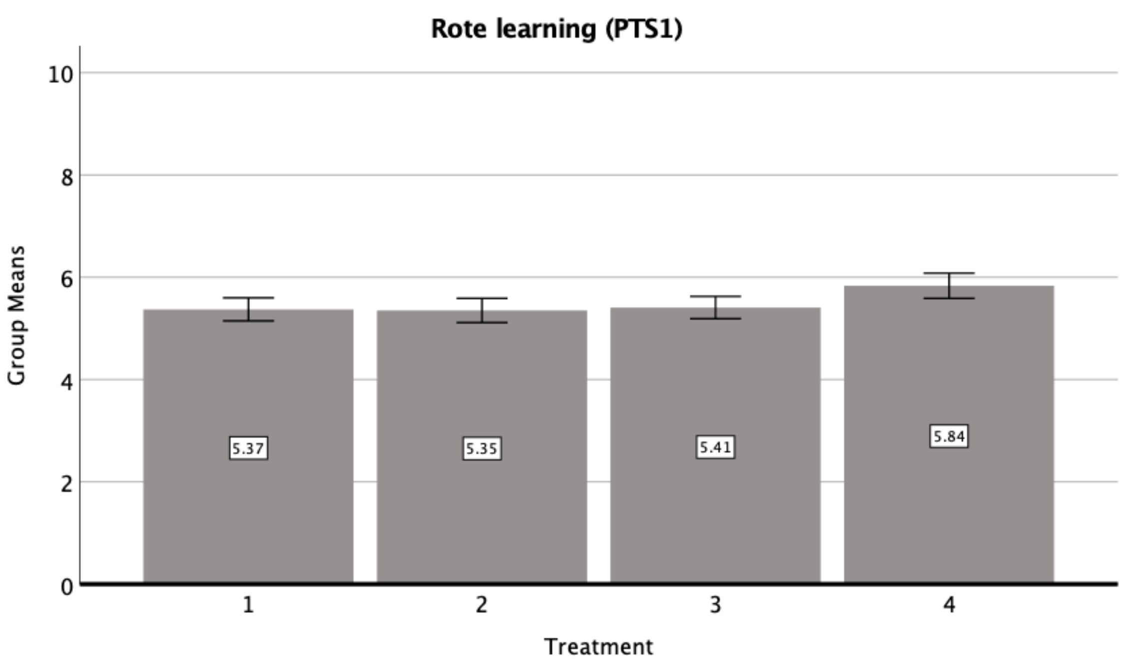

Pretest NRQ
Group means
10
8
6
4
2
0
4.84
4.45
4.68
5.15
1
2
3
4
Treatment

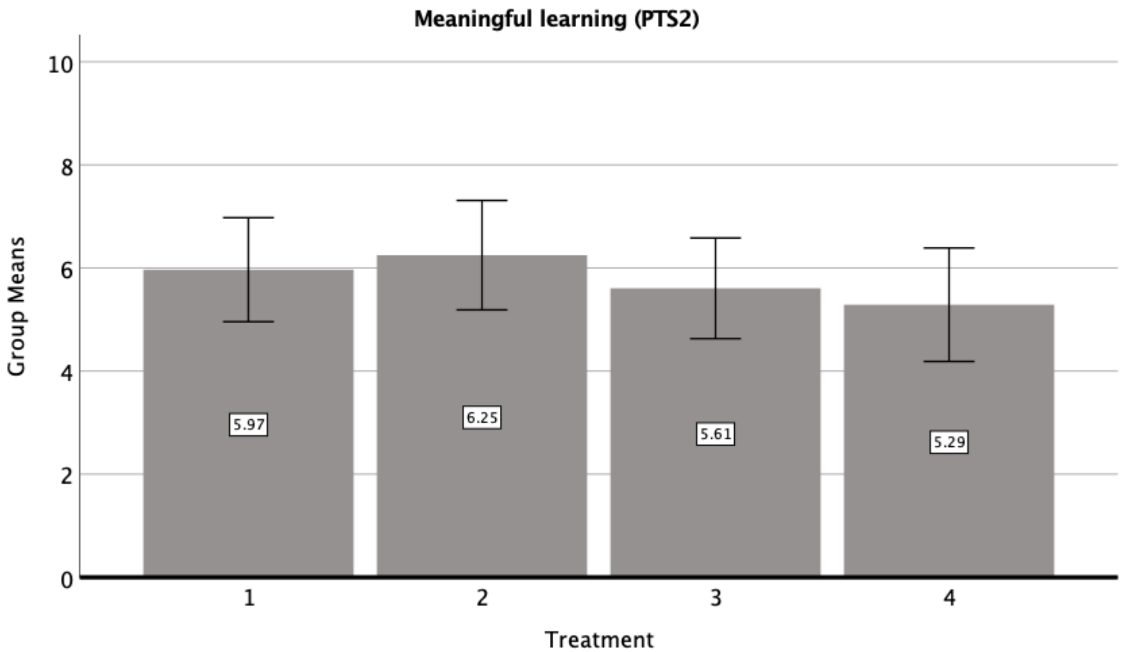

Pretest CRQ
Group Means
10.0
8.0
6.0
4.0
2.0
.0
1.5
1.4
1.5
1.6
1
2
3
4
Treatment

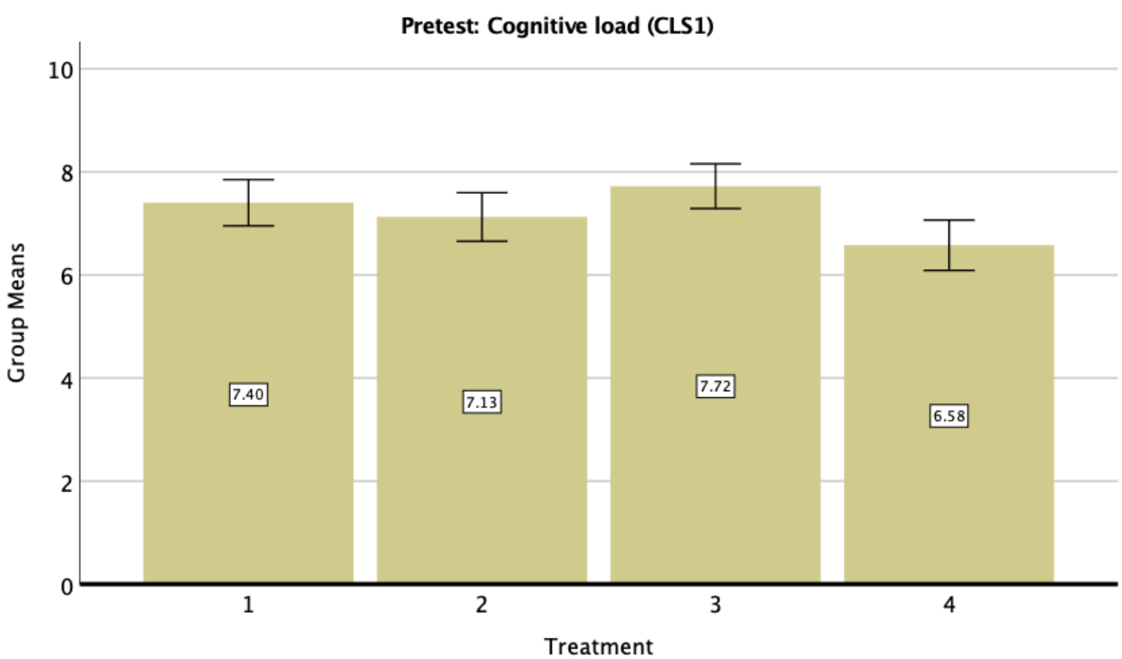

Prior knowledge
10
8
6
4
2
0
Group Means
3.19
2.93
3.10
3.38
1
2
3
4
Treatment

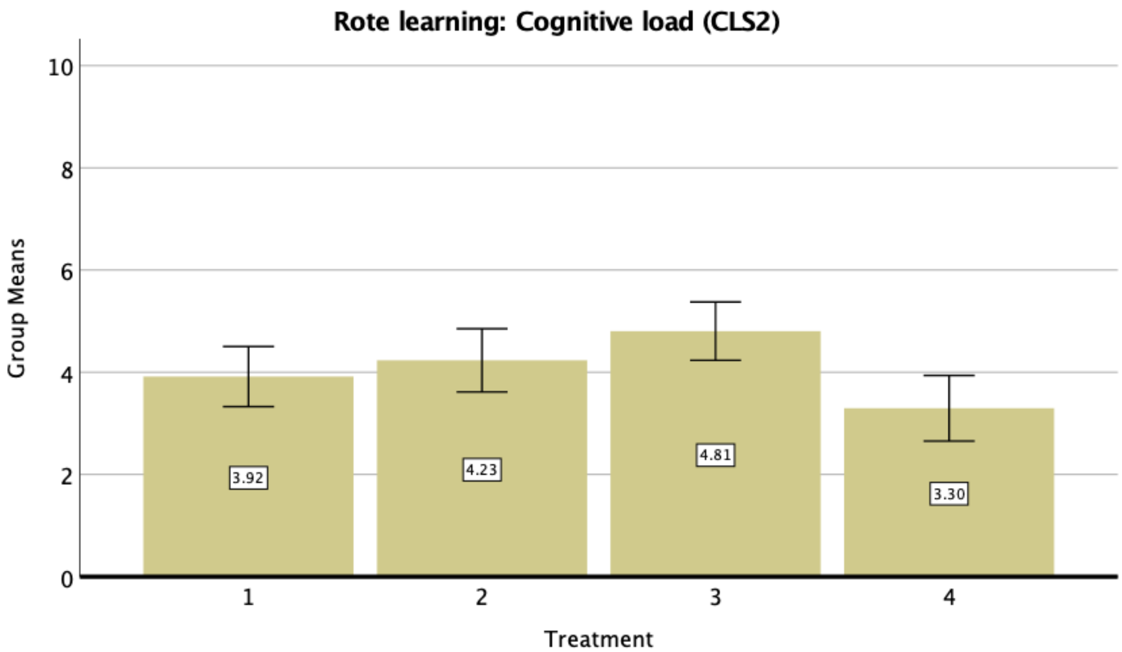

Posttest NRO
10
8
6
4
2
0
Group Means
8.42
8.15
8.13
8.82
1
2
3
4
Treatment

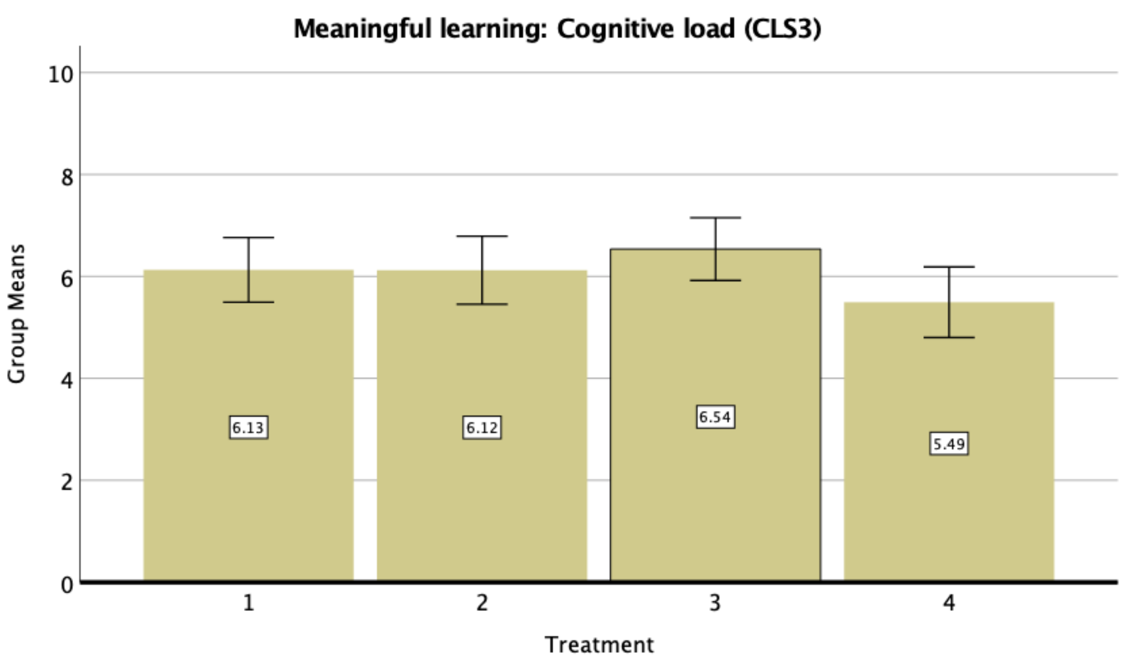

Posttest CRQ
10
8
6
4
2
0
Group Means
2.32
2.55
2.69
2.85
1
2
3
4
Treatment

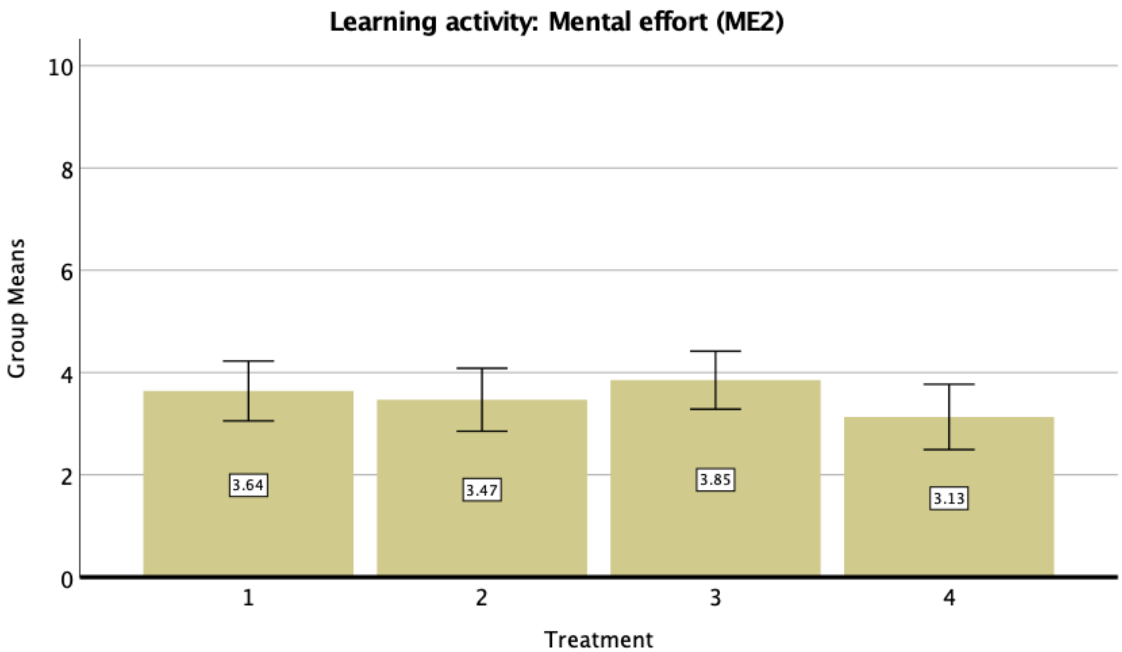

Rote learning (PTS1)
10
8
6
4
2
0
Group Means
5.37
5.35
5.41
5.84
1
2
3
4
Treatment

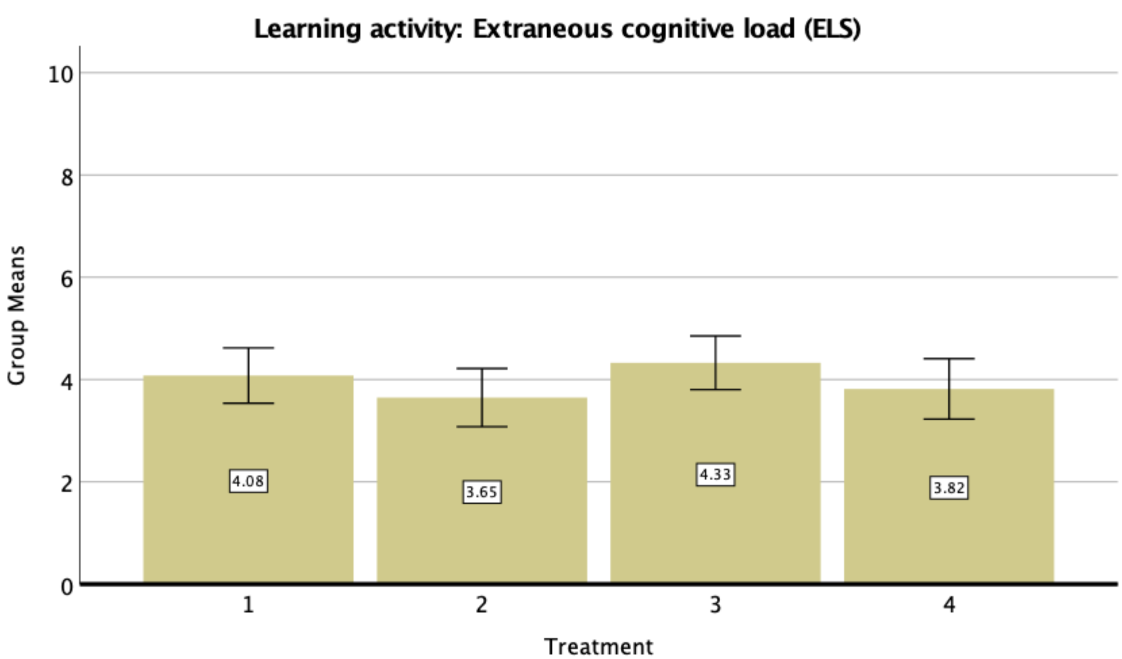

Meaningful learning (PTS2)
Group Means
10
8
6
4
2
0
5.97
6.25
5.61
5.29
1
2
3
4
Treatment

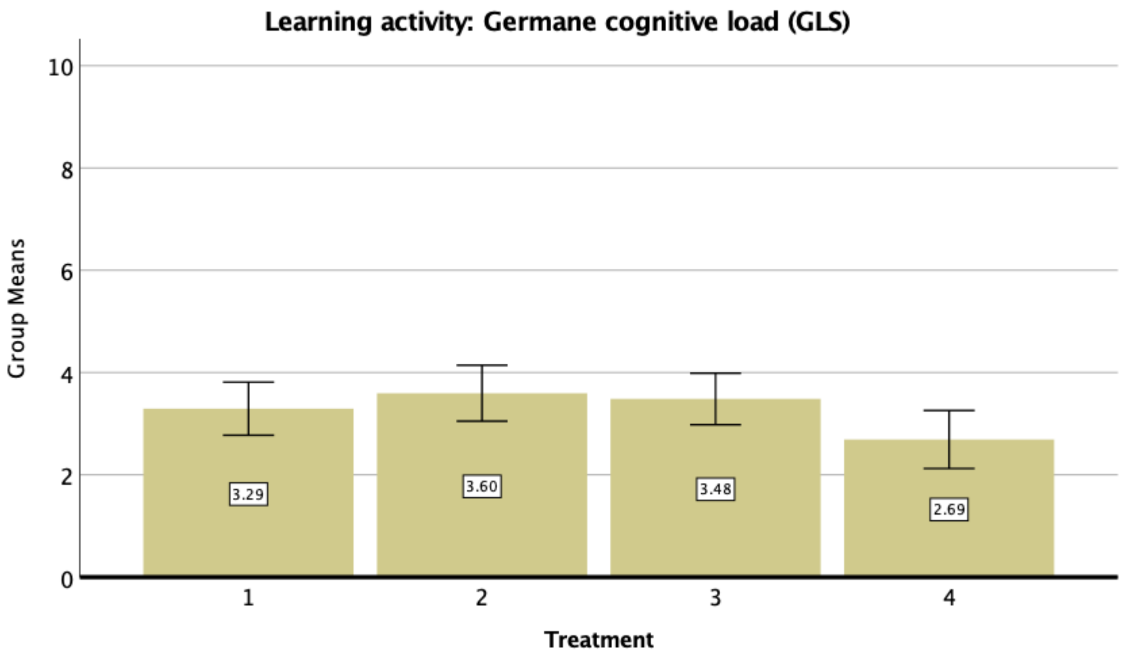

Pretest: Cognitive load (CLS1)
Group Means
10
8
6
4
2
0
7.40
7.13
7.72
6.58
1
2
3
4
Treatment

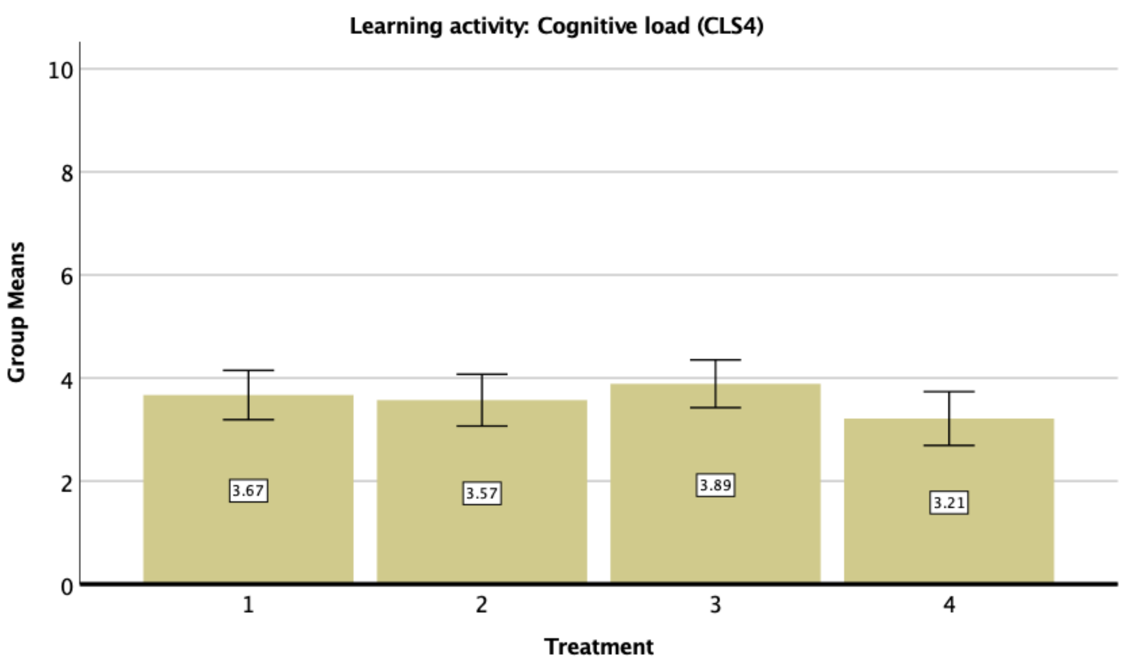

Rote learning: Cognitive load (CLS2)
Group Means
10
8
6
4
2
0
3.92
4.23
4.81
3.30
1
2
3
4
Treatment


Meaningful learning: Cognitive load (CLS3)
Group Means
10
8
6
4
2
0
6.13
6.12
6.54
5.49
1
2
3
4
Treatment


Learning activity: Mental effort (ME2)
Group Means
10
8
6
4
2
0
3.64
3.47
3.85
3.13
1
2
3
4
Treatment


Learning activity: Extraneous cognitive load (ELS)
Group Means
10
8
6
4
2
0
4.08
3.65
4.33
3.82
1
2
3
4
Treatment


Learning activity: Germane cognitive load (GLS)
Group Means
10
8
6
4
2
0
3.29
3.60
3.48
2.69
1
2
3
4
Treatment


Learning activity: Cognitive load (CLS4)
Group Means
10
8
6
4
2
0
3.67
3.57
3.89
3.21
1
2
3
4
Treatment