Learning an Abstract STEM Concept by Constructing a Three-Dimensional Physical Model Compared with a Two-Dimensional Digital Model

Salvatore G. Garofalo

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2022

Abstract

Learning an Abstract STEM Concept by Constructing a Three-Dimensional Physical Model Compared with a Two-Dimensional Digital Model

Salvatore G. Garofalo

This study examined the effectiveness of three instructional and assessment strategies on conceptual understanding of the DNA molecule. Specifically, a model building task was utilized to determine if physical model construction, digital model construction with a touchscreen tablet computer, or paper worksheet activity effected conceptual understanding during the initial exposure to an abstract science concept. The DNA molecule was chosen as an exemplary three-dimensional, abstract concept with physical and digital model building interventions. Conceptual understanding was measured using an objective quiz, a drawing of a DNA molecule, and a hand-written explanation of DNA. Conceptual understanding was measured immediately after intervention and again two months later. The study examined effects to conceptual understanding of model building by comparing physical models constructed using foam pieces and digital models constructed using a touchscreen tablet computer. A control group completed a paper worksheet activity on the topic of DNA. In all conditions, an instructional video about DNA was used to standardize the content taught. To account for the potential covariates of spatial ability and attitudes to scientific inquiry, participants completed a mental rotation test to measure spatial ability and an attitudes to scientific inquiry survey.

A total of 161 students across six intact 9th-grade Living Environment classrooms participated in the study. The results from the three conceptual understanding measures were compared among the three groups at both immediate and delayed post-test timepoints as well as

across the two post-test timepoints. For both immediate and delayed post-test, there were no differences among the groups for the objective quiz measure. However, the physical model group outperformed the digital model and control groups in both the drawing and explanation measures at both timepoints ($p < 0.01$). Across the two timepoints, the control group showed a significant degree of forgetting for the objective quiz measure ($p < 0.001$) and the digital group demonstrated a significant degree of forgetting for the objective quiz measure ($p = 0.03$) and drawing measure ($p < 0.001$). There was a significant difference between the delayed post-test and pre-test of the objective quiz for the physical model group ($p < 0.001$) and no significant difference between the post-test and delayed post-test for the objective quiz for the physical model group suggesting long-term conceptual understanding and retention.

Overall, the physical model group demonstrated greater conceptual understanding at immediate and delayed timepoints for the drawing and explanation measures as well as significant retention of conceptual understanding of DNA as measured by the objective quiz across three timepoints. The digital model group demonstrated a greater degree of forgetting for objective quiz and drawing measures as well as underperformed in the three conceptual understanding measures at both post-test and delayed post-test timepoints. This suggests that the greater degree of physical, haptic manipulation of a three-dimensional model aids in conceptual understanding at all three measures as well as long-term memory when compared with the limited haptic interactions with a two-dimensional touchscreen device.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I believe everything happens for a reason. While there were many events that shaped this journey over the last several years, it is the people that I met, engaged with, and debated with that led me to reach my goals. For all those interactions, I am truly grateful. The privilege of attending Teachers College, Columbia University and working with the top researchers in my field of interest is not lost on me. I will carry the lessons learned both in and out of the classroom with me for the rest of my career.

To Dr. John Black, my dissertation sponsor and mentor, thank you for your encouragement, support, and time on this project as well as all of my research interests. My experience at Teachers College has fundamentally changed the way I think about and conduct research, and I thank you for the opportunity to join this great institution. I am grateful for the chance you took on me and to have been able to work with and learn from a great researcher in the field. I will always remember our conversations that challenged the boundaries of current and future technologies that border and mirror science fiction and look forward to the trajectory of my research agenda. Thank you.

To Dr. Stephen Farenga, my mentor and dear friend, thank you. It is impossible to convey what your support has meant to me. Thank you for seeing potential in me that I wasn't quite sure of myself and for challenging me to continue to reach for and accomplish my goals. I am grateful for our time spent discussing research, debating pedagogical techniques, teaching classes, and presenting at conferences. You have taught me the importance of a team and enjoying the company of the people you work with, which has only enhanced my journey through academia. There is always more to say, but I look forward to our continued partnership and the opportunity to demonstrate what I have learned and what we can produce as a team.

To Dr. James Corter and Dr. Bryan Keller, my dissertation committee members, I want to express my appreciation for the constructive critiques on my statistical analyses. The suggested changes have always strengthened and enhanced this project. I truly appreciate the time spent in meeting with me and reviewing my work. The opportunity to work with you and enhance my statistical knowledge has been invaluable to me and my future career. Thank you.

To Dr. Gary Natriello, my dissertation committee member, for making the dissertation defense a truly enjoyable experience. I will cherish the discussion we had about my dissertation work which felt as a true introduction to academia. Thank you for your thoughtful questions and time in participating on the committee. It is greatly appreciated.

To Dr. Barbara Tversky, for her thoughtful and kind critiques of my research methodology. I am grateful for the opportunity to have been able to learn from a truly great researcher in the field of spatial ability and will use the ideas that developed from our conversations to conduct further research in the field.

To Megan Garofalo, my best friend and wife, I will never be able to thank you enough for all that you have done for me and our family these past few years. When I think about my doctoral studies, I think about our journey together. I started by calling you my fiancé, and I am ending calling you my wife and mother to our son. Thank you for the late nights encouraging me to continue when I didn’t believe I could, and for the late nights caring for our son. Your support is endless, and I will spend my life reciprocating. Thank you for being there every step of the way, building me up when I was down, and cheering me on across the finish line. I love you and appreciate you more than you know.

Finally, to name all the other individuals who have influenced me in the completion of this project would be near impossible. However, even with the risk of making an omission, I must make a special acknowledgement and thank my parents Salvatore and Maryann Garofalo, Dr. David Laurenson, Dr. Gopal Subramaniam, Dr. Eleanor Armour-Thomas, Mark Diercks, Jessica Mills, and to all the students I have taught who have inspired me to question how to best improve their learning and educational experiences.

Dedication

For my family

The one I was born into, married into, and built

For shaping the way I think and providing the motivation to find answers to questions to benefit future generations.

Introduction

James Watson and Francis Crick, who are credited with the discovery of the structure of the DNA molecule, did not stumble onto the correct model. Rather, the process to the correct model required several iterations of a physical construction in the middle of their Cambridge lab. By making these physical constructs, Watson and Crick were easily able to observe the impossibilities of each model they built due to constraints of physics and chemistry. These metal models brought them close to the overall structure of the DNA molecule; however, the interactions of the nucleotide bases seemed to elude them. Frustrated with the solution just out of reach, Watson created model nucleotide bases by cutting them out on paper. Then, as if working with a puzzle, he rotated the bases until they fit together which helped to complete their final model design.

The history of the discovery of the DNA molecule demonstrates the learning process of a three-dimensional, abstract concept in science. In the early 1950s, Watson and Crick did not have access to the technology of today and had to rely on the tools of their time. While technological contributions from Rosalind Franklin's x-ray crystallography aided in the overall shape of the molecule, there was no means other than physical model building to discover the correct model. More importantly, the tools these scientists used to develop their model are tools that are still available today.

Watson and Crick's model building process highlights the importance of model building in the sciences as well as science education. In order to build knowledge, the use of physical constructions is invaluable. Modern science classrooms contain a severe lack of physical models not only to build but even to view. The current curricula in biology and chemistry classes across the United States do not include model building as a criterion. Although the Next Generation

Science Standards does mention the importance of model building, the roll-out of this curriculum has been slow and the number of standards involving model building is relatively low.

Perhaps, even more troubling in the modern classroom is the ubiquitous use of touchscreen tablet devices. These tools appear to be useful, universal devices; however, the limitations in the interactions a young student can make with the device need to be considered. For example, these devices do not aid in the development of handwriting or fine motor skills due to the interactivity of glass screens. Schools often seek new advances in technology to enhance the learning process. While there is a definite need to continue to bring technology into the classroom to support students' growth and potential career opportunities, the technology must be painstakingly examined and tested to be effective before being implemented and disseminated to thousands of students.

The understanding of biology and chemistry are enhanced by the use of models. Models bring the abstract into a concrete form which allows learners to see how these abstract concepts exist and interact in three-dimensional space. When these topics are taught in grades 7-12, the representations that students interact with are found in posters on the walls, images in their textbooks, or images on screens. In other words, flat. An interesting pedagogical shift occurs when students enter college and take an organic chemistry course. One of the materials suggested for success in this class is a molecular model kit which consists of plastic rods and spheres to build three-dimensional models of molecules. This tool aids in students' understanding of the topic of organic chemistry, but it would also aid students' understanding of chemistry at all levels. Perhaps middle and high schools should consider this suggestion made in an upper-level college course to improve on their own curricula.

Overview

The proposed investigation is grounded in the practical application of the science classroom. The investigation has identified three common pedagogical strategies that have been employed in classroom settings: physical models, digital simulations, and paper worksheets. Therefore, the answers to the following questions are relevant in deciding whether students' conceptual understanding can be improved by a specific pedagogy, when introducing an initial science concept. The research is resolved to seek answers to the following pertinent questions that have been recognized as significant by a number of national and international science education movements.

1. How can we best help students develop deep conceptual understanding in science?
2. Does a specific pedagogical strategy improve long-term conceptual knowledge?
3. Does the use of touchscreen tablet devices to learn three-dimensional, abstract science concepts impact appropriate and accurate mental model construction?
4. Is spatial ability a factor in learning and understanding three-dimensional, abstract science concepts?
5. Are attitudes to scientific inquiry a factor in learning and understanding three-dimensional, abstract science concepts?

The answers to the proposed questions are of substantial educational importance for both students and teachers. The investigation is delimited by common factors that examine the initial introduction of the science concept, spatial ability, and conceptual understanding by both short- and long-term assessments.

It is evident that science means different things to different students. To some it means memorizing facts, and to others it means investigating to find solutions. The purpose of this study is to find out whether the initial learning of science concepts and the retention of conceptual knowledge is best taught by one of three pedagogical strategies. Students will be divided into three groups: physical model construction, digital model construction, and traditional learning with no model construction. Student conceptual understanding will be measured by objective-based factual knowledge of the topic, a drawing of the topic, and a learner-generated explanation of the topic. These three components provide a more accurate assessment of conceptual understanding than any one of the components alone.

A potential outcome may be evidenced by a preference of students who prefer to learn initial concepts from a more concrete hands-on perspective than that of a digital method that may require a more abstract formal operational-based thinking (Woolfolk & McCune-Nicolich, 1984). It is also of interest to determine whether spatial ability was a mediating factor in overall conceptual understanding, or whether correlated with a specific method of instruction either the physical model, digital model, or no model. Moreover, the investigation will seek to identify what method was influenced by the attitudes to scientific inquiry by the learner and whether higher satisfaction ratings were reported among students who learned by physical model construction, digital model construction, or no model construction.

Overall, the guiding principle and hypothesis of this study is that the method of instruction has an impact on conceptual understanding. The design of this study is grounded in the modern classroom with the intervention supported by cognitive psychology. This study aims to demonstrate the importance of the use of physical manipulatives for the conceptual understanding of three-dimensional, abstract science concepts while also accounting for prior knowledge, spatial ability, and attitudes to scientific inquiry.

Chapter 1: Literature Review

The literature is replete with studies that demonstrate that conceptual difficulty with science topics is a result of students' initial learning experiences (diSessa, Gillespie, & Esterly, 2004; Perkins, 1998; Treagust, 1988; Tsaparlis & Papaphotis, 2009; Vosniadou, 2008). Middle and high school science students are introduced to biology, chemistry, physics, and the geosciences mostly through two-dimensional visualizations that represent three-dimensional structures and concepts (Wu & Shah, 2004). The concepts in these fields, and the subject of visual representations, are abstract macro-, micro-, and submicroscopic phenomena. Therefore, the only option to observe the concept in totality is through a visual explanation which places the importance and responsibility of conceptual understanding on the representations of the phenomena taught (Tversky, 2019). The Next Generation Science Standards (NGSS), which identifies performance expectations at both the middle school and high school levels, acknowledges the foundational significance of visual models and incorporates terminology that encourages appropriate visualizations and other pedagogical strategies. Sixteen middle school standards and 16 high school standards include the term "develop (or use) a model" as a means of demonstrating an understanding of the science concept. One example of a middle school life science topic as outlined in the NGSS under the topic Heredity: Inheritance and Variation of Traits states "Develop and use a model to describe why structural changes to genes (mutations) located on chromosomes may affect proteins and may result in harmful, beneficial, or neutral effects to the structure and function of the organism." The NGSS performance expectations of this STEM concept, as well as 31 others, compel students to represent inherently abstract concepts in a concrete manner to support their mental models of the topic (NGSS Lead States, 2013).

The question arises whether a lack of experience with appropriate, namely three-dimensional, representations of the concept limits the development of an accurate and useable mental model that demonstrates conceptual understanding (Harle & Towns, 2011). The traditional pedagogical method to teach these concepts is through textbook, poster, and projected images and more modern techniques are on desktop, laptop, and tablet computer screens. Importantly, all of these representations are two-dimensional, seemingly regardless of the fact that STEM concepts such as molecular bonding, geographical elevation, and anatomy are three-dimensional in nature. Learning through non-parallel dimensionalities requires one to examine some of the major components of the content that could be distorted by the limited congruence in the presentation of content and concept (Baddeley, 1986). Where limited congruence exists between the physical nature of the content and the conceptual teaching of that content a confluence of factors such as spatial ability and visualization skills may be required to support correct concept formation (Eichenbaum, 1997; Scherer, Siddiq, & Viveros, 2018). It stands to reason that an optimal learning environment may require the initial instructional experience be provided in the same dimensionality of the concept (Han & Black, 2011; Tulving, 1983).

The research suggests that perception of dimensionality is affected by the medium in which concepts are presented (Preece, Williams, Lam, & Weller, 2013; Richardson, Sammons, & Del-Parte, 2018; Urhahne, Nick, & Schanze, 2009). Cognitively, difficulties involving the ability to transfer and spatial intelligence may arise when students attempt to convert two-dimensional instructional material into three-dimensional mental models due to the increased demand on working memory (Mayer & Moreno, 2003). Developmentally, younger students may find this task even more difficult due to weaker spatial competencies and when there is decreased haptic input (Barrett & Hegarty, 2016; Kalenine, Pinet, & Gentaz, 2011). The concern with touchscreen computers stems from three-dimensional graphic representations displayed behind two-dimensional glass screens which may interfere with the perceptual access to all the properties of the three-dimensions of the object as well as the intricacies of its functional processes. Moreover, the digital representation may not be as effective in fostering conceptual understanding as the physical three-dimensional representation due to the limitations of haptic experiences with the technology and lack of practical experience with science concepts (Garofalo & Farenga, 2021). Theoretically, Gibson's research (1979) on affordance would suggest that experiencing digital content on two-dimensional touchscreen tablet computers impacts the ability to accurately perceive and transfer information for three-dimensional conceptual understanding and practical real-world application.

1.1 Constructivism

The overarching theoretical framework in the design, implementation, and analysis of this study is influenced by Piagetian constructivism. The underlying questions in the study lie on the boundaries of philosophy and psychology as it concerns epistemology (the nature of knowledge) and how individuals learn (general learning theory). The dualism of thought can be explained by Popper (1959) who strategized that empirical truths can be supported by discoveries in the investigation, and that the justifications can be made in relation to the educational importance of knowledge. The nature of this position fits educational settings where one considers a philosophical position of why conceptual understandings develop in a certain way and yet account for individual achievement and the context of the environment (Perkins, 1998).

Piaget's genetic epistemological theory is explained by biological and structural mechanisms. The basis for all knowledge stems from knowledge attained through biological reflexes and gradually moves from concrete experience within the environment to more abstract reasoning (Piaget, 1971; Piaget, 1972). Therefore, knowledge attainment is developmental in nature and organized from simple to complex thought, where reasoning develops from a subjective to a more objective perspective by increasing structure and organization. Constructivism suggests that learners continue to acquire conceptual understanding by adapting to their experiences through a process of continual transformations of structures and reorganization of prior knowledge. For Piaget (1970a), experiences advance structures that are represented by schema which are cognitive maps of the mental and physical actions that involve understanding and knowing. As development proceeds, schemas form hierarchical categories of knowledge that one uses to interpret and understand the world. The process of acquiring information into one's existing schema is known as assimilation. The process of assimilation is both idiosyncratic and subjective because individuals have a tendency to modify the experience or information to fit with their previously held views. It is this reason that possibly explains the difficultly of changing one's misconceptions. To do so would require one to adapt and modify his or her existing schema through a process known as accommodation. Accommodation involves modifying existing schemas, or thoughts, as a result of newly acquired information. Piaget further described a third mechanism equilibration, that balances the dynamic processes of assimilation and accommodation. Equilibration is an attempt to preserve a balance between applying prior knowledge through the process of assimilation and modifying behavior or thought through the process of accommodation to account for new information or experiences. Cognitive development requires that for adaptation to occur the processes of assimilation and accommodation are both necessary and that accommodation is only possible with assimilation (Piaget, 1962).

Piaget (1970b) suggests that experiences and subsequent pieces of new knowledge are cognitively linked by the reciprocal processes of assimilating and accommodating to existing knowledge of a phenomenon. Pedagogically, it is important to interact with the material in similar and meaningful context in order to create appropriate linkages for long-term understanding. The implication for these experiences is that they occur in physical, three-dimensional space that engages the senses. Separate from a behaviorist perspective, constructivism is an idea that is reliant upon the existence of preconceived notions and thoughts of the topic, concept, or information to be learned. The subsequent experiences the learner encounters are incorporated into the existing knowledge to correct and build upon the initially learned material (Sjoberg, 2007).

One consequence of physical, three-dimensional experiences with three-dimensional concepts is the development of imagistic reasoning (Dickmann et al., 2019). Imagistic reasoning encompasses mental rotation, spatial thinking, mental model building and transfer skills that has been demonstrated to support problem solving as well as in conceptual understanding of the STEM fields. This is especially important for complex problem solving that rely on foundational concepts in STEM since the apparent default thought process is through diagrammatic reasoning (Stieff, 2010).

Piaget's constructivism suggests that children formulate empirical abstractions based on their experiences regardless of the realism and transferability of the learned material (Von Glasersfeld, 1991). A point of contention with Piaget's notion of structuralism is that it deals with the logical competence of a participant rather than the participant's actual performance and the impact of concrete content and the environmental press (Wason, 1977). Moreover, the limited functionality of flat surfaces (textbooks, tablets, computer screens) has the potential to create incorrect, non-representative visuospatial mental models of the abstract concepts that exist in three-dimensional, physical space (Baddeley, 1986). In order to alleviate this modality discrepancy, the inclusion of a haptic element may improve learning (Han & Black, 2011). However, the physical experience with three-dimensional concepts must exist in three dimensions to create realistic, usable, and congruent mental models rather than the limited and restrained physical interactions provided by two-dimensional glass screens (Kontra et al., 2015; Preece et al., 2013).

A constructivist framework integrated with the research in embodied cognition and haptics can address the two-fold problem of conceptual understanding of three-dimensional abstract STEM topics with touchscreen technology (Richard et al., 2006). The first is the issue of dimensionality. The presentation of three-dimensional abstract STEM concepts through two-dimensional media can create both misconceptions and incongruent mental models of the concept. The second is the limitations of touchscreen technology. In addition to the problem of dimensionality of the glass screen, the possibility for physical manipulation and therefore an authentic physical experience with the content is limited.

1.2 Embodied Cognition

Arguments for experiential learning involving physical manipulation with the body have pervaded educational research for well over a century by some of the most prominent researchers in the field (Barsalou, 2008; Bloom, Hastings, & Madaus, 1971; Dewey, 1902/1938; Piaget, 1970a/1999; Papert, 1980; Perkins, 1998). Experiential learning, which incorporates hands-on learning, is a general pedagogical strategy that can be examined through grounded cognition. Barsalou (2008) argues that grounded cognition requires that thinking have a basis in simulation, bodily states, and/or situated action. Haptic experiences are foundational to all three of these general categories and include bodily movements, situations, and internalizations to support grounded embodied cognition as a cognitive theory which has neurological and physiological implications.

The ability to physically manipulate objects is essential in developing long-term memory and knowledge acquisition. Black et al. (2012) provide a theoretical understanding of the importance of embodied tasks in learning including through the use of technology. The researchers suggest that utilization of multiple senses including touch, both large and small gestures, and supporting embodied tasks with technology enhance student learning, understanding, and motivation. The incorporation of movement while learning activates neural connections and motor and perceptual areas of the brain during object property verification (Holt & Beilock, 2006; Martin, 2007). These neurological impacts suggest that movements and gestures should be representative of the learned material. For example, swiping across a glass screen may not be enough stimulus to create meaningful connections physiologically and cognitively for complex skills (Garofalo & Farenga, 2021; Stieff, 2010). Alternatively, moving a cursor a certain distance to represent a percentage may be just the right amount of embodiment (Swart et al., 2015). The current research on embodied cognition and technology is varied; however, the overall suggestion is that movements made by the learner must mimic the movements necessary when manipulating the objects in the real, physical world to aid in knowledge acquisition (Black et al., 2012; Han & Black, 2011; Richard et al., 2006). Failure to allow for authentic haptic interactions could lead to misconceptions that would be difficult to correct at a subsequent time.

In addition to conducting accurate and authentic physical movements during model construction and the use of simulations to build procedural knowledge and mental models, embodied tasks can also help ground abstract concepts. Many concepts in the STEM disciplines are abstract and, supported appropriately, can be taught through a variety of media including technology (Leng & Gijlers, 2015; Schott & Marshall, 2018). Mathematical skills of counting and addition tasks and physics concepts such as gravity and trajectory using haptic feedback tools as well as the use of the augmented reality sandbox to teach contour maps has demonstrated success in conceptual understanding when the physical movements are congruent to the digital representation of the gesture (Huang, Vea, & Black, 2011; Richardson, Sammons, & Delparte, 2018; Segal, Tversky, & Black, 2014).

Despite the positive results of the importance of incorporating hands-on and embodied learning with STEM concepts, these forms of active learning are not always prevalent in the classroom (Stains et al., 2018). This is especially concerning when the large majority of content taught in STEM is abstract. Neuroscience research has also demonstrated the benefits of haptic experiences which exemplifies embodied cognition and encourages physical movement to enhance thinking and learning. Learners who embody actions and simulations of abstract concepts are afforded an opportunity to activate both the sensory and motor cortex while acquiring and encoding information (Calvo-Merino et al., 2005; James & Swain, 2011; Kontra et al., 2015). One example of this phenomenon was demonstrated when children aged five to six were presented with actions verbs they had learned through physical motion representing those verbs. The researchers found that the participants' motor cortex was activated while passively responding to the verbs at a later date. The participants who learned the verbs through passive viewing exhibited different regions of brain activation. Therefore, it is possible that enacting the physical motion allowed the participant to imagine the action during subsequent testing which led to motor cortex activation (James & Swain, 2011).

The research on haptic experiences support its usefulness for children of all ages, especially young children (Kalenine et al., 2011; Kontra et al., 2015; Möhring & Frick, 2013). The development of children's thinking can benefit greatly by physical manipulation of objects as it promotes cognitive growth and psychomotor maturation of fine motor skills as well (Kempermann et al., 2010). Three-dimensional concepts are particularly supported through this type of interaction starting at a very early age. Möhring and Frick (2013) observed six-month-old infants exercising mental rotation of physical objects. The results of this study suggest that when participants were allowed to physically manipulate and explore the object, as opposed to passively viewing the object being presented, they were better able to mentally rotate the object and comprehend that the object being presented was indeed the same object in a different position. This was measured by mean looking time which was greater when the infants were confused suggesting they were not familiar with the particular spatial location of the object. This study supports the idea that physical manipulation of an object can instill some level of spatial understanding and long-term memory of the three-dimensional object and that spatial reasoning is a foundational concept that is inherently intertwined with physical manipulation of real-world objects. Continuing this line of research, Kalenine, Pinet, and Gentaz (2011) implemented a visuo-haptic intervention with preschool students for geometric shape recognition. The results of their study mirror that of Möhring and Frick: the young children who were able to manipulate the physical objects were better able to recognize the geometric shapes compared with children who only viewed the objects. In addition to developmentally-appropriate shape recognition, the researchers also noted that the students in the visuo-haptic group were able to recognize the rectangle and triangle which is not a typical result for preschool-aged children. This study provides support for potential cognitive advancements due to embodied cognition.

Concepts in STEM disciplines are spatial (Nagy-Kondor & Esmailnia, 2022), and haptic interactions with activities that embrace these concepts may support conceptual understanding and long-term memory as the physical manipulation of three-dimensional objects with six-month-old and preschool-aged children were demonstrated by Möhring and Frick (2013) and Kalenine et al. (2011). However, haptic experiences of three-dimensional, physical object manipulation are not the only forms of embodied cognition. Kontra et al. (2015) note that enacting certain physical tasks have a neurological impact on subsequent thinking of the physical task. In their study, participants physically experienced a task involving kinetics, which is a concept taught in a college physics course. Participants who had a direct experience with the concept of kinetics demonstrated higher conceptual understanding of the topic when compared with participants who observed the action. In addition, the researchers observed activation of the same sensorimotor regions during a posttest as activated during the physical experience of the topic for the participants who physically experienced the concept. This match in activation of the cerebral cortex aids an individual in their long-term memory and conceptual understanding. The researchers also note that success in this transfer from physical activity to achievement on testing was a result of a match between the sensory and motor experience and learning assessment. It was also noted that concepts that lend themselves well to physical experiences, such as kinetics, are benefited from physical experience with the topic. They also support that hands-on, experiential learning should occur during the initial learning of the concept.

Möhring and Frick (2013), Kalenine et al. (2011), and Kontra et al. (2015) demonstrate the importance of haptic experiences for conceptual and long-term understanding over the early lifespan. In particular, these studies support the importance of physical manipulation of concrete, real-world objects of spatial topics. Despite the perceived and demonstrated importance of haptic experiences of three-dimensional physical objects, learners are provided fewer opportunities to interact with these objects in school and in other aspects of life through their childhood and teenage development (Henderson & Dancy, 2007). Part of the reason for the disengagement with physical objects is the continuous evolution and prevalence of technology. A piece of technology that has found its way overwhelming into secondary schools across the globe is the touchscreen tablet which is inherently a haptic device. Over the past decade, the touchscreen device provided a new opportunity for learners to engage with digital content. Unlike educational television, touchscreen applications allow users the ability to physically interact with images, animations, and texts, in addition to other graphics, which provides an embodied and haptic learning experience. Xie et al. (2018) explored the potential benefits of learning on touchscreen devices through a meta-analysis of empirical studies of participants aged 0 to 5 years old. Some of the results synthesized include embodied motions were more beneficial for learning when compared with no physical interaction with the content, physical experience with STEM content improved learning more than physical content with non-STEM content, and learning improved in the classroom with touchscreens compared with in a laboratory with touchscreens. Xie et al.'s findings support the notion that STEM is a highly spatial and experience-driven content area that requires hands-on activities to improve learning in addition to the benefit of free exploration as a positive learning environment. While touchscreen devices were determined to be useful educational tools for haptic experiences, it is important to note that no other haptic experiences were measured against the touchscreen devices.

Embodied cognition has been demonstrated to be useful for young children especially when they are learning spatial concepts. Developmentally, it is logical that any physical interaction with the material to be learned would be helpful for young children (Piaget, 1971).

However, older individuals who are faced with more complex content may require more intricate, haptic interactions to experience greater conceptual understanding. Pedra, Mayer, and Albertin (2015) examined the levels of interactivity with touchscreen animations utilizing a highly spatial STEM concept to determine if there was an effect. Unlike the results from Xie et al.'s (2018) meta-analysis, Pedra et al. (2015) did not find increased conceptual understanding due to higher levels of interactivity with the touchscreen animation. The participants were engineering and non-engineering college students presented with maintenance procedures for a three-dimensional mechanical device. Despite finding greater interest in the topic, the level of interactivity did not significantly change the result of conceptual understanding. These findings support the argument that even though participants were given the option to physically manipulate objects on screen, the animation presents the three-dimensional engineering concept behind the two-dimensional glass screen of the touchscreen device. Therefore, the transfer from information presentation to recall of conceptual information was not accomplished.

Although modern technology requires user input through movement of a mouse, stroke of a keyboard, or swipe or tap on a screen, the limitations of the extent of physical movement compared to manipulating three-dimensional objects is evident. Pedra et al. (2015) examined the difference between interactivity with a touchscreen and no interactivity, but Schwartz and Plass (2014) examined different types of interactivity with the touchscreen. The different types of interactivity are described as iconic interaction which are dragging functions of the touchscreen device to mimic direction, speed, and force for the movement of the digital object, symbolic interactions which are clicking functions on any digital device in which a simple click of a mouse or tap of touchscreen causes greater movement of the digital object than exerted by the user, and enactive interactions would be considered the most embodied as it requires the user to perform a physical movement that is then mapped on or mirrored by the digital object on-screen. Enactive interactions would require technology such as augmented or virtual reality. Schwartz and Plass (2014) studied the difference between iconic interaction and symbolic interaction. In terms of recall and recognition, the participants who used iconic interactions were more successful than those who used symbolic interaction which demonstrates the benefits of more enriched haptic experiences during learning. Similar results were also found when the participants were tested three weeks later. The results of this study suggest that haptic and embodied experiences can be beneficial for learning with technology provided the physical experiences mirror the movements on-screen.

Russo-Johnson et al. (2017) acknowledge some of the benefits and disadvantages of using touchscreen devices by focusing on the affordances of the touchscreen device as well as the applications presented on the screen. While they support the need for haptic experiences in order to achieve greater learning outcomes, the results from their studies question if the learning and engagement achieved was meaningful. The applications used in the studies were self-directed and therefore gave the participants little choice in their interactions with the device. The use of these applications demonstrated a variety of learning improvements based on several factors such as gender, SES, and age. However, the researchers mention that with limited interactions and lack of intrinsically motivated actions it is difficult to suggest the extent in which the touchscreen device itself is the contributing factor to learning achievements. Recognition that all digital applications do not have the same potential educational benefits is essential for technology integration in classrooms. This notion is further supported by Hirsch-Pasek et al. (2015) who suggest a theoretical framework for evaluating applications to be deemed educational. The four pillars of cognitive affordance necessary in educational applications are active learning, engagement in the learning process, meaningful learning, and social interaction. These pillars align with the necessity for children to physically interact with the learning material being presented. Learning cannot be done passively. It must be done actively.

The uncertainty of the true value of educational technology in conjunction with its ubiquitous nature has led to a continuous search for appropriate uses by teachers, researchers, administrators and other stakeholders. The seemingly important nature of including physical manipulation of three-dimensional concepts at a young age supports the need to continue these experiences. It is possible that interrupting this development by introducing two-dimensional interactions of three-dimensional concepts may stall or prevent natural development (Paulus et al., 2019). However, regardless of the potential negative effects on development, it stands to reason that the disparity in dimensionality will impact conceptual understanding (Baddeley, 1986). While individuals may have haptic experiences with a touchscreen device, the experience is not as rich, nor as physically demanding, as it would be with concrete, physical, three-dimensional objects (Russo-Johnson et al., 2017). Further, cognitive transfer between modalities may prove too difficult to individuals, especially young learners.

1.3 Transfer

Transfer is a cognitive skill that has been discussed since the beginning of formal education. Educators and researchers have questioned the possibility of transferring skills from one context and problem set to others within the discipline or across disciplines (Barnett & Ceci, 2002). However, the concept of transfer in education covers a variety of issues such as knowledge domain, physical context, temporal context, and modality. The abstract nature of STEM concepts requires some foundational context in which to build appropriate and accurate mental models. This implies that all STEM learning will require some form of transfer for conceptual understanding, especially with a constructivist lens. Since concrete models and representations are a foundational prerequisite for learning about these abstract concepts, the context of modality transfer is of particular interest in this study. However, the broad range of physical representations, from textbooks to tablets, presents the question of how far of a transfer is required.

Barnett and Ceci (2002) broadly define far transfer as transfer to a dissimilar context and near transfer as transfer to a more similar context. There are considerations of the varying degrees of distance based on factors which include the content and context. The content factor calls for consideration of the specific or general nature of what is taught, the expected outcome or final performance based on what is taught, and the extent to which what should be remembered based on what was taught. The context factor considers the domain of knowledge, physical and temporal context, and modality. When examining STEM concept learning, the issue of modality is of most importance because of the traditionally two-dimensional method of instruction. Posters, textbooks, and even videos are all used as two-dimensional representations of three-dimensional concepts, and often, these representations do not accurately represent the content to be learned. Therefore, modality transfer should be taken into consideration when using the modern two-dimensional representational tool, the touchscreen device.

The transfer from two-dimensional tablet screen to conceptual understanding of a three-dimensional object brings into question just how far the transfer is. The encoding specificity hypothesis would suggest that the material learned would only be able to be accessed if the representation of the content matches the representation at the point of learning (Tulving, 1983). By this theory, the difference in dimensionality may be too far. However, representational flexibility, the ability to overcome varying representations of learned material for use in novel situations, presents a possibility to use different modalities in learning (Eichenbaum, 1997). While representational flexibility is a promising theory of transfer, it must be developed with age and is also reliant on the task and experience required of the individual. Scherer et al.'s (2018) meta-analysis asks if far transfer is even possible and takes into consideration creative thinking, mathematics skills, metacognition, and spatial skills as primary factors in the ability to allow for far transfer. The answer to this question varies and ultimately relies on cognitive development as a result of age. However, a full developmental scope is not established as many of the studies conducted have focused on young children which does not account for older individuals. These studies also do not often examine STEM concepts. One example of a theory of transfer is the video deficit effect which suggests that children have poor transfer ability of content from television as well as two-dimensional still images when compared with physical interactions. Barr's (2010) literature review suggests the video deficit effect is true of children at least until age three. However, Barr notes that there is a lack of research on older individuals and video deficit effect on skills other than imitation.

The touchscreen tablet is a relatively new technology which prevents our ability to predict how early use of the tablet may impact individuals developmentally: both cognitively and physiologically. As longitudinal research is being conducted, such as by the Adolescent Brain Cognitive Development (ABCD) study (Volkow et al., 2018), an examination of young children's attempts at transfer with tasks involving the touchscreen tablet may provide some insight into the touchscreen device's impact on development.

In previous research, one basic transfer task that was examined was imitation. The imitation task required that participants simply mirrored the physical movement from the two-dimensional tablet to a three-dimensional physical representation of the task that was presented on the screen. Zack et al.'s (2013) study presented 15-month-old infants with the simple task of pushing a button. The infants watched the researchers demonstrate the action on either the screen or the physical button and were able to imitate the action. However, when promoted to push the button on the alternative modality, either screen or physical object, the participants were unable to do so. In a different imitation task, Chen and Siegler (2013) found that young children between age two and two-and-a-half were able to imitate gestures viewed in two-dimensional videos. Although the youngest participants were unable to transfer the skills learned through the video to problem solving where the older children were more successful, this study helps to demonstrate that transfer abilities develop rapidly. When tasks require simple near transfer such as through imitation, children beginning at the age of two are able to successfully transfer the skill.

The next cognitively stimulating task is that of problem solving. Building on the age group studied by Chen and Siegler (2013), Moser et al. (2015) studied the problem-solving transfer for two-and-a-half to three-year olds from touchscreen to three-dimensional puzzles. The puzzle tasks were identical except in dimensionality. When participants practiced and were tested in the same dimensionality, they were successful, but when they practiced on the touchscreen and were tested on the three-dimensional magnetic board, they were not. The far transfer due to modality and dimension change was too large for the participants and caused a transfer deficit. These results could have been due to the general difficulty children three years and younger have with transfer of skills as suggested by Barr (2010).

Tarasuik, Demaria, and Kaufman (2017), however, found promising results in a problem-solving transfer task with children aged 4 to 6. The researchers found no difference in the ability to complete the puzzle in the three-dimensional form regardless if the participants spent time practicing with the two-dimensional touchscreen representation or the actual three-dimensional object. Since the problem and puzzle were the same in different dimensionalities and the physical movements differed only in finger versus hand movements, it is possible that Tarasuik et al. found success with this transfer task due to the relatively near transfer of the task.

These initial studies provide some general context for future research. Namely, the general ability to transfer develops over time and perhaps some of the results found by transfer researchers were a result that the tasks may have been too difficult for young children who are still developing the skills of transfer (Barr, 2010). This study did not measure more complex transfer tasks. Tarasuik et al.'s (2017) study provides the most promise for the transfer from a two-dimensional representation to a three-dimensional representation of the same object or task. However, it is important to consider that more complex tasks as well as conceptual understanding were not measured.

As transfer research continues, a cognitive load perspective may illuminate some of the outcomes for transfer while limiting cognitive load. For example, Dan and Reiner (2017) found that the authenticity of the task in a different modality was a key variable in which to limit cognitive load. These results may be extended to conclude that the transfer is also lessened to aid in conceptual understanding. Despite the positive outlook on these results, the issue remains whether or not manipulation of three-dimensional objects behind a two-dimensional screen can be considered an authentic task. If not, the degree of transfer may still be too great for learners.

1.4 Neuroscience

1.4.1 Haptic & Physiology

The cognitive benefits of haptic experiences have been demonstrated through learning, memory, and representational tasks. It is important to consider the neurological implications of haptic experiences involving physically engaging with an object to provide insight into why these learning opportunities are important at a physiological level. Through the sensation of touch, different parts of the brain are activated in order to comprehend the size, shape, dimensions, and purpose of the object in question. Sathain's 2016 analysis of haptic information in the cerebral cortex provides a map of regions of the brain that are activated due to haptic stimulus. Originally processed by the primary somatosensory cortex, understanding the location of component parts of an object are made possible in conjunction with visual stimulus through the intraparietal sulcus and frontal eye field. However, it is important to note that true understanding of physical objects through haptic interactions is a multisensory experience. As such, tactile spatial acuity, which suggests touch as a means of accurately interpreting the physical properties of an object, involves both tactile inputs as well as visual stimuli.

Grounded embodied cognition is bolstered by the physiological evidence of cortex activation in regions activated by spatial, visual, and tactile responses. This multisensory trifecta is further supported by Stock et al. (2008) who found similar results when observing the brain regions that were activated during visual and haptic activities as well as recall for long-term understanding. The studies conducted by Stock et al. suggest that actions, whether visual or haptic, systematically and regularly activate specific regions of the brain. In addition to location activation at first interaction, the regions activated during long-term recall were the same as when the experience first occurred. Motor and somatosensory cortex regions were activated when participants thought about previous haptically-encoded objects and actions. Calvo-Merino et al. (2005) found this with dance, Beilock et al. (2008) found this with hockey, and James and Swain (2011) found this with action verbs in reading. These repeated results suggest the importance of accurately encoding haptic information to be accessed when necessary.

Haptically-encoding information with incongruent physical motions may be detrimental to long-term understanding if the incorrect regions are activated because of the initial encoding experience.

1.4.2 STEM Physiology

Physical engagement with an object can support long-term understanding of that physical object. Neuroscience research in conjunction with embodied cognition research has separately demonstrated the importance of haptic learning of STEM concepts. Cetron et al. (2019) were interested in identifying brain regions that were activated specifically during STEM conceptual tasks. Participants were asked about forces that are typically taught in physics and engineering classes in addition to their ability to identify where on a given structure forces were being exerted. By examining the results of engineering students and novices, the researchers were able to distinguish the regions of the brain that were activated when a participant was familiar with the STEM concept. The results suggest that engineering students activated a dorsal stream frontoparietal network, which is similar to the results found by Sathian (2016), and a ventral stream occipito-temporal network, which also aligns with the connection to visual input. Cetron et al. (2019) note that the ventral occipito-temporal cortex is also active when mentally visualizing abstract concepts which is useful in reflecting on those concepts in a real-world setting. This is particularly important for STEM concepts because an ultimate goal of teaching and learning STEM is to be applied to real-world problems (NGSS Lead States, 2013). These findings support the notion that activation of prior conceptual knowledge about an object, or STEM three-dimensional concept, involves the ventral occipito-temporal cortex and dorsal stream frontoparietal network.

Mason and Just (2015) found similar cortex activation during STEM concept learning. In their study, participants were presented with different concepts of how certain mechanisms work, such as a scale, brakes, and instruments. The participants were then asked to think about how the objects work before being tested on the mechanics of the system. All mechanisms involved physical input in order to function. Measurements were taken before, twice during, and after the learning treatment to observe cortex activation. The parietal lobe was shown to be active after training and during learning, demonstrating the importance of visual imagery, and the frontal and motor cortex were active when the participant was asked how a person might interact with the mechanism, demonstrating the importance of physical movement in mental models. These results provide another example of motor cortex activation during recall. The inherently three-dimensional, process-driven, mechanical nature of STEM concepts lend themselves to be taught haptically to engage visual, spatial, and tactile responses in the brain.

1.4.3 The Adolescent Brain

The interest in studying adolescents stems from the degree of brain plasticity and myelinogenesis that occurs during this developmental period (Howard-Jones, 2009). As the brain prepares for final large-scale changes, an increase in white matter can be observed as well as impacts on cognition (Arain et al., 2017; Forbes & Gallo, 2017; Geidd et al., 2015; Paulus et al., 2019; Walhovd et al., 2016). This is also a time period of neuron connectivity that is influenced by the actions and thoughts of the individual as well as the environment and impacts several regions in the brain (Alexander-Bloch et al., 2013; Vandekar et al., 2015). This is especially important during adolescence because their interactions in their environment are more complex and therefore may play a larger role in the structural changes in the brain. One aspect of the modern adolescent's environment is the extreme prevalence of touchscreen devices among this age group. Therefore, studying the impacts of this piece of technology on adolescents is extremely prescient in today's educational climate.

Arain et al. (2017) report that myelin synthesis, dendritic pruning, and neurocircuitry strengthening continues throughout adolescence and is particularly vulnerable to changes in hormones and environmental input. They note that this could impact cognitive processes as well. In particular, white matter plasticity is found in the occipital, parietal, frontal and temporal lobe which are all going through maturation and are involved not just in haptic experiences but also in STEM learning and understanding (Stock et al., 2008; Mason & Just, 2015). At this critical period, adolescents have the potential to increase their ability to problem solve, process complex information, and develop talents and lifelong interests. The importance of this period highlights the danger of negative environmental factors that can impact, and possibly diminish, these cognitive skills.

1.4.4 Screen Time Physiology

Myelin is critical for proper brain function as it aids in the conduction of nerve impulses throughout the nervous system, and degradation of myelin can have negative effects on individuals from cognitive deficits to neurological disease. Forbes and Gallo (2017) explain that an enriched environment has the potential to enhance myelin development throughout the lifespan. In the instance of learning, white matter plasticity allows cognitive functioning to take place. Forbes and Gallo particularly note that complex visuomotor skills and motor learning require myelin plasticity. Among the essential components to maintain a healthy and plastic brain are newness and challenge. Therefore, it is in the interest of adolescents, as well as adolescent caregivers, to ensure a healthy, enriched environment for this essential growth.

Myelin development occurs rapidly at a critical period when children are very young followed by a second rapid increase at the onset of puberty continuing to 24 years old (Arian et al., 2017). The American Academy of Pediatrics (AAP) and the World Health Organization (WHO) recommend limiting screen-based media to no more than one hour per day due to concerns that extended media use in replacement of cognitively constructive stimuli could lead to slower or decreased myelination. Hutton et al. (2020) found these suggestions to be warranted by identifying lower microstructural integrity of white matter specifically in regions that support language, executive functions, and emergent literacy skills. Although the researchers questioned the other potential variables that may have led to this decrease in microstructural integrity, they all centered on the increased use of screen-based technology. For instance, one concern was a lesser degree of social interaction due to the solitary nature of screens. Although this study was interested in screen time alone, it is important to note that even the use of the technology for extended periods of time was found to negatively impact white matter and myelination at a critical period of brain development. This can logically be extended to improper uses of technology, other than time, that could potentially have future negative implications.

Further evidence of the neurophysiological implications of increased screen time was measured by Horowitz-Kraus and Hutton (2017). Pre-adolescents participated in a study that sought to observe the degree of brain connectivity with regard to language and visual and cognitive control. The researchers found that participants who spent more time engaging with modern technology such as smartphones and tablets had lower levels of connectivity when compared with participants who spent more time reading. It is important to note, however, that the use of screen-based media was considered passive in comparison to the active behavior of reading and therefore the level of interactivity was largely a factor in data analysis.

When comparing young children and adolescents in terms of screentime, Adelantado-Renau et al. (2019) did not find any significant impacts on academic performance overall. However, their results do suggest that individual screen-based activities such as viewing television and playing video games have negative impacts on adolescents' academic performance. Academic performance was not affected in young children based on the findings of this meta-analysis; however, it is not clear if prolonged exposure to screen-based activities beginning early in life has any long-term effects on academic performance. Although this meta-analysis suggests caution with screen time in adolescents, the use of touchscreen devices or specific academic subjects and subject-specific skills was not measured. It is evident that more research is needed, especially longitudinal studies to observe any effects of modern technology on brain development. However, from these preliminary studies, caution may be warranted when integrating technology into the classroom.

1.4.5 Technology Physiology

Despite the general interest in touchscreen impact on cognition and neurophysiology, the long-term impact of these devices has not been measured largely due to the relatively short time period since its creation. Paulus et al. (2019) have begun to examine the data that is being collected in the Adolescent Brain Cognitive Development (ABCD) longitudinal study. The ABCD study is meant to examine the cognitive, brain, social, and emotional development of adolescents from age 9-10 through early adulthood. Impact factors such as drug and alcohol use, addiction, depression, and smartphone use are among the factors that will be examined. In Paulus et al.'s study, the participants in the ABCD study were asked to self-report their screen-mediated activities and frequency which were measured against fluid and crystallized intelligence. Fluid intelligence involved problem solving and thinking quickly where crystallized intelligence involved experience, and long-term knowledge of skills such as verbal knowledge. While the results varied, it is important to note that social media appeared to have a negative impact on both fluid and crystallized intelligence in addition to physiological impacts to the brain. Screen-based media use is varied which makes it difficult to distinguish which types of activities negatively impact intelligence and cognition specifically. The ABCD study aims to shed some light on the impact of modern technology on neural development.

The neuroscience research has informed the areas of brain activation for haptically-encoded concepts, STEM conceptual understanding, long-term recall, and visuo-spatial concepts as well as the potential impacts of technology on the developing adolescent brain. The relatively contemporary nature of touchscreen technology has prevented the development of research combining the neurological impacts of technology use for learning abstract, three-dimensional STEM concepts. As such, this question remains largely unanswered by neuroscience; however, the neuroscience research provides a necessary background for foundational understanding of how the environment and physical manipulations impact the brain and therefore support the findings of cognitive studies.

1.5 Spatial Thinking

Spatial thinking involves one's ability to perceive, recognize, or conceptualize physical or intellectual constructs in terms of their position or location in both static and dynamic systems (Ness, Farenga, & Garofalo, 2017). Individuals have varying levels of spatial intelligence which can impact their ability to understand and acquire information in the STEM fields (Harle & Towns, 2011). Particularly in the area of the sciences, an individual's ability to exercise spatial thinking is an indicator for success in science learning and perhaps an extension into a science profession (Nagy-Kondor & Esmailnia, 2022; Wai, Lubinski, & Benbow, 2009). The three-dimensional, process-driven nature of the sciences lend themselves to the need for effortless manipulation, construction, deconstruction, rotation, visualization, and perspective-taking on the part of the learner (Uttal et al., 2013). From concrete practices like titration and dissection to abstract concepts like chirality and aerobic respiration, science content is spatial in nature. Therefore, learners' spatial intelligence should be leveraged, science should be taught conceptually as a spatial topic, and hands-on, physical, three-dimensional manipulatives should be used to represent the spatial concepts (Huk, 2006).

The tools to learn and understand spatial concepts support the individual's physical and mental movements through space. Just as a map helps navigation, a molecular model kit unveils the structure, function, and connectivity of molecules to the organic chemist (Clements, 1999; Dickmann et al., 2019). Understanding the importance of the three-dimensional nature of science concepts leads to the understanding that haptic experiences are necessary for accurate mental model representation (Jones et al., 2006). The constructivist lens relies on these experiences to build on preconceived notions of these topics to enhance and correct original thoughts. Therefore, it is important to support spatial thinking with physical manipulatives during learning to allow for conceptual understanding as well as long-term recall (Barrett & Hegarty, 2016).

The concrete nature of some science topics is more palatable for learners because they are able to observe these phenomena in nature. The science topics that are abstract, like gravitational force or planetary orbit, however, are much more difficult for learners to comprehend especially if the models used to support their understanding do not facilitate a spatial understanding of the concept. Moreover, appropriate physical models are needed during the initial learning process of abstract, spatial concepts for long-term knowledge acquisition (Gilbert & Justi, 2016). Without this foundational support, the use of these concepts would not be possible as they require the conceptualization and manipulation of accurate, authentic mental models. Johnson-Laird (2012) proposes that mental models have a structure that matches the construct it is meant to represent, provide adequate possible and logical options for connections, and exist only in a realm of possibility and truth. Models created with these conditions can lead to inferences that aid in thinking and learning new concepts that follow similar guidelines presented in the mental representation. Once constructed, the model can be used as a means for comprehension as well as the ability to generalize. The purpose of the mental model is to be used for later recall and Kosslyn's (2005) theory of mental imagery suggests the similar concept that mental visualization can provide the individual with an image of an object, concept, or action without the presence of the stimuli. Support for this theory is provided by recognizing activation of particular regions of the brain that suggest similar activation in the presence of the physical image (Beilock et al., 2008; Calvo-Merino et al., 2005; James & Swain, 2011). This would be especially beneficial for learning STEM topics or concepts that utilize or incorporate visual images.

Mental imagery in conjunction with mental models can be manipulated in the visuospatial sketchpad of the working memory (Bruyer & Scailquin, 1998). As an image is better understood based on its component parts, the image can be stored in long-term memory. The accuracy of these mental images is dependent upon the propositions that construct the image as well as the spatial information of the parts or the context of the whole. These mental images can support an overall understanding of the visual information (stimuli) and can be quickly recalled. Hegarty and Stull (2012) explain visuospatial thinking as a concept similarly to Kosslyn. Visuospatial thinking is the way in which individuals know and understand relationships between and among persons, places, and things in the spatial plane in thought. An individual can leverage certain spatial competencies through visuospatial thinking to accomplish certain tasks such as mental rotation. Other spatial skills involve cause and effect in a system, navigation through a map, maze or mental landscape, and image transformation. Exercising spatial skills with mental images provide individuals with opportunities to mentally engage with images regardless of their physical presence. Mental images can be a starting point prior to working with a physical object or at a later date as the whole and its component parts can be recalled once stored in long-term memory.

In support of appropriate model construction and visual representation of spatial concepts, Stieff (2010) suggests imagistic reasoning should be taught and used by individuals seeking long-term conceptual understanding. This process would allow the individual to be able to mentally visualize the concept that can then be used for problem solving. Despite the spatial and conceptual benefits of imagistic reasoning, it is often overlooked in the science classroom which could suggest the widespread perceived difficulties of topics. The complex nature of developing appropriate and accurate mental models is also demonstrated by Zhao et al. (2020). These researchers found that when first presented with information that contained text and images, individuals would spend a greater amount of time and attention on text during their initial observation of the content. On a second viewing of the information, mental model construction had a greater emphasis on image processing. Dickmann et al. (2019) further support the importance of appropriate physical and visual representations of science content to support imagistic reasoning and mental model construction. Specifically, the visual model must represent the content and must be comprehended by the individual to be effective. These studies demonstrate the significant amount of time and attention necessary to achieve the essential imagistic reasoning for accurate mental model construction.

Schwartz and Black (1996) tested the effectiveness of mental images in the visuospatial sketchpad with a cognitive task to solve mental rotation problems involving gears. The researchers observed not only mental image and model use but also the importance of spatial ability as participants with varying levels of spatial ability approached and solved the problems differently. One subsequent observation made by the researchers on tasks of mental rotation is that participants often used their hands to mimic turning the gears suggesting a physical aid in the utilization of mental imagery. Context and spatial knowledge are important for mental image and mental model construction and also suggest a link to long-term memory. A general understanding of the topic and its structure and function is critical to creating appropriate mental models and conceptual understanding, without which individuals would struggle with many science concepts and spatial tasks.

While all STEM fields are spatial, it can be argued that the discipline with the highest density of both abstract and spatial concepts is chemistry (Harle & Towns, 2011). Conceptual understanding in chemistry requires spatial intelligence, a strong visuospatial sketchpad, and the appropriate mental models to bridge these two mental constructs. One particular literature review conducted by Wu and Shah (2004) examined research over four decades and determined that spatial ability is an important cognitive skill for learning not only concepts that are spatial in nature, but all chemistry topics. Wu and Shah also noted the transfer difficulties students have between two-dimensional representations in textbooks and three-dimensional conceptual representations of chemistry topics. The research cited supports the need for a foundation in spatial skills and its fostering through the manipulation of three-dimensional physical models especially for students who are considered to have weak spatial abilities.

In addition to science learning, the need for a scientifically literate workforce is essential for the continuing progression of society. Wai, Lubinski, and Benbow (2009) conducted a longitudinal study to demonstrate that spatial ability was a strong predictor for interest and success in the STEM disciplines. Using archived data from Project TALENT and aligning them with data from the Graduate Record Examination and the Study of Mathematically Precocious Youth, the researchers observed relationships between spatial abilities when participants were adolescents with their occupational field 11 years after the Project TALENT study was completed. With a particular focus on STEM occupations and degrees, Wai et al. observed that spatial ability was a greater indicator than mathematical or verbal ability of obtaining a degree in a STEM discipline as well as an occupation in a STEM field. When combined with results from more recent data pools, the researchers made several observations including: spatial ability is indicative of future study and work in STEM, it is important for individuals of all intellectual backgrounds, and it is not limited to mathematics and verbal skills and therefore should be measured in other fields. Similarly, Nagy-Kondor and Esmailnia (2022) found that spatial ability was important for all STEM disciplines with an emphasis on engineering. The researchers specifically measured individual's spatial ability as a tool for answering engineering problems and found that spatial ability had a direct impact on the participants' ability to complete the geometric engineering tasks.

STEM concepts are spatial and spatial thinking is critical for learning, understanding, and working with these concepts. The use of physical, three-dimensional, hands-on manipulatives support the construction of conceptual understanding as well as appropriate mental models (Preece et al., 2013). Research supports these notions and should inform not just modern science teaching but also scientific thinking in general. Simply put, STEM curricula and teaching should consider a shift from diagrammatic reasoning to imagistic reasoning as a means to support mental model construction and spatial thinking. However, the misapplication of touchscreen devices may inhibit an individual's ability for initial STEM concept development. Touted as an "all-in-one" learning device, the haptic screen that can present animations, simulations, three-dimensional graphic images, and static and dynamic images seems to be the perfect tool to support abstract, spatial, science concepts. Arguably, the limited haptic functionality of the touchscreen as well as the increased demands on cognitive load cloud the overwhelming optimism of this device as a teaching tool.

1.6 Attitudes

Attitudes to scientific inquiry have generally been linked to greater achievement, conceptual understanding, and science literacy (Bryan, Glynn, & Kittleson, 2011; Liu et al., 2022). Although achievement and conceptual understanding is important for science content knowledge, students' attitudes have also been used to place students in different courses as well as serve as a predictor for future careers in the sciences (Farenga & Joyce, 1999; Moore & Burrus, 2019; Saw et al., 2019). This correlation establishes attitudes and motivation in science as a critical marker for increasing scientific literacy to ensure that citizens are able to make informed decisions about every aspect of their lives (National Science Board, 2022). As such, studying attitudes towards science has been an area of focus for several decades with researchers such as Fraser (1978), Watson (1963), and Shulman and Tamir (1973), demonstrating not only a need to measure science attitudes because of the projected correlation between attitudes and achievement, but also a need to properly define and measure those attitudes. Fraser's (1978) *Test of Science-Related Attitudes (TOSRA)* was developed for this purpose of accurately measuring science attitudes which has led to greater attention to the topic and has influenced curricular changes such as those seen in the Next Generation Science Standards (NGSS Lead States, 2013).

Many different variables affect students' attitudes towards science. Aside from the general current negative perception of science perpetrated during the length of the Covid-19 Pandemic, attitudes and motivation toward science can be influenced by pedagogical techniques, perceived difficulty of the sciences, academic achievement, and quality of instruction (Almasri et al., 2021; Areepattamannil, Cairns, & Dickson, 2020; Cleveland, Olimpo, & DeChenee-Peters, 2017; Perez et al., 2019; Young et al., 2018). Each of these factors carries potential influence on the attitudes of students towards science and scientific inquiry. Additionally, pre-existing attitudes to science can influence and predict learning outcomes demonstrating the importance of maintaining relatively high attitudes toward science to achieve the goal of a more scientifically literate society.

However, the perceived methods of improving attitudes to scientific inquiry do not always correlate to real-world results. For example, pedagogical differences do not always lead to increases in attitudes and motivation. Inquiry-driven instruction does not necessarily improve attitudes when used exclusively, but rather, a mix of inquiry-driven and teacher-led instruction provided an increase in attitudes (Areepattamannil et al., 2020). Cleveland et al. (2017) also did not find differences in attitudes between two different active-learning environments. These studies provide insight into the complexity of influencing attitudes especially of adolescents. While pedagogical techniques are a factor in impacting attitudes, it is not the only variable.

Considering pre-existing attitudes to predict future success is similarly complex. For example, positive attitudes do not always correlate with career interest. Saw et al. (2019) found a significant correlation between attitude and career interest in mathematics but not in the sciences suggesting that interest in science does not necessarily influence long-term goals. These findings demonstrate the elusiveness of creating a large and persistent STEM workforce. Studies have also shown that one factor, grades, can significantly impact an individual's attitude towards the sciences and even leads to attrition in science courses and undergraduate majors when the desired grades are not achieved (Perez et al., 2019; Young et al., 2019).

Despite the complexities and misconceptions about attitudes, they remain an important cognitive construct. Attitudes can change over time, and attitudes can predict certain outcomes. Attitudes can also influence motivation which provides another layer of potential impact on science literacy and career interest. In order to observe more positive attitudes to science, large scale modifications of pedagogy, assessment, and perception are necessary to elicit the desired outcomes (National Science Board, 2022).

1.7 Conceptual Understanding

Conceptual understanding is a cognitive concept that has many definitions and is quite difficult to measure (Holme, Luxford, & Brandriet, 2015). Perkins (1998), along with his colleagues at Harvard, developed the notion that understanding is demonstrated through "performance perspective." In order to show conceptual understanding, an individual would need to interact with the concept in a variety of ways and "performances" that demonstrate the individual can think about, evaluate, and apply the content knowledge in a manner that extends beyond recall. Perkins notes that this type of assessment is not done in classrooms and true conceptual understanding is not often the goal.

For the majority of educational situations, conceptual understanding is measured by objective-based, multiple choice exams. However, this style of assessment rarely accounts for what an individual knows and understands about the material covered since many of these assessments measure recall. Objective-based, multiple choice exams can account for more than just recall and instead measure an individual's ability to apply knowledge or evaluate information (Putranta & Supahar, 2019). However, the skills required in this advanced exam-making are often difficult to develop and are not widespread (Villafañe et al., 2016). As a result, many exams use different styles of questions within one assessment to better measure conceptual understanding.

Using multiple question types is especially useful in measuring conceptual understanding of science concepts (Perkins, 1998). Since science content is often abstract, visual, and complex, a true measure of conceptual understanding does not rely solely on remembering definitions. Fiorella, Stull, Kuhlmann, and Mayer (2020) examined drawing and explanation assessments to measure students' learning. These two measures were chosen because of the highly-visual nature of the science concept in the study. The researchers found that explanations were a greater indicator of learning compared with drawings. This study is an important example that demonstrates that different assessment tools are necessary for different context and conceptual understanding goals. Fiorella et al. further suggest that for highly-visual content, written explanations support greater learning, understanding, and transfer capabilities.

Cognitive psychologists and science educators would agree that multiple assessment measures, especially the inclusion of learner-generated written answers, are stronger indicators to demonstrate conceptual understanding. For example, to measure conceptual understanding of the science concept of natural selection, Nehm and Schonfeld (2008) used three different measures to capture students' understanding of the topic. The researchers compared objective-based multiple choice questions to written responses and oral interviews. The results of their study determined that the best method for measuring conceptual understanding of this science concept was a new assessment tool that utilized all three styles of questioning. This conclusion was also identified by Villafañe et al. (2016) when designing appropriate measures of conceptual understanding for undergraduate biochemistry students. Rodriguez et al. (2018) make a case for the importance of varying assessment measures, specifically the use of symbols and graphs, to truly grasp conceptual understanding. They believe that responses to multiple choice questions do not adequately demonstrate conceptual understanding in the chemistry classroom but rather the use of open-response, application-based questions provide the appropriate measure. Meanwhile, Babilonia-Rosa, Kuo, and Oliver-Hoyo (2018) make the case for drawings as the appropriate measure of conceptual understanding when the instructional method was the use of three-dimensional physical models of noncovalent interactions in a biochemistry course.

Although there are inconsistencies in the most appropriate measure of conceptual understanding, the pervasive result is that multiple forms of measurement for conceptual understanding are necessary with an emphasis on learner-generated responses. This is especially important in the science disciplines due to their abstract and visual nature. The most effective measure of conceptual understanding will be dictated by the desired definition of conceptual understanding and the level of understanding required by the instructor. As such, each measure has value and should be considered when assessing an individual's conceptual understanding.

1.8 Technology and Cognitive Load

Educators continuously adapt to the changing educational technology landscape often to mediocre results. At fault are unrealistic expectations by administrators, inadequate budgets, unwillingness to change, and the inability to use the same resources annually. Regardless of the myriad of problems plaguing educational technology overall, the cognitive consideration of the students should be the primary concern. Technological hardware and software will always change, but the cognitive development of learners can be severely impacted by appropriate or inappropriate integration. There are a number of considerations to technology use including cognitive load, spatial intelligence, haptic interactions, and visual-auditory integration (Huk, 2006; Mayer & Moreno, 2003; Mayer, 2005). Mayer's cognitive theory of multimedia learning (2009) provides an appropriate framework to consider while designing educational technology. The major points of concern for the user are stimulating both auditory-verbal and visual-pictorial channels, understanding their limited processing capacity, and accessing and building upon prior knowledge (Horz & Schnotz, 2010). These considerations account for the minimization of extraneous load, which contains nonessential and often distracting information from the task at hand, management of intrinsic load, which refers to the necessary and often complex information required to understand the concept, and rationalization of germane load, which refers to motivation on the part of the individual to learn.

According to cognitive load theory, intrinsic cognitive load is affected by the complexity of the material to be learned, the relationship of the new material to prior knowledge, and the amount of material concurrently processed in the working memory. The goal of optimizing intrinsic cognitive load is to develop or extend schema acquisition and automation of activities related to the learning process (Sweller & Chandler, 1994). In doing so, the individual frees working memory. Extraneous cognitive load is considered to have a negative impact on learning that is related to the superfluous cognitive demands of poorly designed instructional materials or disorganized redundant non-integrated instruction (Chandler & Sweller, 1991; Sweller & Chandler, 1994; van Merrienboer & Ayres, 2005). In designing instruction, limiting extraneous cognitive load is imperative when introducing new material. By decreasing the factors relating to extraneous cognitive load, germane load can be increased by perfecting the design and implementation of activities and representations that are used in the learning process. The positive outcomes of increased germane load through streamlining instructional design provide the learner with the opportunity to use cognitive resources in order to increase schema acquisition and automation rather than devoting cognitive resources to other mental activities. Thus, if the germane load is increased it may help to refocus the learner's attention toward relevant information and prior knowledge. Together, the three types of processing demands, intrinsic load, extraneous load, and germane load, comprise the triarchic theory of cognitive load (Mayer & Moreno, 2010).

Cognitive load theory provides an explanation of how representations of physical objects are mentally constructed. In order to conceptualize and manipulate objects mentally, individuals rely on interactions with objects and process those interactions through auditory, visual, and/or haptic channels (Baddeley, 1999; Paivio, 1986). Mayer and Moreno (2010) argue that taking a multimedia approach to teaching abstract concepts may allow for successful knowledge acquisition provided extraneous load is reduced and intrinsic load managed. Understanding abstract concepts inherently have high intrinsic load. To manage intrinsic load, Mayer and Moreno suggest segmenting the material into easily understood chunks (Anderson, 1996), providing foundational knowledge that students can recall while learning more complicated concepts, and using the appropriate medium to relay the specific information required. Further, combining different mediums to support an individual's ability to visualize concepts can be useful as different technologies provide different functions.

It is important to consider, however, that new mediums may not always be applicable to the classroom. Parong and Mayer (2018) used virtual reality to transport users inside a cell. Although the situated environment provided participants with a visual that was previously inconceivable, the digital images that surrounded the visual field of the user were found to provide excessive extraneous load and ultimately led to lower conceptual understanding and knowledge acquisition when compared with a more traditional digital slideshow presentation. Parong and Mayer demonstrate the impact that cognitive load can have on knowledge acquisition, notwithstanding the novel visualization of abstract concepts.

Huk (2007) suggests that even with the introduction of advanced technological hardware, the emphasis on the formation of correct and applicable mental models through appropriate scaffolding and experiences is the primary goal of science education. Huk presents a counterintuitive argument that when learners are exposed to a conceptual model in which they previously have expert knowledge, they may experience an increase in cognitive load and ultimately hinder knowledge acquisition and long-term memory. Alternatively, novice learners benefit from variations of the same conceptual model to help support their overall understanding of the abstract concept. Repeated exposure to a concept across platforms can be beneficial, provided the pedagogical method is appropriately differentiated for the learner.

However, a primary concern is the cognitive load required to transfer from two-dimensions to three-dimensions. Dan and Reiner (2017) suggest that cognitive load, in general, is lower when the change in dimensionality does not exist. In particular, spatial STEM topics should be taught in three-dimensions since they exist in three-dimensions. The researchers did find that authentic experiences with technology have the ability to lessen the cognitive load that individuals experience when transferring experiences with a two-dimensional screen to three-dimensional concept formation. These findings present the potential for the use of two-dimensional technology for three-dimensional concepts at least in lessening cognitive load at certain transition points when engaged in the learning process.

The theoretical considerations of embodied cognition, transfer, spatial thinking, attitudes, conceptual understanding, and cognitive load present a possible path forward for advancing STEM education. The literature is replete with investigations that suggest three-dimensional, abstract concepts should be initially taught in the same dimension as the concept with hands-on manipulatives to lessen cognitive load, create adequate mental models, and bolster conceptual understanding (Kontra et al., 2015; Preece, et al., 2013; Sweller & Chandler, 1994). Technology is seen as a tool for enhancing education and learning. In many ways, technology has been a cognitive support for beneficial curricular enhancements. However, the cognitive, theoretical implications of instituting any change into curricula, technology or otherwise, should be seriously considered rather than blind implementation for the sake of modernity.

In conclusion, this study questions the psychological and cognitive impact of a condensed experiential space which has become more ubiquitous on the educational landscape due to a shift towards digital curriculum. The questions posed examine whether the two-dimensional screens of modern technology gradually distance both the person and the symbol too far from the referent to accurately reflect reality and therefore impact the initial concept developmental sequence in STEM.

1.9 Hypotheses

1. Students who manipulate physical models will demonstrate greater short-term conceptual understanding when compared with students who manipulate digital models when controlling for objective quiz pre-test scores, spatial ability scores, and attitude to scientific inquiry scores.
2. Students who manipulate physical models will demonstrate greater long-term conceptual understanding when compared with students who manipulate digital models when

controlling for objective quiz pre-test scores, spatial ability scores, and attitude to scientific inquiry scores.

3. Students who manipulate digital models will demonstrate greater degree of forgetting of conceptual understanding when compared with students who manipulate physical models when controlling for objective quiz pre-test scores, spatial ability scores, and attitude to scientific inquiry scores.
4. Students who manipulate physical models will demonstrate greater retention of learned material when compared with students who manipulate digital models when controlling for spatial ability scores, and attitude to scientific inquiry scores.

1.9.1 Operational Definitions

Conceptual understanding was measured by post-test objective quiz scores, model drawing scores, and explanation of model scores.

Spatial ability was measured by the score on the *Mental Rotation Test* from Vandenberg and Kuse (1978).

Attitudes to scientific inquiry was measured by the score on the *Attitude to Scientific Inquiry* subscale of the *Test of Science Related Attitudes (TOSRA)* by Fraser (1978).

Degree of forgetting of conceptual understanding was measured by a significant difference between post-test and delayed post-test scores where the means of the post-test scores are higher than the means of the delayed post-test scores on objective quiz scores, model drawing scores, and explanation of model scores.

Retention of learned material was measured by the significant difference between the means of the objective quiz pre-test and delayed post-test and no significant difference between the means of the objective quiz post-test and delayed post-test.

Chapter 2: Pilot Studies

The final design of this study was the result of several iterations of pilot studies that centered around the question of the effectiveness of touchscreen tablet computers to teach abstract STEM concepts. Through these iterations, the final study expanded to include a control group, extra variables of interest, and a more comprehensive short-term analysis. The pilot studies aided in the evolution of the final study and their design and implementation are worth noting.

2.1 Pilot Study 1

The first pilot study was conducted in an in-person, high school setting. Participants included 132 students across four intact classes of a tenth-grade biology course in an all-male high school. The biology course was designed to prepare students for the SAT Subject Test: Biology. The high school that participated in this study was unique in that all students are expected to conduct all schoolwork (notes, tests, studying, and communication) on a touchscreen tablet computer. The hypothesis of this study was that students who engage in physical manipulations using paper construction models would perform better on short-term and long-term assessments on the concept of the DNA molecule than students who engaged in manipulations of digital models.

All participants received equal amounts of instruction time, parallel course content, and were taught by the same instructor. The four classes of participants were divided into two groups with two classes in each group: non-digital instruction and digital instruction. Those two groups were further separated by instruction-assessment dyads: paper instruction with digital assessment, digital instruction with paper assessment, paper instruction with paper assessment, and digital instruction with digital assessment. All groups received the same lecture-style instruction of the DNA molecule prior to the start of the experiment. The digital instruction group participants were provided a digital model of DNA and were asked to create a complementary DNA strand by dragging and dropping nucleotides on the tablet screen. The digital animation, “Nucleic Acids: Build the Complementary DNA Strand!”, was accessed on the website ck12.org via their school-issued touchscreen tablet computer (Appendix A). Students were also allowed time to study information about DNA structure from the CK12 website (Appendix B). Students in the non-digital, paper instruction group were provided with paper image representations of phosphates, sugars, and nucleotide bases printed on it and were asked to cut out and assemble two complementary DNA strands by hand (Appendix C). They were also allowed time to study a paper printout of the online information about DNA structure that the digital group read (Appendix B).

Immediately after completion of their models, an objective-based quiz was given to assess short-term understanding of the concepts. The digital assessment groups completed the quiz on their tablets and the non-digital, paper assessment groups completed the quiz on paper. The objective-based quiz consisted of modified questions on the topic of DNA form and function from the textbook *Life: The Science of Biology 12th ed* (Appendix D). This textbook is a representative biology textbook used in introductory college biology courses. The objective-based quiz consisted of 10 multiple choice questions and was scored out of 10. A long-term assessment of deep knowledge was conducted five months after the experiment which asked participants to draw and label a model of DNA. The digital assessment groups drew a model on their tablets, and the non-digital, paper assessment groups drew a model on paper. The drawings were measured on a 15-point scale using a model drawing rubric (Appendix E). The components of the score consisted of sequence, ratio, two-dimensional representation, accuracy, and conceptual understanding. Sequence was scored based on appropriate matching of the nucleotide bases. Ratio was scored based on consistent distance between the DNA strands for the length of the model and an understanding that there are two backbones. Two-dimensional representation was scored based on the accuracy of a two-dimensional drawing to represent the three-dimensional concept. Accuracy was scored based on how accurately the model was constructed. Conceptual understanding was scored based on a demonstration of knowledge of the DNA molecule by the component parts of the model and a reasonable explanation of the model. Two raters provided scores for the long-hand response score. Consistency between the raters was high and a third rater was used if a score difference existed greater than one point.

An independent-samples t-test was conducted to compare digital method of instruction with paper method of instruction as measured by the results of the objective-based quiz. The short-term assessment revealed no significant difference between the digital group (M = 7.79, SD = 1.59) and the paper group (M = 7.71, SD = 1.46); $t(134) = 0.34$, $p = 0.37$. An ANOVA measuring the difference between the four dyads also did not show significance for the short-term assessment: $F(3, 132) = 0.76$, $p = 0.52$. However, the long-term drawing assessment demonstrated a significant difference in the scores for paper (M = 12.16, SD = 2.66) and digital (M = 10.84, SD = 3.18) conditions; $t(130) = 2.56$, $p = 0.006$. Further, Cohen's effect size value (d=.78) suggested a moderate to high practical significance in that non-digital instructional models demonstrated greater conceptual understanding of the DNA molecule and that non-digital models provided greater conceptual understanding over time.

An additional one-way ANOVA was used to measure the differences among the four dyads. There was a significant effect of the dyad on conceptual understanding as measured by the long-term drawing assessment: $F(3) = 3.47$, $p = 0.01$. A Tukey HSD identified a significant difference between paper instruction with paper assessment group (M = 12.54, SD = 2.89) and the digital instruction with digital assessment group (M = 10.33, SD = 2.74).

2.1.1 Lessons Learned

This pilot study provided a foundation for which to further explore how touchscreen tablet computers may impact conceptual understanding of three-dimensional abstract STEM concepts. The school in which the pilot study was conducted was specific in their emphasis on the use of tablets and therefore may not have provided results that were fully generalizable. The use of a single gender, however, was a recommendation from NSF research guidelines for selecting a purposeful sample as a means to remove the potential variable of gender (Institute of Education Sciences, 2013).

The questions posed in the objective-based quiz did not require any deep-level thinking which could have been a factor as to why there were no differences between the groups. Also, the quiz was only given as a short-term assessment after the experiment was completed. The quiz could have also been completed as a pre-test to measure the potential effectiveness of digital and non-digital instruction as well as to measure any prior knowledge on the topic.

The digital model used from the CK12 website was very limited. While this is partly a result of the limitations of touchscreen devices, the simulation required basic physical movements in comparison to the non-digital instruction group which constructed the paper models. A more comprehensive digital model that resembles the paper model was needed to better account for any differences between the digital and non-digital group.

Additionally, the primary reason for concern that the touchscreen devices may not be an appropriate medium in which to learn abstract concepts is the three-dimensional spatial component of the concept. As such, it became apparent that an individual’s spatial ability may impact their ability to learn and understand these concepts. The use of a spatial ability measure was needed.

This first pilot study provided preliminary evidence that learning with hands-on physical manipulatives may aid in greater long-term conceptual understanding compared with interactions with digital models on a touchscreen device as measured by an objective-based assessment and drawing measure. However, better models and more variables need to be considered to confidently support these results. The differences found among the dyads also provided support for similar method of instruction and assessment. In future studies, there will not be mixing of instruction and assessment.

2.2 Pilot Study 2

The second pilot study was conducted in a very different environment when compared with the first pilot study. The sample selected for this study consisted of 39 graduate students (8 males and 31 females) enrolled in a Master's-level psychology class which required students to participate in research for course credit. Additionally, this study took place during the fall of 2020 which required the format of the study occur by exclusively virtual means due to the Covid-19 Pandemic. As such, the decision was made to conduct this experiment asynchronously which allowed participants to complete the steps of the experiment at their leisure following a timeline for completion.

The first hypothesis of this study was that participants who built and manipulated physical, paper models would have better conceptual understanding than participants who built and manipulated digital models. Conceptual understanding was measured by objective-based test scores and model drawing scores. The second hypothesis was that regardless of the type of model built, participants would demonstrate greater conceptual understanding as measured by objective-based post-test scores. The final hypothesis was that participants who built and manipulated physical, paper models would have better conceptual understanding than participants who built and manipulated digital models when controlling for spatial ability.

The participants were first asked to complete a demographic information survey that provided information about their science course history as well as the materials they had at their respective locations to better assign groups (Appendix F). The limiting factor in determining if participants would be in a paper model or digital model group was the availability of a printer. Once the two groups were formed, each participant was asked to complete an objective-based quiz with questions on the topic of DNA form and function from the textbook *Life: The Science of Biology 12th ed* (Appendix G). This textbook is a representative biology textbook used in introductory college biology courses. The objective-based quiz consisted of nine questions and was scored out of nine. All participants also completed the *Mental Rotation Test* (Vandenberg & Kuse, 1978) to assess their spatial ability (Appendix H). The *Mental Rotation Test* consists of 20 items and is scored out of 20.

To begin the experiment, all participants were asked to read background information about the DNA molecule (Appendix B). The paper group was provided a document with image representations of phosphates, sugars, and nucleotide bases to print, cut out, and use to build their model (Appendix I). The digital model group was directed to the Gizmos "Building DNA" simulation where they would build their DNA model and take a screenshot to record their participation (Appendix J). After building their models, all participants completed the post-test objective-based quiz, which was the same as the pre-test (Appendix G). Three weeks after the experiment, the participants were asked to draw a DNA molecule. The drawings were measured on a 15-point scale using a model drawing rubric (Appendix E). The components of the score consisted of sequence, ratio, two-dimensional representation, accuracy, and conceptual understanding. Sequence was scored based on appropriate matching of the nucleotide bases. Ratio was scored based on consistent distance between the DNA strands for the length of the model and an understanding that there are two backbones. Two-dimensional representation was scored based on the accuracy of a two-dimensional drawing to represent the three-dimensional concept. Accuracy was scored based on how accurately the model was constructed. Conceptual understanding was scored based on a demonstration of knowledge of the DNA molecule by the component parts of the model and a reasonable explanation of the model. Two raters provided scores for the long-hand response score. Consistency between the raters was high and a third rater was used if a score difference existed greater than one point.

A dependent t-test was conducted to compare objective pre-test and post-test scores. There was a statistically significant difference in pre-test scores (M = 3.46, SD = 2.25) and post-test scores (M = 5.31, SD = 2.08); $t(38) = 5.53$, $p < .001$. Further, Cohen's effect size value (d = .84) suggested a moderate to high practical significance. While these findings suggest that learning occurred in both groups, there were no significant differences found between model drawing scores of the digital group (M = 6.89, SD = 3.97) and the paper group (M = 6.44, SD = 2.86); $t(36) = 0.42$, $p = 0.68$. Additionally, no significant difference was found between post-test scores of the digital group (M = 5.09, SD = 2.33) and the paper group (M = 5.63, SD = 1.75); $t(36) = -0.82$, $p = 0.42$.

A one-way ANCOVA was used to measure if spatial ability was a covariate for the conceptual understanding as measured by the drawing score. Spatial ability was not significantly related to the model drawing score: $F(1, 36) = 0.79$, $p = 0.33$. There was also no significant effect of the group on the model drawing score after controlling for the effect of the spatial score:

$F(1, 36) = 0.18$, $p = 0.68$. An additional one-way ANCOVA was used to measure if spatial ability was a covariate for the conceptual understanding as measured by the objective-based post-test score. Spatial ability was significantly related to the objective-based post-test score: $F(1, 36) = 4.87$, $p = 0.03$. However, there was no significant effect of the group on objective-based post-test score after controlling for the effect of the spatial score: $F(1, 36) = 0.79$, $p = 0.37$.

Overall, the results of this pilot study suggest that model type did not impact conceptual understanding through the objective-based quiz or drawing. However, model building did have an impact on conceptual understanding through the objective-based quiz as the mean for the post-test score was significantly greater than the mean of the pre-test score. Spatial ability was found to be related to objective-based post-test score but not the model drawing scores. Spatial ability was also not a significant covariate for either model drawing score or objective-based post-test score.

2.2.1 Lessons Learned

This study continued the methodology from the first pilot study with some clear differences, both planned and unexpected. Conducting this pilot study virtually and asynchronously posed challenges to the implementation of the planned design; however, the experience provided confidence in the changes made from the first pilot study. The pre-test/post-test repeated measures design is helpful in measuring if learning was achieved. The new digital model used in this study is a better match with the paper model than the digital model used in the first pilot study. The digital model was also an improvement over the digital model in the first pilot study due to the inclusion of component parts and greater flexibility of model construction. Also, the spatial ability test used in this iteration provided some interesting results, although not necessarily the expected outcomes.

The major concern with this study is the sample. The participants have had much more experience with science topics than adolescent science learners, and they were fairly homogenous in their spatial ability scores. Adolescents may have more variability in their spatial ability scores which may make this measure a more salient part of the study. Additionally, it seems apparent that this study is best carried out in-person and in a school setting. This would allow for more accountability in terms of completion of each part of the experiment. Although spatial ability was not found to be a significant covariate for all measures in this pilot study, the *Mental Rotation Test* will continue to be used as these results may vary based on sample.

The long-term response of drawing the DNA molecule was carried over from the initial pilot study. While this has provided greater depth to the question of measuring conceptual understanding, it was evident from this pilot study that a repeated measure of a drawing may prove helpful in determining conceptual growth and retention. Additionally, a written explanation describing the drawing will also be added. This change stems from Fiorella et al. (2020) who suggest that drawing diagrams based on diagram-heavy material may not provide an appropriate measure of learning but rather the use of learner-generated explanations to demonstrate conceptual understanding.

2.3 Pilot Study 3

The third pilot study was also conducted through asynchronous digital delivery. The sample selected for this study consisted of 15 graduate students (2 males and 13 females) enrolled in a Master's-level psychology class which required students to participate in research for course credit. Additionally, this study took place during the spring of 2021 which required the format of the study occur by exclusively virtual means due to the Covid-19 Pandemic. As such, the decision was made to conduct this experiment asynchronously which allowed participants to complete the steps of the experiment at their leisure following a timeline for completion. The lessons learned from the first two pilot studies helped to streamline the delivery of the materials and collection of the data. While the final design of the study does not incorporate asynchronous work, this pilot study provided support for the changes made to the final study as a result of the first two studies.

The first hypothesis of this study was that participants who built and manipulated physical, paper models would have better short-term conceptual understanding than participants who built and manipulated digital models controlling for objective-based pre-test score and spatial ability. Conceptual understanding was measured by short-term objective-based test scores, model drawing scores, and explanation of model scores. The second hypothesis was that participants who built and manipulated physical, paper models would have better long-term conceptual understanding than participants who built and manipulated digital models controlling for objective-based pre-test score and spatial ability. Conceptual understanding was measured by long-term objective-based test scores, model drawing scores, and explanation of model scores. The third hypothesis was that participants who manipulated physical models would demonstrate greater retention of conceptual understanding when compared with students who manipulate digital models when controlling for objective-based pre-test scores and spatial ability scores. Retention of conceptual understanding is measured by the difference between short-term and long-term scores on objective-based test scores, model drawing scores, and explanation of model scores. The final hypothesis was that participants who manipulated physical models would demonstrate greater learning gains when compared with students who manipulate digital models. Learning gains are measured by the average differences among the pre-test, short-term, and long-term objective-based test scores.

The experiment was split into three parts: before the experiment, the experiment, and after the experiment. Before the experiment, participants completed a demographic questionnaire (Appendix F), the pre-test objective-based quiz (Appendix G), and the *Mental Rotation Test* (Vandenberg & Kuse, 1978) (Appendix H). The *Mental Rotation Test* consists of 20 items and is scored out of 20.

The demographic information was used primarily to split the participants into paper model or digital model groups. The determining factor was the participant's ability to print the paper model worksheet to cut out the component parts. The objective-based quiz consisted of questions on the topic of DNA form and function from the textbook *Life: The Science of Biology 12th ed* (Appendix G). This textbook is a representative biology textbook used in introductory college biology courses. The objective-based quiz consisted of nine questions and was scored out of nine. The experiment portion asked all participants to read background information about the DNA molecule (Appendix B). The digital model group was directed to the Gizmos "Building DNA" simulation where they would build their DNA model and take a screenshot to record their participation (Appendix J). After building their models, all participants completed the post-test objective-based quiz, which was the same as the pre-test, and completed a drawing of the DNA molecule with an explanation of the drawing. Three weeks after the experiment, the participants were asked to once again draw a DNA molecule and provide an explanation of the drawing. The drawings were scored based on the Model Drawing Rubric (Appendix K). The scores were based on inclusions of component parts (nucleotide bases, deoxyribose sugar, phosphate, two strands, and pairing of bases) and spatial representation (accurate two-dimensional representation of three-dimensional concept, consistent spacing of strands, and double helix morphology of molecule). The model drawings were scored out of 48. The explanations of the model were scored based on the Learner-Generated Written Explanation Response Rubric (Appendix L). The scores were based on a scale from attempted to exemplary in which the explanations provided evidence of conceptual understanding through the use of academic language. The explanations of the model were scored out of 8. Two raters provided scores for the long-hand response score for inter-rater reliability.

Comparisons were made between the two groups in terms of objective-based test scores as well as their model drawings and written explanations of their models. In addition, spatial ability and objective-based pre-test scores were measured as a covariate for objective-based scores, model drawing scores, and explanation of model scores. Descriptive statistics can be found in Appendix M.

For the first hypothesis, the following three ANCOVAs were conducted to measure short-term conceptual understanding. The covariate, spatial ability, was significantly related to the objective-based post-test score: $F(1, 11) = 5.45$, $p = 0.04$. The objective-based pre-test covariate was also significantly related to the objective-based post-test score: $F(1, 11) = 10.47$, $p = 0.008$. However, there was no significant effect of the type of model on the objective-based post-test score after controlling for the effect of spatial ability and objective-based pre-test score: $F(1, 11) = 0.47$, $p = 0.51$. The covariate, spatial ability, was not significantly related to the short-term model drawing score: $F(1, 11) = 1.70$, $p = 0.22$. The objective-based pre-test covariate was also not significantly related to the short-term model drawing score: $F(1, 11) = 2.11$, $p = 0.17$. There was also no significant effect of the type of model on the short-term model drawing score after controlling for the effect of spatial ability and objective-based pre-test score: $F(1, 11) = 3.45$, $p = 0.09$. The covariate, spatial ability, was not significantly related to the short-term explanation of model score: $F(1, 11) = 3.75$, $p = 0.08$. The objective-based pre-test covariate was also not significantly related to the short-term explanation of model score: $F(1, 11) = 3.25$, $p = 0.10$. However, there was a significant effect of the type of model on the short-term explanation of model score after controlling for the effect of spatial ability and objective-based pre-test score: $F(1, 11) = 6.19$, $p = 0.03$.

For the second hypothesis, the following three ANCOVAs were conducted to measure long-term conceptual understanding. The covariate, spatial ability, was not significantly related to the long-term objective-based score: $F(1, 11) = 1.12$, $p = 0.31$. The objective-based pre-test covariate was significantly related to the long-term objective-based score: $F(1, 11) = 8.46$, $p = 0.01$. However, there was no significant effect of the type of model on the long-term objective-based score after controlling for the effect of spatial ability and objective-based pre-test score: $F(1, 11) = 0.39$, $p = 0.54$. The covariate, spatial ability, was not significantly related to the long-term model drawing score: $F(1, 11) = 0.49$, $p = 0.49$. The objective-based pre-test covariate was also not significantly related to the long-term model drawing score: $F(1, 11) = 1.61$, $p = 0.23$. There was also no significant effect of the type of model on the long-term model drawing score after controlling for the effect of spatial ability and objective-based pre-test score: $F(1, 11) = 3.80$, $p = 0.08$. The covariate, spatial ability, was not significantly related to the long-term explanation of model score: $F(1, 11) = 0.08$, $p = 0.78$. The objective-based pre-test covariate was also not significantly related to the long-term explanation of model score: $F(1, 11) = 4.39$, $p = 0.06$. There was also no significant effect of the type of model on the long-term explanation of model score after controlling for the effect of spatial ability and objective-based pre-test score: $F(1, 11) = 0.0003$, $p = 0.99$.

For the third hypothesis, the following three ANCOVAs were conducted to measure retention of conceptual understanding. The covariate, spatial ability, was not significantly related to the difference in short-term and long-term objective-based score: $F(1, 11) = 1.73$, $p = 0.21$. The objective-based pre-test covariate was also not significantly related to the difference in short-term and long-term objective-based score: $F(1, 11) = 0.08$, $p = 0.78$. There was no significant effect of the type of model on the difference in short-term and long-term objective-based score after controlling for the effect of spatial ability and objective-based pre-test score: $F(1, 11) = 0.002$, $p = 0.96$. The covariate, spatial ability, was not significantly related to the difference in short-term and long-term model drawing score: $F(1, 11) = 0.70$, $p = 0.42$. The objective-based pre-test covariate was also not significantly related to the difference in short-term and long-term model drawing score: $F(1, 11) = 0.40$, $p = 0.54$. There was also no significant effect of the type of model on the difference in short-term and long-term model drawing score after controlling for the effect of spatial ability and objective-based pre-test score: $F(1, 11) = 0.36$, $p = 0.56$. The covariate, spatial ability, was not significantly related to the difference in short-term and long-term explanation of model score: $F(1, 11) = 1.71$, $p = 0.22$. The objective-based pre-test covariate was also not significantly related to the difference in short-term and long-term explanation of model score: $F(1, 11) = 0.10$, $p = 0.75$. There was also no significant effect of the type of model on the difference in short-term and long-term explanation of model score after controlling for the effect of spatial ability and objective-based pre-test score: $F(1, 11) = 3.85$, $p = 0.08$.

For the final hypothesis, the following repeated measures ANOVAs were conducted. The objective-based test score was statistically significantly different at the three different time points: $F(2, 28) = 6.20$, p = 0.006, generalized eta squared = 0.102. Post-hoc analyses with a Bonferroni adjustment revealed that the pairwise difference between pre-test and long-term objective-based test scores was statistically significantly different (p = 0.016). The comparison between pre-test and post-test (p = 0.211) and post-test and long-term (p = 0.381) were not significant. A two-way repeated measures ANOVA was performed to evaluate the effect of type of model building over time on objective-based test score. There was not a statistically significant interaction between type of model and time on objective-based test score: $F(2, 39) = 0.24$, $p = 0.79$, generalized eta squared = 0.012. There were no statistically significant main effects of type of model ($F(1,39) = 0.282$, $p = 0.6$) and time ($F(2,39) = 2.15$, $p = 0.13$).

The results of this study suggest the following conclusions. The covariates of spatial ability and objective-based pre-test are related to short-term objective-based post-test. Additionally, the covariate of objective-based pre-test was related to long-term objective-based test. Participants in the paper model group were better able to explain their model in the short term when controlling for spatial ability and objective-based pre-test scores. Finally, participants in both groups scored higher on the objective-based test at the long-term time point when compared with the pre-test time.

2.3.1 Lessons Learned

This third pilot study provided confidence in the final design of the study. The directions disseminated were clear and concise. Participants were able to easily access all materials and participate in a timely fashion. Due to the asynchronous nature, the clarity in instructions was gauged by the limited number of email correspondences with the participants. Although the pilot study ran without difficulty, there were a few considerations that led to differences between this third pilot study and the final study design.

The four hypotheses as well as the subsequent data analysis supported the development of the hypotheses and data analyses of the dissertation study. This pilot study provided an opportunity to run the data as separate ANCOVAs for each of the measures of conceptual understanding: objective-based test score, model drawing score, and explanation of model score. A repeated measures ANOVA was also used to measure the differences in objective-based test scores at three different time periods across the type of model groups. The limitations of the sample size (N = 15) may have impacted the results of this study and may suggest the limited significance between groups for the different dependent variables.

As with the second pilot study, it was evident that the participants' prior experiences were a factor in the results of the current study based on their responses to the demographic survey. As graduate students, many of the participants had extensive knowledge of the topic while others relied on knowledge attained several years prior to the study and therefore were not actively engaged in learning STEM topics. The participants of the dissertation study were adolescents to better reflect the population that is impacted by the large number of digital instruction instances in a STEM classroom. The adolescent population is also helpful in measuring the impact of model building on initial learning experiences since participation in the study should have been the first in-depth lesson on DNA the participants had experienced.

Additionally, a need for a control group arose when considering if model building in general is a potential factor in conceptual understanding. Since the traditional method of initial STEM teaching does not typically incorporate models, a control group that follows traditional instruction will be used. Moreover, the question of the measurement of conceptual understanding was determined to be three-fold: objective-based quiz score, model drawing score, and explanation of model score. In order to measure conceptual understanding at both a short-term and long-term interval, a third repeated measure of the objective-based quiz will be given at the long-term measure.

Paper models were used in the three pilot studies. Although a different paper model was considered after the first pilot study, the use of the paper model continued in the second and third pilot studies due to the experimental limitations imposed by the virtual environment. However, the dissertation study was conducted in person and in schools. The decision was made to use a three-dimensional, physical model made of foam rather than paper to provide students with a more haptic, three-dimensional experience with the model. One's modality is suggested to be beneficial for student conceptual understanding.

Finally, it became apparent that participants who had interests in science, as observed by the number of science courses they have taken during their academic careers, performed better on the pre-test, post-test, drawings, and explanations. Since the participants in the dissertation study had a narrow history of science courses to observe, the inclusion of an attitudes survey was decided to determine if positive attitudes towards science inquiry is a potential covariate for the adolescent population in the final study.

Chapter 3: Methodology

3.1 Participants

The sample for the present study consisted of six in-tact classes of 9^{th}-grade Living Environment students. Each group of three classes are part of the same high school and taught by the same Living Environment teacher. Two schools with approximately equivalent demographics and equivalent number of males and females were recruited for this experiment. The schools are located in a large, diverse, suburban area and follow the state's curriculum and graduation requirements. The sample selected for the study consisted of 161 9^{th}-grade students. Due to attrition, a total of 148 participants participated throughout the entirety of the study. Each of the three classes in each school were randomly assigned to one of the three groups of the study. The Living Environment scope and sequence follows the Next Generation Science Standards (NGSS).

The content of the study was designed to cover and align with the content taught in the "MS-LS3 Heredity: Inheritance and Variation of Traits" as outlined by the NGSS as well as Key Idea 2 "Organisms inherit genetic information in a variety of ways that result in continuity of structure and function between parents and offspring" from the Living Environment Core Curriculum as outlined by New York State. This alignment is to ensure that the participants' learning of the curriculum would not be impacted by the study.

Two treatment groups and one control group were used to measure differences in conceptual understanding. The physical model group received all aspects of the experiment with a physical model. The digital model group received all aspects of the experiment with a digital model. The control group received all aspects of the experiment without the support of a physical or digital model.

3.2 Instruments

The instruments used in this study included an objective-based question survey, mental rotation test, attitudes to scientific inquiry survey, constructed response questions, physical and digital models, worksheet, and video.

3.2.1 Objective Quiz

The objective-based survey instrument, or objective quiz, contained multiple-choice questions developed from past, standardized Living Environment exams given to demonstrate completion and understanding of the Living Environment curriculum (Appendix N). Questions were chosen to demonstrate conceptual understanding of the shape and structure of the DNA molecule. Participants were expected to answer the questions based on the content presented in the informational video at the beginning of the experiment that was reinforced through the model building exercise. The video provided the academic language and specific vocabulary while the model construction allowed interaction with component parts, complementary pairing, and spatial arrangement to provide the knowledge necessary to answer the questions. This objective quiz was used as the pre-test as well as post-test and delayed post-test measures of objective knowledge. The objective quiz contained 10 multiple choice questions and was scored out of 10.

3.2.2 Mental Rotation Test

The *Mental Rotation Test* is an instrument developed by Vandenberg and Kuse (1978). The *Mental Rotation Test* has been extensively field tested and demonstrates significant internal consistency (Kuder-Richardson 20 = .88), a test-retest reliability of .83, and consistent results based on gender and age. The instrument presents participants with an initial image of a continuous string of 10 cubes bent and arranged in different orientations in three-dimensional space. This initial image is then matched by the participant to two other images that are exactly the same except for being presented at a different angle. The other two images in the series of four are not the same. There are three practice questions followed by two parts of 10 questions each. Participants had three minutes to complete part one and three minutes to complete part two. A point was given for a question when both of the correct responses were marked. The *Mental Rotation Test* contained 20 questions and was scored out of 20 (Appendix H).

3.2.3 Test of Science-Related Attitudes (TOSRA)

The *Test of Science-Related Attitudes (TOSRA)*, developed by Fraser (1978), measures seven distinct science-related attitudes among secondary school-aged students. The current study used one of the 10-question scales, *Attitude to Scientific Inquiry* (Appendix O). The *TOSRA* has been extensively field tested and reports high levels of reliability for each of its scales, and low intercorrelations among its scales. Specifically, the *Attitude to Scientific Inquiry* scale demonstrated high reliability in international and United States cross-validation studies reporting alpha reliabilities of .82 and .84. Further, statistical analysis identified test-retest reliability of .79 and a mean correlation with other scales of .13. The scale contains 10 items scored one to five with a minimum and maximum score possible from five to 50. There are five positively worded items and five negatively worded items, where responses strongly agree (SA), agree (A), not sure (N), disagree (D), strongly disagree (SD) are scored 5, 4, 3, 2, 1. The negative item responses SA, A, N, D, SD are scored 1, 2, 3, 4, 5, respectively.

3.2.4 Constructed Response

The constructed response consisted of the model drawing and written model explanation (Appendix P). The model drawing task asked participants to hand draw a model of DNA. The explanation of the model task asked participants to write an explanation of their model. The constructed response was collected both immediately after treatment and two months later. The drawings were scored based on the Model Drawing Rubric (Appendix K). The overall model drawing score consisted of the factors of component parts of the molecule and spatial representation of the DNA molecule. Component parts was measured by the inclusion of the nucleotide bases, deoxyribose sugars, phosphates, two strands of DNA, and appropriate pairing of the nucleotide bases. Each component part accounted for up to six points for a total of 30 points for component parts. Spatial representation was measured by the inclusion of an accurate double helix shape, equivalent and consistent distance between the two strands of DNA, and representation of the component parts in a manner that demonstrates understanding of a three-dimensional structure. Each spatial representation measure accounted for up to six points for a total of 18 points. The model drawing score was measured out of a combined score of component parts and spatial representation for a total of 48. The explanations of the model were scored based on the Learner-Generated Written Explanation Response Rubric (Appendix L). The scores were based on a scale from attempted to exemplary in which the explanations provided evidence of conceptual understanding through the use of academic language. The explanations of the model were scored out of 8. Two raters provided scores for both the model drawing and written explanations for inter-rater reliability. Differences in more than one point were assigned to a third rater to increase reliability.

3.2.5 Physical Model

The physical model instrument is the model that participants in the physical model group created. The DNA model used was acquired from 3D Molecular Designs. The model consists of foam representations of phosphates, sugars, and nucleotide bases that participants used to create a model of DNA by snapping them together in the appropriate sequence (Appendix Q). The physical model was able to be twisted into the double helix shape of the DNA molecule once the foam pieces were appropriately joined together.

3.2.6 Digital Model

The digital model instrument is the model that participants in the digital model group created. The “Building DNA” animation is provided by the website Gizmos. Accessed on internet-enabled touchscreen devices, the simulation provides digital phosphates, sugars, and nucleotide bases. Participants used these components to create a digital model of DNA by dragging and dropping them together in the appropriate sequence (Appendix J). Correct alignment of the component parts was required to build the model.

3.2.7 Worksheet

The worksheet instrument was used by the participants in the control group. The worksheet provided questions about the shape and component parts of the DNA molecule. The worksheet also provided a DNA sequence in which the participants needed to write the complementary DNA strand by appropriately writing the letter representation for the nucleotide base that pairs with the strand given (Appendix R). The worksheet was framed as an investigation to determine the difference between DNA and RNA viruses based on the component parts and spatial arrangement of the DNA molecule.

3.2.8 Video

An instructional video was created to provide information about the DNA molecule to the participants (Appendix S). The video contains both voice-over and live-action sequences to provide auditory and visual explanations. Images chosen for the video were accurate representations of the DNA molecule with regard to component parts and spatial orientation and representation. The content covered in the video is content that is covered in a typical DNA lesson in the Living Environment curriculum. Additionally, all of the questions on the objective quiz measure could have been answered from the information in the video. All participants watched the same video to ensure equal and consistent instruction among the three groups.

3.3 Procedure

At the beginning of the class period, participants in this experiment were told that they would be learning about the DNA molecule in class on the day the experiment took place. All participants were initially given the objective quiz to measure their prior knowledge (Appendix N), the *Mental Rotation Test* to measure their spatial ability (Appendix H), and the *Attitudes to Scientific Inquiry* survey (Appendix O). Participants were randomly assigned a number based on the class roster which was pre-written on the objective quiz, the *Mental Rotation Test*, and the *Attitude to Scientific Inquiry* survey. Recordings of numbers on all parts of the experiment were used to provide anonymity for the participants and to match the data across pre-test, post-test, and delayed post-test. All three measures were completed on paper.

After the initial baseline surveys, all participants watched a four-minute instructional video consisting of a verbal and visual explanation of the DNA molecule's shape and function (Appendix S). This video was presented on the SMART Board® at the front of the classroom where all participants could watch and listen to the video. At the end of the video, participants began working on the activity as dictated by their treatment group for the next 20 minutes.

The physical model group began to construct the physical model of DNA with the component foam pieces of phosphates, sugars, and nucleotide bases (Appendix Q). The participants were presented with a DNA sequence on an instructional sheet and created a model of a double-strand of DNA based on that sequence. The foam components were fit together. After building the model, participants were able to twist the model to visualize the double helix structure of the DNA molecule. Participants stopped their model construction at the end of the 20-minute period.

The digital model group utilized Apple iPads® to access the Gizmos simulation "Building DNA" for DNA molecule construction (Appendix J). Participants were presented with a DNA sequence on an instructional sheet and created a model of a double-strand of DNA based on that sequence. The component pieces of the model, phosphates, sugars, and nucleotide bases, were dragged and dropped from the right side of the screen to the center of the screen. Participants took a screenshot of their completed model at the end of the 20-minute period.

Participants in the control group did not construct any model. Rather, they completed a worksheet that asks questions about the DNA molecule based on the information presented in the video to investigate the differences between DNA and RNA viruses (Appendix R). Participants were given a DNA sequence on the worksheet and asked to write the complementary base pairings for each of the bases in the given sequence. A word search of DNA vocabulary was also included on the sheet. The worksheet was completed by the end of the 20-minute period.

Following the treatment, all participants were asked to complete the objective quiz they completed at the beginning of the class period. Participants' randomly assigned numbers were pre-written on the paper. In addition to answering the objective quiz, participants were asked to draw a molecule of DNA and explain their drawing on the paper provided (Appendix P). Participants' randomly assigned numbers were pre-written on this paper. The constructed responses were collected at the end of the class period. Together, the objective quiz, drawing, and explanation completed immediately after treatment comprise the post-test timepoint.

Two months later without any additional formal teaching about DNA, the participants were asked once again to complete the objective quiz, draw a molecule of DNA, and explain their drawing on the paper provided. Participants' randomly assigned numbers were pre-written on this paper. The responses on these measures were collected. Together, the objective quiz, drawing, and explanation completed two months after treatment comprise the delayed post-test timepoint.

3.4 Statistical Analysis

A cluster randomized controlled design was used for the randomly assigned groups since there were six intact classes of participants across two schools. Conceptual understanding is a complex compound construct. In order to assess the impact method of instruction had on conceptual understanding, it is essential to consider the three components of conceptual understanding as measured in this study: objective quiz score, DNA model drawing score, and explanation of DNA model score. One of these measures does not adequately cover the complexities of conceptual understanding and therefore they must be considered collectively although as separate measures.

One-way ANCOVAs were used to measure if there are differences in conceptual understanding among the physical model, digital model, and control groups when controlling for pre-test objective quiz scores, spatial ability scores, and attitude to scientific inquiry scores. ANCOVAs were used for each of the three measures of conceptual understanding: objective quiz scores, model drawing scores, and explanation of model scores. ANCOVAs were conducted for all post-test and delayed post-test measures.

A two-way mixed ANOVA was used to measure degree of forgetting for the treatment and control groups across post-test and delayed post-test timepoints. The two-way mixed ANOVA was conducted for each of the three measures of conceptual understanding (objective quiz, model drawing, and explanation of model) and controlled for pre-test objective quiz scores, spatial ability scores, and attitude to scientific inquiry scores.

A final analysis to measure retention of the learned material across pre-test, post-test, and delayed post-test used a two-way mixed measures ANOVA to assess differences in objective-based conceptual understanding across the treatment and control groups. Spatial ability and attitude to scientific inquiry were measured as covariates.

Chapter 4: Results

Overall descriptive statistics can be found in Appendix T. A correlation matrix of the covariates, post-test measures, and delayed post-test measures can be found in Appendix U.

For the first hypothesis, three ANCOVAs were conducted to model short-term conceptual understanding. For the first ANCOVA, objective quiz scores were compared across the three groups. The adjusted means and standard errors of the objective quiz scores of the three groups at post-test can be found in Figure 1 and Table 1. The results of the ANCOVA suggest there was no significant effect of the group on the objective quiz post-test score after controlling for the effect of spatial ability, objective quiz pre-test score, and attitude to science inquiry: $F(2, 152) = 2.79$, $p = 0.06$ (see Table 2 and Table 3). The partial eta squared effect size value (.04) suggested a small practical significance (see Table 2). The covariate, spatial ability, was significantly related to the objective quiz post-test score: $F(1, 152) = 13.40$, $p < 0.001$ (see Table 2). The objective quiz pre-test covariate was also significantly related to the objective quiz post-test score: $F(1, 152) = 22.94$, $p < 0.001$ (see Table 2). The attitudes to science inquiry covariate was not significantly related to the objective quiz post-test score: $F(1, 152) = 0.67$, $p = 0.41$ (see Table 2).

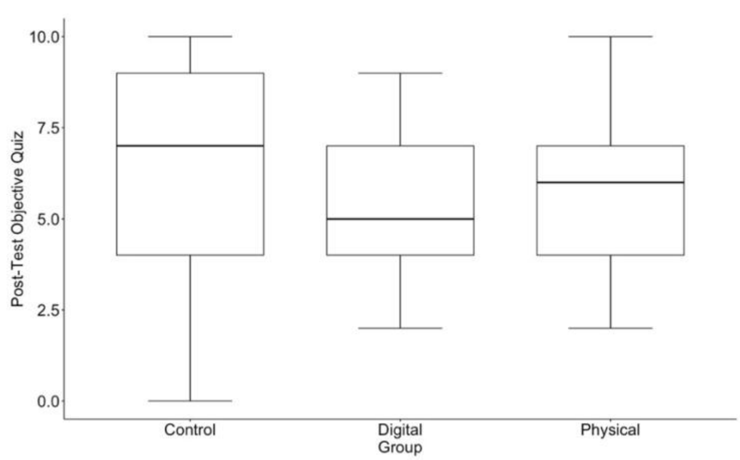


Figure 1: Box Plot of Post-Test Objective Quiz Scores

Table 1: Table of Adjusted Means for Post-Test Objective Quiz Scores across Treatment Groups

	Adjusted Mean	SE	n
Physical	5.89	0.27	60
Digital	5.59	0.28	53
Control	6.40	0.29	48

Table 2: ANCOVA of Post-Test Objective Quiz Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

	Sum of Squares	*df*	*F*	*p*	*Partial eta squared*
Group	23.1	2	2.79	0.06	0.04
Classroom	16.8	3	1.36	0.26	0.03
Pre-Test	94.7	1	22.94	< 0.001	0.13
Spatial	55.3	1	13.40	< 0.001	0.08
Attitudes	2.8	1	0.67	0.41	0.004
Residuals	627.7	152			

Table 3: Bonferroni Corrections of Post-Test Objective Quiz Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

	Estimate	*Std. Error*	*Statistic*	*p*
Control – Digital	0.81	0.41	2.00	0.05
Control – Physical	0.52	0.40	1.30	0.196
Digital – Physical	-0.29	0.39	-0.76	0.45

The adjusted means and standard errors of the drawing scores of the three groups at post-test can be found in Figure 2 and Table 4. The results of the ANCOVA suggest a significant effect of group on the post-test model drawing score after controlling for the effect of spatial ability, objective quiz pre-test score, and attitudes to science inquiry: $F(2, 151) = 7.53$, $p < 0.001$. The partial eta squared effect size value (.09) suggested a moderate practical significance (see Table 5). The covariate, spatial ability, was significantly related to the post-test model drawing score: $F(1, 151) = 6.63$, $p = 0.01$ (see Table 5). The objective quiz pre-test covariate was also significantly related to the post-test model drawing score: $F(1, 151) = 4.91$, $p = 0.03$ (see Table

5). The attitudes to science inquiry covariate was not significantly related to the post-test model drawing score: $F(1, 151) = 0.00$, $p = 0.99$ (see Table 5). Post hoc comparisons using Bonferroni corrections, as shown in Table 6, indicated that the adjusted mean score for the physical group (M = 11.4, SE = 0.74) was significantly different than the digital group (M = 7.96, SE = 0.77) ($p = 0.002$). Additionally, the physical group (M = 11.4, SE = 0.74) was significantly different than the control group (M = 7.06, SE = 0.81) ($p < 0.001$). The digital group (M = 7.96, SE = 0.77) did not significantly differ from the control group (M = 7.06, SE = 0.81) ($p = 0.424$).

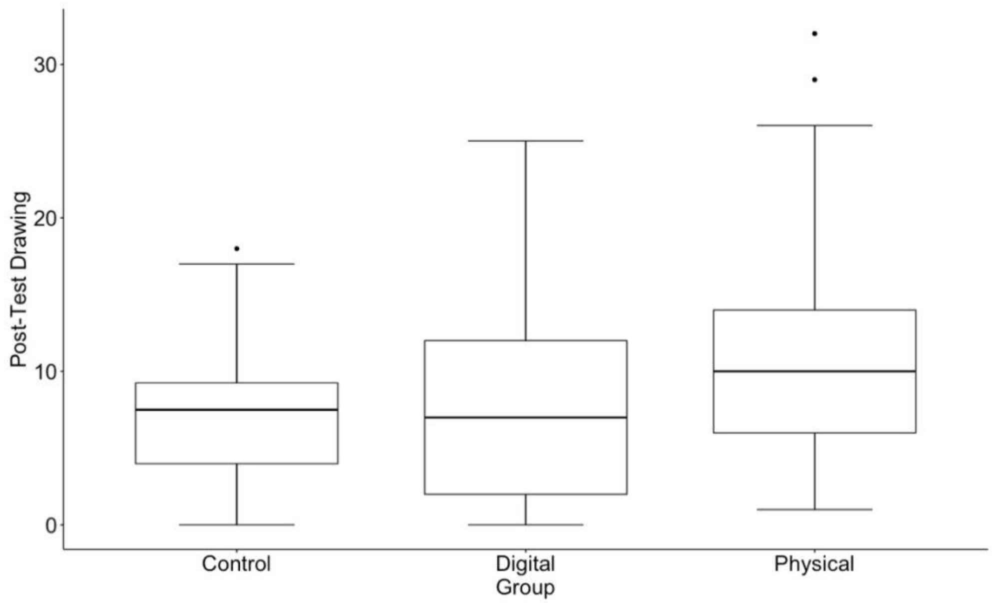


Figure 2: Box Plot of Post-Test Drawing Scores

Table 4: Table of Adjusted Means for Post-Test Drawing Scores across Treatment Groups

	Adjusted Mean	SE	n
Physical	11.4	0.74	59
Digital	7.96	0.77	53
Control	7.06	0.81	48

Table 5: ANCOVA of Post-Test Drawing Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

	Sum of Squares	*df*	*F*	*p*	*Partial Eta Squared*
Group	476	2	7.53	< 0.001	0.09
Classroom	98	3	1.033	0.38	0.02
Pre-Test	155	1	4.91	0.03	0.03
Spatial	209	1	6.63	0.01	0.04
Attitudes	0.0	1	0.00	0.99	0.00
Residuals	4768	151			

Table 6: Bonferroni Corrections of Post-Test Drawing Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

	Estimate	*Std. Error*	*Statistic*	*p*
Control – Digital	-0.90	1.12	-0.80	0.424
Control – Physical	-4.36	1.10	-3.96	< 0.001
Digital – Physical	-3.46	1.07	-3.23	0.002

The adjusted means and standard errors of the explanation scores of the three groups at post-test can be found in Figure 3 and Table 7. The results of the ANCOVA suggest a significant effect of group on the post-test explanation of model score after controlling for the effect of spatial ability, objective quiz pre-test score, and attitudes to science inquiry: $F(2, 151) = 8.51$, $p < 0.001$. The partial eta squared effect size value (.10) suggested a moderate practical significance (see Table 8). The covariate, spatial ability, was not significantly related to the post-test explanation of model score: $F(1, 151) = 3.46$, $p = 0.06$ (see Table 8). The objective quiz pre-test covariate was significantly related to the post-test explanation of model score: $F(1, 151) = 11.77$, $p < 0.001$ (see Table 8). The attitudes to science inquiry covariate was not significantly related to the post-test explanation of model score: $F(1, 151) = 2.22$, $p = 0.14$ (see Table 8). Post hoc comparisons using Bonferroni corrections, as shown in Table 9, indicated that the adjusted mean score for the physical group (M = 2.37, SE = 0.17) was significantly different than the digital group (M = 1.41, SE = 0.18) ($p < 0.001$). Additionally, the physical group (M = 2.37, SE

= 0.17) was significantly different than the control group (M = 1.57, SE = 0.19) (*p* = 0.002). The digital group (M = 1.41, SE = 0.18) did not significantly differ from the control group (M = 1.57, SE = 0.19) (*p* = 0.55).

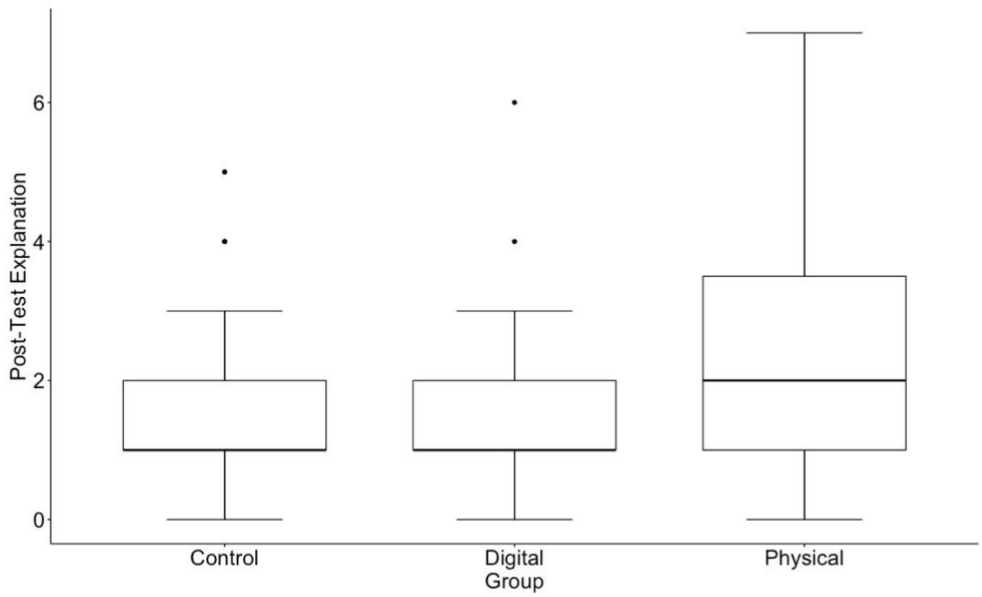


Figure 3: Box Plot of Post-Test Explanation Scores

Table 7: Table of Adjusted Means for Post-Test Explanation Scores across Treatment Groups

	Adjusted Mean	SE	n
Physical	2.37	0.17	59
Digital	1.41	0.18	53
Control	1.57	0.19	48

Table 8: ANCOVA of Post-Test Explanation Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

	Sum of Squares	*df*	*F*	*p*	*Partial eta squared*
Group	26.81	2	8.51	< 0.001	0.10
Classroom	24.36	3	5.16	0.002	0.09
Pre-Test	18.53	1	11.77	< 0.001	0.07
Spatial	5.44	1	3.46	0.06	0.02
Attitudes	3.49	1	2.22	0.14	0.01
Residuals	237.74	151			

Table 9: Bonferroni Corrections of Post-Test Explanation Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry Tukey Contrasts

	Estimate	*Std. Error*	*Statistic*	*p*
Control – Digital	0.15	0.26	0.60	0.55
Control – Physical	-0.81	0.26	-3.15	0.002
Digital – Physical	-0.96	0.25	-3.86	< 0.001

For the second hypothesis, three ANCOVAs were conducted to model long-term conceptual understanding. For the first ANCOVA, objective quiz scores were compared across the three groups. The adjusted means and standard errors of the objective quiz scores of the three groups at delayed post-test can be found in Figure 4 and Table 10. The results of the ANCOVA suggest there was no significant effect of the group on the objective quiz delayed post-test score after controlling for the effect of spatial ability, objective-based pre-test score, and attitude to science inquiry: $F(2, 140) = 1.27$, $p = 0.28$ (see Table 11 and Table 12). The partial eta squared effect size value (.02) suggested a small practical significance (see Table 11). The covariate, spatial ability, was significantly related to the objective quiz delayed post-test score: $F(1, 140) = 9.46$, $p = 0.003$ (see Table 11). The objective quiz pre-test covariate was also significantly related to the objective quiz delayed post-test score: $F(1, 140) = 29.17$, $p < 0.001$ (see Table 11). The attitudes to science inquiry covariate was not significantly related to the objective quiz delayed post-test score: $F(1, 140) = 0.10$, $p = 0.75$ (see Table 11).

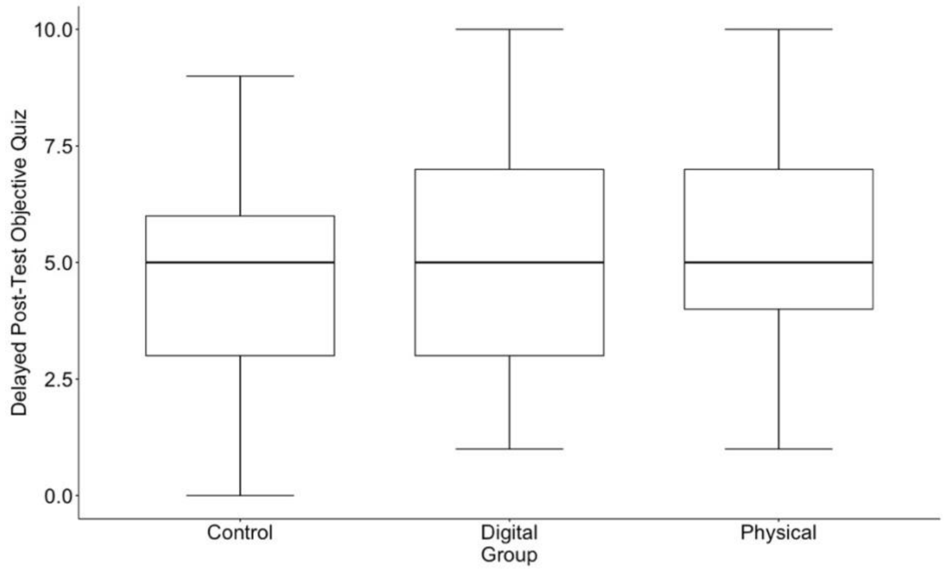


Figure 4: Box Plot of Delayed Post-Test Objective Quiz Scores

Table 10: Table of Adjusted Means for Delayed Post-Test Objective Quiz Scores across Treatment Groups

	Adjusted Mean	SE	n
Physical	5.49	0.27	56
Digital	5.03	0.30	47
Control	4.61	0.30	46

Table 11: ANCOVA of Delayed Post-Test Objective Quiz Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

	Sum of Squares	*df*	*F*	*p*	*Partial eta squared*
Group	10.3	2	1.27	0.28	0.02
Classroom	8.4	3	0.69	0.56	0.01
Pre-Test	119.1	1	29.17	< 0.001	0.17
Spatial	38.6	1	9.46	0.003	0.06
Attitudes	0.4	1	0.10	0.75	0.00
Residuals	1571.3	140			

Table 12: Bonferroni Corrections of Delayed Post-Test Objective Quiz Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

	Estimate	*Std. Error*	*Statistic*	*p*
Control – Digital	-0.42	0.42	-0.99	0.32
Control – Physical	-0.87	0.40	-2.17	0.032
Digital – Physical	-0.46	0.40	-1.14	0.26

The adjusted means and standard errors of the drawing scores of the three groups at delayed post-test can be found in Figure 5 and Table 13. The results of the ANCOVA suggest a significant effect of group on the delayed post-test model drawing score after controlling for the effect of spatial ability, objective quiz pre-test score, and attitudes to science inquiry: $F(2, 139) = 19.27$, $p < 0.001$. The partial eta squared effect size value (.22) suggested a large practical significance (see Table 14). The covariate, spatial ability, was not significantly related to the delayed post-test model drawing score: $F(1, 139) = 3.78$, $p = 0.05$ (see Table 14). The objective quiz pre-test covariate was significantly related to the delayed post-test model drawing score: $F(1, 139) = 8.78$, $p = 0.004$ (see Table 14). The attitudes to science inquiry covariate was not significantly related to the delayed post-test model drawing score: $F(1, 139) = 0.12$, $p = 0.73$ (see Table 14). Post hoc comparisons using Bonferroni corrections, as shown in Table 15, indicated that the adjusted mean score for the physical group (M = 10.8, SE = 0.66) was significantly different than the digital group (M = 4.58, SE = 0.71) ($p < 0.001$). Additionally, the physical group (M = 10.8, SE = 0.66) was significantly different than the control group (M = 7.07, SE = 0.72) ($p < 0.001$). The control group (M = 7.07, SE = 0.72) was significantly different from the digital group (M = 4.58, SE = 0.71) ($p = 0.015$).

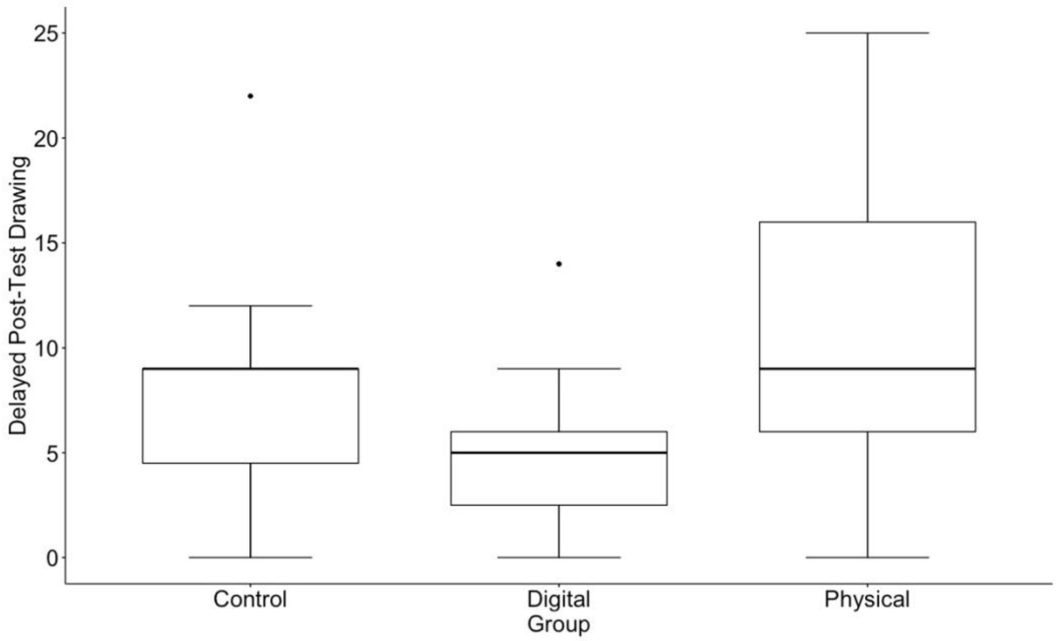


Figure 5: Box Plot of Delayed Post-Test Drawing Scores

Table 13: Table of Adjusted Means for Delayed Post-Test Drawing Scores across Treatment Groups

	Adjusted Mean	SE	n
Physical	10.8	0.66	55
Digital	4.58	0.71	47
Control	7.07	0.72	46

Table 14: ANCOVA of Delayed Post-Test Drawing Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

	Sum of Squares	*df*	*F*	*p*	*Partial eta squared*
Group	920	2	19.27	< 0.001	0.22
Classroom	21	3	0.29	0.83	0.006
Pre-Test	210	1	8.78	0.004	0.06
Spatial	90	1	3.78	0.05	0.03
Attitudes	3	1	0.12	0.73	0.00
Residuals	3320	139			

Table 15: Bonferroni Corrections of Delayed Post-Test Drawing Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

	Estimate	*Std. Error*	*Statistic*	*p*
Control – Digital	2.49	1.01	2.47	0.015
Control – Physical	-3.71	0.98	-3.80	< 0.001
Digital – Physical	-6.20	0.97	-6.37	< 0.001

The adjusted means and standard errors of the explanation scores of the three groups at post-test can be found in Figure 6 and Table 16. The results of the ANCOVA suggest a significant effect of group on the delayed post-test explanation of model score after controlling for the effect of spatial ability, objective quiz pre-test score, and attitudes to science inquiry: $F(2, 139) = 14.38$, $p < 0.001$. The partial eta squared effect size value (.17) suggested a large practical significance (see Table 17). The covariate, spatial ability, was not significantly related to the delayed post-test explanation of model score: $F(1, 139) = 2.14$, $p = 0.15$ (see Table 17). The objective quiz pre-test covariate was significantly related to the delayed post-test explanation of model score: $F(1, 139) = 25.52$, $p < 0.001$ (see Table 17). The attitudes to science inquiry covariate was not significantly related to the delayed post-test explanation of model score: $F(1, 139) = 2.89$, $p = 0.09$ (see Table 17). Post hoc comparisons using Bonferroni corrections, as shown in Table 18, indicated that the adjusted mean score for the physical group (M = 1.68, SE = 0.13) was significantly different than the digital group (M = 0.67, SE = 0.14) ($p < 0.001$). Additionally, the physical group (M = 1.68, SE = 0.13) was significantly different than the control group (M = 1.08, SE = 0.14) ($p = 0.003$). The digital group (M = 0.67, SE = 0.14) was significantly different from the control group (M = 1.08, SE = 0.14) ($p = 0.04$).

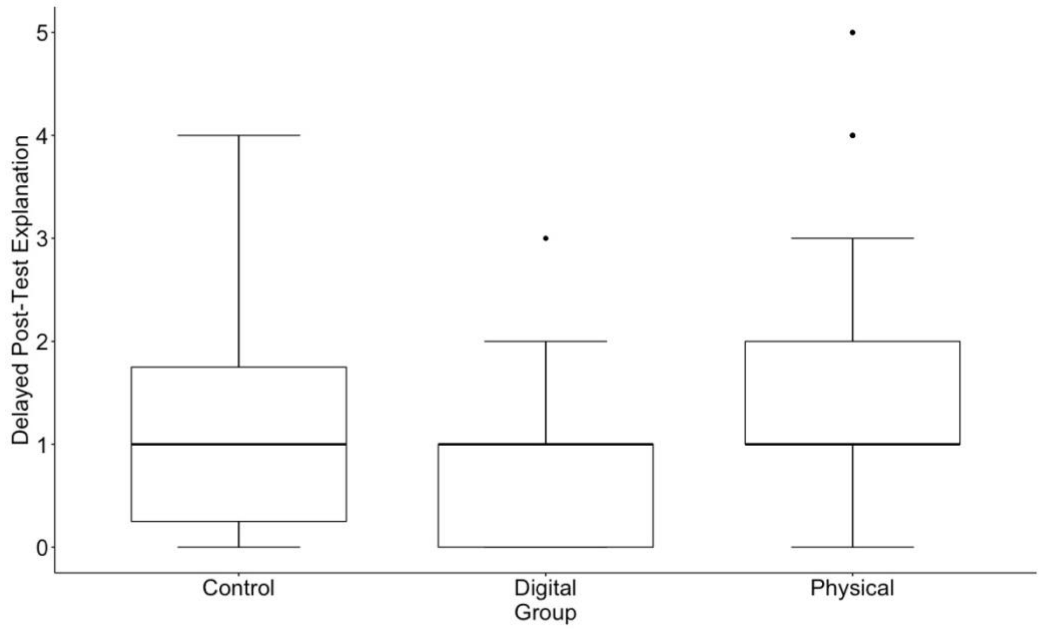


Figure 6: Box Plot of Delayed Post-Test Explanation Scores

Table 16: Table of Adjusted Means for Delayed Post-Test Explanation Scores across Treatment Groups

	Adjusted Mean	SE	n
Physical	1.68	0.13	55
Digital	0.67	0.14	47
Control	1.08	0.14	46

Table 17: ANCOVA of Delayed Post-Test Explanation Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

	Sum of Squares	*df*	*F*	*p*	*Partial eta squared*
Group	25.33	2	14.38	< 0.001	0.17
Classroom	8.15	3	3.09	0.03	0.06
Pre-Test	22.47	1	25.52	< 0.001	0.16
Spatial	1.88	1	2.14	0.15	0.02
Attitudes	2.54	1	2.89	0.09	0.02
Residuals	122.40	139			

Table 18: Bonferroni Corrections of Delayed Post-Test Explanation Scores Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

	Estimate	*Std. Error*	*Statistic*	*p*
Control – Digital	0.41	0.20	2.08	0.04
Control – Physical	-0.60	0.19	-3.09	0.003
Digital – Physical	-1.01	0.19	-5.26	< 0.001

For the third hypothesis, mixed ANOVAs were conducted to model degree of forgetting over time. The adjusted means and standard errors of the objective quiz scores of the three groups across post-test and delayed post-test can be found in Table 19. The mixed ANOVA suggested a statistically significant interaction between model building group and time in explaining the objective quiz scores: $F(2, 145) = 5.44$, $p = 0.005$. The partial eta squared effect size value (.07) suggested a moderate practical significance (see Figure 7 and Table 20). The covariate, spatial ability, was significantly related to objective quiz degree of forgetting: $F(1, 139) = 12.39$, $p < 0.001$ (see Table 20). The objective quiz pre-test covariate was also significantly related to the objective quiz degree of forgetting: $F(1, 139) = 41.41$, $p < 0.001$ (see Table 20). The attitudes to science inquiry covariate was not significantly related to the objective quiz degree of forgetting: $F(1, 139) = 1.88$, $p = 0.17$ (see Table 20). Pairwise comparisons using Bonferroni corrections, as shown in Figure 21, suggest that the adjusted mean score was significantly different between post-test (M = 6.53, SE = .297) and delayed post-test (M = 4.66, SE = .297) for the control group ($p < 0.001$). There was a statistically significant difference between post-test (M = 5.73, SE = .296) and delayed post-test (M = 5.01, SE = .296) objective quiz scores for the digital group ($p = 0.03$). There was no statistically significant difference between post-test (M = 5.89, SE = .275) and delayed post-test (M = 5.51, SE = .275) objective quiz scores for the physical group ($p = 0.19$).

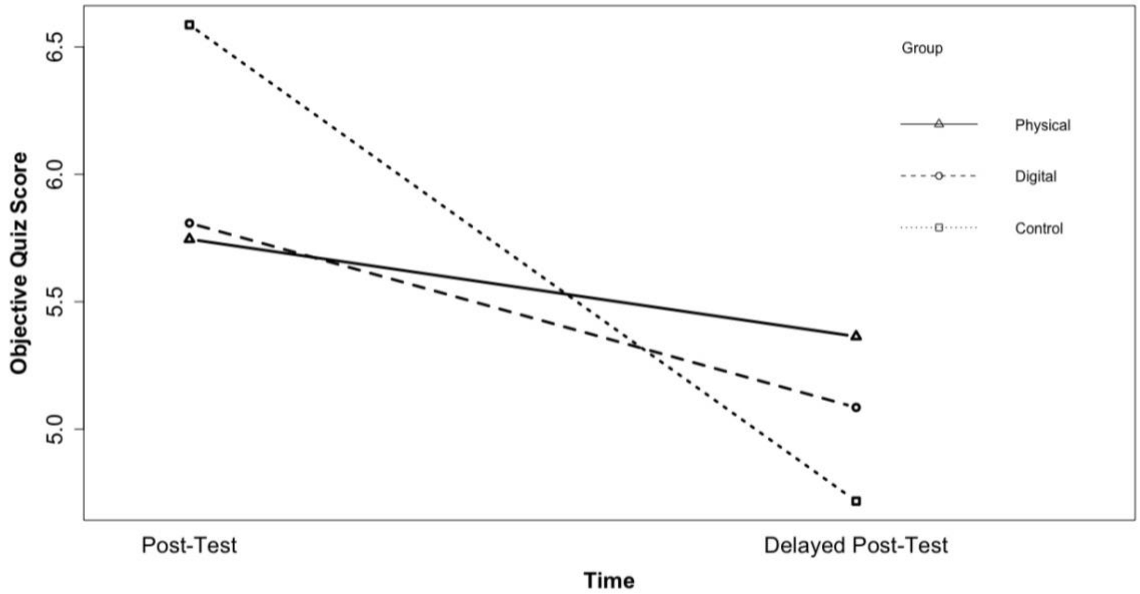


Figure 7: Interaction Plot of Objective Quiz Scores

Table 19: Table of Adjusted Means for Degree of Forgetting Objective Quiz Scores across Treatment Groups

	Post-Test			Delayed Post-Test		
	Adjusted Mean	SE	n	Adjusted Mean	SE	n
Physical	5.89	.275	55	5.51	.275	55
Digital	5.73	.296	47	5.01	.296	47
Control	6.53	.297	46	4.66	.297	46

Table 20: Mixed ANOVA of Objective Quiz Scores by Time and Group Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

Effect	*Sum Sq*	*Mean Sq*	*NumDF*	*DenDF*	*F*	*p*	*Partial eta squared*
Group	12.53	6.27	2	139	2.31	0.10	0.03
Classroom	14.27	4.76	3	139	1.75	0.16	0.04
Time	72.30	72.30	1	145	26.62	< 0.001	0.16
Pre-Test	112.47	112.47	1	139	41.41	< 0.001	0.23
Spatial	33.66	33.66	1	139	12.39	< 0.001	0.08
Attitudes	5.11	5.11	1	139	1.88	0.17	0.01
Group:Time	29.53	14.77	2	145	5.44	0.005	0.07

Table 21: Bonferroni Corrections of Objective Quiz Scores by Time and Group Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

Group	*Time 1*	*Time 2*	*n1*	*n2*	*statistic*	*df*	*p*
Control	Post-Test	Delayed Post-Test	46	46	4.70	45	< 0.001
Digital	Post-Test	Delayed Post-Test	47	47	2.31	46	0.025
Physical	Post-Test	Delayed Post-Test	55	55	1.32	55	0.193

The adjusted means and standard errors of the drawing scores of the three groups across post-test and delayed post-test can be found in Table 22. There was a statistically significant interaction between model building group and time in explaining the drawing scores: $F(2, 145) = 3.79$, $p = 0.025$. The partial eta squared effect size value (.05) suggested a moderate practical significance (see Figure 8 and Table 23). The covariate, spatial ability, was significantly related to drawing score degree of forgetting: $F(1, 139) = 5.75$, $p = 0.018$ (see Table 23). The objective quiz pre-test covariate was also significantly related to the drawing score degree of forgetting: $F(1, 139) = 9.35$, $p = 0.003$ (see Table 23). The attitudes to science inquiry covariate was not significantly related to drawing score degree of forgetting: $F(1, 139) = 0.11$, $p = 0.74$ (see Table 23). Pairwise comparisons using Bonferroni corrections, as shown in Table 24, suggest that the adjusted mean score was significantly different between post-test (M = 7.80, SE = .781) and delayed post-test (M = 4.61, SE = .781) for the digital model group ($p < 0.001$). There was no statistically significant difference between post-test (M = 11.50, SE = .727) and delayed post-test (M = 10.90, SE = .727) drawing scores for the physical group ($p = 0.49$). There was also no statistically significant difference between post-test (M = 7.24, SE = .783) and delayed post-test (M = 7.08, SE = .783) drawing scores for the control group ($p = 0.84$).

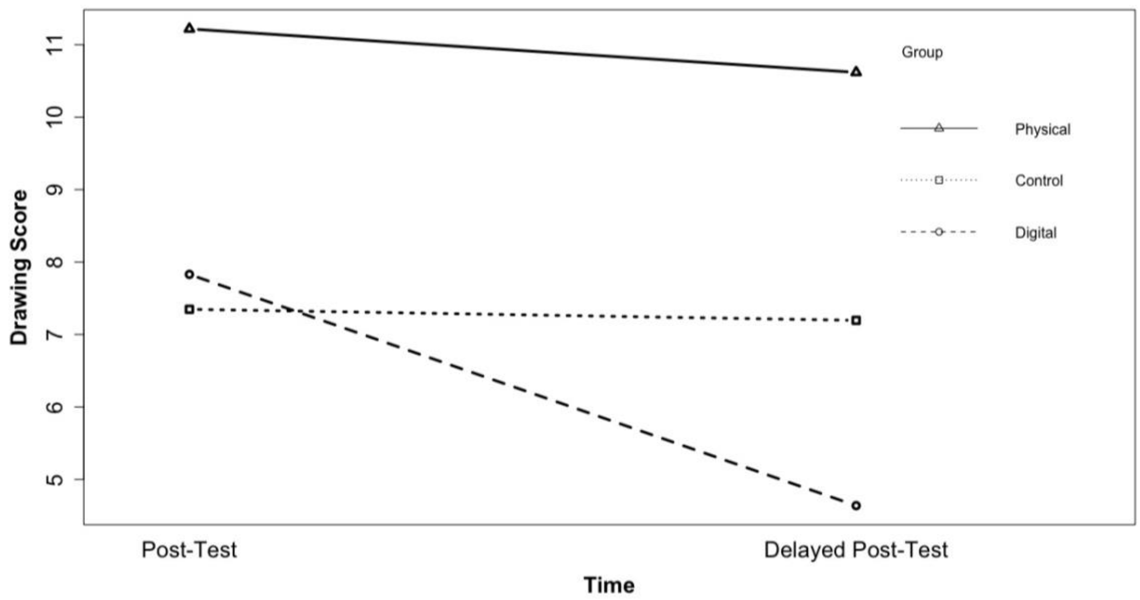


Figure 8: Interaction Plot of Drawing Scores

Table 22: Table of Adjusted Means for Degree of Forgetting Drawing Scores across Treatment Groups

	Post-Test			Delayed Post-Test		
	Adjusted Mean	SE	n	Adjusted Mean	SE	n
Physical	11.50	.727	55	10.90	.727	55
Digital	7.80	.781	47	4.61	.781	47
Control	7.24	.783	46	7.08	.783	46

Table 23: Mixed ANOVA of Drawing Scores by Time and Group Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

Effect	*Sum Sq*	*Mean Sq*	*NumDF*	*DenDF*	*F*	*p*	*Partial eta squared*
Group	58.64	29.32	2	139	1.74	0.18	0.02
Classroom	25.81	8.60	3	139	0.51	0.68	0.01
Time	127.07	127.07	1	145	7.54	0.007	0.05
Pre-Test	157.51	157.51	1	139	9.35	0.003	0.06
Spatial	96.75	96.75	1	139	5.75	0.018	0.04
Attitudes	1.82	1.82	1	139	0.11	0.74	0.00
Group:Time	127.84	63.92	2	145	3.79	0.025	0.05

Table 24: Bonferroni Corrections of Drawing Scores by Time and Group Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

Group	*Time 1*	*Time 2*	*n1*	*n2*	*statistic*	*df*	*p*
Control	Post-Test	Delayed Post-Test	46	46	0.208	45	0.84
Digital	Post-Test	Delayed Post-Test	47	47	3.73	47	< 0.001
Physical	Post-Test	Delayed Post-Test	55	55	0.70	54	0.49

The adjusted means and standard errors of the explanation scores of the three groups across post-test and delayed post-test can be found in Table 25. The mixed ANOVA suggested there was not a statistically significant interaction between model building group and time in explaining the explanation scores: $F(2, 145) = 0.40$, $p = 0.67$ (see Figure 9 and Table 26). The partial eta squared effect size value (.00) suggested a low practical significance (see Figure 9 and Table 26). The covariate, spatial ability, was significantly related to explanation of model score degree of forgetting: $F(1, 139) = 5.54$, $p = 0.02$ (see Table 26). The objective quiz pre-test covariate was significantly related to explanation of model score degree of forgetting: $F(1, 139) = 25.70$, $p < 0.001$ (see Table 26). The attitudes to science inquiry covariate was significantly related to explanation of model score degree of forgetting: $F(1, 139) = 4.21$, $p = 0.04$ (see Table 26). There was a statistically significant main effect of group on explanation score: $F(2, 139) = 7.80$, $p < 0.001$ (see Table 26). Pairwise comparisons using Bonferroni corrections, as shown in Table 27, suggest that the adjusted mean score was significantly different between physical model group (M = 2.06, SE = .127) and control group (M = 1.35, SE = .136) ($p < 0.001$). Additionally, the adjusted mean score was significantly different between physical model group (M = 2.06, SE = .127) and digital model group (M = 0.99, SE = .136) ($p < 0.001$). The digital group (M = 0.99, SE = .136) did not significantly differ from the control group (M = 1.35, SE = .136) ($p = 0.07$). There was also a statistically significant main effect of time on explanation score: $F(1, 145) = 35.85$, $p < 0.001$. Pairwise comparisons using Bonferroni corrections, as shown in Table 27, suggest that the adjusted mean score was significantly different between post-test explanation (M = 1.79, SE = .094) and delayed post-test explanation (M = 1.14, SE = .094) ($p < 0.001$).

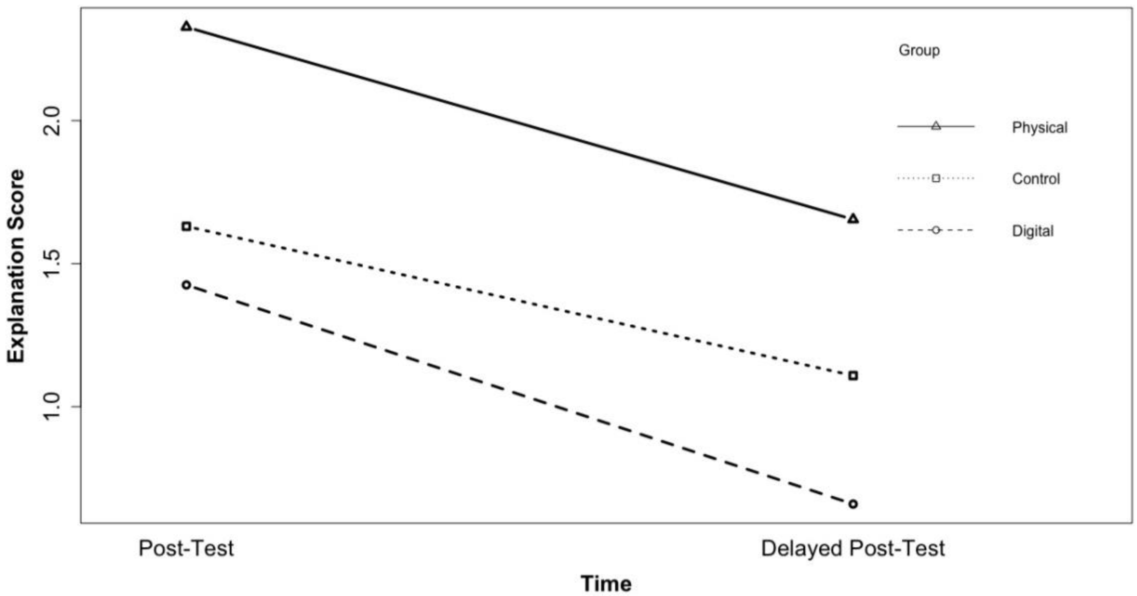


Figure 9: Interaction Plot of Explanation Scores

Table 25: Table of Adjusted Means for Degree of Forgetting Explanation Scores across Treatment Groups

	Adjusted Mean	SE	n
Post-Test	1.79	.094	148
Delayed Post-Test	1.14	.094	148
Physical	2.06	.127	110
Digital	0.99	.136	94
Control	1.35	.136	92

Table 26: Mixed ANOVA of Explanation Scores by Time and Group Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

Effect	*Sum Sq*	*Mean Sq*	*NumDF*	*DenDF*	*F*	*p*	*Partial eta squared*
Group	13.67	6.83	2	139	7.80	< 0.001	0.10
Classroom	7.65	2.55	3	139	2.91	0.04	0.06
Time	31.40	31.40	1	145	35.85	< 0.001	0.20
Pre-Test	22.51	22.51	1	139	25.70	< 0.001	0.16
Spatial	4.85	4.85	1	139	5.54	0.02	0.04
Attitudes	3.69	3.69	1	139	4.21	0.04	0.03
Group:Time	0.71	0.35	2	145	0.40	0.67	0.00

Table 27: Main Effects of Explanation Score by Time and Group Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

Time 1	*Time 2*	*n1*	*n2*	*statistic*	*df*	*p*
Explain Post-Test	Explain Delayed Post-Test	148	148	6.05	147	< 0.001

Group 1	*Group 2*	*n1*	*n2*	*statistic*	*df*	*p*
Control	Digital	92	94	1.77	293	0.07
Control	Physical	92	110	-3.49	293	< 0.001
Digital	Physical	94	110	-5.36	293	< 0.001

For the fourth hypothesis, a mixed ANOVA was conducted to model retention of learned material over time. The adjusted means and standard errors of the objective quiz scores of the three groups across pre-test, post-test, and delayed post-test can be found in Table 28. There was a statistically significant interaction between model building group and time in explaining the objective quiz scores: $F(4, 290) = 2.80$, $p = 0.026$. The partial eta squared effect size value (.04) suggested a moderate practical significance (see Figure 10 and Table 29). The covariate, spatial ability, was significantly related to objective quiz retention: $F(1, 140) = 9.97$, $p = 0.002$ (see Table 29). The attitudes to science inquiry covariate was not significantly related to the objective quiz retention: $F(1, 140) = 0.26$, $p = 0.61$ (see Table 29). Pairwise comparisons using Bonferroni corrections, as shown in Table 30, suggest that the mean score was significantly different between pre-test (M = 4.18, SE = .316) and post-test (M = 6.57, SE = .316) for the control group ($p < 0.001$), post-test (M = 6.57, SE = .316) and delayed post-test (M = 4.70, SE = .316) for the control group ($p < 0.001$), pre-test (M = 4.00, SE = .314) and post-test (M = 5.75, SE = .314) for the digital group ($p < 0.001$), pre-test (M = 4.00, SE = .314) and delayed post-test (M = 5.02, SE = .314) for the digital group ($p = 0.006$), post-test (M = 5.75, SE = .314) and delayed post-test (M = 5.02, SE = .314) for the digital group ($p = 0.025$), pre-test (M = 4.10, SE = .292) and post-test (M = 5.87, SE = .292) for the physical group ($p < 0.001$), and pre-test (M = 4.10, SE = .292) and delayed post-test (M = 5.48, SE = .292) for the physical group ($p < 0.001$). There was no statistically significant difference between pre-test (M = 4.18, SE = .316) and delayed post-test (M = 4.70, SE = .316) for the control group ($p = 0.16$), and post-test (M = 5.87, SE = .292) and delayed post-test (M = 5.48, SE = .292) for the physical group ($p = 0.19$).

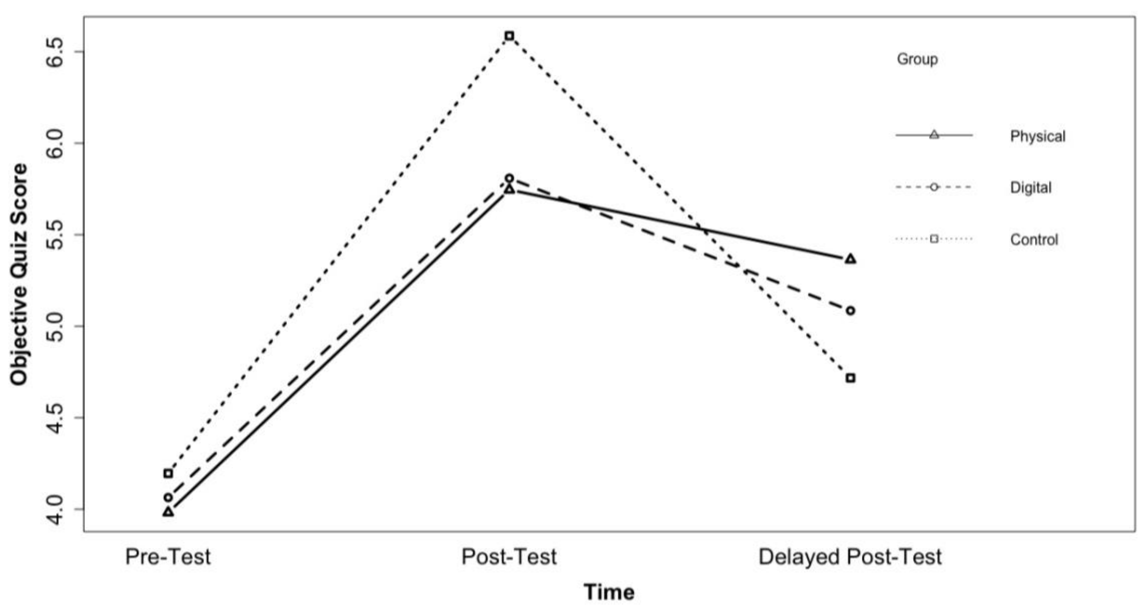


Figure 10: Interaction Plot of Objective Quiz Pre-Test, Post-Test, Delayed Post-Test Scores

Table 28: Table of Adjusted Means for Retention of Objective Quiz Scores across Treatment Groups

	Pre-test Adjusted Mean	SE	n	Post-Test Adjusted Mean	SE	n	Delayed Post-Test Adjusted Mean	SE	n
Physical	4.10	.292	55	5.87	.292	55	5.48	.292	55
Digital	4.00	.314	47	5.75	.314	47	5.02	.314	47
Control	4.18	.316	46	6.57	.316	46	4.70	.316	46

Table 29: Mixed ANOVA of Objective Quiz Scores by Time and Group Controlling for Spatial Ability, and Attitudes to Science Inquiry

Effect	*Sum Sq*	*Mean Sq*	*NumDF*	*DenDF*	*F*	*p*	*Partial eta squared*
Group	4.61	2.31	2	140	0.86	0.43	0.01
Classroom	7.73	2.58	3	140	0.96	0.42	0.02
Time	284.38	142.19	2	290	52.76	< 0.001	0.27
Spatial	26.87	26.87	1	140	9.97	0.002	0.07
Attitudes	0.71	0.71	1	140	0.26	0.61	0.00
Group:Time	30.16	7.54	4	290	2.80	0.026	0.4

Table 30: Bonferroni Corrections of Objective Quiz Scores by Time and Group Controlling for Pre-Test, Spatial Ability, and Attitudes to Science Inquiry

Group	*Time 1*	*Time 2*	*n1*	*n2*	*statistic*	*df*	*p*
Control	Pre-Test	Post-Test	46	46	-5.39	45	< 0.001
Control	Pre-Test	Delayed Post-Test	46	46	-1.43	45	0.16
Control	Post-Test	Delayed Post-Test	46	46	4.70	45	< 0.001
Digital	Pre-Test	Post-Test	47	47	-6.79	46	< 0.001
Digital	Pre-Test	Delayed Post-Test	47	47	-2.88	46	0.006
Digital	Post-Test	Delayed Post-Test	47	47	2.31	46	0.025
Physical	Pre-Test	Post-Test	55	55	-6.08	54	< 0.001
Physical	Pre-Test	Delayed Post-Test	55	55	-5.21	54	< 0.001
Physical	Post-Test	Delayed Post-Test	55	55	1.32	54	0.19

Chapter 5: Discussion

The purpose of this study was to measure the effectiveness of different pedagogical techniques on conceptual understanding of abstract science concepts at the initial learning stage. The topic of DNA was chosen as the exemplary concept to demonstrate the impact of learning with touchscreen devices. For true conceptual understanding of the DNA molecule, an individual must exhibit understanding of the form and function of the molecule in the context of both biological and chemical sciences. Chemical concepts of bonding and polarity integrate with concepts of gene expression and heredity when learning about this molecule at a basic level. In the Living Environment curriculum sequence, DNA appears in three separate units that are separated by material that is unrelated to DNA. Teachers and students of Living Environment cite DNA as one of the more difficult concepts to create accurate and long-term conceptual understanding. In addition to its curricular importance, DNA is an example of a three-dimensional, abstract science concept in which the three-dimensional shape is essential for true understanding. Due to the integration of biology and chemistry principles, the perceived difficulty of understanding the concept, and the importance of the three-dimensional shape, the results of this study provide a degree of generalizability to other three-dimensional science concepts.

The different pedagogical techniques used to help teach the concept of DNA were physical model construction, digital model construction through the use of touchscreen devices, and a paper worksheet typical in many classrooms as a teaching tool. The focus of this study was to measure the impact of touchscreen devices due to the two-dimensional representations of three-dimensional objects as well as the limited haptic interactions that are possible with a digital simulation presented on a touchscreen device. A physical model comparison was used due to the three-dimensional nature of physical models as well as the expanded range of haptic interactions when building a physical model. A paper worksheet was used for the control group to model traditional pedagogical techniques where there were no touchscreen devices or physical models present. The worksheet does not allow for model construction outside of matching a nucleotide sequence and therefore haptic interactions are not considered.

Conceptual understanding is often measured with a multiple-choice exam. However, demonstration of science knowledge is not limited to choosing a correct option from a set of preconceived statements. Therefore, the current study measured conceptual understanding with an objective quiz, a drawing of the DNA molecule, and a written explanation of the DNA molecule. The addition of the drawing and explanation as measures allowed participants an opportunity to demonstrate their knowledge of the concept of DNA by generating their own images and written explanations which are more robust indicators of understanding since they are learner generated (Babilonia-Rosa et al., 2018; Fiorella et al., 2020). In addition to different measures exhibiting conceptual understanding, true conceptual understanding is demonstrated if an individual is able to create an accurate mental model that can be used during recall. Therefore, conceptual understanding was measured at two timepoints: immediately after treatment and again two months later. It is important to consider that the extended two-month, delayed timepoint included a two-week school break as well as midterm exams. The break and exams act as distractors from the content presented at the initial experience two months prior and places greater importance on the delayed post-test results that physical model construction at initial learning supports accurate mental models and long-term conceptual understanding.

To account for possible covariates, an objective quiz was completed as well as a spatial ability survey and an attitude to scientific inquiry survey at baseline. The participants in this study had not been formally taught about the topic of DNA prior to their participation in the study; however, due to the ubiquitous nature of DNA, the objective pre-test provided a measure of prior knowledge of the topic. Since DNA is a three-dimensional molecule and its shape is an essential component to understand its function, a mental rotation test was given to measure individual's spatial ability. Additionally, the perceived relationship between attitudes to science and science achievement led to the inclusion of the *Attitudes to Scientific Inquiry* survey. Controlling for these three covariates allowed for a clearer interpretation of the differences among the three treatment groups at and across the two timepoints.

5.1 Hypothesis 1 Analysis

The results of the three ANCOVAs to measure the first hypothesis principally support the premise that the physical model group demonstrates greater conceptual understanding compared with the digital model and control groups while controlling for pre-test, spatial ability, and attitudes to scientific inquiry at immediate post-test. Although there was no significant difference among the groups for the objective quiz measure of conceptual understanding, these results were anticipated based on the results from the pilot studies. For both the drawing and explanation measures of conceptual understanding, the physical model group showed significantly higher scores than both the digital and the control groups. Further analyses of the other two measures of conceptual understanding revealed that there was no difference between the digital and the control groups for drawing and explanation scores at post-test.

The physical model group's difference in the drawing and explanation scores suggests that model type plays a factor in students' ability to acquire information for abstract concepts in the short-term. Similar results have been reported in the research that discuss the importance of haptic interactions with the material being taught yields greater understanding (Preece et al., 2013). Combined with the current finding from this study, these measures of conceptual understanding are generally held to demonstrate a fuller understanding compared with recall usually associated with a multiple-choice format of objective questions. Although there was no difference among model type for the objective quiz, the difference in the other two measures support the hypothesis.

The significance of the covariates is also important at post-test. For all three measures of conceptual understanding, the pre-test significantly influenced the outcome measures. Additionally, spatial ability score was a significant covariate for both the objective quiz and the drawing measure. This suggests that an individual's spatial ability may influence their ability to initially learn about concepts that are abstract and inherently spatial. It is also possible that spatial ability may support the creation of mental models in the visuospatial sketchpad to be able to produce more complete and accurate drawings (Hegarty & Stull, 2012; Johnson-Laird, 2012).

5.1.1 Hypothesis 1 Implications

The results from this analysis holds implications for the method of assessment of abstract concepts, specifically that objective questions may not be a complete form of assessment to measure understanding since method of instruction does not influence the outcome. Because there was a difference with the drawing and explanation conceptual understanding measures, science educators may want to consider incorporating more physical model building in their pedagogy. It is important to note that there was no difference between the digital model group and the control group at post-test for both drawing and explanation scores. This implies that the use of digital model building does not serve a functional improvement in conceptual understanding over more traditional paper worksheet methods of instruction. This is particularly interesting since there has been a recent surge in educational technology and one-to-one device implementation in schools where each student has their own Google Chromebook® or Apple iPad® (Gray & Lewis, 2021). Teachers and administrators should consider the potential impact of digital hardware and software before spending large percentages of the school's budget on these tools (Farenga & Joyce, 2001).

Additionally, the potential positive influence of an individual's spatial ability on conceptual understanding of abstract science concepts suggests that more time and attention should be spent on increasing students' spatial ability (NGSS Lead States, 2013; Wei et al., 2009). Spatial ability is a construct that can be improved over time provided the appropriate tools are used (Harle & Towns, 2011).

5.2 Hypothesis 2 Analysis

The results of the three ANCOVAs to measure the second hypothesis principally support that the physical model group demonstrates greater conceptual understanding compared with the digital model and control groups while controlling for pre-test, spatial ability, and attitudes to scientific inquiry at delayed post-test. The analysis for the delayed post-test is similar to the immediate post-test. For the objective quiz, there was no significant difference among the groups. For both the drawing and explanation measures of conceptual understanding, the physical model group showed significantly higher scores than both the digital and the control groups. For the drawing and explanation measures, the control group had higher scores compared with the digital group.

The physical model group's difference in the drawing and explanation scores suggests that model type plays a factor in students' ability to acquire information for abstract concepts in the long-term. This result supports prior research on the importance of using accurate, physical models and visuals to support mental model construction, memory, and conceptual understanding specifically in the long-term (Bruyer & Scailquin, 1998; Huk, 2007; Kontra et al., 2015; Stieff, 2010). Additionally, the difference between the drawing and explanation scores of control group and the digital group suggests that the digital model did not adequately support the individual's ability to recreate a drawing or provide an explanation from their experience two months prior and may have even been a detriment to their long-term conceptual memory. The studies conducted by Mayer and Moreno (2003), Pedra et al. (2015), and Russo-Johnson et al. (2017) similarly question the limited physical interactions and affordances of the touchscreen to support accurate mental model construction and long-term conceptual understanding. Although there was no difference among model type for the objective quiz, the difference in the other two measures supports the hypothesis.

The significance of the covariates is also important at delayed post-test. For all three measures of conceptual understanding, the pre-test significantly influenced the long-term outcome measures. Additionally, spatial ability score remained a significant covariate for the objective measure. This finding supports the suggestion that an individual's spatial ability may influence their ability to learn about concepts that are abstract and inherently spatial (Harle & Towns, 2011; Wai et al., 2009).

5.2.1 Hypothesis 2 Implications

The results at the delayed post-test hold implications for how model techniques aid memory after two months. Physical model construction once again demonstrated greater conceptual understanding for the drawing and explanation measures of conceptual understanding. This implies that students who build physical models may have greater understanding and memory of the topic compared with building digital models or using paper worksheets. Additionally, the drawing and explanation scores of the control group were greater than the drawing and explanation scores of the digital group. This suggests that the digital model construction may have negatively impacted the students' understanding of the visual representation of the DNA molecule as well as their ability to adequately explain and describe the concept. This finding, in conjunction with the analysis of the first hypothesis, should serve as a warning that digital simulations may not be as beneficial as educators may think when teaching abstract science concepts at initial instruction. While spatial ability was only significantly related to objective quiz at delayed post-test, it is still important to consider the potential positive impact strengthening spatial ability may have on other abstract spatial science concepts (NGSS Lead States, 2013; Wai et al., 2009).

5.3 Hypothesis 3 Analysis

The results of the two-way mixed ANOVA did not support the hypothesis that the digital model group would have greater degree of forgetting over time compared with the physical model group and the control group at all three measures of conceptual understanding. However, the digital model group demonstrated a significant degree of forgetting for the objective and drawing measures of conceptual understanding. The control group experienced a significant degree of forgetting for the objective quiz measure. The physical model group did not demonstrate a significant degree of forgetting at all three measures of conceptual understanding.

Forgetting was measured as a significant difference between post-test and delayed post-test scores where the means of the delayed post-test scores were lower than the means of the post-test scores. For the objective quiz, the digital group and the control group demonstrated a significant difference between post-test and delayed post-test scores with the delayed post-test score lower than the post-test score. The physical model group did not have significant differences between these two timepoints for the objective quiz. This suggests that physical model construction aided in retention of conceptual understanding as measured by the objective quiz rather than forgetting. It is possible that the added haptic interactions required for the physical model building activity aided in long-term memory as demonstrated in prior studies (Black et al., 2012; Kontra et al., 2015; Möhring & Frick, 2013; Schwartz & Plass, 2014).

The drawing measure of conceptual understanding demonstrated a significant degree of forgetting with the digital group. This suggests that the digital model group was not able to retain a spatially accurate visual representation of the DNA molecule to reproduce at the delayed post-test timepoint. Both the physical model group and the control group did not have significant differences between the two timepoints. While the physical model group's retention was expected, the control group's ability to retain the visual representation is interesting. Perhaps the drawing scores remained constant at the delayed post-test measure because of the visual representations presented to the control group. The control group had only experienced the DNA molecule in the video that all groups viewed prior to treatment. The video's representation of DNA was accurate for both component parts and spatial construction, which were measures for the drawing score as outlined in the Rubric for Model Drawing (Appendix K). It is possible that the digital model group's experiences with the visual representation of the DNA molecule in the digital simulation influenced their long-term understanding and memory of an accurate image of the DNA molecule (Hegarty & Stull, 2012; Kosslyn, 2005; Stieff, 2010).

The explanation measure of conceptual understanding did not show a significant interaction between group and time. This suggests that while considering these two measures together, there were no significant differences between the post-test and delayed post-test measures across the three groups. As such, all three groups demonstrated retention of the conceptual understanding as measured by written explanations of the DNA molecule rather than forgetting. However, the results of the main effects show where differences lie across time and among the groups. The post-test explanation scores were greater than the delayed post-test explanation scores. Although there are no differences of these measures across the groups, there was overall conceptual understanding loss over time. Additionally, the physical group's explanation scores across time were greater than both the digital model group and control group.

As for the covariates, degree of forgetting was impacted by the pre-test objective quiz score and spatial ability score for all three measures of conceptual understanding. This suggests that prior knowledge and mental rotation had an impact on participants' conceptual understanding across time. This relationship is similar to the results at the individual time points. Therefore, spatial ability is an important factor for these assessment measures. This is expected since these two measures require the use of the visuospatial sketchpad (Hegarty & Stull, 2012). Additionally, attitudes to scientific inquiry impacted degree of forgetting for the explanation measure of conceptual understanding. This demonstrates the importance of considering attitudes in terms of more complex conceptual understanding assessment measures such as learner-generated explanations (Fiorella, et al., 2020).

Overall, the physical model group did not demonstrate a significant degree of forgetting for conceptual understanding of the DNA molecule as measured by the objective quiz, model drawing, and model explanation. The digital group experienced a significant degree of forgetting for the objective quiz and drawing measures of conceptual understanding, and the control group experienced a significant degree of forgetting for the objective quiz measure of conceptual understanding. The differences in where the significant degree of forgetting was found demonstrate that the methods of assessment of conceptual understanding are different and measure different cognitive functions. Additionally, the pedagogical technique dictates the degree of probability of retention of conceptual understanding as assessed by a specific measure.

5.3.1 Hypothesis 3 Implications

The results for this analysis on degree of forgetting is especially important when considering appropriate pedagogical techniques at initial learning experiences for conceptual understanding. This study measured ability to recall information after a period of time where participants did not interact with the concept of DNA. With only the initial treatment during experimentation, the results of this study demonstrate the impact of physical and digital model building as well as the impact of the traditional paper worksheet. A primary implication is that physical model construction should be considered as a helpful tool for retention of abstract science concepts. The haptic functionality of model building in three-dimensional space may have aided in long-term memory as well as mental model construction as measured by the objective quiz, drawing score, and explanation score (Barsalou, 2008; Kontra et al., 2015; Mason & Just, 2015).

An additional important implication about degree of forgetting involves the perceived usefulness of digital simulations. For the digital model group, there was a significant degree of forgetting for the objective quiz and drawing measures but not for the explanation measure. This implies that the images presented in the digital simulation may have influenced the participants' understanding of an appropriate visual representation of the concept. Additionally, the limited haptic interaction with the model as well as building the model in two-dimensional space are cognitive considerations for the decrease in model drawing score which explicitly measured component and spatial accuracy of the drawing (Babilonia-Rosa et al., 2018; Preece et al., 2013). Although the digital model construction showed a significant degree of forgetting for the objective quiz and drawing measures, it should be taken into consideration that the physical model construction did not lead to forgetting at any of the measures of conceptual understanding when initially presenting abstract science concepts. This would suggest there should be some hesitation with the implementation of a digital simulation at initial learning especially when the visual representations of the concept differ between instruction and assessment.

5.4 Hypothesis 4 Analysis

The fourth hypothesis specifically measured the differences of the objective quiz at the three timepoints across the three groups. The results of the two-way mixed ANOVA did support the hypothesis that the physical model group provided greater retention of the learned material over time compared with the digital model group and the control group. Retention of the learned material was measured by no significant difference between post-test and delayed post-test as well as a significantly greater score at delayed post-test timepoint in comparison with the pre-test score.

The physical model group and digital model group demonstrated significantly higher scores on the objective quiz at the delayed post-test compared with the pre-test measure. However, while all three groups experienced an increase at post-test and a decrease at delayed post-test, only the physical model group maintained significantly higher scores at delayed post-test compared with pre-test while experiencing no significant difference between the post-test and delayed post-test timepoints. This suggests that the use of physical models supported greater long-term conceptual understanding and mental model construction (Bruyer & Scailquin, 1998; Kontra et al., 2015).

The control group's mean score was higher than both the physical and digital groups at post-test and was significantly higher than the control group's pre-test scores and delayed post-test objective scores. This suggests that even though the control group scored well immediately after treatment, they were not able to retain the information since there was no significant difference between the pre-test objective score and the delayed post-test objective score. It is possible that the control group's mean score at post-test was greater than the physical and digital groups because the objective quiz was a similar format to the paper worksheet the control group worked on. Since the method of instruction and the method of assessment were similar, there was more limited, near transfer involved for this group (Barnett & Ceci, 2002; Tulving, 1983). However, at delayed post-test, the significant drop in the control group's mean score suggests that the work with the paper worksheet did not have a lasting impact on mental model construction.

Both the physical model group and the digital model group had post-test and delayed post-test objective scores greater than their pre-test scores. However, it was only the physical model group that demonstrated no significant difference between post-test and delayed post-test. This suggests retention of the learned material. The digital model group did experience a significant degree of forgetting between the post-test and delayed post-test timepoints. Even though the delayed post-test score was greater than the pre-test score, the difference between the post-test and delayed post-test does not support a significant level of retention. Although it is important to consider that some level of conceptual understanding was achieved by the digital group since the delayed post-test scores were greater than pre-test scores. These findings support that model construction during initial learning positively impacts mental model construction. However, it is important to note that the physical group's delayed post-test mean scores for the objective measure was higher than both the control and digital groups. Additionally, spatial ability scores were significantly related to the objective quiz scores over time. This finding supports the importance of spatial ability with regard to abstract science conceptual understanding similar to the previous three hypotheses' results (Nagy-Kondor & Esmailnia, 2022; NGSS Lead States, 2013; Wai et al., 2009; Wu & Shah, 2004).

5.4.1 Hypothesis 4 Implications

The results from the objective quiz over three timepoints provides several implications of pedagogical technique for this type of conceptual understanding measurement. First, there was no significant difference between the control group's pre-test and delayed post-test scores. This implies that several weeks after the participants completed the worksheet assignment their conceptual understanding was no greater than it was before the study began. While their post-test objective scores were higher than both pre-test and delayed post-test objective scores, this implies that the control treatment may only be useful for immediate recall. Additionally, while there was significant difference between the digital group's pre-test and delayed post-test objective scores there was also a significant degree of forgetting between post-test and delayed post-test. This suggests that digital model building did not support retention of learned material. The implication is that for initial learning digital simulations on touchscreen devices should not be used if retention of learned material is the desired outcome (Scherer et al., 2018). Rather, when possible, physical models should be used for long-term retention since it was the only treatment group that demonstrated significantly greater scores at the delayed post-test measure compared with the pre-test.

Second, there was no significant difference in delayed post-test objective scores for the physical model group compared with the post-test objective scores. This implies that the participants who built physical models demonstrated greater retention between the short-term and long-term timepoints. Physical model building was demonstrated as a useful tool for long-term retention when assessing with solely objective means. The implication is that of the physical model construction supports long-term retention of learned material and should be the preferred method of instruction for initial learning.

Finally, spatial ability was significantly related to the objective quiz outcome measure across time. This implies that spatial ability is important not only at initial learning, but also beneficial for long-term recall and memory. The implication is that spatial ability's importance to science learning should result in opportunities for students to enhance their spatial skills since it may be an asset to their overall understanding of science concepts.

5.5 Historical Significance

The historical impact of the Covid-19 Pandemic must be considered in the analysis and interpretation of this study. All of the safety protocols outlined by the participating schools were followed as well as the safety protocols and guidelines required by Teachers College, Columbia University. These safety protocols included students and their teachers wearing masks at all times as well as maintaining social distance when possible. Students had been following these protocols for several months prior to the start of this study and therefore may have had minimal impact on the outcome of the study.

The primary focus of this study was to measure the impact of touchscreen devices on students' conceptual understanding when compared with physical learning opportunities. It is important to consider that beginning in March 2020 the participants of this study had received exclusively remote, digital instruction for several months followed by a hybrid mix of digital and in-person instruction until June 2021. The study began in December 2021 which was over three months after exclusively in-person instruction for the study's participants. The impact of the pandemic on students' social, emotional, and cognitive well-being is not yet fully understood.

Additionally, it is not clear if the months students spent in digital instruction may have impacted their opinions of digital or physical instructional tools. The inclusion of the *Attitudes to Scientific Inquiry* survey was meant to account for participants' attitudes towards instructional tools in the science classroom. The participants' attitudes were not significantly related to any of the measures of conceptual understanding across either timepoint. Therefore, it is possible that, in terms of preference for digital or physical instruction, the pandemic did not impact the study's instrument choices nor the results from the use of those instruments.

This study did not intend to measure the impacts of the pandemic and therefore the results cannot be used to measure any influence of the disjointed curriculum delivery prior to the academic schoolyear in which the study took place. Although it is undeniable that the pandemic disrupted the academic lives of the participants, the disruption impacted all of the participants of this study.

Moreover, it was important to use a science concept that would be presented at initial learning. Since the disruptions to the science curriculum occurred prior to the academic year in which the study took place, the DNA molecule was a concept that had not previously been covered in the science curriculum and would not be formally covered until the latter part of the school year. These considerations supported that the DNA molecule would be an appropriate topic for the study and the pandemic may have helped prevent any contamination of prior formal lessons.

5.6 Overall Analysis and Literature Integration

The results of this study indicate that when students create physical, three-dimensional models they are better able to construct accurate mental models when compared with students who create digital, two-dimensional models as demonstrated by immediate and long-term conceptual understanding measures. The theoretical framework of Piagetian constructivism largely supports this finding. Since this was the initial learning experience with the DNA molecule, students assimilated the image of DNA through viewing the video and accommodated the new information about the DNA molecule through model building (Sjoberg, 2007). When considering assimilation and accommodation, students who built a three-dimensional model of DNA were better able to construct schema of a mental model that represented the three-dimensional double-helix shape of the DNA. This was demonstrated at delayed post-test for the objective quiz and post-test and delayed post-test of the drawing and explanation measures. Additionally, the physical model affords an accurate referent for empirical and reflective abstractions because of the component and spatial accuracy of the physical model with the abstract concept. At a foundational level, constructivism supports the importance of the use of physical models for mental model construction (Baddeley, 1986; Piaget, 1970a). In addition to the overarching framework of constructivism, the result of the physical model construction as a better method for fostering appropriate and accurate mental models and conceptual understanding is supported by theories of working memory, transfer, cognitive load, embodied cognition, and spatial reasoning. These theories and psychological processes are further bolstered by neuroscience research that demonstrates physiological impacts of physical interactions when learning science concepts.

Both physical model and digital model construction required haptic interactions on the part of the student. However, it is clear that the construction of the physical model necessitated a greater degree of interactivity than the touchscreen digital model since the digital model only required tapping and swiping on a touchscreen and the physical model required fitting together foam pieces to create a larger structure. Embodied cognition supports the need for physical interactions with the learned content to achieve greater conceptual understanding and long-term memory. Since the advent of educational technology, researchers have demonstrated the importance of incorporating embodied cognition and haptic interactions with digital tools (Black et al., 2012; Russo-Johnson et al., 2017; Swart et al., 2015; Xie et al., 2018). However, the amount and complexity of the physical interactions and manipulations with the material was demonstrated to be positively correlated with conceptual understanding (Schwartz & Plass, 2014). The need for physical interactions and stimulating embodiment is further supported by studies conducted by Kontra et al. (2015) and Preece et al. (2013) which suggest that physical interactions with physical models outperform other pedagogical methods including digital tools. These prior studies lay the foundation for the current study's findings of the benefits of physical model construction due to the extended level of interactivity and, by extension, embodiment of the concept when compared with interactions with a two-dimensional touchscreen device.

In addition to the theoretical support of embodied cognition, neuroscience research provides empirical physiological evidence of the impact of haptic interactions on recall and long-term memory. Stock et al. (2008), Calvo-Merino et al. (2005), Beilock et al. (2008), and James and Swain (2011) all found that the regions of the brain that are activated during physical interactions and movements are also activated when thinking about the action at a later date whether through reading or viewing pictures or videos. Similarly, Cetron et al. (2019), Sathain (2016), and Mason and Just (2015) found similar brain activation of the occipital and parietal lobes, which are responsible for auditory and visual stimuli, as well as the motor cortex at learning and recall specifically with STEM concepts. Although brain activity was not measured in the current study, the results suggest the potential that similar brain regions were activated during recall which led to more accurate drawings, better written explanations, and higher objective achievement scores due to the physical model building activity.

Theories on mental model construction also support the current study's findings. The concept of transfer can explain several of the results due to the difficulty of far transfer and the relative ease of near transfer (Barnett & Ceci, 2002). The near transfer of the control group's worksheet to paper objective test may explain the high objective test score at post-test, while near transfer of three-dimensional physical models to three-dimensional mental models may explain why physical model construction was beneficial at several of the measures. Alternatively, the far transfer of two-dimensional model building with three-dimensional mental model construction may explain the digital model group's lower means as suggested by the work of Dan and Reiner (2007). This is further supported by Tulving's (1983) encoding specificity hypothesis that suggests the material presented at initial learning must accurately represent the content to be accessed in the long-term. Additionally, qualitative analysis of the drawings constructed by the digital group supports this assumption. At the delayed post-test timepoint, the digital group's drawings overwhelmingly resembled a flat ladder-like structure. The digital simulation that was used for this group resulted in a flat ladder-like structure for the final DNA molecule model. This inaccurate spatial representation of the model resulted in lower drawing scores as specified in the model drawing rubric. Interestingly, the control group's drawings more frequently resembled the three-dimensional shape of the double helix. This resulted in higher drawing scores as specified by the Rubric for Model Drawing (Appendix K). Since the control group did not use the digital simulation or the physical model, it is possible that their mental model construction was based on the only full visual image of the DNA molecule they experienced during the study, which was a virtual, three-dimensional image of a double helix model of DNA presented in the instructional video. Even though all participants viewed the instructional video, the images presented in the video may not have persisted with the digital group. A possible explanation is that, during accommodation, the mental model representation of the DNA molecule created while watching the video was replaced with the digital simulation's visual representation. This influence has significant implications for presenting three-dimensional concepts with a two-dimensional medium such that the two-dimensional representation should not be used at initial learning experience. As supported by theory of transfer, the far transfer of dimensionality requires more advanced knowledge of the concept to effectively execute the change (Barnett & Ceci, 2002; Barr, 2010).

In addition to transfer, cognitive load also influences mental model construction (Baddeley, 1999; Paivio, 1986; Sweller & Chandler, 1994). Cognitive load theory is especially important to consider when presenting science concepts. The topic of DNA requires an understanding of component parts, spatial arrangement, as well as microbiological functions. Therefore, the topic itself has high intrinsic load which leads to a need to limit extraneous and germane load. Dan and Reiner (2017) suggest utilizing authentic tasks, in this case physical model building, to help limit cognitive load, and Huk (2007) found that for initial learning appropriate scaffolding and experiences can lead to appropriate mental model construction. Physical model building allows for limited extraneous load since the only point of focus is the component pieces that fit together and also limits germane load through the near transfer of dimensionality which in turn would require less of the working memory (Chandler & Sweller, 1991; van Merrienboer & Ayres, 2005). Digital model construction does not limit extraneous load due to other images and functions of the simulation and also does not limit germane load due to the learning curve of using the simulation on a touchscreen and the far transfer to a three-dimensional mental model.

The images presented at initial learning have a lasting impact on conceptual understanding because they directly impact mental model construction (Dickmann et al., 2019; Stieff, 2010). The importance of the mental model and images is demonstrated in Kosslyn's (2005) theory of mental imagery which suggests the ability to visualize an object or concept without anything physically present. Bruyer and Scailquin (1998) describe that the accuracy of a mental model is dependent upon component parts of the image as well as spatial information about the image which can later be recalled. The process of physical model building provides an accurate physical image of component parts as well as spatial arrangement to allow for mental imagery and mental models. Hegarty and Stull (2012) further explain the importance of visuospatial thinking in regard to mental imagery and the ability to manipulate and rotate spatial objects and concepts without the concept present, which is especially important for three-dimensional, abstract science concepts.

The DNA molecule is inherently spatial due to the three-dimensional nature of the double-helix shape, and therefore, the interaction of spatial concepts and working memory is necessary to consider. For example, the two-dimensional digital model may not have produced the highest conceptual understanding scores due to the greater degree of working memory required to create a three-dimensional mental model as suggested by Mayer and Moreno (2003). The results of this study support the concept of the interaction of spatial ability and working memory since the participants' mental rotation test score was a significant covariate for objective quiz and drawing measures. Not only does this suggest that strong spatial reasoning skills support the learning of three-dimensional concepts but also supports that spatial ability is an important factor in science learning and achievement (Nagy-Kondor & Esmailnia, 2022; Wai et al., 2009; Wu & Shah, 2004).

Finally, the *Attitudes to Scientific Inquiry* survey did not demonstrate significance as a covariate at any of the conceptual understanding measures or timepoints with the exception of degree of forgetting of explanation scores. This finding is similar to other research that expected attitudes to be correlated with career interest or determined by a specific pedagogical technique and found no significance (Areepattamannil et al., 2020; Cleveland et al., 2017; Saw et al., 2019). While there was just one instance of significance in the present study, it is possible that a longer intervention or an intervention that was associated with academic performance that held weight on grade point average may have exposed relationships of attitudes and conceptual understanding (Perez et al., 2019; Young et al., 2019).

5.7 Implications

5.7.1 Theoretical Implications

This study joins the body of research in the areas of constructivism, embodied cognition, transfer, cognitive load, and spatial ability. It also pushes forward several of these theories by taking a novel approach to the methodology of this study. While studies have examined the difference between physical tools and digital tools, there have been few if any studies that examined physical models and digital models on a touchscreen device. Arguably, the touchscreen requires more haptic interactions than the keyboard and mouse on more traditional computers which adds to the literature on haptic interactions with digital tools. Additionally, measuring spatial ability as a potential covariate for the impact on the difference of physical and digital instruction has not extensively been examined. Through the use of three measures of conceptual understanding, the results imply the importance of spatial ability for objective and drawing measures but not explanations. This opens an avenue for further investigation on the impact of spatial ability on physical or digital pedagogical tools as well as in the science classroom.

Additional methodological considerations include location and time. This study did not take place in a lab but rather in schools demonstrating a true field-based experiment. The results must be considered in this context with implications of generalizability in other classroom settings. Another consideration is the inclusion of a long-term measure of two months. While this is not the first study to occur in schools or over the course of two months, combined and in the context of measuring the difference between physical and digital models on a topic that had not been formally covered in class adds a nuance to the field.

Importantly, this study provides data on this topic during a global pandemic. Compared with other studies that showed similar results, this study demonstrates both the resilience of adolescents as well as the strength of theories such as embodied cognition, transfer, and cognitive load. Even through unusual circumstances and interruptions to traditional schooling and curriculum, similar results were found.

5.7.2 Practical Implications

The practical implications are a direct result of the theoretical implications and data analysis. Any modifications to real-world implementation must consider the theoretical and cognitive impacts and should be based on empirical evidence. This study specifically measured the impact of the use of touchscreen devices on conceptual understanding of an abstract science concept at initial learning experience. Through the measurement of conceptual understanding with an objective quiz, model drawing, and written model explanation, the overall conclusions drawn suggest that the limitations of the touchscreen device impact conceptual understanding at the short-term and long-term timepoints as well as across time. While this study does not overtly suggest removing digital simulations as a method of teaching and learning, it does suggest that digital simulations should not be used at initial concept development when a relevant physical model is available. The digital simulations' usefulness as an opportunity to haptically interact with the material is an improvement over the traditional paper worksheets but is limited when considering different assessment measures. Therefore, the use of digital simulations on a touchscreen device should be seriously considered before implementation.

While the digital model construction demonstrated some improvement over paper worksheets, the results of this study support the use of physical, three-dimensional models as a more beneficial tool in conceptual understanding of abstract science concepts at initial learning experience. At nearly all of the measures of conceptual understanding and across time, the physical model construction was demonstrated to provide the greatest conceptual understanding at the immediate post-test and delayed post-test timepoints as well as for retention.

The primary implication of this study is the incorporation of more physical, hands-on interactions with content in the science classroom at initial learning. Over the past several years, physical models have been replaced with digital models which were seen as equivalent tasks. However, the results of this study demonstrates that physical model construction and digital model construction do not produce equivalent results at any of the three measures of conceptual understanding. The theoretical considerations of difference in haptic interactions, embodied cognition, and mental model construction support this comparison (Han & Black, 2011; Johnson-Laird, 2012; Kosslyn, 2005). This study does not suggest that there is no place for digital simulations with touchscreen devices in the science classroom. Rather, the use of touchscreens as a haptic tool is important and should be used, provided it is not at the expense of a physical model during the initial learning experience (Russo-Johnson et al., 2017; Schwartz & Plass, 2014; Tarasuik et al., 2017). An additional consideration is the appropriate use, timing, and scaffolding of digital models. While the digital model construction on touchscreen devices did not support conceptual understanding at initial learning, digital models may be beneficial after initial concept development with the use of other pedagogical tools such as physical models.

Science classrooms should be places for exploration and invention where students use technology to enrich their learning. This study examined touchscreen devices due to their ubiquitous use in modern classrooms. The haptic limitations of touchscreen devices, which are specific to touchscreens, may have been a primary factor in the digital model group's relatively poor performance in demonstrating conceptual understanding when compared with the physical model group. What may have been lacking in the digital treatment was the disconnect of the physical and virtual interaction. Technology, however, is constantly advancing and evolving. The results of this study illuminate the requirements of useful pedagogical technological tools. Namely, the technology that is being developed for educational use must incorporate more meaningful haptic interactions that engage students in fuller sensory perception. Devices that provide individuals with access to augmented reality may support more complex and intricate haptic experiences. Augmented reality relies on a digital overlay of information onto the physical world. Presently, the primary device for augmented reality use is the touchscreen smartphone or tablet computer to view digital content on the real world through a combination of the rear facing camera and the screen. However, as augmented reality technology evolves, the possibility of engaging with stimuli that accurately represents the physical three-dimensional world will be more widespread. The results from the current study suggest that increasing the degree of sensory interactivity can support both short- and long-term conceptual understanding. Moreover, greater embodied experiences could personalize learning and offer a more salient involvement with the subject matter. Therefore, providing opportunities to physically engage with content that can be adequately and appropriately expressed with digital representations, such as submicroscopic molecules, can enhance the learning experience. Whether technology innovation involves manipulating a physical object that has a digital consequence in augmented reality or manipulating a digital environment with the use of haptic feedback gloves in a physical environment, the future of pedagogical technology should be encouraged. Science classrooms should be places for exploration and invention. However, an important consideration for technology implementation is that it is equitable, transformative, and sustainable. The touchscreen tablet has been a popular device for over a decade, and they are still not available in every classroom. Educational technology comes at a steep price, and an education student benefit-based cost analysis should be considered before a new technological tool is implemented.

The results of this study also have implications on assessment measures and the importance of spatial ability. Conceptual understanding was measured using three different tools: objective quiz, model drawing, and model explanation. These three measures were used because one of those measures alone does not accurately assess an individual's conceptual understanding (Nehm & Schonfeld, 2008; Perkins, 1998; Rodriguez et al., 2018). Despite this assessment, the overwhelming majority of student achievement is measured by responses to objective tests, including standardized exams. The additional measure of drawing to demonstrate conceptual understanding of an abstract, spatial topic taught with visuals bolsters the arguments made by Babilonia-Rosa et al. (2018) to use drawings to measure understanding of the sciences concept when taught with physical models. Furthermore, explanations as a measure of conceptual understanding is supported by Fiorella et al. (2020) who suggest explanations are a more appropriate measure of learning when the content is highly-visual.

The results of this study demonstrate that there are differences among these measures as supported by the different treatment groups exhibiting greater conceptual understanding at different timepoints based on the different assessments of conceptual understanding. The use of traditional objective questions, drawings, and written explanations provided an opportunity to examine the differences of these measures based on pedagogical technique and conceptual understanding outcome. The implication is that educators and high-stakes test creators should consider what they are trying to measure as well as the nature of the content being measured since not all assessment tools deliver the same results of conceptual understanding. For example, producing a drawing requires different cognitive skills than producing a written explanation as demonstrated by the significance of spatial ability as a covariate for drawing score and not for the written explanation score. Since there is a direct link between instruction and assessment, the instructional tools and the assessment tools must match (Kontra et al., 2015). The results of this study may serve as a guide to appropriate instructional tools for assessment measures. For example, for purely objective assessments, both physical and digital model construction would be beneficial for long-term assessment when compared with traditional paper worksheets.

Finally, an individual's spatial ability was significantly related to the impact of treatment for all of the objective quiz and drawing measures of conceptual understanding at the two timepoints and across time as well as the explanation measure of conceptual understanding across time. These results suggest the importance in spatial ability when learning about abstract science concepts especially for objective quiz and drawing assessment measures. Therefore, a direct practical implication of this observation should be to increase spatial ability training for students enrolled in science classes. Science is inherently spatial and requires the exercise of one's spatial ability to manipulate three-dimensional concepts both physically and mentally (Ness et al., 2017; Wai et al., 2009; Wu & Shah, 2004). Therefore, if spatial ability can influence conceptual understanding, it is imperative that every student has the opportunity to practice with and increase their spatial ability skills.

5.8 Limitations

There were several limitations in this study that will be addressed in future studies. The current study only measures conceptual understanding for one abstract science concept. In order to make more generalizable claims about the importance of physical modeling, more abstract science concepts should be tested. Additionally, only one age group was measured. The results of this study should be considered for ninth-grade Living Environment students and may not be applicable to other age groups enrolled in other science courses. Finally, the pandemic was also a limitation in terms of obtaining schools' approval for participation in the study as well as attendance and attrition in the study. There were several participants who were unable to participate in either the immediate post-test or delayed post-test outcome measures due to Covid-related absence from school.

5.9 Future Studies

The results of this study demonstrate a need for further research on the topic of touchscreen device use at initial learning of abstract science concepts. The immediate future study is to measure conceptual understanding at an even later timepoint after formal instruction of the topic. The DNA molecule is typically not covered until the latter part of the academic school year. It would be interesting to measure if there was a difference in unit test scores after formal learning of DNA based on model treatment group at initial learning. This study may support the importance of physical model construction even further than the current study with the expected outcome that students who used physical models at the initial learning stage would outperform the students who learned with digital models.

While DNA was chosen as an exemplary abstract science concept, future studies involving different abstract science concepts may help support the present study's findings. A future study with the topic of topographic maps would be beneficial to measure the effectiveness of physical manipulation compared with digital touchscreen manipulation on a difficult earth science topic. Topics in chemistry and physics can also be explored to adequately survey the four main science disciplines covered in high school.

In addition to measuring different science concepts, it would also be beneficial to sample different age groups. While the participants in the current study were high school students, this study may yield different results in the elementary or middle school setting. Other than age serving as a potential factor, a similar study with younger students would further ensure that the topics would be presented at an initial learning experience since there would have been fewer opportunities for the concept to have been taught in a formal learning environment.

Finally, the study should be replicated in a time beyond the pandemic. Although schools are attempting to operate as they were before the pandemic, there are many protocols in place that may influence individuals' participation in a study. A future study in a classroom that more closely resembles the classrooms prior to the pandemic may yield interesting results.

References

Adelantado-Renau, M., Moliner-Urdiales, D., Cavero-Redondo, I., Beltran-Valls, M. R., Martinez-Vizcaino, V., & Alvarez-Bueno, C. (2019). Association between screen media use and academic performance among children and adolescents: A systematic review and meta-analysis. *Journal of American Medical Association Pediatrics*, *173*(11), 1058-1067.

Alexander-Bloch, A., Raznahan, A., Bullmore, E., & Giedd, J. N. (2013). The convergence of maturational change and structural covariance in human cortical networks. *Journal of Neuroscience, 33*, 2889-2899.

Almasri, F., Hewapathirana, G. I., Ghaddar, F., Lee, N., & Ibrahim, B. (2021). Measuring attitudes towards biology major and non-major: Effect of students' gender, group composition and learning environment. *PLoS ONE*, *16*(5), e0251453. https://doi.org/10.1371/journal.pone.0251453.

Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., Sandhu, R., & Sharma, S. (2013). Maturation of the adolescent brain. *Neuropsychiatric Disease and Treatment, 9*, 449-461.

Areepattamannil, S., Cairns, D., & Dickson, M. (2020). Teacher-directed versus inquiry-based science instruction: Investigating links to adolescent students' science dispositions across 66 countries. *Journal of Science Teacher Education*, *31*(6), 675-704.

Babilonia-Rosa, M. A., Kuo, H. K., & Oliver-Hoyo, M. T. (2018). Using 3D printed physical models to monitor knowledge integration in biochemistry. *Chemistry Education Research and Practice*, *19*(4), 199-1215.

Baddeley, A. (1986). Working memory. New York, NY: Clarendon/Oxford University Press.

Baddeley, A. (1998). Human memory. Boston: Allyn and Bacon.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612-637.

Barr, R. (2010). Transfer of learning between 2D and 3D sources during infancy: Informing theory and practice. *Developmental Review, 30*, 128-154.

Barrett, T. J., & Hegarty, M. (2016). Effects of interface and spatial ability on manipulation of virtual models in a STEM domain. *Computers in Human Behavior*, *65*, 220-231.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 1-21.

Black, J. B., Segal, A., Vitale, J., & Fadjo, C. (2012). Embodied cognition and learning environment design. In D. Jonassen and S. Lamb (Eds.). *Theoretical foundations of student-centered learning environments*. New York: Routledge.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.

Bruyer, R., & Scailquin, J-C. (1998). The visuospatial sketchpad for mental images: Testing the multicomponent model of working memory. *Acta Psychological, 98*(1), 17-36.

Bryan, R. R., Glynn, S. M., & Kittleson, J. M. (2011). Motivation, achievement, and advanced placement intent of high school students learning science. *Science Education, 95*(6), 1049-1065.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex, 15*, 1243-1249.

Cetron, J. S., Connolly, A. C., Diamond, S. G., May, V. V., Haxby, J. V., & Kraemer, D. J. M. (2019). Decoding individual differences in STEM learning from functional MRI data. *Nature Communications, 10,* 1-10.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Learning, 8*(4), 293-332.

Chen, Z., & Siegler, R. S. (2013). Young children's analogical problem solving: Gaining insights from video displays. *Journal of Experimental Child Psychology, 116*, 904-913.

Clements, D. H. (1999). Geometric and spatial thinking in young children. In J. Copley (Ed.), *Mathematics in the early years* (pp. 66-79). Reston, VA: NCTM Press.

Cleveland, L. M., Olimpo, J. T., & DeChenne-Peters, S. E. (2017). Investigating the relationship between instructors' use of active-learning strategies and students' conceptual understanding and affective changes in introductory biology: A comparison of two active-learning environments. *CBE - Life Science Education, 16*(2). https://doi.org/10.1187/cbe.16-06-0181.

Dan, A., & Reiner, M. (2017). Reduced mental load in learning a motor visual task with virtual 3D method. *Journal of Computer Assisted Learning, 34*, 84-93.

Dewey, J. (1902). *The child and the curriculum*. Chicago: University of Chicago Press.

Dewey, J. (1938). *Experience and education*. New York: Colliers Book.

diSessa, A. A., Gillespie, N., & Esterly, J. (2004). Coherence vs. fragmentation in the development of the concept of force. *Cognitive Science, 28*, 843-900.

Dickmann, T., Opfermann, Dammann, E., Lang, M., & Rumann, S. (2019). What you see is what you learn? The role of visual model comprehension for academic success in chemistry. *Chemistry Education Research and Practice, 20*, 804-820.

Eichenbaum, H. (1997). Declarative memory: Insights from cognitive neurobiology. *Annual Review of Psychology, 48*, 547-572.

Farenga, S. J., & Joyce, B. A. (1999). Intentions of young students to enroll in science courses in the future: An examination of gender differences. *Science Education, 83*(1), 55-75.

Farenga, S. J., & Joyce, B. A. (2001). Hardware versus brainware: Where are technology dollars being invested?. *Journal of Technology and Teacher Education, 9*(3), 313-319.

Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2020). Fostering generative learning from video lessons: Benefits of instructor-generated drawings and learner-generated explanations. *Journal of Educational Psychology, 112*(5), 895-906.

Forbes, T. A., & Gallo, V. (2017). All wrapped up: Environmental effects on myelination. *Trends in Neuroscience, 40*(9), 572-587.

Fraser, B. J. (1978). Development of a Test of Science-Related Attitudes. *Science Education, 62*(4), 509-515.

Garofalo, S. G., & Farenga, S. F. (2021). Cognition and spatial concept formation: Comparing non-digital and digital instruction using three-dimensional models in science. *Technology, Knowledge, and Learning, 26*(1), 231-241.

Gibson J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Giedd, J. N., Raznahan, A., Alexander-Bloch, A., Schmitt, E., Gogtay, N., & Rapoport, J. L. (2015). Child psychiatry branch of the National Institute of Mental Health longitudinal structural magnetic resonance imaging study of human brain development. *Neuropsychopharmacology, 40*, 43-49.

Gilbert, J. K., & Justi, R. (2016). Towards authentic learning in science education. In J. K. Gilbert & R. Justi (Eds.). *Modelling-based teaching in science education.* Switzerland: Springer.

Gray, L., & Lewis, L. (2021). Use of educational technology for instruction in public schools: 2019–20 (NCES 2021- 017). *U.S. Department of Education*. Washington, DC: National Center for Education Statistics.

Han, I., & Black, J. B. (2011). Incorporating haptic feedback in simulations for learning physics. *Computers and Education, 57*, 2281-2290.

Harle, M., & Towns, M. (2011). A review of spatial ability literature, its connection to chemistry, and implications for instruction. *Journal of Chemical Education, 88*(3), 351-360.

Hegarty, M., & Stull, A. T. (2012). Visuospatial thinking. *The Oxford Handbook of Thinking and Reasoning*, 606-630.

Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics- Physics Education Research, 3*(2), 020102.

Hirsch-Pasek, K., Zosh, J. M., Michnick Golinkoff, R., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest, 16*(1), 3-34.

Holme, T. A., Luxford, C. J., & Brandriet, A. (2015). Defining conceptual understanding in general chemistry. *Journal of Chemical Education, 92*, 1477-1483.

Holt, L. E., & Beilock, S. L. (2006). Expertise and its embodiment: Examining the impact of sensorimotor skill expertise on the representation of action-related text. *Psychonomic Bulletin & Review, 13*(4), 694-701.

Horowitz-Kraus, T., & Hutton, J. S. (2017). Brain connectivity in children is increased by the time they spend reading books and decreased by the length of exposure to screen-based media. *Acta Paediatrica, 107*, 685-693.

Horz, H., & Schnotz, W. (2010). Cognitive load in learning with multiple representations. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory*. New York: Cambridge University Press.

Howard-Jones, P. (2009). *Introduction to neuroeducational research: Neuroscience, education and the brain from contexts to practice.* Routledge.

Huang, S. C. D., Vea, T., & Black, J. B. (2011). Learning abstract physics system with a 3-D force feedback joystick. *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2011*, pp. 1618-1624. Chesapeake, VA: AACE.

Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of Computer Assisted Learning, 22*, 392-404.

Hutton, J. S., Dudley, J., Horowitz-Kraus, T., DeWitt, T., & Holland, S. K. (2020). Associations between screen-based media use and brain white matter integrity in preschool-aged children. *Journal of American Medical Association Pediatrics, 174*(1), 1-10.

Institute of Education Sciences. (2013). Common guidelines for education research and development: A report from the Institute of Education Sciences and the National Science Foundation. Washington, DC: U.S. Department of Education.

James, K. H., & Swain, S. N. (2011). Only self-generated actions create sensori-motor systems in the developing brain. *Developmental Science, 14*(4), 673-678.

Johnson-Laird, P. N. (2012). Inference with mental models. In K. J. Holyoak & R. G. Morrison (Eds.), *The oxford handbook of thinking and reasoning* (pp. 134-145). Oxford University Press.

Jones, M. G., Minogue, J., Tretter, T. R., Negishi, A., & Taylor, R. (2006). Haptic augmentation of science instruction: Does touch matter? *Science Education, 90*(1), 111-123.

Kalenine, S., Pinet, L., & Gentaz, E. (2011). The visual and visuo-haptic exploration of geometrical shapes increases their recognition in preschoolers. *International Journal of Behavioral Development, 35*(1), 18-26.

Kempermann, G., Fabel, K., Ehninger, D., Babu, H., Leal-Galicia, P., Garthe, A., & Wolf, S. A. (2010). Why and how physical activity promotes experience-induced brain plasticity. *Frontiers in Neuroscience, 4*, 1-9.

Kontra, C., Lyons, D. J., Fischer, S. M., & Beilock, S. L. (2015). Physical experience enhances science learning. *Psychological Science, 26*(6), 737-749.

Kosslyn, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology, 22*(3/4), 333-347.

Leng, B. D., & Gijlers, H. (2015). Collaborative diagramming during problem based learning in medical education: Do computerized diagrams support basic science knowledge construction? *Medical Teacher, 37*(5), 450-456.

Liu, C-C., Wen, C-T., Chang, H-Y., Chang, M-H., Lai, P-H., Chiang, S-H., …, Hwang, F-K. (2022). Augmenting the effect of virtual labs with “teacher demonstration’ and “student critique” instructional designs to scaffold the development of scientific literacy. *Instructional Science.* https://doi.org/10.1007/s11251-021-09571-4.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58*(1), 25-45.

Mason, R. A., & Just, M. A. (2015). Physics instruction induces changes in neural knowledge representation during successive stages of learning. *NeuroImage, 111*, 36-48.

Mayer, R. E. (2009). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E., & Moreno, R. E. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

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

Möhring, W., & Frick, A. (2013). Touching up mental rotation: Effects of manual experience on 6-month-old infants’ mental object rotation. *Child Development, 84*(5), 1554-1565.

Moore, R., & Burrus, J. (2019). Predicting STEM major and career intentions with the theory of planned behavior. *The Career Development Quarterly, 67*, 139-155.

Moser, A., Zimmermann, L., Dickerson, K., Grenell, A., Barr, R., & Gerhardstein, P. (2015). They can interact, but can they learn? Toddlers' transfer learning from touchscreens and television. *Journal of Experimental Child Psychology, 137,* 137-155.

Nagy-Kondor, R., & Esmailnia, S. (2022). Development of Spatial Ability Extra Tasks (SAET): Problem solving with spatial intelligence. *Quality & Quantity.* https://doi.org/10.1007/s11135-021-01284-7.

National Science Board, National Science Foundation. (2022). *Science and Engineering Indicators 2022: The State of U.S. Science and Engineering*. NSB-2022-1. Alexandria, VA. Available at https://ncses.nsf.gov/pubs/nsb20221.

Nehm, R. H., & Schonfeld, I. S. (2008). Measuring knowledge of natural selection: A comparison of CINS, an open-response instrument, and an oral interview. *Journal of Research in Science Teaching, 45*(10), 1131-1160.

Ness, D., Farenga, S. J., & Garofalo, S. G. (2017). *Spatial intelligence: Why it matters from birth through the lifespan.* New York: Routledge.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books.

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology, 110*(6), 785-797.

Paulus, M. P., Squeglia, L. M., Bagot, K., Jacobus, J., Kuplicki, R., Breslin, F. J., Bodurka, J., Sheffield Morris, A., Thompson, W. K., Bartsch, H., & Tapert, S. F. (2019). Screen media activity and brain structure in youth: Evidence for diverse structural correlation networks from the ABCD study. *NeuroImage, 185,* 140-153.

Pedra, A., Mayer, R. E., & Albertin, A. L. (2015). Role of interactivity in learning from engineering animations. *Applied Cognitive Psychology, 29,* 614-620.

Perez, T., Wormington, S. V., Barger, M. M., Schwartz-Bloom, R. D., Lee, Y-K., & Linnenbrink-Garcia, L. (2019). Science expectancy, value, and cost profiles and their proximal and distal relations to undergraduate science, technology, engineering, and math persistence. *Science Education, 103,* 264-286.

Perkins, D. (1998). What is understanding? In M. S. Wiske (Ed.) *Teaching for Understanding: Linking Research with Practice*. San Francisco: Josey-Bass Publishers.

Piaget, J. (1970a). *Science of education and the psychology of the child.* New York: Orion Press.

Piaget, J. (1970b). *Structuralism*. Basic Books, New York

Piaget, J. (1971). *Biology and knowledge*. Edinburgh University Press.

Piaget, J. (1972). *Psychology and epistemology*. Penguin Press, Harmondworth.

Piaget, J. (1999). *The psychology of intelligence.* London: Routledge.

Popper, K. (1959). *The logic of scientific discovery*. Hutchinson, London.

Preece, D., Williams, S. B., Lam, R., & Weller, R. (2013). "Let's get physical": Advantages of the physical model over three-dimensional computer models and textbooks in learning imaging anatomy. *Anatomical Sciences Education, 6*(4), 216-224.

Putranta, H., & Supahar. (2019). Development of Physics-Tier Tests (PsyTT) to measure students' conceptual understanding and creative thinking skills: A qualitative synthesis. *Journal for the Education of Gifted Young Scientists, 7*(3), 647-775.

Richard, E., Tijou, A., Richard, P., Ferrier, J. L. (2006). Multi-modal virtual environments for education with haptic and olfactory feedback. *Virtual Reality, 10*, 207–225.

Richardson, R. T., Sammons, D., & Del-Parte, D. (2018). Augmented affordances support learning: Comparing the instructional effects of the augmented reality sandbox and conventional maps to teach topographic map skills. *Journal of Interactive Learning Research, 29*(2), 231-248.

Rodriguez, J-M. G., Santos-Diaz, S., Bain, K., & Towns, M. H. (2018). Using symbolic and graphical forms to analyze students' mathematical reasoning in chemical kinetics. *Journal of Chemical Education, 95*, 2114-2125.

Russo-Johnson, C., Troseth, G., Duncan, C., & Mesghina, A. (2017). All tapped out: Touchscreen interactivity and young children's word learning. *Frontiers in Psychology, 8*, 1-15.

Sathain, K. (2016). Analysis of haptic information in the cerebral cortex. *Journal of Neurophysiology, 116*, 1795-1806.

Saw, G. K., Swagerty, B., Brewington, S., Change, C-N., & Culbertson, R. (2019). Out-of-school time STEM program: Students' attitudes toward and career interests in mathematics and science. *International Journal of Evaluation and Research in Education, 8*(2), 356-362.

Scherer, R., Siddiq, F., & Viveros, B. S. (2018). The cognitive benefits of learning computer programming: A meta-analysis of transfer effects. *Journal of Educational Psychology, 111*(5), 764-792.

Schott, C., & Marshall, S. (2018). Virtual reality and situated experiential education: A conceptualization and exploratory trial. *Journal of Computer Assisted Learning, 34*, 843-852.

Schwartz, D. L., & Black, J. B. (1996). Analog imagery in mental model reasoning: Depictive models. *Cognitive Psychology, 30*, 154-219.

Schwartz, R. N., & Plass, J. L. (2014). Click versus drag: User-performed tasks and the enactment effect in an interactive multimedia environment. *Computers in Human Behavior, 33*, 242-255.

Segal, A., Tversky, B., & Black, J. B. (2014). Conceptually congruent actions can promote thought. *Journal of Applied Research in Memory and Cognition, 3*(3), 124-130.

Shulman, L. S., & Tamir, P. (1973). Research on teaching in the natural sciences. In R. M. W. Travers (Ed.), *Second Handbook of Research on Teaching*. Chicago: Rand McNally.

Sjoberg, S. (2007). Constructivism and Learning. In E. Baker, B. McGaw and P. Peterson, (eds.), International Encyclopedia of Education (3rd edition.). Oxford, Elsevier.

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., Dechenne-Peters, S. E., …, & Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science, 359*(6383), 1468-1470.

Stieff, M. (2010). When is a molecule three dimensional? A task-specific role for imagistic reasoning in advanced chemistry. *Science Education, 95*(2), 310-336.

Stock, O., Röder, B., Burke, M., Bien S., & Rösler, F. (2008). Cortical activation patterns during long-term memory retrieval of visually or haptically encoded objects and locations. *Journal of Cognitive Neuroscience, 21*(1), 58-82.

Swart, M., Friedman, B., Kornakasem, S., Black, J., & Vitale, J. (2015). M3-situating embodied learning: Embedding gestures in narratives to learn mathematical fraction sin a digital tablet environment. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.) *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185-233.

Tarasuik, J., Demaria, A., & Kaufman, J. (2017). Transfer of problem solving skills from touchscreen to 3D model by 3- to 6-year-olds. *Frontiers in Psychology, 8*, 1-6.

Treagust, D. (1988). Teaching science for conceptual change: Theory and practice. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 629-646). New York: Routledge.

Tsaparlis, G., & Papaphotis, G. (2009). High-school students' conceptual difficulties and attempts at conceptual change: The case of basic quantum chemical concepts. *International Journal of Science Education, 31*(7), 895-930.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

Tversky, B. (2019). *Mind in motion: How action shapes thought*. New York: Basic Books.

Urhahne, D., Nick, S., & Schanze, S. (2009). The effect of three-dimensional simulations on the understanding of chemical structures and their properties. *Research in Science Education, 39*(4), 495-513.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352-402.

van Merriënboer, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning, educational technology, research and development. *Educational Technology Research and Development, 53*(3), 5-13.

Vandekar, S. N., Shinohara, R. T., Raznahan, A., Roalf, D. R., Ross, M., DeLeo, N., … & Satterthwaite, T. D. (2015). Topologically dissociable patterns of development of the human cerebral cortex. *Journal of Neuroscience, 35*, 599-609.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations: A group test of three dimensional objects. *Science 171*, 701-703.

Villafañe, S. M., Heyen, B. J., Lewis, J. E., Loertscher, J., Minderhout, V., & Murray, T. A. (2016). Design and testing of an assessment instrument to measure understanding of protein structure and enzyme inhibition in a new context. *Biochemistry and Molecular Biology Education, 44*(2), 179-190.

Von Glasersfeld, E. (1991). Abstraction, re-presentation, and reflection: An interpretation of experience and Piaget’s approach. In *Epistemological foundations of mathematical experience* (pp. 45-67). Springer, New York, NY.

Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., … Weiss, S. (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration. *Developmental cognitive neuroscience*, *32*, 4–7.

Vosniadou, S. (2008). Conceptual change research: An introduction. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. xii-xxviii). New York: Routledge.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817-835.

Wason, P. C. (1977). The theory of formal operations: A critique. In B. A. Gerber, (Ed.). *Piaget and Knowing: Studies in Genetic Epistemology* pp. 119-135. Routledge & Kegan Paul, London.

Watson, F. G. (1963). Research on teaching science. In N. L. Gage (Ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally.

Woolfolk, A. E., & McCune-Nicolich, L. (1984). *Educational psychology for teachers: Study guide.* Englewood Cliffs, NJ: Prentice-Hall.

Wu, H-K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, *88*(3), 465-492.

Xie, H., Peng, J., Qin, M., Huang, X., Tian, F., & Zhou, Z. (2018). Can touchscreen devices be used to facilitate young children's learning? A meta-analysis of touchscreen learning effect. *Frontiers in Psychology, 9*, 1-14.

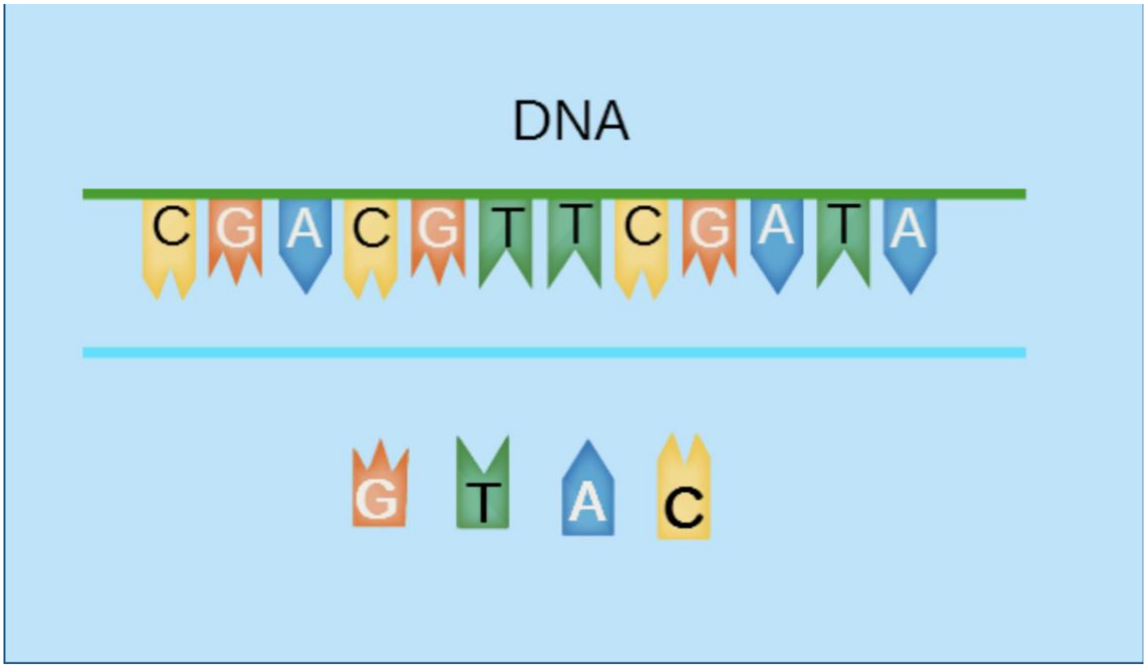
Young, A. M., Wendel, P. J., Esson, J. M., & Plank, K. M. (2018). Motivational decline and recovery in higher education STEM courses. *International Journal of Science Education, 40*(9), 1016-1033.

Zack, E., Gerhardstein, P., Meltzoff, A. N., & Barr, R. (2013). 15-month-olds' transfer of learning between touch screen and real-world displays: Language cues and cognitive loads. *Scandinavian Journal of Psychology*, *54*, 20-25.

Zhao, F., Schnotz, W., Wagner, I., & Gaschler, R. (2020). Texts and pictures serve different functions in conjoint mental model construction and adaptation. *Memory & Cognition, 48*, 69-82.

Appendix A

Digital Model for Pilot Study 1



Appendix B

DNA Information for Pilot Studies 1, 2, and 3

Nucleic Acids

Organic compounds that consist of carbon, hydrogen, oxygen, nitrogen, and phosphorus. Examples: DNA, RNA.

Objectives

- Describe the composition of a nucleic acid.
- List the components of a nucleotide.
- Identify significant differences between DNA and RNA.
- Describe the shape of DNA.
- Summarize significant roles of nucleic acids.

You may have heard that something is "encoded in your DNA." What does that mean?

Nucleic acids. Essentially the "instructions" or "blueprints" of life. Deoxyribonucleic acid, or DNA, is the unique blueprints to make the proteins that give you your traits. Half of these blueprints come from your mother, and half from your father. Therefore, every person that has ever lived - except for identical twins - has his or her own unique set of blueprints - or instructions - or DNA.

Nucleic Acids

A **nucleic acid** is an organic compound, such as DNA or RNA, that is built of small units called **nucleotides**. Many nucleotides bind together to form a chain called a **polynucleotide**. The nucleic acid **DNA** (deoxyribonucleic acid) consists of two polynucleotide chains. The nucleic acid **RNA** (ribonucleic acid) consists of just one polynucleotide chain.

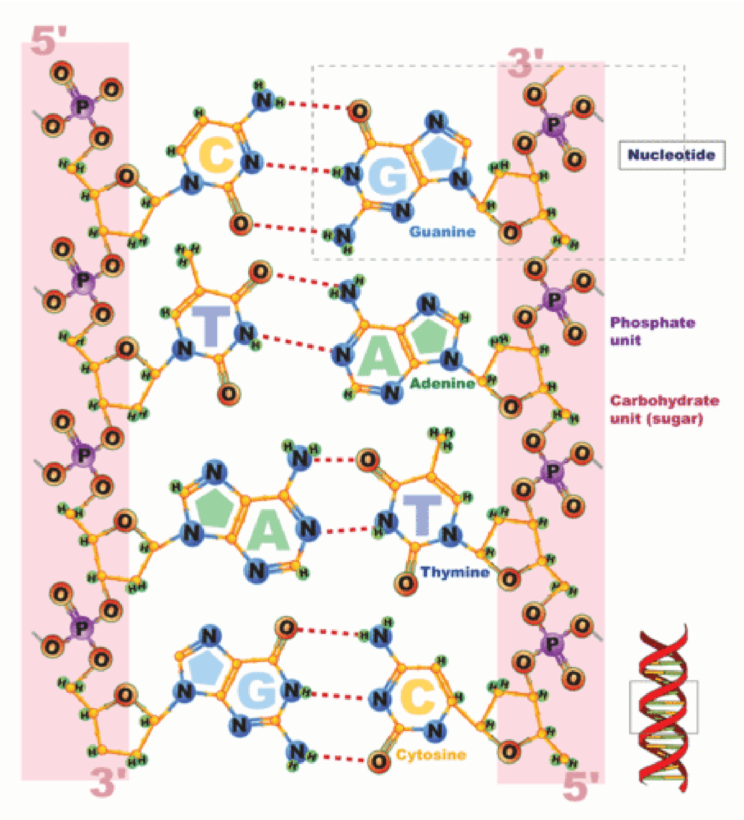
Structure of Nucleic Acids

Each nucleotide consists of three smaller molecules:

1. sugar
2. phosphate group
3. nitrogen base

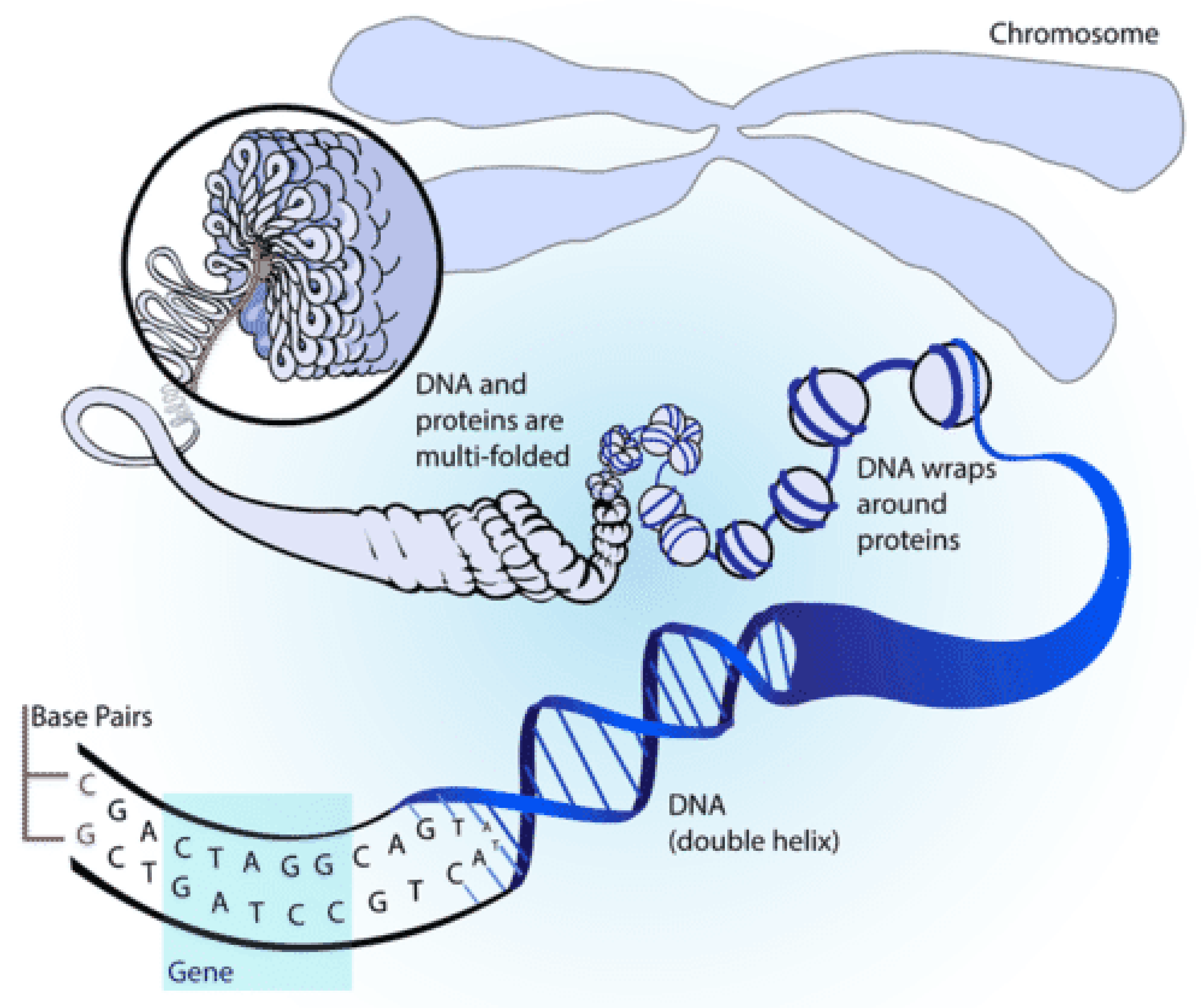
If you look at **Figure** below, you will see that the sugar of one nucleotide binds to the phosphate group of the next nucleotide. These two molecules alternate to form the backbone of the nucleotide chain. This backbone is known as the *sugar-phosphate backbone.*

The nitrogen bases in a nucleic acid stick out from the backbone. There are four different types of bases: cytosine (C), adenine (A), guanine (G), and either thymine (T) in DNA, or uracil (U) in RNA. In DNA, bonds form between bases on the two nucleotide chains and hold the chains together. Each type of base binds with just one other type of base: cytosine always binds with guanine, and adenine always binds with thymine. These pairs of bases are called **complementary base pairs**.



Nucleic Acid. Sugars and phosphate groups form the backbone of a polynucleotide chain. Hydrogen bonds between complementary bases hold two polynucleotide chains together.

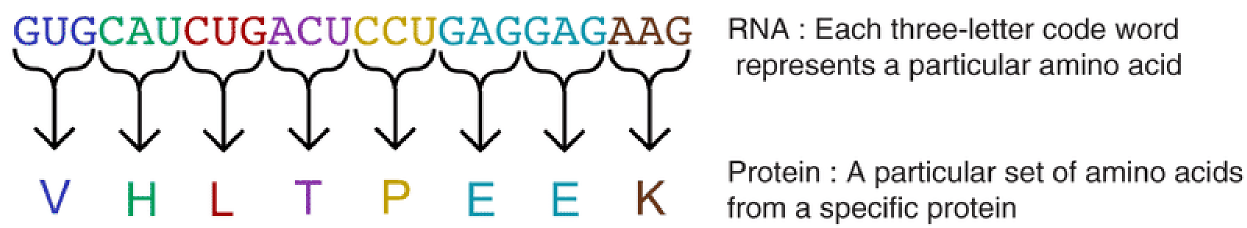
The binding of complementary bases allows DNA molecules to take their well-known shape, called a **double helix**, which is shown in **Figure** below. A double helix is like a spiral staircase. The double helix shape forms naturally and is very strong, making the two polynucleotide chains difficult to break apart.



DNA Molecule. Bonds between complementary bases help form the double helix of a DNA molecule. The letters A, T, G, and C stand for the bases adenine, thymine, guanine, and cytosine. The sequence of these four bases in DNA is a code that carries instructions for making proteins. Shown is how the DNA winds into a chromosome.

Roles of Nucleic Acids

DNA is also known as the hereditary material or genetic information. It is found in genes, and its sequence of bases makes up a code. Between "starts" and "stops," the code carries instructions for the correct sequence of amino acids in a protein (see **Figure** below). DNA and RNA have different functions relating to the genetic code and proteins. Like a set of blueprints, DNA contains the genetic instructions for the correct sequence of amino acids in proteins. RNA uses the information in DNA to assemble the correct amino acids and help make the protein. The information in DNA is passed from parent cells to daughter cells whenever cells divide. The information in DNA is also passed from parents to offspring when organisms reproduce. This is how inherited characteristics are passed from one generation to the next.



The letters G, U, C, and A stand for the bases in RNA. Each group of three bases makes up a code word, and each code word represents one amino acid (represented here by a single letter, such as V, H, or L). A string of code words specifies the sequence of amino acids in a protein.

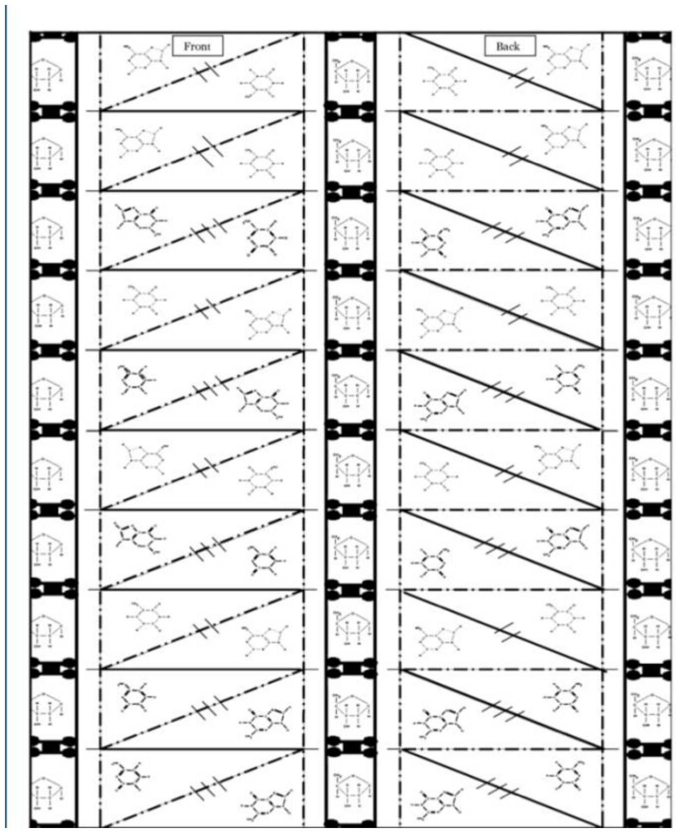
Summary

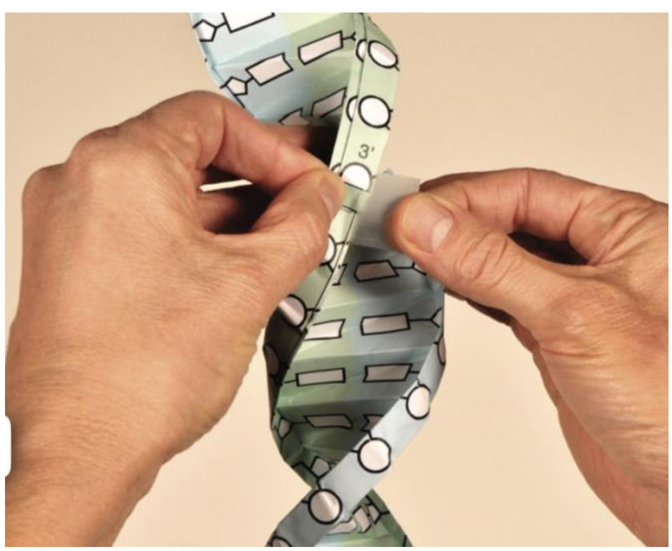
- DNA and RNA are nucleic acids. Nucleic acids are built of small units called nucleotides.
- The bases of DNA are adenine, guanine, cytosine and thymine. In RNA, thymine is replaced by uracil.
- In DNA, A always binds to T, and G always binds to C.
- The shape of the DNA molecule is known as a double helix.

DNA contains the genetic instructions for the correct sequence of amino acids in proteins. RNA uses the information in DNA to assemble the correct amino acids and help make the protein.

Appendix C

Paper Model for Pilot Study 1





Appendix D

Objective-based Test for Pilot Study 1

DNA Objective-Based Questions

Please complete the following questions after you build your model.

1. Which physical feature of DNA is affected by the pairing of purines with pyrimidines?

   Mark only one oval.

   - Its length
   - Its width
   - Its parallel nature
   - Its helical structure

2. Which molecular model describes the structure of the DNA molecule?

   Mark only one oval.

   - Single-stranded and antiparallel
   - Double-stranded and antiparallel
   - Single-stranded and parallel
   - Double-stranded and parallel

3. Double-stranded DNA looks like a ladder that has been twisted into a helix or spiral. The side supports of the ladder are

   Mark only one oval.

   - alternating bases and sugars.
   - alternating bases and phosphate groups.
   - alternating sugars and phosphates.
   - alternating bases, sugars, and phosphates.

4. Thymine binds with

Mark only one oval.

- adenine
- thymine
- guanine
- cytosine

5. Guanine binds with

Mark only one oval.

- adenine
- thymine
- guanine
- cytosine

6. If a sequence of one strand of DNA is ATGATGT, the complementary strand must be ______. (Include correct orientation.)

Mark only one oval.

- ATGATGT
- CGTCGTG
- TACTACA
- GCAGCAC

7. If a sequence in one strand of DNA is 5′-AGCTGCTGA-3′, what is the sequence in the complementary strand? (Include correct orientation.)

Mark only one oval.

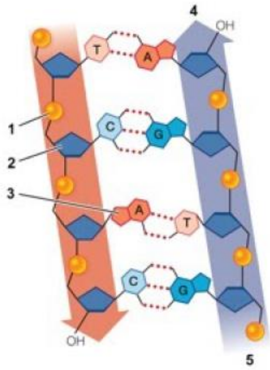
- AGCTGCTGA
- AGCTGCTGT
- TCGATGACT
- TCGACGACT

8. DNA is held together in a double helix by the force of

Mark only one oval.

- the twists.
- covalent bonds.
- ionic bonds.
- hydrogen bonds.

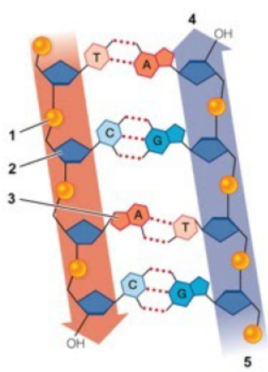
9. Which label corresponds to a phosphate group?



Mark only one oval.

- ⬭ 1
- ⬭ 2
- ⬭ 3
- ⬭ 4
- ⬭ 5

10. Which label corresponds to a deoxyribose sugar?



Mark only one oval.

- ⬭ 1
- ⬭ 2
- ⬭ 3
- ⬭ 4
- ⬭ 5

Appendix E

Scoring Rubric for Model Drawing for Pilot Study 1 and 2

Category	Emerging 1 point	Proficient 2 points	Exemplary 3 points
Sequence	Some nucleotide bases are appropriately paired	Most nucleotide bases are appropriately paired	All nucleotide bases are appropriately paired throughout the entire model
Ratio	The two strands of the DNA molecule does maintain the same distance throughout more than half of the model	The two strands of the DNA molecule maintain the same distance throughout more than half of the model	The two strands of the DNA molecule maintain the same distance throughout the entire model
Two-Dimensional Representation	Unclear that the drawing is meant to represent three-dimensional molecule (looks like a ladder and component parts are missing shape)	Drawing is an acceptable representation of the three-dimensional concept (drawing is twisted but not a double helix and some component parts are missing shape)	Drawing is an exemplary two-dimensional representation of the three-dimensional concept (looks like double helix with appropriate shapes representing component parts)
Accuracy	Model only shows one strand or unclear representation of two strands	Model accurately represent two strands of DNA that do not show a sugar and phosphate in the backbone	Model accurately represent two strands of DNA that consist of a sugar-phosphate backbone attached to nucleotide bases
Conceptual Understanding	Model demonstrates emerging conceptual understanding of the DNA molecule. Less than half of the component parts are included and labeled.	Model demonstrates general conceptual understanding of the DNA molecule. More than half of the component parts are included and labeled.	Model demonstrates overall conceptual understanding of the DNA molecule. All component parts are included and labeled.

Total Score: ___________

Appendix F

Demographic Information Survey for Pilot Studies 2 and 3

Demographic Information

The demographic information collected is used to supply data regarding research participants and is necessary for the determination of whether the individuals in the investigation are a representative sample of the population at-large for generalization purposes. All information will be coded and kept confidential. No individual will be identified in any way in the investigation. All coded information will be destroyed at the end of the project.

1. Email *

   ____

2. Gender

   Mark only one oval.

   - ( ) Female
   - ( ) Male
   - ( ) Prefer not to say
   - ( ) Other: ______________________________

3. Age

   Mark only one oval.

   - ( ) 19-22
   - ( ) 23-25
   - ( ) 26-28
   - ( ) 29-31
   - ( ) > 32

4. Current Occupation

5. Current Location (Country)

6. First Language

7. Undergraduate Major

8. How many credits of biology have you completed to date?

9. How many credits of chemistry have you completed to date?

10. How many credits of physics have you completed to date?

11. How many credits of mathematics have you completed to date?

12. How many credits of psychology have you completed to date?

13. How long has it been since you last took a biology and/or chemistry course?

14. Which of the following materials do you have access to? Please select all that apply.

Check all that apply.

- [ ] Printer (black and white is fine)
- [ ] Scissors
- [ ] Tape
- [ ] Tablet Computer (ex. iPad)
- [ ] Touchscreen laptop
- [ ] Laptop (no touchscreen)

Appendix G

Objective-based Test for Pilot Study 2 and 3

DNA Objective-Based Questions

Please complete the following questions before you build your model as well as after you build your model.

1. Email *

2. Which physical feature of DNA is affected by the pairing of purines with pyrimidines?

   Mark only one oval.

   - ◯ Its length
   - ◯ Its width
   - ◯ Its parallel nature
   - ◯ Its helical structure
   - ◯ Its helix length

3. Which molecular model describes the structure of the DNA molecule?

   Mark only one oval.

   - ◯ Single-stranded and antiparallel
   - ◯ Double-stranded and antiparallel
   - ◯ Single-stranded and parallel
   - ◯ Double-stranded and parallel
   - ◯ Triple-stranded and parallel

4. Double-stranded DNA looks like a ladder that has been twisted into a helix or spiral. The side supports of the ladder are

Mark only one oval.

- individual nitrogenous bases.
- alternating bases and sugars.
- alternating bases and phosphate groups.
- alternating sugars and phosphates.
- alternating bases, sugars, and phosphates.

5. If 30 percent of the bases in a double-stranded DNA molecule are T, _______ percent must be G.

Mark only one oval.

- 20
- 30
- 40
- 50
- 60

6. If a sequence of one strand of DNA is 5′-ATGATGT-3′, the complementary strand must be _______. (Include correct orientation.)

Mark only one oval.

- 5'-ATGATGT-3'
- 5'-TACTACA-3'
- 3'-ATGATGT-5'
- 3'-TACTACA-5'
- 3'-GCAGCAC-5'

7. If a sequence in one strand of DNA is 5′-AGCTGCTGA-3′, what is the sequence in the complementary strand? (Include correct orientation.)

Mark only one oval.

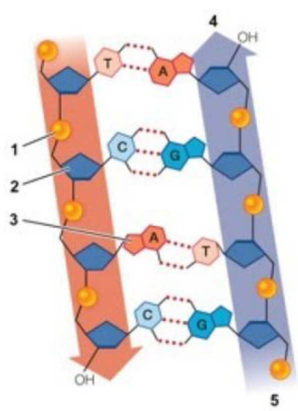
- 5'-AGCTGCTGA-3'
- 3'-AGCTGCTGA-5'
- 5'-TCGACGACT-3'
- 3'-TCGATGACT-5'
- 3'-TCGACGACT-5'

8. DNA is held together in a double helix by the force of

Mark only one oval.

- the twists.
- covalent bonds.
- ionic bonds.
- ionic interactions.
- hydrogen bonds.

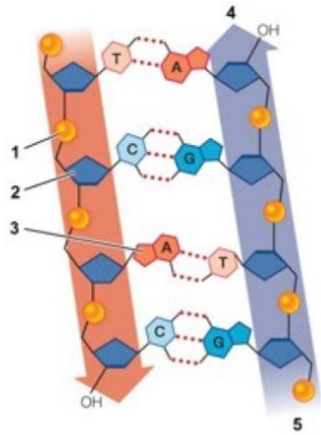
9. Which label corresponds to a phosphate group?



Mark only one oval.

- 1
- 2
- 3
- 4
- 5

10. Which label corresponds to a deoxyribose sugar?



Mark only one oval.

- ( ) 1
- ( ) 2
- ( ) 3
- ( ) 4
- ( ) 5

Appendix H

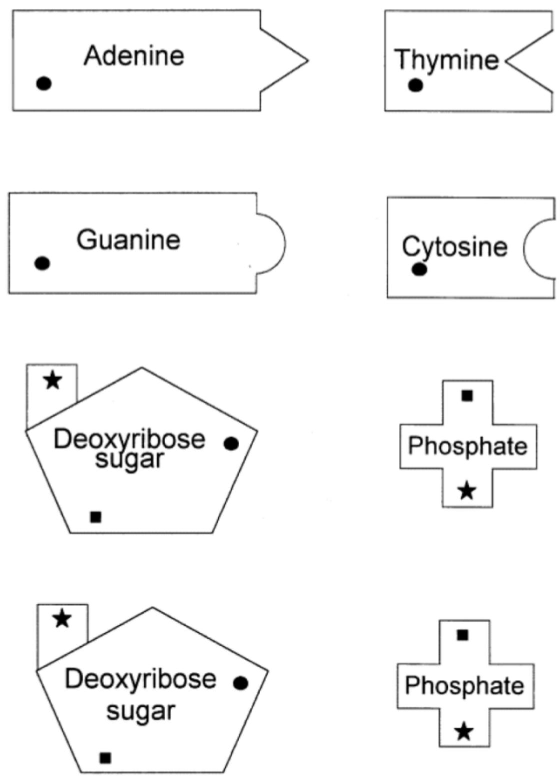
Mental Rotation Test (Vandenberg & Kuse, 1978)

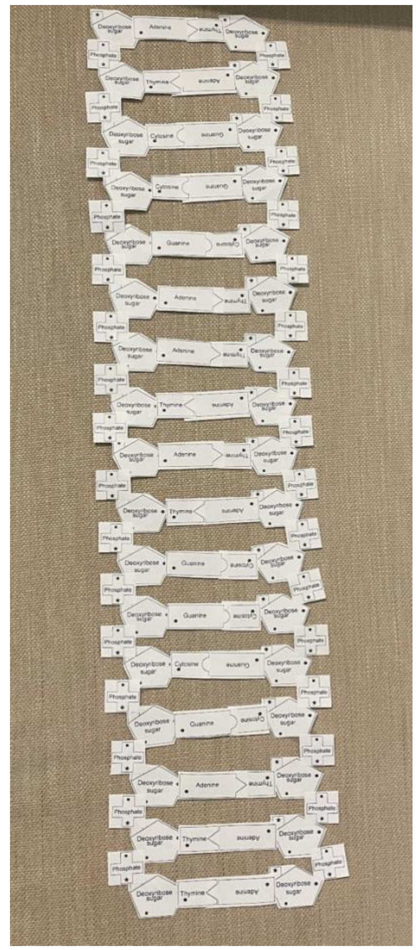
Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations: A group test of three dimensional objects. *Science 171*, 701-703.

Mental Rotation Test
(Vandenberg & Kuse, 1978)

Appendix I

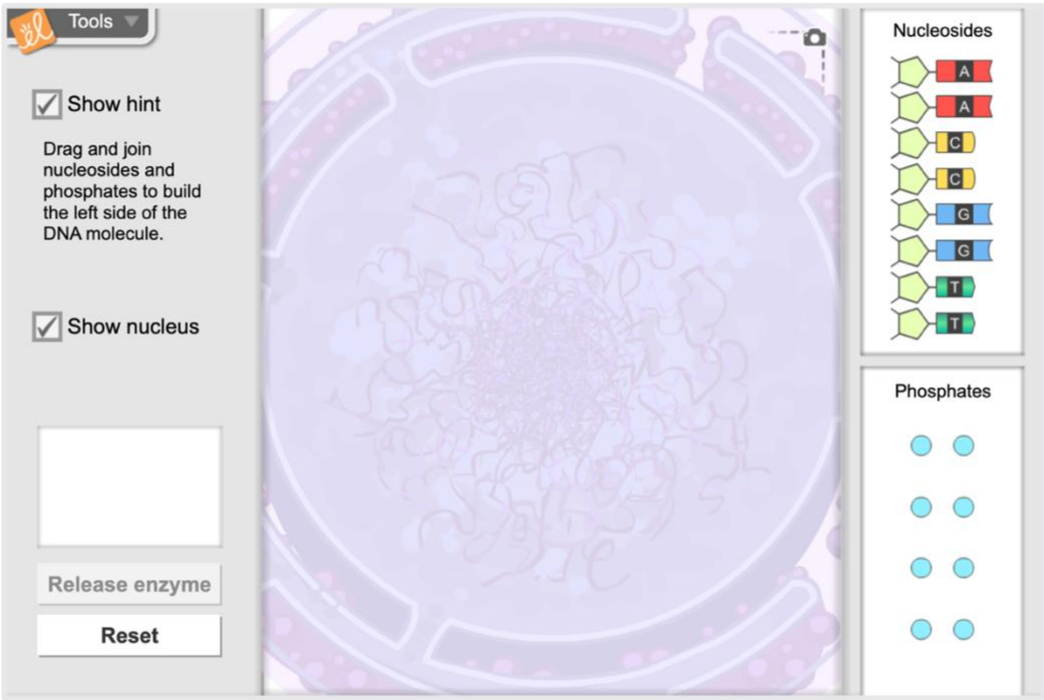
Paper Model for Pilot Study 2 and 3

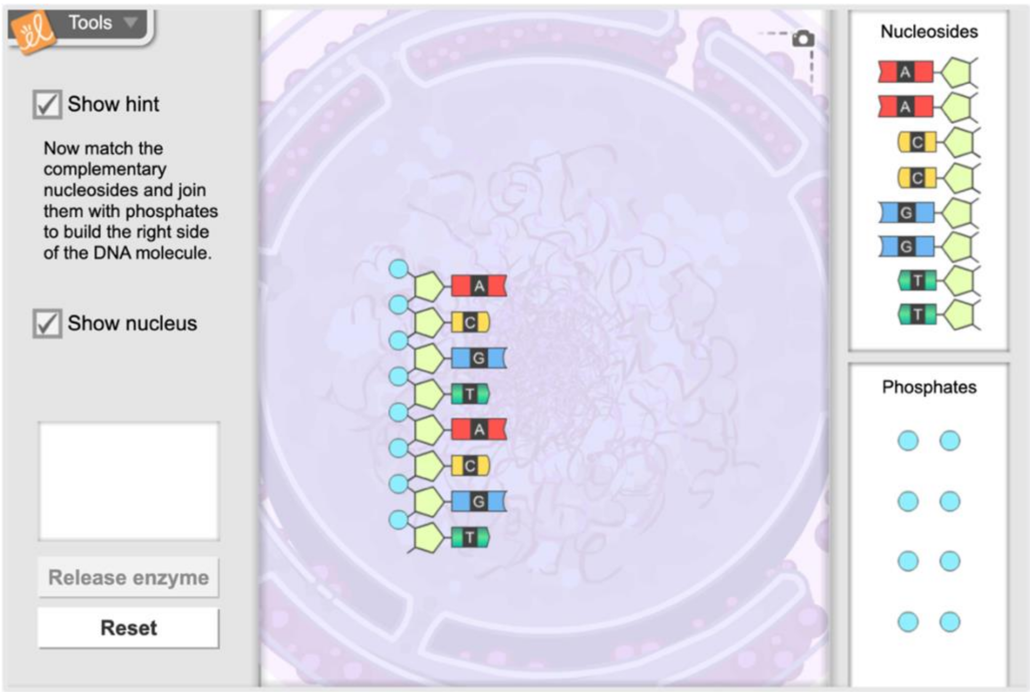


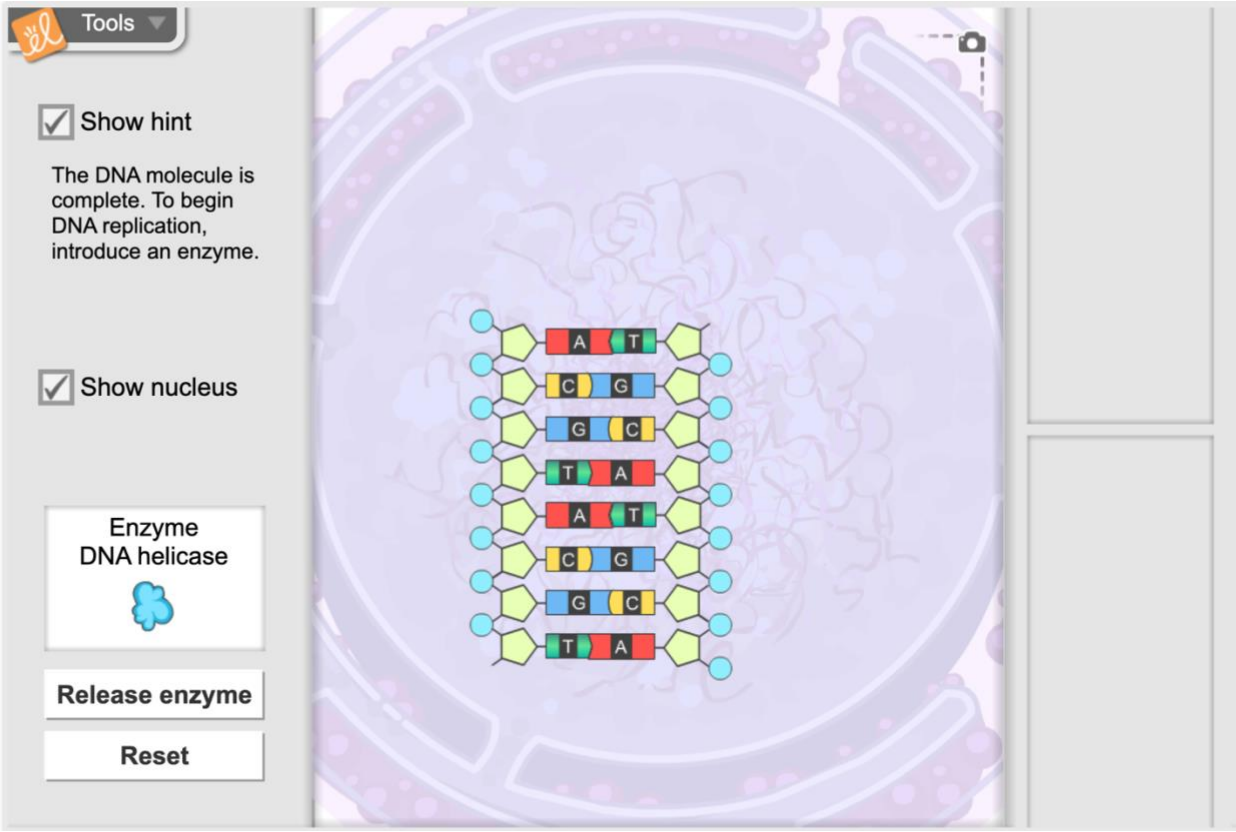


Appendix J

Digital Model for Pilot Study 2, Pilot Study 3, and Dissertation Study






Tools
Show hint
The DNA molecule is complete. To begin DNA replication, introduce an enzyme.
Show nucleus
Enzyme
DNA helicase
Release enzyme
Reset
A
T
C
G
G
C
T
A
A
T
C
G
G
C
T
A

Appendix K

Scoring Rubric for Model Drawing for Pilot Study 3 and Dissertation Study

	Emerging	Proficient	Exemplary
Component Parts	**1-2 points**	**3-4 points**	**5-6 points**
Nucleotide Bases	Some nucleotide bases are shown	Four nucleotide bases are shown without adequate labeling	All four nucleotide bases are shown and labeled correctly and attached by hydrogen bonds
Deoxyribose Sugar	Improper shape and inadequate labeling of deoxyribose sugar	Deoxyribose sugar is present in the backbones	A pentagon is used to represent the deoxyribose sugar in the backbones
Phosphate	Unclear or no use of phosphate in the backbones	Inclusion but no labeling of phosphates in the backbones	Inclusion and labeling of phosphate in the backbones
Two Strands	Model only shows one strand or unclear representation of two strands	Model accurately represent two strands of DNA that do not show a sugar and phosphate in the backbone	Model accurately represent two strands of DNA that consist of a sugar-phosphate backbone attached to nucleotide bases
Pairing of Bases	Some nucleotide bases are appropriately paired	Most nucleotide bases are appropriately paired	All nucleotide bases are appropriately paired throughout the entire model
Spatial Representation	**1-2 points**	**3-4 points**	**5-6 points**
Accurate 2D Representation of 3D Concept	Unclear that the drawing is meant to represent three-dimensional molecule (looks like a ladder and component parts are missing shape)	Drawing is an acceptable representation of the three-dimensional concept (drawing is twisted but not a double helix and some component parts are missing shape)	Drawing is an exemplary two-dimensional representation of the three-dimensional concept (looks like double helix with appropriate shapes representing component parts)

Consistent Spacing of Strands	The two strands of the DNA molecule does maintain the same distance throughout more than half of the model	The two strands of the DNA molecule maintain the same distance throughout more than half of the model	The two strands of the DNA molecule maintain the same distance throughout the entire model
Double Helix (Morphology of Molecule)	Model is not twisted and does not represent double helix shape	Shape of the model is twisted but not a double helix	Shape of the model is a double helix

Total Score: ____________

Appendix L

Learner-Generated Written Explanation Response Rubric

Category	Score	Description
No Response	0	Either the work was not attempted or the work is incorrect, irrelevant, or off task.
Attempted	1-2	The response demonstrates only a minimal understanding of the DNA molecule and a reasonable explanation of the model is not suggested. The response is incomplete, contains major errors, or reveals serious flaws in reasoning.
Emerging	3-4	The response contains evidence of an emerging conceptual understanding of the DNA molecule as demonstrated by use of appropriate academic language that includes some (2-4) of the following terms: nucleotide bases, deoxyribose sugar, phosphate, complementary strand, double helix, antiparallel, hydrogen bonds, and representative shapes (ie. pentagon). However, on the whole, the explanation is not well developed. Although there may be serious flaws in reasoning, the response is somewhat correct.
Proficient	5-6	The response contains evidence of a proficient conceptual understanding of the DNA molecule as demonstrated by use of appropriate academic language that includes the majority (5-7) of the following terms: nucleotide bases, deoxyribose sugar, phosphate, complementary strand, double helix, antiparallel, hydrogen bonds, and representative shapes (ie. pentagon). The response also is generally well developed, but contains some omissions or minor errors.
Exemplary	7-8	The response contains evidence of exemplary conceptual understanding of the DNA molecule as demonstrated by use of appropriate academic language that includes all (8) of the following terms: nucleotide bases, deoxyribose sugar, phosphate, complementary strand, double helix, antiparallel, hydrogen bonds, and representative shapes (ie. pentagon). The response is logically sound, clearly written, and does not contain any significant errors.

Total Score: ____________

Appendix M

Descriptive Statistics for Pilot Study 3

	N	Min	Max	Mean	SD
Spatial Ability	15	2	19	13.53	5.45
Pre-Test	15	0	9	4.07	2.99
Short-Term Post-test	15	2	9	5.4	2.52
Short-Term Drawing	15	1	44	26.33	15.67
Short-Term Explanation	15	0	8	3.73	3.2
Long-Term Post-test	15	2	9	6.13	5.4
Long-Term Drawing	15	11	40	30.87	9.47
Long-Term Explanation	15	0	8	4.27	2.71
Paper Short-Term Post-test	8	2	9	5.38	2.39
Paper Short-Term Drawing	8	1	44	31.63	14.4
Paper Short-Term Explanation	8	0	8	5	2.62
Paper Long-Term Post-test	8	4	9	6.13	1.55
Paper Long-Term Drawing	8	21	40	34.5	6.72
Paper Long-Term Explanation	8	0	8	4	2.39
Digital Short-Term Post-test	7	2	9	5.43	2.88
Digital Short-Term Drawing	7	1	43	20.29	15.85
Digital Short-Term Explanation	7	0	8	2.29	3.35
Digital Long-Term Post-test	7	2	9	6.14	3.13
Digital Long-Term Drawing	7	11	37	26.71	10.9
Digital Long-Term Explanation	7	0	8	4.57	3.21

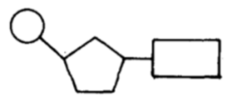
Appendix N

Objective Quiz for Dissertation Study

1. What are the basic structural units of a DNA molecule?

(A) glucose molecules
(B) amino acids
(C) lipids
(D) nucleotides

2. The diagram below represents the building block of a large molecule known as a



(A) protein
(B) fatty acid
(C) carbohydrate
(D) nucleic acid

3. Which molecule has the shape of a double-stranded helix?

(A) RNA
(B) DNA
(C) ADP
(D) ATP

4. Thymine binds with

(A) adenine
(B) thymine
(C) guanine
(D) cytosine

5. Guanine binds with

(A) adenine
(B) thymine
(C) guanine
(D) cytosine

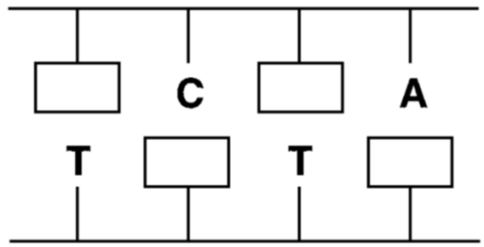
6. If a sequence of one strand of DNA is ATGATGT, the complementary strand must be __________. (Include correct orientation.)

(A) ATGATGT
(B) CGTCGTG
(C) TACTACA
(D) GCAGCAC

7. DNA is held together in a double helix by the force of

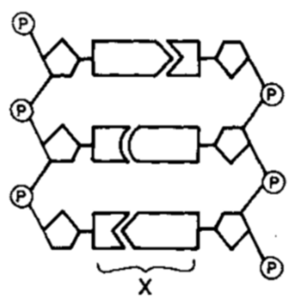
(A) the twists.
(B) covalent bonds.
(C) iconic bonds.
(D) hydrogen bonds.

8. The diagram below represents an incomplete section of a DNA molecule. The boxes represent unidentified bases. When the boxes are filled in, the total number of bases represented by the letter A (both inside and outside the boxes) will be



(A) 1
(B) 2
(C) 3
(D) 4

9. The diagram below represents a portion of a DNA molecule. The letter X represents two nitrogenous bases that are



(A) identical and joined by hydrogen bonds
(B) complementary and joined by hydrogen bonds
(C) identical and joined by iconic bonds
(D) complementary and joined by ionic bonds

10. The nitrogenous bases found in DNA are represented by the letters
(A) A, U, G, and C
(B) A, T, G, and C
(C) T, A, P, and C
(D) T, U, G, and C

Appendix O

Attitudes to Scientific Inquiry subscale from the *TOSRA*

Fraser, B. J. (1978). Development of a Test of Science-Related Attitudes. *Science Education, 62*(4), 509-515.

TOSRA

(Fraser, 1978)

Appendix P

Drawing and Explanation Prompt

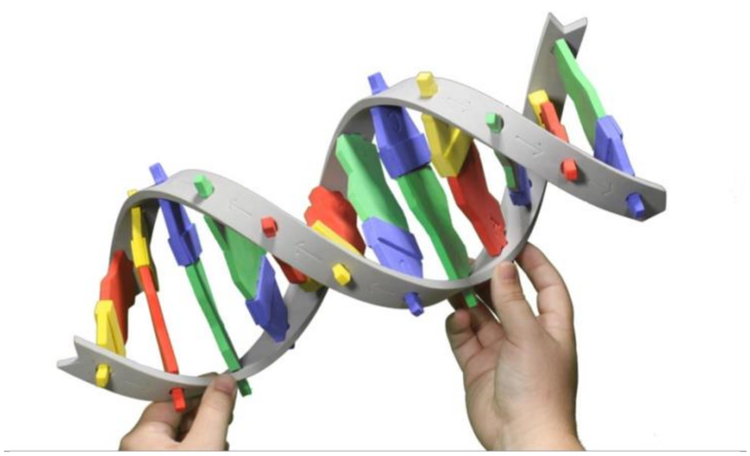
You have learned about the DNA molecule.
From your memory using no other information, please **draw** and **label** a DNA molecule.

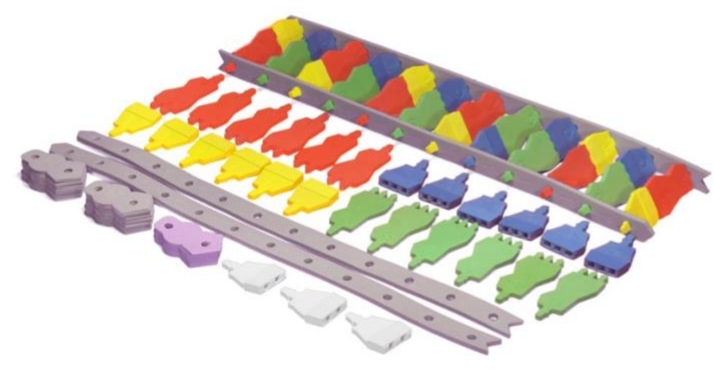
Now, on the lines below, please explain the shape, structure, and parts of the DNA molecule that you drew above.

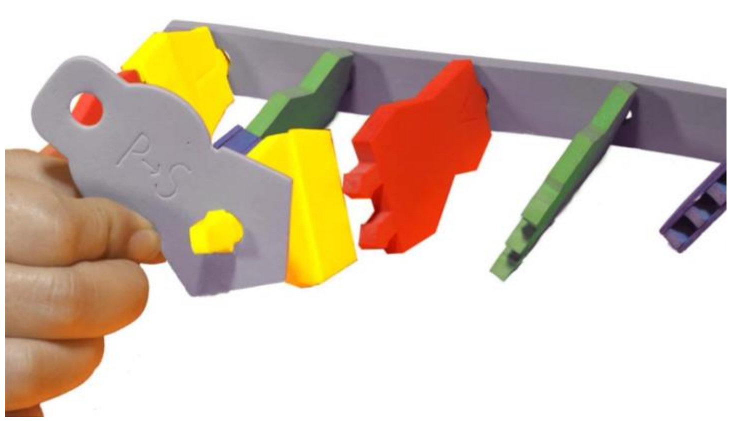
Please use the back of the paper if you need extra room.

Appendix Q

Physical Model for Dissertation Study







Appendix R

Control Group Worksheet Questions

The Covid-19 virus is an RNA retrovirus. Retroviruses are particularly difficult to treat because they contain RNA instead of DNA for their genetic material. For example, HIV is a retrovirus that has caused illness for over four decades with no current cure. Today, you are working as an epidemiologist trying to get ahead of a new, potentially pandemic-inducing virus. In order to assess the threat of the virus, you must determine the type of genetic information the virus contains, either RNA or DNA, and the genome of the virus. Answer the questions below to guide your understanding of this new virus.

1. What does DNA look like?

   ____

   ____

2. What are the parts that make up a DNA molecule?

   ____

   ____

3. How many strands does DNA have? __________
4. How are the DNA strands held together?

   ____

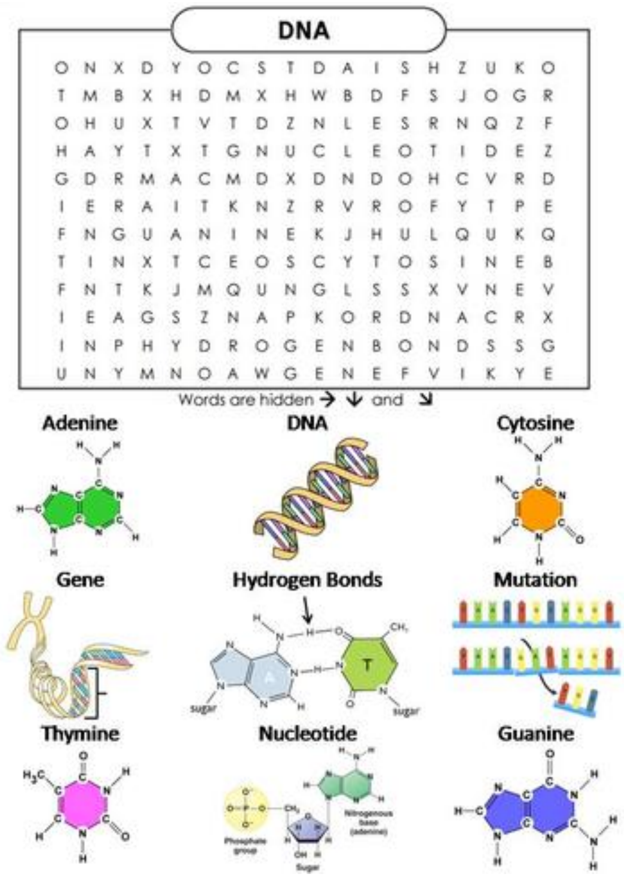
   ____

5. What DNA nucleotide base does Adenine bind with? ____________
6. What DNA nucleotide base does Thymine bind with? ____________
7. What DNA nucleotide base does Guanine bind with? ____________
8. What DNA nucleotide base does Cytosine bind with? ____________
9. Write the complementary DNA strand to the DNA strand you find in the viral sample:

   A T G C C T G T A G G

   ____

10. Complete the following word search on the back of this page:



11. The following question can be answered after viewing the image on the screen. Is the new virus a retrovirus or a DNA virus?

____

Appendix S

Link to Instructional Video

https://youtu.be/dl9W9i2e6G4

Appendix T

Descriptive Statistics for Dissertation Study

	N	Min	Max	Mean	SD
Spatial Ability	161	0	20	5.98	4.32
Attitudes to Science Inquiry	161	18	50	36.57	6.09
Pre-Test Objective Quiz	161	0	10	4.06	1.96
Post-Test Objective Quiz	161	0	10	5.94	2.26
Post-Test Drawing	160	0	32	8.97	5.99
Post-Test Explanation	160	0	7	1.81	1.41
Delayed Post-Test Objective Quiz	149	0	10	5.07	2.25
Delayed Post-Test Drawing	148	0	25	7.66	5.57
Delayed Post-Test Explanation	148	0	5	1.17	1.12
Physical Pre-Test Objective Quiz	60	0	10	3.93	2.11
Physical Post-Test Objective Quiz	60	2	10	5.75	1.98
Physical Post-Test Drawing	59	1	32	11.19	6.63
Physical Post-Test Explanation	59	0	7	2.34	1.52
Physical Delayed Post-Test Objective Quiz	56	1	10	5.36	2.11
Physical Delayed Post-Test Drawing	55	0	25	10.62	6.74
Physical Delayed Post-Test Explanation	55	0	5	1.65	1.29
Digital Pre-Test Objective Quiz	53	1	8	4.09	1.68
Digital Post-Test Objective Quiz	53	2	9	5.64	2
Digital Post-Test Drawing	53	0	25	8.06	5.99
Digital Post-Test Explanation	53	0	6	1.42	1.18
Digital Delayed Post-Test Objective Quiz	47	1	10	5.09	2.3
Digital Delayed Post-Test Drawing	47	0	14	4.64	3.05
Digital Delayed Post-Test Explanation	47	0	3	0.66	0.76
Control Pre-Test Objective Quiz	48	0	10	4.19	2.07
Control Post-Test Objective Quiz	48	0	10	6.52	2.76

Control Post-Test Drawing	48	0	18	7.25	4.17
Control Post-Test Explanation	48	0	5	1.60	1.33
Control Delayed Post-Test Objective Quiz	46	0	9	4.72	2.35
Control Delayed Post-Test Drawing	46	0	22	7.2	4.11
Control Delayed Post-Test Explanation	46	0	4	1.11	0.95

Appendix U

Correlation Matrix

	Spatial	Attitudes	Pre-Test	Objective Post-Test	Draw Post-Test	Explain Post-Test	Objective Delayed Post-Test	Draw Delayed Post-Test	Explain Delayed Post-Test
Spatial	1	0.01	0.09	0.27	0.15	0.09	0.21	0.08	0.05
Attitudes	0.01	1	0.03	-0.06	-0.01	-0.12	-0.03	-0.06	-0.14
Pre-Test	0.09	0.03	1	0.36	0.17	0.24	0.41	0.21	0.35
Objective Post-Test	0.27	-0.06	0.36	1	0.31	0.4	0.44	0.21	0.33
Draw Post-Test	0.15	-0.01	0.17	0.31	1	0.59	0.26	0.49	0.34
Explain Post-Test	0.09	-0.12	0.24	0.4	0.59	1	0.44	0.45	0.48
Objective Delayed Post-Test	0.21	-0.03	0.41	0.44	0.26	0.44	1	0.19	0.4
Draw Delayed Post-Test	0.08	-0.06	0.21	0.21	0.49	0.45	0.19	1	0.6
Explain Delayed Post-Test	0.05	-0.14	0.35	0.33	0.34	0.48	0.4	0.6	1

n

	Spatial	Attitudes	Pre-Test	Objective Post-Test	Draw Post-Test	Explain Post-Test	Objective Delayed Post-Test	Draw Delayed Post-Test	Explain Delayed Post-Test
Spatial	161	161	161	161	160	160	149	148	148
Attitudes	161	161	161	161	160	160	149	148	148
Pre-Test	161	161	161	161	160	160	149	148	148
Objective Post-Test	161	161	161	161	160	160	149	148	148
Draw Post-Test	160	160	160	160	160	160	148	148	148
Explain Post-Test	160	160	160	160	160	160	148	148	148
Objective Delayed Post-Test	149	149	149	149	148	148	149	148	148
Draw Delayed Post-Test	148	148	148	148	148	148	148	148	148
Explain Delayed Post-Test	148	148	148	148	148	148	148	148	148

P	Spatial	Attitudes	Pre-Test	Objective Post-Test	Draw Post-Test	Explain Post-Test	Objective Delayed Post-Test	Draw Delayed Post-Test	Explain Delayed Post-Test
Spatial									
Attitudes	0.8711								
Pre-Test	0.2831	0.6903							
Objective Post-Test	0.0006	0.4685	0						
Draw Post-Test	0.0595	0.8622	0.033	0					
Explain Post-Test	0.2538	0.1265	0.002	0	0				
Objective Delayed Post-Test	0.009	0.7407	0	0	0.0016	0			
Draw Delayed Post-Test	0.317	0.4969	0.009	0.011	0	0	0.019		
Explain Delayed Post-Test	0.5137	0.0822	0	0	0	0	0	0	