EXAMINATION OF MULTIMEDIA LEARNING PRINCIPLES IN AUGMENTED REALITY AND VIRTUAL REALITY LEARNING ENVIRONMENTS

BURÇ ÇEKEN

BOĞAZİÇİ UNIVERSITY

2023

EXAMINATION OF MULTIMEDIA LEARNING PRINCIPLES IN AUGMENTED REALITY AND VIRTUAL REALITY LEARNING ENVIRONMENTS

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Management Information Systems

by

Burç Çeken

Boğaziçi University

2023

DECLARATION OF ORIGINALITY

I, Burç Çeken, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature…………………………………………….....

Date …………………………………………..............

ABSTRACT

Examination of multimedia learning principles in

augmented reality and virtual reality learning environments

This study examined the effects of augmented reality (AR), virtual reality (VR), and traditional learning environments on university students' cognitive load, motivation, and learning outcomes. The efficacy of segmenting, pre-training, and modality principles was also investigated in reducing cognitive load and improving learning outcomes within these contexts. A 3x4 factorial design was implemented with 383 participants, assessing retention, transfer, cognitive load, and motivation scores. Results showed that AR significantly improved retention for cell structure compared to traditional learning, while no significant differences were found for lightning formation. The modality effect was observed for lightning formation in AR but not for cell structure. The segmenting effect was present for retention in both subjects but absent for transfer and cognitive load scores. The pre-training effect was observed for retention in VR and AR for lightning formation but not for cell structure, with inconsistent results for transfer scores. These findings suggest that various factors, including subject matter complexity, learning environment characteristics, instructional design, and individual learner differences, influence the presence or absence of these effects. This highlights the importance of considering these factors when designing educational interventions to optimize learning outcomes. Further research must identify conditions for effectively leveraging these effects across different subject matters, learning environments, and outcomes.

ÖZET

Artırılmış gerçeklik ve sanal gerçeklik öğrenme ortamlarında çoklu ortam öğrenme ilkelerinin incelenmesi

Bu araştırmada, artırılmış gerçeklik (AG), sanal gerçeklik (SG) ve geleneksel öğrenme ortamlarının üniversite öğrencilerinin bilişsel yüklenmesine, motivasyonuna ve öğrenme çıktılarına olan etkileri incelenmiştir. Ayrıca bu öğrenme ortamlarında bilişsel yükü azaltma ve öğrenme çıktılarını iyileştirme konusunda bölümleme, ön hazırlık ve kanal ilkelerinin etkinliği araştırılmıştır. 383 katılımcıyla gerçekleştirilen 3x4 faktöriyel tasarımda, kalıcılık, transfer, bilişsel yük ve motivasyon puanları değerlendirilmiştir. Sonuçlar, AG'nin hücre yapısı konusunda geleneksel öğrenmeye kıyasla kalıcılığı önemli ölçüde artırdığını gösterirken, yıldırım oluşumu için anlamlı farklılıklar bulunmamıştır. AG öğrenme ortamında kanal etkisi, yıldırım oluşumu konusu için gözlemlenirken hücre yapısı için gözlemlenmemiştir. Bölümleme etkisi, kalıcılık puanlarına göre her iki konuda mevcutken, transfer ve bilişsel yük puanlarına göre bölümleme etkisi görülmemiştir. Ön hazırlık etkisi, yıldırım oluşumu konusundaki kalıcılık puanlarına göre SG ve AG öğrenme ortamlarında gözlemlenirken transfer puanları için tutarsız sonuçlar elde edilmiştir. Bu bulgular, konu karmaşıklığı, öğrenme ortamı özellikleri, eğitim tasarımı ve bireysel öğrenen farklılıkları dahil olmak üzere çeşitli faktörlerin, bu etkilerin varlığını veya yokluğunu etkilediğini öne sürmektedir. Bulgular, öğrenme çıktılarını optimize etmek için eğitsel müdahaleler tasarlanırken bu faktörlerin göz önünde bulundurulmasının önemini vurgulamaktadır. Sonraki araştırmalar, bu etkilerin farklı konular, öğrenme ortamları ve sonuçlar arasında etkili bir şekilde kullanılması için gerekli koşulları belirlemelidir.

CURRICULUM VITAE

NAME: Burç Çeken

DEGREES AWARDED:

PhD in Management Information Systems, 2023, Boğaziçi University

MA in Educational Technology, 2016, Boğaziçi University

BA in Computer Education and Educational Technology, 2013, Boğaziçi University

AREAS OF SPECIAL INTEREST

E-Learning, multimedia learning, cognitive load, augmented reality, virtual reality, management information systems, information systems, educational data mining

PROFESSIONAL EXPERIENCE

Business Analyst Chapter Lead, Architecht, 2018 – present

Research Assistant, Dept. of Comp. Ed. & Ed. Tech., Boğaziçi University, 2014 - 2018

ICT Teacher, Terakki Foundation Schools, 2013-2014

AWARDS AND HONORS

Highest Honors List, Boğaziçi University, 2016

Ranked 2nd in the Educational Technology Class of 2016

Highest Honors List, Boğaziçi University, 2013

Ranked 3rd in the Computer Education & Educational Technology Class of 2013

GRANTS

Scientific Research Projects, Boğaziçi University, Grant Number 21N03D1 – 2021

PUBLICATIONS

Journal Articles

Çeken, B., & Taşkın, N. (2022). Multimedia learning principles in different learning environments: a systematic review. *Smart Learning Environments*, *9*(1), 1-22.

Özturan, M., Gürsoy, F., & Çeken, B. (2019). An empirical analysis on the effects of investment assessment methods on IS/IT project success. *International Journal of Information Systems and Project Management*, *7*(4), 33-52.

Özturan, M., Mutlutürk, M., Çeken, B., & Sarı, B. (2019). Evaluating the information systems integration maturity level of travel agencies. *Information Technology & Tourism*, *21*(2), 237-257.

Ceken, B. & Akpınar, Y. (2017). The effect of tutorial feedback type on the choice of feedback type in pre-service teachers' development of learning objects. *G. Journal of Information Technology: Emerging Technologies. 7*(3), 71-85.

Conference Proceedings

Kutlu, B., Ceken, B., Mutluturk, M., Turkmen, C. (2018). "A Meta-Analysis of Social Media Learning Studies in Educational Research", Fifth International Management Information Systems Conference, Ankara, Turkey, 24-26 October.

Çeken, B. & Akpınar, Y. (2106). "Investigating multimedia knowledge representation in learning object feedback", Sixth *World Conference on Educational Technology Research*. Antalya, Turkey, 5-6 May.

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to my current thesis advisor, Assist. Prof. Nazım Taşkın and my former thesis advisor Prof. Birgül Kutlu Bayraktar, for their invaluable guidance, support, and mentorship throughout my doctoral studies. Their expertise, enthusiasm, and dedication have been a constant source of inspiration and have greatly contributed to my intellectual and personal growth.

I am also grateful to my thesis committee members, Prof. Yavuz Akpınar, Prof. Meltem Özturan, Prof. Zuhal Tanrıkulu, and Assoc. Prof. Duygu Mutlu Bayraktar for their insightful comments, constructive criticisms, and invaluable suggestions that have helped refine my research.

My sincere thanks go to my friends, especially Meltem Mutlutürk and Ekrem Kutbay, for their camaraderie, intellectual discussions, and invaluable feedback on my research. Their friendship has been a source of motivation and comfort during this journey.

I'm greatly thankful for the financial support from Boğaziçi University Research Fund Grant Number 18441, which is crucial in my research's progression. This assistance has been a powerful force in driving my study and academic pursuits.

I want to express my heartfelt appreciation to my wife, Büşra Öz Çeken, for her unwavering love, understanding, and support throughout my academic journey. Her belief in me has been a constant source of strength during the highs and lows of this journey.

Finally, I express my gratitude to all those who have contributed directly or indirectly to my research and development as a scholar.

DEDICATION

To my loving wife, Büşra Öz Çeken, and my beautiful daughter, Erva Çeken,

This thesis is dedicated to both of you, the pillars of strength and love that have stood by me throughout this arduous journey. Your unwavering support, understanding, and patience have been the foundation upon which I have built my dreams.

To my dearest Büşra, thank you for being my confidante, best friend, and life partner. You have been my rock during the most challenging moments, and your faith in me has inspired me to persevere. Your love and companionship have transformed the darkest days into rays of hope and motivation.

To my precious Erva, you are the light of my life and the source of immeasurable joy. Your laughter, innocence, and curiosity have filled my heart with love and have reminded me of the importance of balance in life. Your presence has been a constant reminder of the impact that my work and actions can have on future generations.

I am eternally grateful to both of you for the sacrifices you have made and the endless love and encouragement you have given me. This achievement would not have been possible without you by my side.

With all my love,

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDIX TABLES

LIST OF FIGURES

LIST OF APPENDIX FIGURES

CHAPTER 1

INTRODUCTION

Advancements have influenced the evolution of education in technology, transforming how students acquire and process information. These innovations have redefined traditional teaching methods and expanded the possibilities for learning experiences. Learning is described as a transformative process in which experience brings about change, resulting in improved performance and increased potential for future learning (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010). Modern technologies have made this process more effective by providing advanced equipment and tools that promote interactivity and facilitate student learning (Raja & Nagasubramani, 2018).

Multimedia technology, such as computers and tablets, has supported learning and knowledge acquisition. Akkoyunlu and Yılmaz (2005) suggest that digital technology can provide multiple benefits in educational environments, including creating appropriate teaching materials for learners with different characteristics, enriching teaching environments, and improving accessibility to teaching materials. There has been an increasing trend among instructional designers and teachers to use interactive multimedia in education, which combines five essential elements, including text, video, sound, graphics, and animation, to create an interactive learning environment (Asthana, 2008).

The impact of modern digital technologies and online resources has expanded the role of the learning environment, altering the methods by which students acquire knowledge and educators impart instruction (Manzoor, 2016; Vinales, 2015). To optimize the learning experience, it is essential to consider learners' needs and learning styles when designing learning environments (Erden & Altun, 2006). This requires appropriate flexibility in the learning environment and careful selection of multimedia technologies to enhance the effectiveness of the learning experience.

Multimedia is the concurrent presentation of information using visual and verbal elements (Richter, Scheiter, & Eitel, 2016). Multimedia learning, as defined by Mayer (1997), is the process of acquiring knowledge or skills through the presentation of instructional messages that combine both visual and verbal information, such as pictures (e.g., videos, animations) and words (e.g., written or spoken language). A substantial body of research has been dedicated to exploring the influence of multimedia in education (Mayer, 2017; Wang, Sundararajan, Adesope, & Ardasheva, 2017; Weng, Otanga, Weng, & Cox, 2018), and the number of such studies has increased significantly since the 1990s (Li, Antonenko, & Wang, 2019). Building on these earlier findings, Mayer (2009) formulated the Cognitive Theory of Multimedia Learning (CTML), a framework that elucidates the cognitive mechanisms underlying meaningful learning experiences.

Mayer and his collaborators established multimedia learning principles, initially validated through written text and illustrations that accompanied verbal or auditory presentations (Mousavi, Low, & Sweller, 1995; Mayer, Bove, Bryman, Mars, & Tapangco, 1996; Moreno & Mayer, 2000). Numerous scholars have investigated the efficacy of multimedia learning principles under various circumstances and contexts (Akçayır & Akçayır, 2017; Ginns, 2005; Mutlu-Bayraktar, Cosgun, & Altan, 2019; Rey et al., 2019). The swift progression of multimedia technologies has led to the expansion of these principles across various learning settings, including computer-based learning systems (Park, Flowerday, & Brünken, 2015; Kutbay & Akpinar, 2020), online learning platforms (Chen & Yang, 2020; Sung & Mayer, 2012), virtual learning spaces (Baceviciute, Mottelson, Terkildsen, & Makransky, 2020; Klingenberg, Fischer, Zettler, & Makransky, 2023; Parong & Mayer, 2018; Petersen, Klingenberg, Mayer, & Makransky, 2020), and augmented learning environments (Küçük, Kapakin, & Göktaş, 2016; Lai, Chen, & Lee, 2019).

As multimedia technologies have rapidly evolved, applying these principles has expanded to encompass diverse learning settings. For instance, these principles have been implemented in computer-based learning platforms and successfully enhanced educational results (Park et al., 2015; Kutbay & Akpinar, 2020). Similarly, web-based learning environments have become increasingly popular, with studies showing that multimedia elements can enhance student engagement and motivation (Chen & Yang, 2020; Sung & Mayer, 2012). Virtual learning settings, which simulate real-world experiences, have also been shown to effectively support the learning process (Baceviciute et al., 2020; Klingenberg et al., 2023; Parong & Mayer, 2018; Petersen et al., 2020). Finally, the use of augmented reality in learning environments has the potential to provide students with immersive and interactive experiences that can enhance their learning outcomes (Küçük et al., 2016; Lai et al., 2019). Overall, applying multimedia principles in various learning environments can significantly improve the quality of education and promote better learning outcomes for students.

Aside from conducting empirical studies in various learning settings, meta-analysis and systematic review studies play an important role in advancing our understanding of the current situation of the multimedia learning field. For example, most research exploring the segmenting, pre-training, and modality principles has predominantly been carried out in computer-based slideshows or animations, with limited studies in AR and VR settings (Avcı, Coklar, & İstanbullu, 2019; Mutlu-Bayraktar et al., 2019).

The recent review study by Çeken and Taşkın (2022) found that most studies investigating the effectiveness of multimedia learning principles have been carried out within conventional learning settings (96.3%). This was followed by studies conducted in virtual reality settings, which accounted for 3.7% of the studies. Surprisingly, the review found that no studies have been conducted to investigate the impact of multimedia learning principles in AR learning settings. This gap in the research indicates that investigating the effects of multimedia learning principles in AR learning environments could provide valuable insights into optimizing educational experiences, promoting deeper understanding, and potentially revolutionizing the way learners interact with and process information in this emerging technology.

In addition to the limited research on multimedia learning in AR and VR environments, there are contradictory findings in computer-based learning environments. A study by Mutlu-Bayraktar et al. (2019) found some inconsistency in the results of studies conducted in computer-based environments, highlighting the need for further investigation. The study also revealed that most studies focus on the effects of extraneous load on multimedia learning outcomes, while only a small proportion of studies (15%) have examined the impact of intrinsic cognitive load.

Each of the chosen principles inherently addresses this gap. The segmenting principle, which proposes breaking down complex information into manageable parts, can alleviate the intrinsic load by simplifying the inherent complexity of the content (Mayer & Moreno, 2003). The pre-training principle, offering learners foundational information before delving into intricate subjects, can help reduce the intrinsic load by familiarizing learners with core concepts and thus easing the cognitive demands of subsequent learning (Mayer, 2005). Lastly, the modality principle, advocating the optimal use of auditory and visual channels for information processing, has the potential to manage the intrinsic load by leveraging the dual-processing capacity of human cognition (Mayer & Moreno, 1998). Therefore, by exploring these principles in the context of multimedia learning, this thesis not only addresses an identified research gap but also contributes to a deeper understanding of how to manage the intrinsic cognitive load, a less studied but equally significant aspect of cognitive load theory.

Moreover, despite the significant body of research on motivation and learning outcomes, there needs to be a greater understanding of how motivation interacts with multimedia learning principles. Mutlu-Bayraktar et al. (2019) highlights this gap in literature, indicating the need for more research in this area. Additionally, as previously noted, there are conflicting findings in applying multimedia learning principles, specifically in segmenting, pre-training, and modality principles. This underscores the importance of my research in addressing these gaps and contradictions, providing a comprehensive perspective on the effectiveness of multimedia learning principles and their potential implications in emerging educational technologies.

Given the significant gaps in research in AR and VR learning environments and the conflicting findings regarding the effectiveness of multimedia learning principles, it is crucial to conduct further studies to investigate these principles in AR and VR contexts. Such studies can provide insights into the effectiveness of multimedia learning principles in these new and emerging learning environments and enable researchers to compare the results to those obtained in traditional learning environments. This comparison is necessary to ensure the findings' validity and identify whether the same principles can be applied across different learning environments. Additionally, such studies can help identify any challenges or opportunities unique to AR and VR environments that must be considered when implementing multimedia learning principles. By filling these gaps in knowledge, researchers can help to optimize the use of multimedia technology in education and promote better learning outcomes for students. The current study investigates the effect of segmenting, modality, and pre-training principles on learning outcomes, cognitive load, and motivation across different learning environments (augmented reality, virtual reality, and traditional).

The study aims to answer the following research questions:

i. How does the learning environment (AR, VR, and Traditional) affect university students' cognitive load, motivation, and learning outcomes (Retention and Transfer)?
   - i.i. How does the learning environment affect university students' cognitive load?
   - i.ii. How does the learning environment affect university students' motivation?
   - i.iii. How does the learning environment affect university students' retention test scores?
   - i.iv. How does the learning environment affect university students' transfer test scores?

ii. To what extent do segmenting, pre-training, and modality effects influence cognitive load and learning outcomes for university students across different learning environments?
   - ii.i. To what extent do segmenting, pre-training, and modality effects influence retention test scores for university students across different learning environments?

ii.ii. To what extent do segmenting, pre-training, and modality effects influence transfer test scores for university students across different learning environments?

ii.iii. To what extent do segmenting, pre-training, and modality effects influence cognitive load for university students across different learning environments?

iii. Is there a significant difference between the cognitive load scores of the students using the lightning formation learning material and those using the cell learning material?

iii.i. Is there a significant difference between the intrinsic cognitive load scores of the students using the lightning formation learning material and those using the cell learning material?

iii.ii. Is there a significant difference between the extraneous cognitive load scores of the students using the lightning formation learning material and those using the cell learning material?

iii.iii. Is there a significant difference between the germane cognitive load scores of the students using the lightning formation learning material and those using the cell learning material?

CHAPTER 2

LITERATURE REVIEW

2.1 Augmented reality and virtual reality learning environments

The educational setting encompasses not only physical spaces but also the contexts and cultural aspects in which students acquire knowledge (Bakhshialiabad, Bakhshi, & Hassanshahi, 2015). This setting is a multifaceted and evolving system where educators employ targeted approaches and leverage available resources to attain predefined educational objectives (Wang & Kinuthia, 2004). The educational setting plays a crucial role in education (Vinales, 2015) by fostering the growth of students' abilities, comprehension, mindset, and conduct (Ozerem & Akkoyunlu, 2015).

Although "learning environment" has conventionally referred to physical classrooms, recent digital technologies, methods, and approaches have revolutionized this concept, paving the way for enhanced and more efficient learning experiences (Baeten, Kyndt, Struyven, & Dochy, 2010). Incorporating technology into educational processes is commonly called technology-enhanced learning (Law, Niederhauser, Christensen, & Shear, 2016). Utilizing technology to enrich and elevate educational experiences is often called technology-enhanced learning. This notion of leveraging technology for improved learning outcomes has been identified through various terms in the literature, including virtual reality-based (Hamilton, McKechnie, Edgerton, & Wilson, 2021), augmented reality-based (Cubillo et al., 2014), mobile-based learning (Küçük et al., 2016), computer-assisted learning (Moos & Azevedo, 2009), web-based instruction (Chen & Yang, 2020). Numerous contemporary technologies, such as Web 2.0, mobile devices, VR, and AR, are progressively incorporated into educational processes to enhance learning experiences by capitalizing on their distinct characteristics (Cubillo et al., 2014). For example, augmented reality and virtual reality technologies have been employed in an overwhelming majority of UK universities (96%) and a significant proportion of UK colleges (79%) to deliver high-quality, immersive learning experiences for students, as reported by UK Authority (2019).

Adopting augmented reality as a cutting-edge technology has increased within educational settings, leveraging its distinct capabilities, such as creating interactive and immersive learning experiences to enhance the learning experience. AR has been defined differently in literature. Azuma (1997) characterized augmented reality as a technological innovation that enriches the real-world environment by seamlessly integrating virtual elements, thereby providing supplementary information without compromising the authenticity of the tangible world. Gonzato, Arcila and Crespin (2008) defined it as a technology that makes the real world seem augmented by using virtual elements (sound, text, video, etc.). AR facilitates the connection between the real and digital worlds by overlaying and manipulating virtual components on top of physical objects, as described by Wu, Hwang, Yang, and Chen (2018). Even though there are different definitions, AR has three common essential features: (a) the fusion of virtual and tangible elements within an authentic setting, (b) real-time interactive capabilities, and (c) being three-dimensional (Azuma et al., 2001).

AR technology is divided into image- and location-based categories (Cheng & Tsai, 2013). Image-based AR systems utilize image recognition methods to identify the location of the 3D model to be integrated into the environment. These systems can be marker-based or non-marker-based. In marker-based AR setups, pre-defined markers (such as labels or square barcodes) are required within the environment to prompt or initiate the content, enabling accurate placement of 3D models on display. The position of the 3D model is determined by referencing this pointer. In non-marker-based AR systems, there is tracking existing physical objects in the environment instead of adding markers. Location-based AR systems employ GPS or wireless networks as recognition techniques to activate digital content rather than utilizing markers (Cheng & Tsai, 2013).

VR is characterized as a synthetic environment, crafted by software, which transports users into an alternate setting distinct from the real world. VR presents an immersive simulation that evokes a sense of presence within the environment while offering opportunities for interaction (Schroeder, 1996; Makransky & Lilleholt, 2018). According to Burdea and Coiffet (2003), VR is an imitation of a real environment that makes it possible to solve more complex problems with electronic equipment, such as gloves and headsets involving a screen inside. It also allows people to observe cases or places with limited access or impossible to explore in the tangible world (Freina & Ott, 2015).

Virtual reality technologies can be classified into two main categories in terms of immersion, making users feel like they are part of the virtual reality environment (Lee & Wong, 2014). These are immersive and non-immersive VR. The essential differences arise from the methods employed for controlling and presenting the virtual setting (Cheng & Tsai, 2020). Non-immersive VR also referred to as desktop VR, provides virtual content to users via computer monitors controlled by a keyboard or mouse (Sally Wu & Alan Hung, 2022). The real world is accessible simultaneously in the non-immersive VR while experiencing the virtual world (Onyesolu & Eze, 2011). On the other hand, users equipped with a headset and controllers only interact with the 360°

virtual environment in immersive VR, also called a head-mounted display (Çoban, Bolat, & Göksu, 2022).

Even though Azuma (1997) considers AR a subset of VR, they have some differences. Milgram and Kishino (1994) aimed to clarify the concepts between virtual and real environments with a “Reality-Virtuality Continuum” as in Figure 1. There is an environment people perceive with the naked eye without using any hardware in the "Real Environment" section and an environment created by only virtual objects in the "Virtual Environment" section. The transitions between the two worlds are defined as “Mixed Reality”, in which objects from the real and virtual environments are displayed concurrently (Milgram & Kishino, 1994). AR constitutes a real-world setting enhanced with virtual content, whereas Augmented Virtuality (AV) is defined as a virtual world with some overlaid real-world content (Ludwig & Reimann, 2005). As a result, progressing from left to right along the continuum leads to a greater prevalence of virtual imagery and a diminished association with the natural world (Billinghurst, Kato & Poupyrev, 2001). Augmented reality helps increase one's perception of reality using technology (Graham, Zook, & Boulton, 2022).

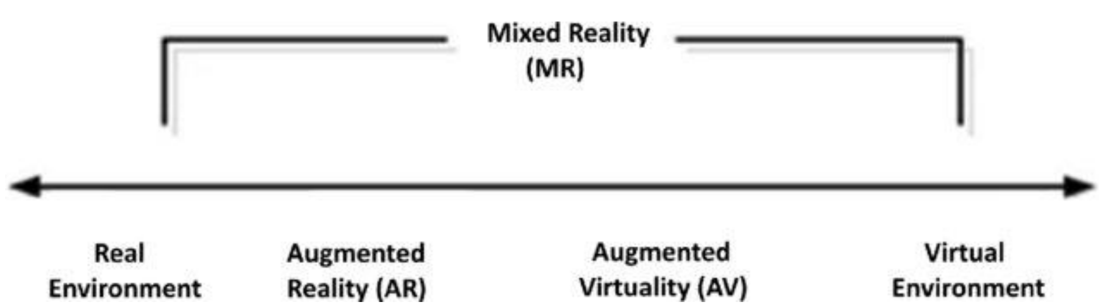

Figure 1. Milgram's reality–virtuality continuum (Milgram & Kishino, 1994)

2.1.1 Augmented reality in education

AR has gained significant attention in educational research in recent years, as demonstrated by various research (Akçayır & Akçayır, 2017; Ibáñez & Delgado-Kloos, 2018; Yu, Denham, & Searight, 2022; Mystakidis, Christopoulos, & Pellas, 2022; Sırakaya & Alsancak Sırakaya, 2022). This technology is implemented across various educational levels, from early childhood to higher education (Sırakaya & Alsancak Sırakaya, 2022). Additionally, research on AR spans a wide range of disciplines, including biology (Hwang, Wu, Chen, & Tu, 2016; Küçük et al., 2016), chemistry (Yang, Mei, & Yue, 2018), physics (Akçayır, Akçayır, Pektaş, & Ocak, 2016; Cai, Chiang, Sun, Lin, & Lee, 2017), mathematics (Estapa & Nadolny, 2015; Sommerauer & Müller, 2014), geometry (Hwang & Hu, 2013; Lin, Chen, & Chang, 2015), language (Godwin-Jones, 2016; Küçük, Yılmaz, & Göktaş, 2014), arts (Chang et al., 2014; Di Serio, Ibáñez, & Kloos, 2013), medicine (Calle-Bustos, Juan, García-García & Abad, 2017; Jamali, Shiratuddin, Wong, & Oskam, 2015), astronomy (Yen, Tsai, & Wu, 2013; Chen & Wang, 2015), and engineering (Henderson & Feiner, 2010; Potkonjak et al., 2016).

AR technology in educational settings has been found to have numerous advantages (Cheng & Tsai, 2013). In a recent review by Sırakaya and Alsancak Sırakaya (2022), the benefits of AR in science, technology, engineering, and mathematics (STEM) education were examined and grouped into four main categories: benefits for the student, educational benefits, interactive benefits, and other advantages. The review shows that AR can enhance student performance (Chen & Wang, 2015; Lin et al., 2015), boost motivation (Chang & Hwang, 2018; Ibáñez, Di Serio, Villarán, & Kloos, 2014), foster positive attitudes toward learning (Hsiao, Chen, & Huang, 2012), heighten attention and eagerness to learn (Chen, Chou, & Huang, 2016), make learning more manageable (Martin-Gonzalez, Chi-Poot, & Uc-Cetina, 2016), decrease mental effort (Küçük et al., 2016), improve student satisfaction (Ibáñez et al., 2014), enhance spatial skills (Gun & Atasoy, 2017), enhance cognitive capabilities (Yoon, Elinich, Wang, Steinmeier, & Tucker, 2012), promote long-term retention (Sommerauer & Müller, 2014), and raise classroom participation (Moro, Štromberga, Raikos, & Stirling, 2017). Moreover, incorporating Augmented Reality in education has yielded numerous advantages for learning outcomes, including the capacity to visualize phenomena imperceptible to the naked eye (Yoon, Anderson, Lin, & Elinich, 2017), making learning more enjoyable (Gun & Atasoy, 2017), promoting teamwork (Bressler & Bodzin, 2013), allowing for personalized instruction (Kamarainen et al., 2013), providing more thorough understanding (Chiang, Yang, & Hwang, 2014), making abstract ideas more concrete (Laine, Nygren, Dirin, & Suk, 2016), allowing learning to take place at any time and place (Küçük et al., 2016), putting the focus on the student (Kamarainen et al., 2013), and supporting informal learning opportunities (Yoon et al., 2017).

One of the most cited benefits of AR is that it improves students' learning outcomes (Akçayır & Akçayır, 2017). Using AR technologies leads to more significant learning gains than other technology resources, traditional lectures, and traditional teaching materials (Garzón, Pavón, & Baldiris, 2019). AR can make it easier for students to comprehend complex topics, such as chemical reactions that are challenging to witness, by providing visual aids (Akçayır et al., 2016). Additionally, AR allows students to connect their learning to real-world situations by overlaying virtual elements on physical objects (Wu et al., 2018). For example, AR can assist students who struggle with geometry by allowing them to interact with 3D geometric shapes. Due to these features, AR can enhance the learning process by promoting student engagement in learning activities (Di Serio et al., 2013; Akçayır & Akçayır, 2017; Garzón et al., 2019). While AR has been shown to positively affect student achievement, the magnitude of this impact may vary for students at different levels of achievement (Sırakaya & Alsancak Sırakaya, 2022) and levels of education (Batdi & Talan, 2019). For example, Lin et al. (2015) found that AR proved beneficial at small and medium scales for learners who demonstrated average and low academic performance but not for students with high academic achievements. Additionally, a comprehensive meta-analysis examining AR technologies in education concluded that across various disciplines, AR demonstrated a significant impact on students pursuing bachelor's degrees or equivalent levels of education (Garzón & Acevedo, 2019).

Another key benefit of AR in education is the decrease in cognitive load (e.g., Bressler & Bodzin, 2013; Cheng, 2017; Küçük et al., 2016; Lai et al., 2019; Liou, Yang, Chen, & Tarng, 2017; Thees et al., 2020; Turan, Meral, & Sahin, 2018; Wu et al., 2018). AR can help enhance instructional design and learning materials, diminishing germane and extraneous cognitive load (Singh, Mantri, Sharma, Dutta, & Kaur, 2019). For example, Cheng (2017) conducted a study involving 153 students and found that their perceptions of cognitive load were lower while reading an AR book. Nevertheless, some studies have also suggested that using AR can cause increasing cognitive load (Cheng & Tsai, 2013; Wu, Lee, Chang, & Liang, 2013) or that there is no difference in the usage of AR in education (Altmeyer et al., 2020; Çoban, 2020; Tugtekin & Odabasi, 2022; Wu et al., 2018). A recent review study conducted by Lai et al. (2019) found that only a limited number of studies have examined the application of multimedia learning principles in AR environments or assessed the impact of AR on cognitive load.

Students, motivated to learn, tend to exhibit greater engagement, perseverance, and dedication to accomplishing tasks than their less motivated counterparts (Keller, 1979; Schmidt, 2007). Empirical research demonstrates that students employing AR technology exhibit heightened motivation levels (e.g., Chang & Hwang, 2018; Chen, 2020; Chiang et al., 2014; Di Serio et al., 2013). The current meta-analyses also highlighted that using AR applications in educational settings increases motivation (Garzón et al., 2019; Lai et al., 2019). For instance, Chang et al. (2016) pointed out that students' motivation was negatively impacted by the extended time necessary for plant growth and the scarcity of opportunities to witness alterations in plant growth within laboratory settings. The results revealed that motivation was enhanced by AR by permitting students to observe and engage with these changes (e.g., sunlight exposure).

While AR has many benefits, it also poses certain challenges that must be considered. Students may find AR challenging to use and encounter technical difficulties interacting with it (Lin, Hsieh, Wang, Sie, & Chang, 2011). An example of a challenge in using AR is that location-based AR applications may have issues with GPS accuracy (Chiang et al., 2014). Squire and Jan (2007) pointed out that a poorly designed AR interface for students and an application that needs more guidance on the usage process can be quite challenging for students. Furthermore, Kaufmann and Dünser (2007) emphasized that it is essential to design AR technologies used in educational applications to reduce students' focus on the learning task and reduce their cognitive load levels. They also stated that more efforts should be made on interaction and user interface design to develop AR technologies in a way that will ensure user satisfaction.

Lin et al. (2011) pointed out that students often perceive AR applications as complicated and commonly face technical issues while utilizing this technology. A review by Akçayır and Akçayır (2017) emphasizes an array of challenges related to the implementation of AR in educational settings. These challenges include students having a hard time using it, increased time requirements to complete tasks, recognition problems, GPS errors leading to frustration, unsuitability for large group instruction, technical issues involving cameras, internet, and indoor usage, mental strain, the distraction of students' focus, high costs, limited sharing due to substantial file sizes, ergonomic issues, difficulty in designing AR experiences, and a deficiency in teacher expertise with the technology. Despite these challenges, more efficacious applications of AR for science education can be devised by taking them into account during the design and implementation phases.

To sum up, the literature above provides detailed information about AR's widespread application and benefits in various educational fields. AR technology has been shown to improve student performance, motivation, and cognitive abilities while reducing cognitive load. However, challenges in implementing AR, such as technical difficulties, design issues, and teacher expertise, are also noted. It is also presented evidence from various studies and meta-analyses, showcasing the potential advantages of AR in different educational settings and for students with varying levels of achievement.

The provided literature is highly relevant to our research, which examines the effect of different learning environments (AR, VR, and Traditional) on learning outcomes, cognitive load, and motivation. Thus, considering the literature, it can be understood better how AR technology can be effectively integrated into various learning environments and how it compares to virtual reality and traditional learning methods regarding learning outcomes, cognitive load, and motivation.

2.1.2 Virtual reality in education

Nowadays, it is possible to find VR technologies in many different fields, such as tourism, sports, cinema, architecture, education, health, engineering, space sciences, military, economics, natural sciences, social sciences, trade, law, and art (Akman, 2019; Avcı et al., 2019; Luo, Li, Feng, Yang, & Zuo, 2021). This technology can offer unique opportunities for education by creating a simulated or altered environment for users to engage with (Makransky & Lilleholt, 2018). Thus, VR is increasingly being explored and has shown significant positive outcomes in the educational field (Brown & Green, 2016; Curcio, Dipace, & Norlund, 2016; Yu & Xu, 2022). Additionally, VR particularly appeals to educators, as they constantly seek ways to create an engaging and enjoyable learning environment for their students (Lumby, 2011). Though VR has been commonly utilized for training purposes, research on its application in education is still relatively new and novel (Ferdig, Gandolfi, & Immel, 2018; Kavanagh, Luxton-Reilly, Wuensche, & Plimmer, 2017).

Virtual technologies provide several advantages in education, such as concretizing abstract concepts, performing experiments that are not safe in real life, experiencing the feeling of being in physically impossible or hard-to-reach environments, making boring classes more fun, practicing an activity from daily life, and visually presenting historical events and environments during history classes (Freina & Ott, 2015). In addition, it increases student motivation and engagement, a constructivist approach to learning, affordability and accessibility, and increased interaction compared to conventional learning materials. (Parong & Mayer, 2018; Zhao, Lin, Sun, & Liao, 2020). The immersive aspect of VR is a key advantage as it allows students to experience and interact with complex concepts, making the learning process more engaging and memorable (Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014). It also allows students to experience different environments where they can participate, providing them with hands-on experiences and encouraging their active involvement in various learning activities (Di Natale, Repetto, Riva, & Villani, 2020).

A multitude of research has explored the impact of VR technology on various aspects of learning, including educational performance (Chang, Hsu, Kuo, & Jong, 2020), retention of knowledge (Buttussi & Chittaro, 2017; Meyer, Omdahl, & Makransky, 2019), motivation for learning (Chang et al., 2020; Cheng & Tsai, 2020), and knowledge acquisition (Chittaro, Corbett, McLean, & Zangrando, 2018; Jimenez et al., 2018). Wu et al. (2013) state that virtual reality in the educational process makes learning meaningful, facilitates the transfer of information, concretizes abstract concepts, and presents complex topics understandably. Additionally, VR can enhance learners' analytical skills and ability to discover new concepts (Pan, Cheok, Yang, Zhu, & Shi 2006). A current review by Radianti, Majchrzak, Fromm, and Wohlgenannt (2020) indicates that VR can effectively promote student learning, particularly in acquiring various kinds of knowledge (e.g., procedural, practical, and declarative) in higher education. However, there is still much to learn about the potential of VR in education, and further research is needed to identify best practices for designing and implementing effective VR interventions.

Using VR technology in education can increase learners' engagement and motivation by providing a more interactive and immersive learning experience (Huang, Zou, Cheng, & Xie, 2021; Lee, Sergueeva, Catangui, & Kandaurova, 2017; Makransky, Terkildsen, & Mayer, 2019). Learners are more motivated and interested in learning when using a virtual environment than other learning methods (Akçayır et al., 2016;

Parong & Mayer, 2018). For example, nursing students improved their medical knowledge and learning motivation after participating in a VR-based learning activity (Dubovi, Levy, & Dagan, 2017). Moreover, a current review by Di Natale et al. (2020) found that immersive VR positively impacts students' motivation toward learning, indicating that the ability to navigate and interact within these virtual environments was thought to contribute to a deeper understanding of course material.

Despite the positive outcomes reported in many studies, some found no impact or adverse effects of VR-based learning (Baceviciute, Lucas, Terkildsen, & Makransky, 2022). One significant challenge in the virtual reality learning environment is the potential for excessive visual information, which can cause unnecessary cognitive strain and obstruct the learning process (Makransky et al., 2019; Meyer et al., 2019; Parong & Mayer, 2021). For example, Richards and Taylor (2015) found that learners who received a 2D presentation performed better than those who received a 3D presentation, which they attribute to the increased cognitive load caused by the visually intensive 3D animation in immersive VR. Additionally, as previously mentioned, the study by Makransky et al. (2019) showed adverse effects of VR compared to a desktop-based presentation regarding recall but not transfer.

According to Christou (2010), there are several disadvantages to using VR as an educational tool, including the expensive acquisition cost of a system, graphics that lack realism and issues with the transfer of skills, and physical effects on users, such as simulator sickness. The cost of developing and purchasing a VR system can be high because it involves software, computers, and display technology (Mantovani, Castelnuovo, Gaggioli, & Riva, 2003). The limitations in generating realistic graphics and maintaining interactivity also contribute to the lack of realism. Additionally, using head-mounted displays in immersive environments can result in physical side effects such as simulator sickness caused by a mismatch between visual perception and sense of movement (Mantovani et al., 2003; Riva, 2003).

To summarize, the provided literature above explores the use of VR technologies across various fields and their implications for education. It highlights VR's benefits, such as concretizing abstract concepts, facilitating hands-on experiences, and increasing student motivation and engagement. Several studies are mentioned to support these positive outcomes. However, it also acknowledges the challenges and disadvantages of using VR in education, such as excessive visual information causing cognitive strain, expensive acquisition costs, and physical side effects like simulator sickness. The provided literature provides valuable insights related to current research, which examines the effect of different learning environments (AR, VR, and traditional) on learning outcomes, cognitive load, and motivation. By exploring the implications of VR in education, our research can contribute to developing best practices for designing and implementing VR interventions, thereby maximizing the benefits and addressing potential drawbacks in various learning environments.

2.2 Cognitive load theory

Sweller, van Merriënboer, and Paas (2019) assert that Cognitive Load (CL) plays a pivotal role in the learning process, pertaining to the strain or burden placed on the working memory while engaged in cognitive tasks such as problem-solving or reasoning. This perspective aligns with the study by Akbulut (2011), who underscores that cognitive load involves the complex interplay of mental activities occurring simultaneously within the short-term memory, all vying for the learner's attention.

Leppink, Gog van, Paas, and Sweller (2015) further clarify that CL is a nuanced construct, encompassing not only the pressure exerted on individuals' cognitive processes while undertaking a specific task but also accounting for factors such as prior knowledge, intrinsic task complexity, and the strategies utilized to manage these demands. Consequently, understanding and managing cognitive load effectively is crucial for optimizing learning experiences and ensuring long-term information retention.

Cognitive Load Theory (CLT) was first introduced by Sweller (1988) and has since been considerably developed, now acknowledging the limitations of cognitive processing during learning and offering a more extensive comprehension of cognitive load and its implications for instructional design and learning outcomes (Sweller et al., 2019). A key distinction made in CLT is the categorization of cognitive load into three types: intrinsic, extraneous, and germane load. This typology has been instrumental in the evolution of the theory and our understanding of how different types of cognitive load influence learning (Sweller et al., 2019).

Central to CLT is the idea that human cognitive architecture consists of two key components: working memory and long-term memory (Sweller et al., 2019). The theory argues that for learning to be effective, cognitive resources must be managed efficiently, particularly within the constraints of working memory, to facilitate the successful integration and storage of information in long-term memory.

The working memory functions as a critical cognitive component that manages information. It obtains data from the sensory register, retrieves related schemas from long-term memory, processes and refines them, and then transfers any updated or altered schemas into long-term memory (Sweller, Ayres, & Kalyuga, 2011). Moreover, the cognitive load theory underscores the process by which new information is processed in the working memory before being consolidated and stored in the long-term memory, highlighting the importance of managing cognitive resources to facilitate effective learning and retention (Sweller et al., 2019). The acquisition of new knowledge becomes challenging when the working memory, with its finite capacity, is overwhelmed by information and processing demands. As such, instructional materials should be designed to minimize extraneous cognitive load, ensure more efficient use of working memory, and foster more effective learning experiences (Mutlu-Bayraktar et al., 2019).

The CLT delineates three types of cognitive load impacting the working memory during learning: intrinsic, extraneous, and germane (Sweller et al., 2019). Intrinsic cognitive load pertains to the inherent complexity of the content and its interplay with learners' pre-existing knowledge and experience (van Merrienboer & Sweller, 2005). The level of intrinsic cognitive load can fluctuate depending on the number of elements requiring simultaneous processing in the working memory (Leppink et al., 2015), emphasizing the importance of considering individual learners' capabilities and adjusting instructional strategies accordingly. Indeed, certain educational tasks possess a higher intrinsic complexity than others, regardless of the teaching method employed. For instance, analyzing a Shakespearean sonnet is more complex than identifying the main idea of a simple paragraph. The intrinsic cognitive load rises as the task's difficulty increases or the learner's expertise decreases (Bruder, 2018). Consequently, the inherent difficulty of a learning task and the learners' prior knowledge are critical factors in determining the intrinsic cognitive load (Sweller et al., 2019), emphasizing the need for tailored instruction that accounts for individual learners' backgrounds and abilities.

Conversely, the extraneous cognitive load does not pertain to the difficulty of a task but rather focuses on the presentation of learning material (Kalyuga, 2009) and the impact of educational materials and teaching techniques on working memory, which are unrelated to the subject matter (Clark, Nguyen, & Sweller, 2011). This cognitive load arises from suboptimal instructional design, such as insufficient explanations or poor integration of instructional materials (Mutlu-Bayraktar et al., 2019), emphasizing the importance of effective design in reducing extraneous load and facilitating learning. By employing suitable design principles, cognitive load can be effectively minimized, leading to more efficient learning experiences (Akçayır et al., 2016). Consequently, it is crucial to reduce the use of weakly designed learning materials as much as possible to optimize learning outcomes (Haji, Rojas, Childs, de Ribaupierre, & Dubrowski, 2015; Paas & Sweller, 2014).

Germane cognitive load represents the degree to which learners invest mental effort in building schemas by linking new information to existing knowledge in their long-term memory, thereby facilitating deeper understanding and meaningful learning (Mayer, 2009). Germane cognitive load can be influenced by other factors, such as learner motivation or interest (Whelan, 2007). After accounting for extraneous and intrinsic loads, the remaining capacity determines the extent to which germane cognitive load can increase or decrease (Paas & Sweller, 2014). Allocating working memory resources to germane cognitive load rather than extraneous cognitive load is essential because schema formation, which supports meaningful learning, occurs through the interaction of elements associated with germane cognitive load (Sweller, 2010). Effective and efficient learning occurs when a substantial portion of working memory is dedicated to germane cognitive load. The sum of intrinsic, extraneous, and germane cognitive loads constitutes the total cognitive load an individual experiences (Kalyuga, 2009).

While CLT primarily focuses on cognitive aspects of learning, recent insights point to a crucial relationship between cognitive load and affective states in the learning process (Sweller et al., 2019). Emotional states such as anxiety, frustration, and boredom can affect the cognitive load experienced by learners, thereby influencing learning outcomes. For instance, high levels of anxiety or stress can inadvertently increase cognitive load and impede the learning process. This is particularly evident in high stakes testing scenarios, where the pressure to perform well can introduce an additional, unnecessary cognitive burden. On the other hand, positive emotional states like interest and enjoyment can contribute to reducing perceived cognitive load and enhancing learning. Hence, it is essential to consider learners' affective states when designing instructional materials to optimize cognitive load and facilitate learning effectively.

There exists a lack of consensus among researchers regarding accurate and effective methods for measuring cognitive load (Sweller, van Merrienboer, & Paas, 1998). However, a scale created by Bratfisch, Borg, and Dornic (1972) and subsequently revised by Paas and van Merrienboer (1993) based on ratings is widely used for measuring cognitive load. Various techniques are also discussed for measuring cognitive load, including task-based techniques that simultaneously present different tasks and physiological techniques that assume that cognitive activity during the learning process is reflected in physiological measurements. Brünken, Plaas, and Leutner (2003) assessed cognitive load measures using a framework based on two dimensions: objectivity (subjective and objective) and causal relationship (direct and indirect).

In subjective-direct measurements, learners rate the difficulty of learning materials using rating scales. The scores obtained from the rating are assumed to be directly related to the cognitive load experienced by the learner. On the other hand, learners are asked to declare the amount of mental effort they exerted to comprehend the learning material. However, Brünken et al., (2003) pointed out that these measurements reflected the subjective perception of the effort exerted and did not provide clear indications of the actual mental effort expended on cognitive load.

Objective-indirect methods for measuring cognitive load involve evaluating learners' performance on tests or observing their behavioral patterns and psychological states during the learning process (Timar, 2020). These methods include indicators such as the time spent on different pages, navigation errors, and eye-tracking analysis. Physiological measurements, such as heart rate and pupil size, have also been employed to gauge cognitive load (Paas & van Merrienboer, 1994). The resulting data from these methods are considered indicators of the cognitive load experienced by learners during the learning process.

There are two primary methods for objective-direct measurement of cognitive burden. One is neuroimaging techniques used to measure brain activity during task performance. However, this method is not preferred due to the complexity of the learning process, the unclear relationship between prefrontal cortex activity and cognitive load, and the intricate nature of the devices used for measurement (Braver et al., 1997). These factors can make obtaining accurate, reliable data challenging and may limit the practicality and feasibility of using physiological measures in diverse educational settings.

The second method for measuring cognitive load is the dual-task approach. This approach assumes that limited cognitive resources can be shared across different task dimensions (Brünken et al., 2003). If a learner has two tasks that require the same visual or auditory resources, the appropriate distribution of resources is required between the two tasks. Therefore, for accurate measurements of dual-task performance, the primary and secondary tasks must use the same cognitive resources (Brünken et al., 2003).

The dual-task approach is the second method for measuring cognitive load, based on the assumption that limited cognitive resources can be shared across different task dimensions (Brünken et al., 2003). When a learner engages in two tasks that demand the same visual or auditory resources, allocating resources between the two tasks is necessary. The dual-task methodology involves assessing the learner's performance on a secondary task while performing a primary learning task. Any decline in the secondary task performance may indicate a higher cognitive load experienced during the primary task, allowing researchers and educators to gauge the cognitive load imposed by different instructional designs or learning activities.

Moreover, the nature of cognitive load can change dynamically over time and in different contexts, making the assessment of cognitive load a complex undertaking (Sweller et al., 2019). For instance, as learners gain expertise in a certain area, the intrinsic cognitive load associated with tasks in that domain typically decreases. Furthermore, learning environments that adapt and respond to learners' needs can modulate cognitive load in real-time, providing scaffolding when the cognitive load is high and gradually reducing support as learners become more proficient. Hence, modern approaches to measuring cognitive load need to account for this dynamic, context-dependent nature of the cognitive load. Technological advancements are providing novel ways to achieve this, such as the use of machine learning algorithms to analyze real-time performance data and adjust learning materials accordingly. However, further research is needed to refine these techniques and evaluate their efficacy.

The learning environment can significantly impact cognitive load and learning outcomes (Cheryan, Ziegler, Montoya, & Jiang, 2017). A well-designed physical environment can help minimize extraneous cognitive load. For instance, a quiet and distraction-free environment can allow learners to devote their full cognitive resources to the learning task at hand, thereby facilitating deeper understanding and better retention of the material. Similarly, a clutter-free workspace can help learners stay organized and focused, reducing the cognitive load associated with searching for resources or managing multiple tasks simultaneously.

All in all, a comprehensive overview of CLT and its relevance in the learning process is provided in this section. It is discussed the different types of cognitive load (intrinsic, extraneous, and germane) and their impact on learning outcomes. It also examines the various methods of measuring subjective and objective cognitive load, as well as their advantages and limitations. The provided information is closely related to our research, which examines the effect of different learning environments and multimedia learning principles on learning cognitive load. Understanding the intricacies of CLT and the different forms of cognitive load is crucial for designing effective instructional materials in these learning environments. By identifying the factors contributing to cognitive load, the presentation of multimedia learning materials can be managed better to minimize extraneous load and maximize germane load, ultimately enhancing learning outcomes. By investigating the impact of these principles on cognitive load and learning outcomes in various learning environments, valuable insights into educational technology and instructional design can be contributed.

2.3 Cognitive theory of multimedia learning

The Cognitive Theory of Multimedia Learning (CTML), formulated by Mayer (2005), describes the mental processes that happen as learners acquire significant knowledge from multimedia educational materials. CTML is founded on three key assumptions: dual-channel processing, limited capacity, and active processing (Mayer, 2005). These assumptions help explain how learners effectively acquire and retain information when interacting with multimedia instructional materials.

The dual-channel assumption posits that information is managed through visual and auditory channels. The visual channel processes information received by the eyes, such as words on a monitor, while the auditory channel processes information received by the ears (Mayer, 2009). This notion of separated information processing is supported by Paivio's (1991) dual coding theory and Baddeley's (1986) working memory theory, both of which emphasize distinct channels for managing different types of information.

In alignment with the principles of Cognitive Load Theory (Chandler & Sweller, 1991) and Working Memory (Baddeley, 1986), the limited capacity assumption posits that each processing channel has a restriction for handling information concurrently. This constraint emphasizes the need for efficient instructional design that avoids overloading learners' cognitive resources while they engage with multimedia materials. Miller (1956) postulates that individuals can typically simultaneously maintain up to seven chunks of information in their working memory. However, proficient metacognitive strategies can enable learners to manage their restricted cognitive resources better (Mayer, 2009).

The third assumption, active processing, involves the learner actively participating in the learning process. This process encompasses three stages. Initially, the learner selects words and pictures through sensory input (e.g., ears, eyes). Subsequently, the chosen data (texts and pictures) is organized into coherent mental representations. Finally, these mental representations are integrated with existing knowledge from long-term memory (Mayer, 2009). This highlights the importance of learner engagement in effectively processing and retaining information from multimedia instructional materials.

Given the limited capacity of working memory, as stated in the assumption, learning becomes impaired when this capacity is exceeded (De Jong, 2010). This excess demand for cognitive resources can result in cognitive overload, which hampers the learner's ability to process and retain new information effectively. To prevent overloading memory demand and reduce cognitive load, instructional designs should be tailored to accommodate individuals' cognitive processing abilities. Mayer (2009) proposed 12 multimedia learning principles, organized into three categories reflecting different kinds that correspond to various forms of learner processing: extraneous processing, essential processing, and generative processing. These categories parallel the concepts of intrinsic, extraneous, and germane cognitive load, offering practical guidelines for creating multimedia instructional materials that optimize learning outcomes by effectively managing cognitive load.

There are five principles designed to minimize extraneous cognitive load: coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles.

The coherence principle posits that the most effective learning from multimedia materials occurs when captivating,

but unrelated content is avoided, as it fails to support the learning process (Mayer & Jackson, 2005). These principles provide valuable guidance for creating instructional materials that reduce unnecessary cognitive demands, thereby improving learning outcomes. Including irrelevant content in instructional materials may hinder students from constructing accurate mental models that effectively represent the information being taught. By eliminating unnecessary distractions, students can better focus on the essential elements, thus facilitating the formation of coherent mental representations that promote understanding and long-term retention.

The signaling principle posits that learners benefit when cues are incorporated into educational materials to direct their attention to the critical aspects of the content (e.g., Van Gog, 2014). Using visual aids such as highlighting, arrows, or other indicators, instructional designers can effectively capture learners' interest and guide them toward the most relevant information, thus facilitating better comprehension and retention.

The redundancy principle highlights the importance of carefully selecting the modes of information presented in multimedia learning materials. By avoiding unnecessary duplication of information across different formats, such as animation, narration, and on-screen text, instructional designers can prevent cognitive overload and optimize the use of learners' working memory. This approach enables learners to process, understand, and retain the content, ultimately leading to improved learning outcomes.

The spatial contiguity principle focuses on the physical proximity between text and images in multimedia learning materials. It posits that presenting words and images close together enhances learning (Mayer & Fiorella, 2014). When related text and visuals are separated, learners need to expend additional cognitive effort to locate and establish connections, increasing cognitive load. By adhering to the spatial contiguity principle, instructional designers can facilitate the efficient processing of information and improve learning outcomes.

The segmenting, pre-training, and modality principles are designed to manage intrinsic cognitive load. In the context of the current study, the aim is to investigate the effects of these principles across various learning environments, including AR, VR, and Traditional learning settings. The segmenting principle posits that learners can absorb information more effectively in smaller, manageable segments (Mayer & Pilegard, 2014). This principle also suggests that allowing learners to control the pace of multimedia instruction results in better learning outcomes, as it enables a "user-paced" approach. Conversely, "system-paced" instruction may hinder comprehension and the ability to discern causal relationships between steps, as learners may struggle to keep up or fully understand the content presented at a predetermined pace.

The pre-training principle suggests that learning can be enhanced when introducing core ideas and primary features before delving into the main content (Mayer, 2009). This approach allows learners to build a mental framework, especially for complex subject matter, to facilitate understanding during multimedia instruction. Pre-training effectively helps manage essential processing demands by providing foundational knowledge of crucial elements and features. The modality principle posits that learners benefit more from the information presented as narration rather than on-screen text, as it utilizes both auditory and visual channels in the learning process (Moreno & Mayer, 1999). The cognitive load is distributed by leveraging both channels, allowing more efficient processing and comprehension of the instructional content.

The final four principles aim to minimize germane cognitive load: multimedia, personalization, voice, and image principles. The multimedia principle asserts that learners gain a deeper understanding when presented with both words and images, as this facilitates mental connections (Mayer, 2009). To prevent redundancy, words should be presented in either written or spoken form, but not concurrently in both formats.

The personalization principle suggests that adopting an informal style promotes learning more effectively than a formal one (Mayer, Fennell, Farmer, & Campbell, 2004). Therefore, instructional designers should minimize academic language use in favor of more conversational tones. The voice principle emphasizes that learners respond more favorably to narrations delivered in a human voice than a computer-generated one (Mayer, 2009).

Lastly, the image principle posits that incorporating images of speakers in learning materials does not necessarily enhance learning outcomes. Instead, using relevant animations and visuals is more beneficial than merely displaying the instructor's talking head, as these elements can better illustrate key concepts and support understanding.

To conclude, CTML is the foundation for understanding how learners effectively acquire and retain information through multimedia instructional materials. This theory is based on three key assumptions: dual-channel processing, limited capacity, and active processing. To optimize learning outcomes and effectively manage cognitive load, Mayer (2009) proposed 12 multimedia learning principles organized into three categories: extraneous processing, essential processing, and generative processing. These principles are crucial to our research as we investigate modality, segmenting, and pre-training effects on learning outcomes, cognitive load, and motivation in various learning environments, such as augmented reality, virtual reality, and traditional settings. By applying these principles, we aim to understand better how different learning environments can be adapted to facilitate efficient instructional design, ultimately improving the learners' ability to process and retain new information without cognitive overload. This research seeks to contribute valuable insights to multimedia learning by examining the applicability and effectiveness of Mayer's multimedia principles in diverse learning environments, offering practical guidelines for instructional designers and educators in optimizing learning outcomes and managing cognitive load across various contexts.

2.3.1 Modality principle

At first glance, it may appear that there is no difference in learning outcomes between multimedia instruction delivered through on-screen text or narration when the content is the same. However, as Mayer (2009) suggested, the modality principle indicates that learners can improve their retention and comprehension of information when presented through narration instead of on-screen text due to the utilization of two channels simultaneously. This is because narration engages visual and auditory channels, allowing for more effective cognitive processing and reducing the likelihood of cognitive overload. Thus, although the information may be the same, how it is presented can significantly impact learning outcomes.

For example, Mayer and Moreno (1998) conducted two experiments investigating the modality effect on learning. The first experiment focused on the lightning formation, while the second explored a car's braking system. In each study, participants viewed multimedia presentations featuring animation and written text or animation and spoken narration. The results from both experiments revealed that participants exposed to spoken narration performed better on transfer tests, signifying enhanced comprehension of the material. These findings underscore the importance of the modality effect, indicating that pairing animation with spoken narration leads to more effective learning than combining animation and written text.

This principle has been supported by several other studies (Baceviciute et al., 2020; Clark & Mayer, 2016; Harskamp, Mayer, & Suhre, 2007; Mayer, 2017; Moreno, Mayer, Spires, & Lester, 2001; Tabbers, 2002). Moreover, this has been demonstrated through various measures such as reduced mental effort with narration (Oliveira, Huang, Azevedo, 2015), less cognitive load (Izmirli & Kurt, 2016), better decision-making performance (Fiorella, Vogel-Walcutt & Schatz, 2012), shorter task completion times (Mousavi et al., 1995), and better scores on retention and transfer assignments (Craig, Gholson, Driscoll, 2002; Greenberg, Zheng, Gardner & Orr, 2021; Harskamp et al., 2007; Kutbay & Akpınar, 2020). In addition, the modality effect has proven to be an effective instructional design technique in various domains, such as geometry (Jeung, Chandler, & Sweller, 1997), engineering (Tindall-Ford, Chandler, & Sweller, 1997), meteorology (Craig et al., 2002), mechanics (Mayer & Moreno, 1998), botany (Moreno, Mayer, & Lester, 2000), algebra (Atkinson, 2002), statistic (Leahy, Chandler, & Sweller, 2003), anatomy (Brünken, Steinbacher, Plass, & Leutner, 2002) and history (Brünken, Plass, & Leutner, 2004), evidenced by better transfer and problem-solving performances (Moreno, 2006).

Numerous studies have shown a positive impact of the modality principle, as evidenced by significant effect sizes. However, some studies do not demonstrate a positive effect size, particularly in specific conditions (Chen & Yang, 2020; Dousay & Trujillo, 2019; Inan et al., 2015, Mayer, Wells, Parong & Howarth, 2019; Liu, Jang & Roy-Campbell, 2018; Oberfoell & Correia, 2016; Schüler, Scheiter, & Gerjets, 2013; Wang, Li, Mayer, & Liu, 2018). For instance, Baceviciute et al. (2020) investigated the modality principle in virtual reality by contrasting presentations that combined audio and visual elements with only visual components. The findings indicate an opposite modality effect: reading leads to more significant learning gains than listening. Moreover, Ginns (2005) and Reinwein (2012) identified moderator variables, including the pace of presentation (e.g., Crooks, Cheon, Inan, Ari, & Flores, 2012; Schnotz, 2005; Tabbers, Martens, & Van Merriënboer, 2004), length of a verbal segment, and type of visualization, which influence the effectiveness of implementing the modality principle. The meta-analyses suggest that employing system-paced presentations, shorter verbal segments, and dynamic visualizations can lead to a more pronounced modality effect. These findings emphasize the importance of considering these factors when designing instructional materials to optimize learning outcomes.

A more recent systematic review by Çeken and Taşkın (2022) examined the modality regarding learning environments such as AR, VR, and traditional learning environments. Interestingly, augmented and virtual reality have not been utilized as learning environments in the examined studies. Moreover, the findings of this review can help educators and instructional designers to make informed decisions about the use of AR and VR in the design of learning environments. The review also emphasizes the importance of considering multimedia learning principles when designing educational materials for traditional and emerging learning environments. By adhering to these principles, educators and designers can optimize students' learning experiences and improve their learning outcomes. Overall, this review underscores the need for further exploration of the potential of AR and VR in educational contexts and the importance of incorporating multimedia learning principles in the design of learning materials.

2.3.2 Segmenting principle

The Segmenting Principle posits that individuals are more likely to learn effectively when information is presented in smaller sections (Mayer, 2009). For example, studies on multimedia instruction have investigated the efficacy of breaking up dynamic instructional animations into smaller segments instead of presenting them continuously (Almasseri & AlHojailan, 2019; Clark & Mayer, 2010; Lusk et al., 2009; Soicher & Becker-Blease, 2020). This principle also suggests that if learners can regulate the pace of multimedia instruction, known as "user-paced", they will experience improved learning outcomes (Cheon, Crooks, & Chung, 2014; Hasler, Kersten, & Sweller, 2007; Mayer et al., 2019). Conversely, when multimedia instruction is "system-paced," learners may struggle to grasp the material completely and recognize the causal links between various steps.

Several other studies have corroborated the segmenting principle (Chen & Yen, 2021; Hassanabadi, Robatjazi, & Savoji, 2011; Mayer & Chandler, 2001; Mayer et al., 2019; Mayer, Howarth, Kaplan, & Hanna, 2018; Mayer & Moreno, 2003; Moreno, 2007). Additionally, this has been evidenced by different indicators like decreased mental effort (Mayer et al., 2018; Schüler et al., 2013), reduced perceived difficulty (Hassanabadi et al., 2011; Mayer et al., 2019), lower cognitive load (Hasler et al., 2007; Mayer & Moreno, 2003), improved comprehension (Boucheix & Schneider, 2009), increase satisfaction (Altinpulluk, Kilinc, Firat, & Yumurtaci, 2020) better achievement score (Fong, Lily, & Por, 2012; Hasler et al., 2007), and superior performance on retention and transfer tasks (Cheon et al., 2014; Klingenberg et al., 2023; Schüler et al., 2013). Furthermore, a research review on segmenting by Mayer and Pilegard (2014) found that, in all 10 experimental comparisons, students demonstrated better performance on transfer tests when they engaged with a segmented version of a multimedia lecture rather than a continuous one.

Lightning, brakes, and pumps subject matters were initially utilized more often than other topics during the development of CTML (Mayer, 2017), which has led many researchers to employ these subjects in various contexts or scenarios to examine multimedia learning principles. Regarding current literature, the segmenting principle has been demonstrated to be a successful method for designing instruction in diverse subject matters, including but not limited to astronomy (Hasler et al., 2007), authoring tool (Chen & Yang, 2020), cell division (Fong et al., 2012; Schüler et al., 2013), teaching skills (Moreno, 2007), history (Lusk et al., 2009), kidney function (Soicher & Becker-Blease, 2020), geography (Mayer et al., 2018; Mayer et al., 2019), and bloodstream (Parong & Mayer, 2018) as demonstrated by improved problem-solving abilities and transfer of knowledge.

Despite the numerous studies that have demonstrated the beneficial impact of the segmenting principle, certain studies have failed to show positive effects, especially in particular circumstances, on mental effort (Cheon et al., 2014; Hassanabadi et al., 2011), perceived difficulty (Cheon et al., 2014; Stiller & Zinnbauer, 2011), cognitive load (Chen & Yen, 2021), motivation (Parong & Mayer, 2018), enjoyment (Mayer et al., 2019), achievement (İzmirli, 2012), retention (Schüler et al., 2013; Stiller & Zinnbauer, 2011), and transfer (Chen & Yang, 2020; Soicher & Becker-Blease, 2020). For instance, a recent study by Klingenberg et al. (2023) revealed that incorporating segmentation into an immersive VR lesson resulted in a better learning transfer than the control condition. However, it did not lead to the acquisition of more factual knowledge. Similarly, segmenting a multimedia presentation of healthcare-related information did not result in a notable decrease in cognitive load or provide any advantages for remembering fundamental facts or transferring the information to new scenarios (Soicher & Becker-Blease, 2020).

2.3.3 Pre-training principle

The pre-training principle suggests that learners can more effectively comprehend new content when familiar with the terminology and attributes of the primary ideas involved. By providing a foundation of essential knowledge beforehand, learners are better prepared to engage with new information and more likely to experience successful learning outcomes. Pre-training helps to address the demands of essential processing by introducing important elements and features of the material, thereby facilitating learning.

For instance, in their experiment, Mayer, Mathias, and Wetzell (2002) sought to investigate the effectiveness of multimedia learning principles by teaching a group of students about a car's braking system. In the experiment, students were separated into two distinct groups: the first group underwent pre-training on individual system components prior to viewing a narrated animation, whereas the second group did not engage in any pre-training activities. The pre-training consisted of presenting descriptions or illustrations of each component's various states, aiming to familiarize students with the terminology and attributes of each part. The study's findings demonstrated that students who underwent pre-training outperformed those who did not in terms of transfer and retention test scores. This indicates that providing learners with a foundational knowledge of system components can lead to more effective learning and better long-term information retention.

Pollock, Chandler and Sweller (2002) investigated the effectiveness of pre-training in apprentices of electrical engineering who underwent multimedia lessons about the safety testing of electrical applications. These results suggest that pre-training can be an effective strategy for improving problem-solving transfer in learners with low prior knowledge. Learners may be better equipped to transfer this knowledge to new problem-solving situations by clearly explaining how each component works. However, pre-training may offer little advantages for learners with high prior knowledge as they may already possess the necessary knowledge and skills to solve problems in new contexts.

Several other studies have provided support for the pre-training principle. For instance, Mayer, Mautone and Prothero's (2002) study showed that pre-training improved students' performance in a simulation game that focused on geology. In another study, Clarke, Ayres and Sweller (2005) found that pre-training enhanced the performance of high school students with low spreadsheet skills in a lesson on graphic representations of linear functions. Kester and colleagues conducted four experiments and found that pre-training was helpful for one study involving electrical circuit problems (Kester, Kirschner, & van Merriënboer, 2006) but not for other studies involving statistics problems (Kester, Kirschner, & van Merriënboer, 2004) or neural network problems (Kester, Lehnen, van Gerven, & Kirschner, 2006). Furthermore, Eitel, Scheiter and Schüler (2013) demonstrated that pre-training with a self-paced diagram improved performance on a comprehension test about the structure and functioning of a pulley system.

The study by Meyer et al. (2019) showed that the impact of pre-training varied depending on the media and method used. These findings revealed that pre-training could improve knowledge acquisition, transfer, and self-efficacy immediately after the immersive VR (I-VR) environment intervention. These results suggest that pre-training can have a lasting impact on self-efficacy in I-VR environments, even after the intervention has passed. However, pre-training had no significant effects on knowledge acquisition, transfer, and self-efficacy in video-based learning.

The study by Petersen et al. (2020) provides valuable insights into the effectiveness of pre-training in VR by having participants take a virtual tour of Greenland to learn about climate change. They found that participants who received pre-training before the VR experience had significantly better transfer scores than those who received the same training material integrated into the VR experience. The study proposes that pre-training can improve the efficiency of VR learning experiences. Although some studies have confirmed the benefits of pre-training for enhancing learning outcomes, some studies indicate no significant effect or suggest that the effectiveness of pre-training can depend on factors such as the type of pre-training or learners' prior knowledge level (Mayer et al., 2002; Mayer & Pilegard, 2014; Meyer et al., 2019).

Mayer and Pilegard (2014) and Mayer (2017) reviewed previous literature on the pre-training principle and discovered a medium effect. Most research on the pre-training principle has focused on slideshows or video games with low immersion (Meyer et al., 2019). Additionally, Çeken and Taşkın (2022) discovered that only one investigation had been conducted on the pre-training principle in virtual reality learning environments, and no studies have been carried out on utilizing this principle in augmented learning environments. Therefore, there is a substantial gap in the literature concerning the pre-training principle in both VR and AR learning environments.

CHAPTER 3

METHODOLOGY

This research investigates the effect of modality, segmenting, and pre-training principles on learning outcomes, cognitive effort, and motivation in various learning settings such as traditional, virtual reality, and augmented reality. This chapter will detail the research methodology, including research design, participants, instrumentation, learning materials, data collection procedures, data scoring, and data analysis.

3.1 Research design

A 3 X 4 factorial design was used in the study. Participants were assigned to groups randomly to minimize internal validity threats such as selection, history, regression, and mortality. There are four dependent variables in the study. These are retention, transfer, cognitive load, and motivation scores. There are two independent variables: learning environments (Virtual Reality, Augmented Reality, and Traditional), and learning materials designed based on multimedia learning principles (N, NS, T, NP). There are twelve treatment groups, as in Table 1. Each group used two learning materials about the lightning formation and cell subject matter to examine the complexity effect.

Table 1. Research Design of the Study

Groups	Intervention 1	Post Tests 1	Intervention 2	Post Tests 2
VR - Narration	$X_{1.1}$	* Retention * Transfer * Cognitive Load * Motivation	$X_{1.2}$	* Retention * Transfer * Cognitive Load * Motivation
VR – Narration + Segmenting	$X_{2.1}$		$X_{2.2}$	
VR - On-screen Text	$X_{3.1}$		$X_{3.2}$	
VR – Narration + Pre-training	$X_{4.1}$		$X_{4.2}$	
AR - Narration	$X_{5.1}$		$X_{5.2}$	
AR – Narration + Segmenting	$X_{6.1}$		$X_{6.2}$	
AR - On-screen Text	$X_{7.1}$		$X_{7.2}$	
AR – Narration + Pre-training	$X_{8.1}$		$X_{8.2}$	
TR – Narration	$X_{9.1}$		$X_{9.2}$	
TR – Narration + Segmenting	$X_{10.1}$		$X_{10.2}$	
TR - On-screen Text	$X_{11.1}$		$X_{11.2}$	
TR – Narration + Pre-training	$X_{12.1}$		$X_{12.2}$	

Intervention 1: Using learning material specially designed for each treatment group about the lightning formation

Intervention 2: Using learning material specially designed for each treatment group about the cell structure

3.2 Participants

Convenience sampling was used as a method (Creswell & Creswell, 2017). This method was chosen due to its practicality and ease of access to potential participants within the university setting. Initially, professors from different departments at the Boğaziçi University were reached out and asked for their support to disseminate information about the study to their student groups. Then, students from these departments were invited to participate in the study. After that, interested students who expressed a willingness to participate in the study were included. Efforts were made to ensure diverse representation by involving professors from various departments and academic disciplines. However, it is important to note that the final composition of the sample was influenced by the voluntary participation of students.

In the spring semester of the 2021-2022 academic year, 393 students from various departments at Boğaziçi University participated in the experiment. They participated in the research voluntarily. A prior knowledge test was carried out on all students just before the experiment session to ensure that all participants had low prior knowledge about the subject matter of the learning material. After administering the prior knowledge questionnaire, among the potential participants, 10 students who were found to have a high level of prior knowledge were excluded from the sample. Therefore, the final sample for the study consisted of 383 university students with low prior knowledge. The students who participated in the study were between 19 to 34 years old, with an average age of 23.31 years old (Std. Dev. = 2.191). Further information about the distribution of participants across different departments and their demographic characteristics are presented in Tables 2 and 3.

Table 2. Gender and Age Distribution of the Participants

Gender	*f*	%
Male	214	55,9%
Female	165	43,1%
Other	4	1,0%
Age	*f*	%
19	10	2,6%
20	23	6,0%
21	53	13,8%
22	56	14,6%
23	66	17,2%
24	56	14,6%
25	71	18,5%
26	25	6,5%
27	11	2,9%
28	6	1,6%
29	4	1,0%
32	1	0,3%
34	1	0,3%

Table 3. Main Fields and Department Distribution of the Participants

Main Fields	Department	*f*	%	*f*	%
Social Science	Computer Education and Educational Technology	123	32,1%	190	49,6%
	Guidance and Psychological Counseling	27	7,0%		
	Tourism Administration	13	3,4%		
	Mathematics Education	9	2,3%		
	Economics	5	1,3%		
	International Trade	4	1,0%		
	Chemistry Education	3	0,8%		
	Management	2	0,5%		
	Philosophy	2	0,5%		
	Physics Education	1	0,3%		
	Primary Education	1	0,3%		
Applied Science	Management Information Systems	93	24,3%	156	40,8%
	Computer Engineering	47	12,3%		
	Industrial Engineering	9	2,3%		
	Electrical and Electronic Engineering	2	0,5%		
	Environmental Science	2	0,5%		
	Chemical Engineering	1	0,3%		
	Civil Engineering	1	0,3%		
	Mechanical Engineering	1	0,3%		
Humanities	Foreign Language Education	25	6,5%	26	6,8%
	Turkish Language and Literature	1	0,3%		
Natural Science	Physics	6	1,6%	11	2,9%
	Mathematics	3	0,8%		
	Molecular Biology and Genetics	2	0,5%		

3.3 Instrumentation

Before the main study, a pilot study was conducted to evaluate and refine the AR, VR, and traditional learning materials used in this research. The materials were tested on a small group of students, and their feedback was collected regarding comprehensibility, usability, and any technical issues. Based on the results of the pilot study, several improvements were made to enhance the effectiveness of the materials and to ensure the smooth conduct of the main study. In addition, the following instruments were utilized to collect the data needed to answer the research questions in the study: demographic information questionnaire, prior knowledge test, retention test, transfer test, cognitive load questionnaire, and motivation questionnaire. The following section will explain the pilot study and the details of each of these instruments.

3.3.1 Demographic information questionnaire

The demographic information questionnaire consists of the following items: gender, age, department, learning environment, and version of the learning material. Each item was mandatory. Participants were informed about which version of the learning environment was used in the experiment and the learning environment they were involved. Age, gender, and department information were only intended to obtain and report descriptive information about the sample participating in the study. The learning environment and learning material version information were considered independent variables.

3.3.2 Prior knowledge test

The study included two prior knowledge tests, which aimed to assess the initial knowledge level of the participants about meteorology and cell structure. The meteorology prior knowledge test was adopted from the study by Moreno and Mayer (1999). The knowledge checklist used in the study consisted of 7 items, which included a prompt for participants to place a checkmark next to the items that applied to them. One example of an item on the checklist is "I regularly read the weather maps in the newspaper." The remaining items on the checklist can be found in Appendix A. Each checkmark was graded one point. Participants in the study were asked to rate their level of knowledge about meteorology (weather) using a 5-point self-rating scale, where one indicated "very little" and 5 indicated "very much". The prompt for the question was, "Please put a checkmark indicating your knowledge of meteorology (weather):". The maximum number of points that could be obtained on the prior knowledge test for meteorology was 12. Individuals who scored above six were excluded to ensure that only participants with a low level of prior knowledge about meteorology were included in the study.

The prior knowledge test about cell structure was adopted from the study by Meyer et al. (2019). This test included a 9-item knowledge checklist and a 7-item self-rating scale, which assessed participants' level of knowledge about cell biology. One of the items on the knowledge checklist was "How much do you know about cell biology?" The remaining items on the checklist can be found in Appendix B. On the 7-item self-rating scale ranging from 1 (very little) to 7 (a lot), the following question was asked to the participants: “How much do you know about cell biology?”. Participants were not allowed to proceed to the next stage in the research without responding to the prior knowledge tests, including meteorology and cell structure.

3.3.3 Retention test

There were two retention tests about the lightning formation and cell structure. The retention test about lightning formation was as in the study by Moreno and Mayer (1999). In the lightning formation retention test, participants were instructed to write down everything they could remember about how lightning works. The answers were graded based on the rubric (see Appendix C, Table C1). The maximum number of points was 50 in total. The retention test about cell structure was as in the study by Meyer et al. (2019). The retention test question for cell structure subject matter was as follows:

"Please write down all you can remember about cells". The answers were graded based on the rubric (see Appendix C, Table C2). The maximum number of points was 80 in total. The participants were asked to write their answers in Turkish for both subjects (in their native language).

3.3.4 Transfer test

There were two transfer tests about lightning formation and cell structure. The transfer test about lightning formation was adopted from the study by Moreno and Mayer (1999). One example of an item on the test is: "What could you do to reduce the intensity of lightning?”. The remaining items on the test can be found in Appendix D. The answers were graded based on the rubric (see Appendix E, Table E1). The maximum number of points was 8 in total.

The transfer test about cell structure was adopted from the study by Meyer et al. (2019). One example of an item on the test is: "Why does the cell not get flooded with all the various elements surrounding the cell?”. The remaining items on the test can be found in Appendix D. The answers were graded based on the rubric (see Appendix E, Table E2). The maximum number of points was 8 in total. The participants were asked to write their answers in Turkish for both subjects (in their native language).

3.3.5 Cognitive load questionnaire

The cognitive load questionnaire was initially developed by Leppink, Paas, Van der Vleuten, Van Gog, and Van Merriënboer (2013). Then, Andersen and Makransky (2021) enhanced the instrument by dividing extraneous cognitive load into sub-dimensions stemming from instructions, noise, and devices. The version by Andersen and

Makransky (2021) was used in the study (see Appendix F). According to the meta-analysis by Mutlu-Bayraktar et al. (2019), the instrument is the second most frequently used subjective measure. The instrument contains 16 items to measure the intrinsic load (three items, e.g., "The topics covered in the activity were very complex"), extraneous load for instruction (three items, e.g., "The instructions and explanations during the activity were very unclear"), extraneous load for noise (three items, e.g., "Students talking to me during the activity made it hard to learn"), extraneous load for the device (three items, e.g., "Message and notifications from my phone/computer made learning unclear"), and germane load (four items, e.g., "The activity really enhanced my understanding of the topics covered"). The Likert scale ranges from 0 = “not at all the case” to 10 = “completely the case”.

3.3.6 Motivation questionnaire

The "Instructional Materials Motivation Survey" (IMMS), first created by Keller in 2009, originally comprised 36 items. However, Loorbach, Peters, Karreman, and Steehouder, (2015) condensed it to 12 items. This shortened version was employed in the research (see Appendix G). The 12-item scale uses a 5-point range, with 0 representing "not true" and five indicating "very true." It measures four sub-dimensions with three items each: attention (e.g., "The text's quality assisted in maintaining my focus"), relevance (e.g., "The lesson's content will be beneficial to me"), confidence (e.g., "While working on this lesson, I felt sure I could grasp the material"), and satisfaction (e.g., "Studying this lesson was genuinely enjoyable").

3.3.7 Pilot study

The aim of the pilot study is to determine whether there are any technical or usability shortcomings in the AR, VR, and traditional learning materials, and to see if there will be any issues encountered during the research process. For this reason, a pilot study was conducted. There are twelve treatment groups in the study. For each treatment group, five participants attended the pilot study. The total number of participants in the pilot study was 60.

The results of the pilot study indicated generally positive feedback on the comprehensibility and usability of the AR and VR materials. However, some students encountered technical issues (e.g., lags when using the material), which negatively impacted on the learning process. Traditional learning materials, on the other hand, were generally found to be understandable and no technical issues were encountered. In addition, some participants stated the need for an introduction process (5-10 seconds) to focus on the learning material. It was determined how long each version of learning materials took in the pilot study to provide information for the participants in the main study.

Based on these findings, additional work was decided to be done to resolve the technical issues in the AR and VR materials. Also, it was planned to further improve the user interface to increase the usability of these materials. The pilot study has made a significant contribution, particularly in early detection and resolution of technical issues. Therefore, the pilot study has played a critical role in the applicability and effectiveness of the main research.

3.3.8 Learning materials

The study has three learning environments (VR, AR, and Traditional). Learning materials with four versions, namely Narration (N), Narration + Segmenting (NS), On-screen Text (T), and Narration + Pre-training (NP) on two different topics (lightning formation and cell structure), were developed for each learning environment specifically for this study. There are slight differences between versions of the learning materials. For instance, the N version delivers the instruction as narrated and system-paced, whereas instruction is provided as narrated but user-paced in the NS version. The only difference between the N and T versions is that instruction is presented as on-screen text in the T version. When comparing the N version, there is an extra screen in the NP version. The extra screen explains the key terms to enhance the learning process.

When the content and scenario of the learning material were developed, the opinions of three faculty members with doctorate degrees in the educational technology field were taken. The content and scenario were shaped in line with expert opinions, and each subsequent stage was presented for expert opinion. Then, it was delivered to the software development company as an outsourcing service to develop learning materials for AR and VR learning environments. The development process took six months. Figure 2, 3, 4, 5, 6, 7, 8,9,10, and 11 shows sample screenshots of each subject material for the VR and AR learning environment. The participants used learning material in the virtual reality environment via Oculus Quest 2 headset, whereas an Android mobile phone was used for the learning material in the augmented reality environment.

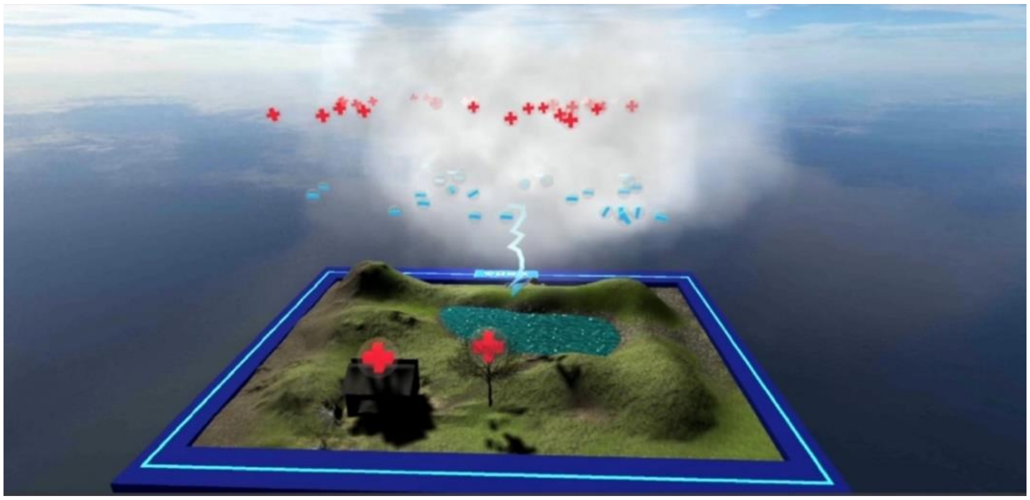

Figure 2. Sample screenshot of learning material about lightning formation in the VR learning environment – N version

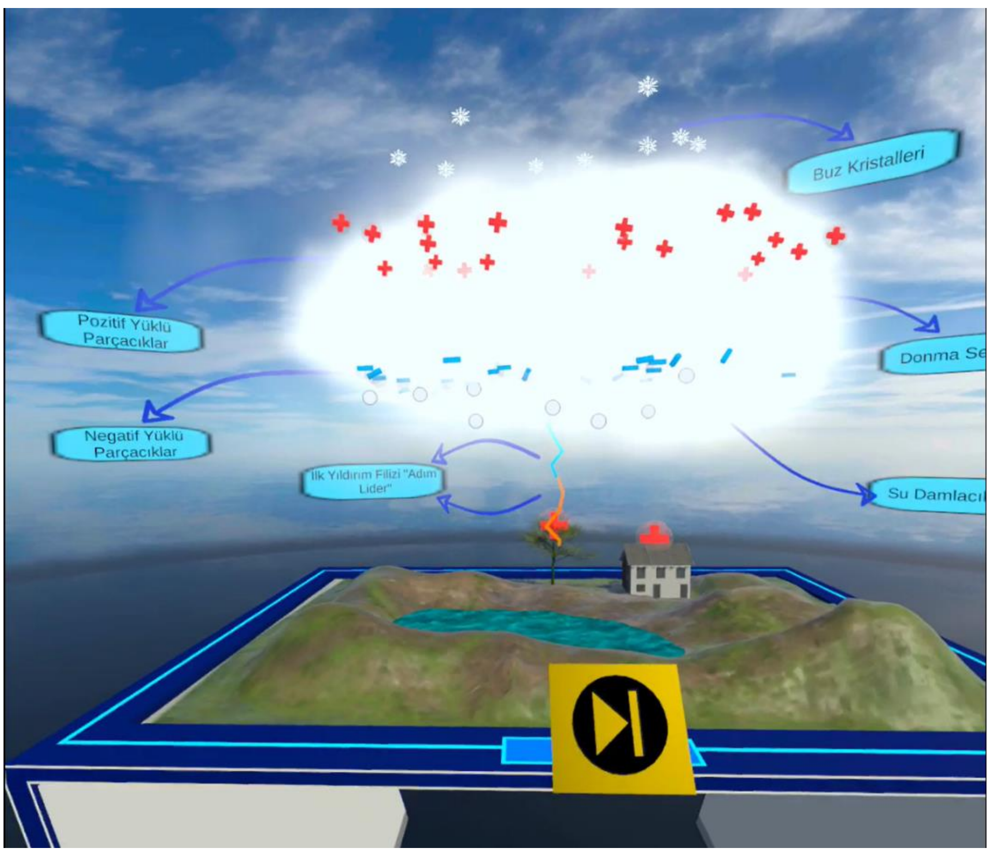

Figure 3. Sample screenshot of learning material about lightning formation in the VR learning environment – NP version

Figure 4. Sample screenshot of learning material about lightning formation in the VR learning environment – T version

Figure 5. Sample screenshot of learning material about lightning formation in the VR learning environment – NS version

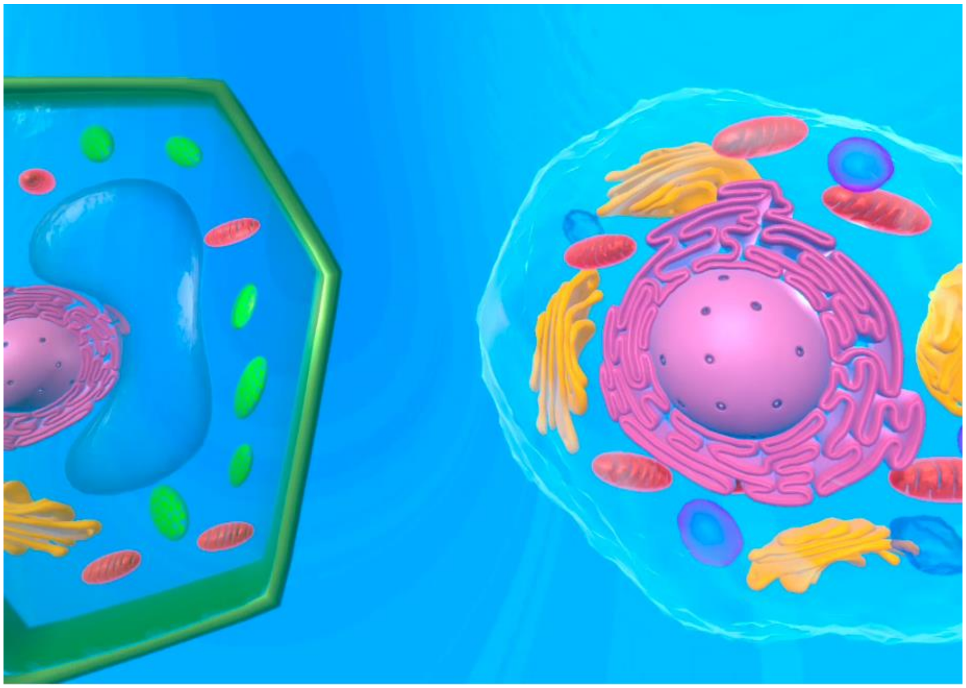

Figure 6. Sample screenshot of learning material about cell structure in the VR learning environment – N version

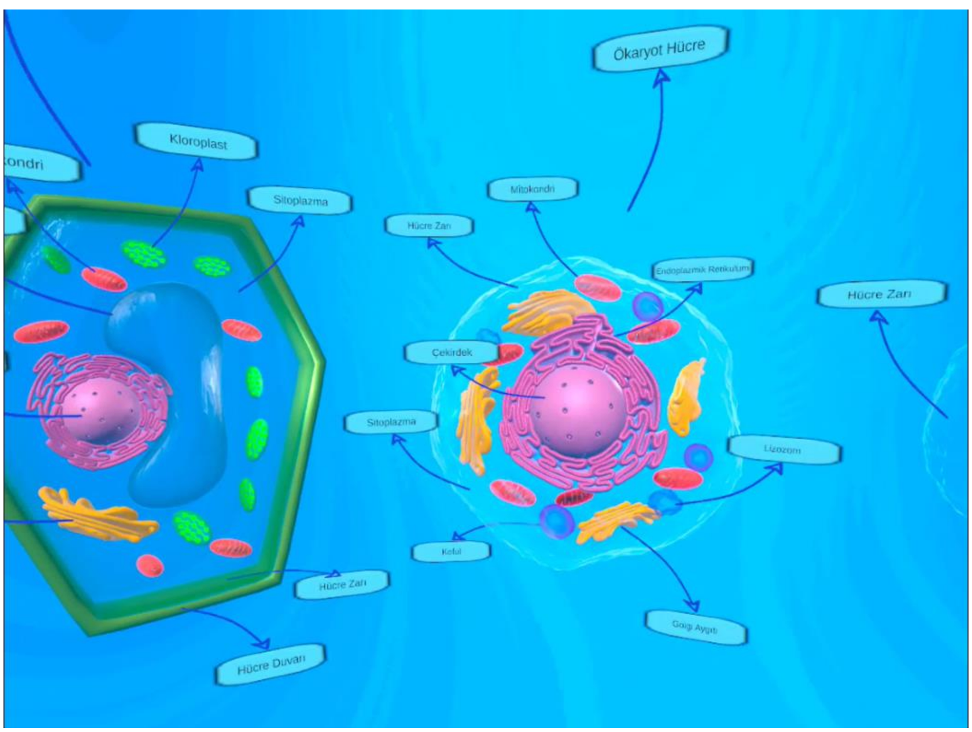

Figure 7. Sample screenshot of learning material about cell structure in the VR learning environment – NP version

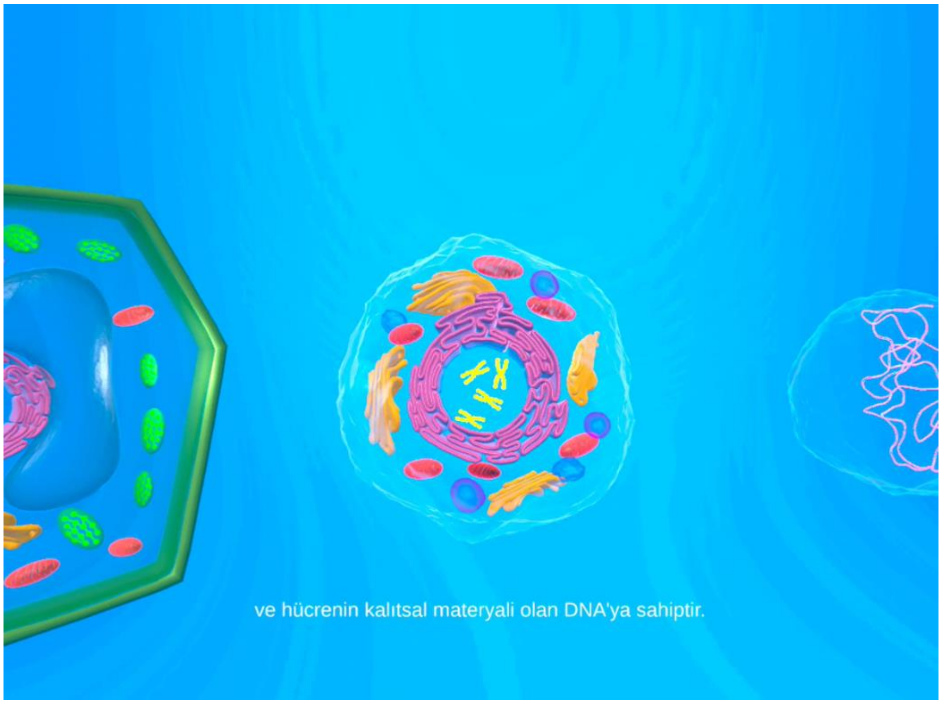

Figure 8. Sample screenshot of learning material about cell structure in the VR learning environment – T version

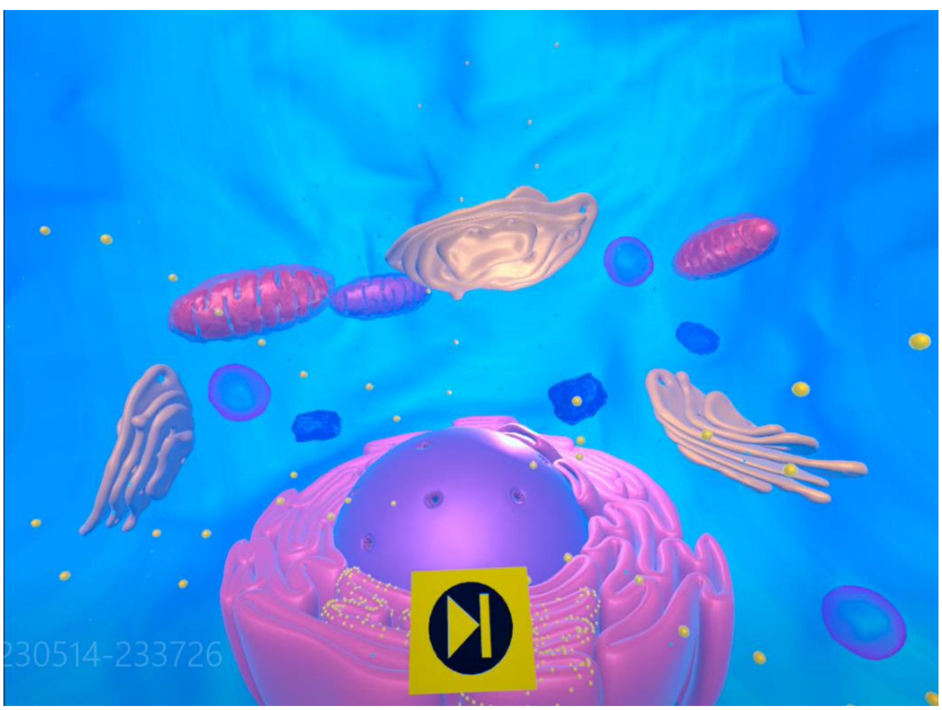

Figure 9. Sample screenshot of learning material about cell structure in the VR learning environment – NS version

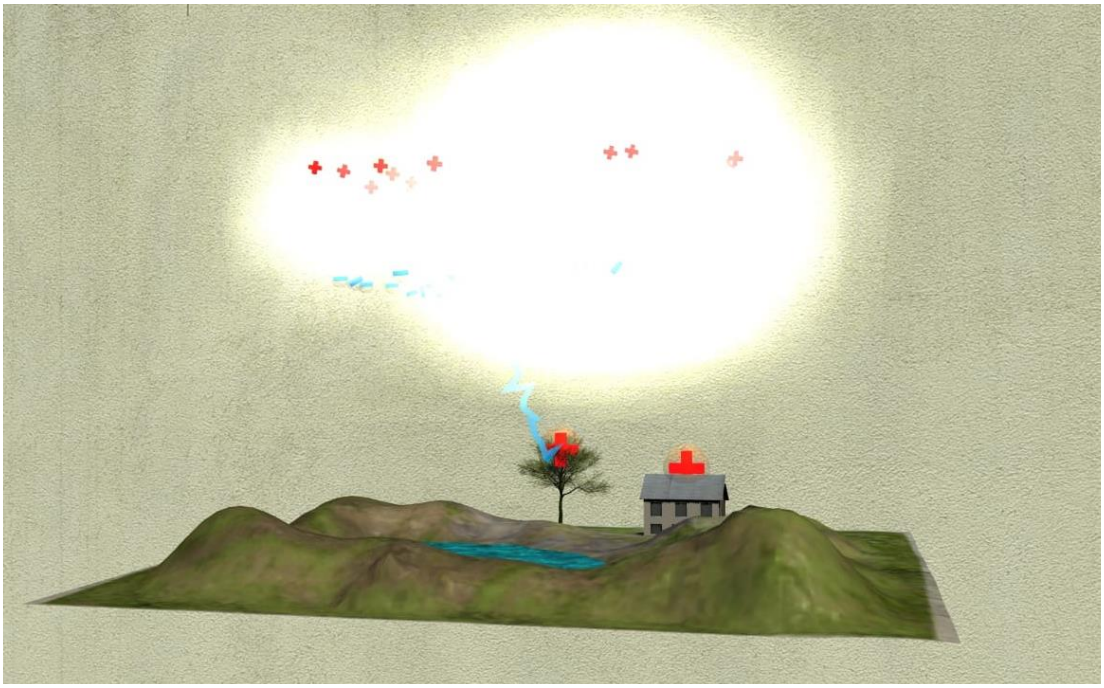

Figure 10. Sample screenshot of learning material about lightning formation in the AR learning environment – N version

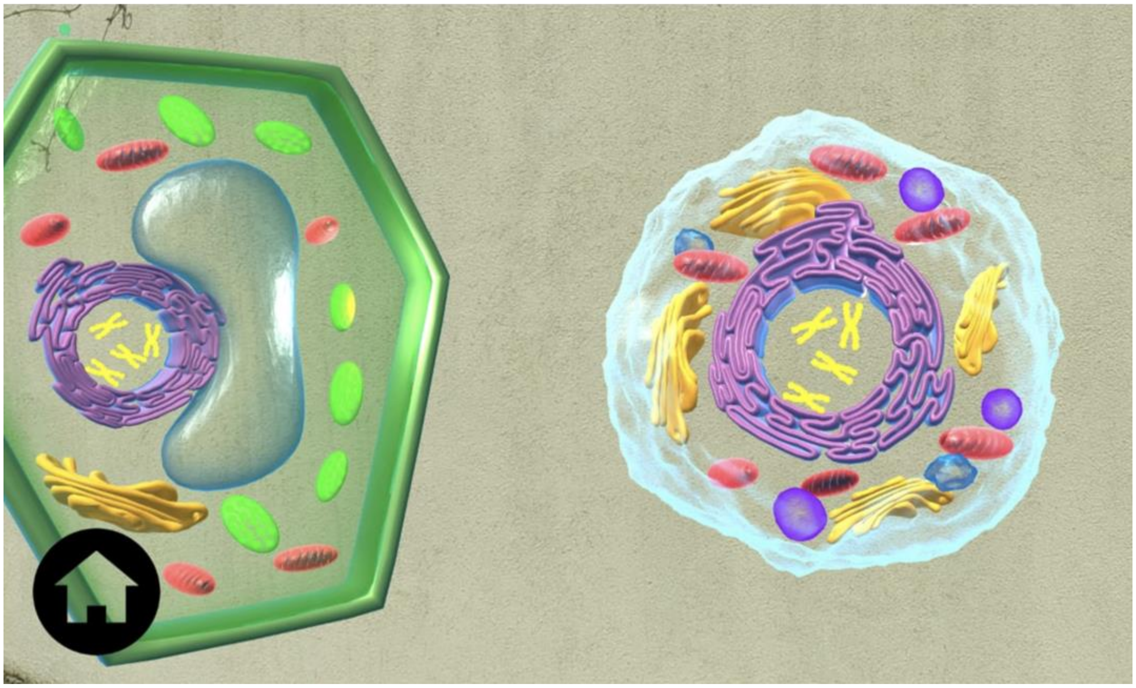

Figure 11. Sample screenshot of learning material about cell structure in the AR learning environment – N version

The researcher developed learning materials for traditional learning environments. Designing and creating the traditional learning environment took around two months to complete. A domain and hosting service was purchased to serve learning material online. Figures 12, 13, 14, and 15 show sample screenshots of each subject material for the traditional learning environment. The participants used the learning material via their mobile phones or computers.

Figure 12. Sample screenshot of learning material about lightning formation in the traditional environment – N version

Figure 13. Sample screenshot of learning material about lightning formation in the traditional environment – T version

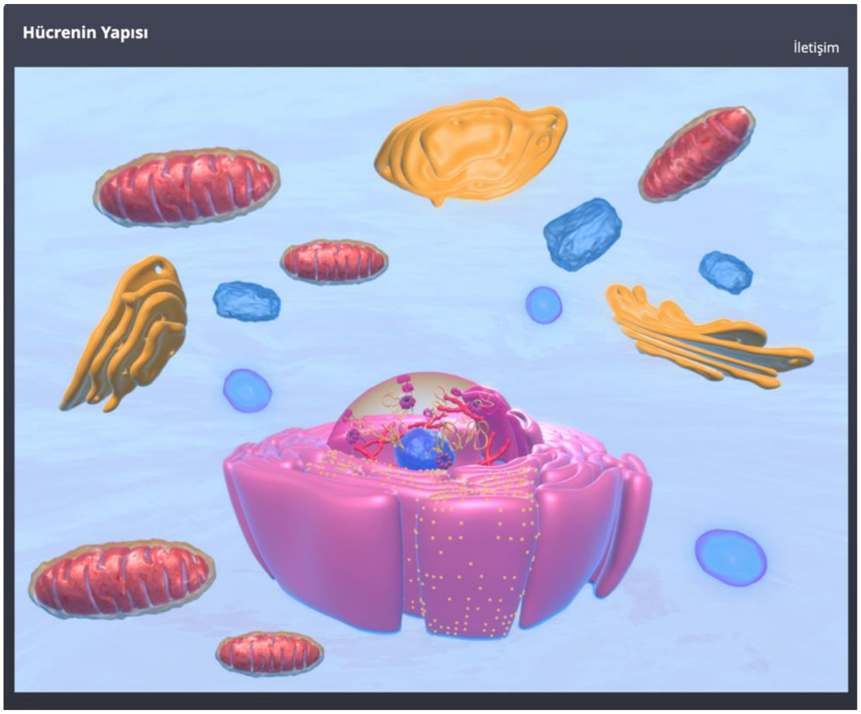

Figure 14. Sample screenshot of learning material about cell structure in the traditional environment – N version

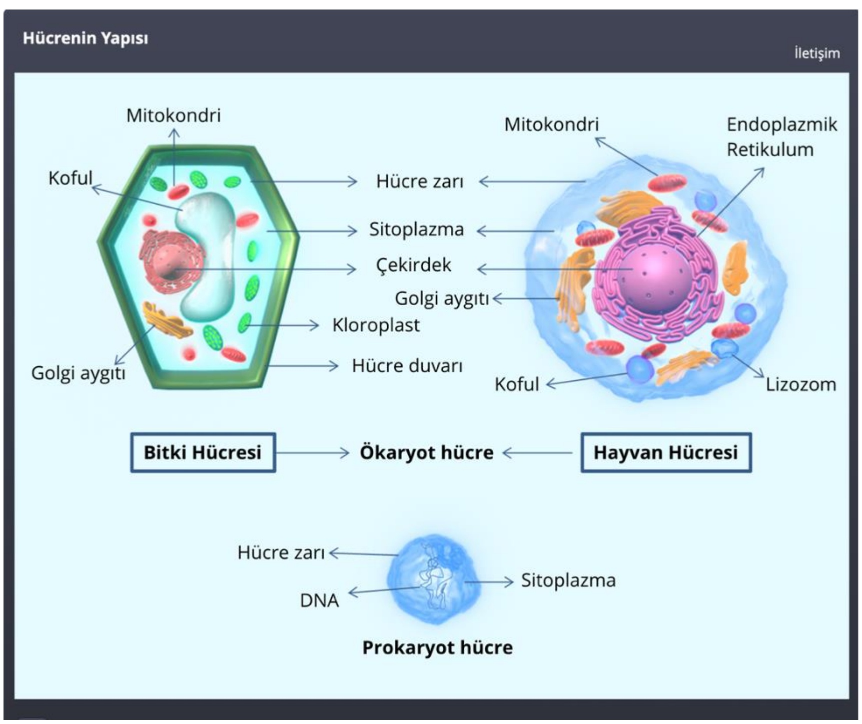

Figure 15. Sample screenshot of learning material about cell structure in the traditional environment – extra screen for NP version

3.4 Data collection procedure

The research data collection process was conducted in different places at Boğaziçi University, such as IT laboratories and classes. Since the study is experimental, an application was made to obtain approval from the Ethics Committee for Master and Ph.D. Theses in Social Sciences and Humanities (SOBETİK). The permission with protocol number 11776 was obtained on 15.04.2021 (see Appendix H).

Before the data collection process, the professors at Boğaziçi University were asked to invite their students to study, and the research summary and contact information were shared with the students. Some professors allocated some of their class time for the data collection. In addition, some students contacted the researcher via email and the meetings were set for them separately. Volunteer students were briefed on the experimental research procedure before initiating the data collection process. The stages of data collection can be seen in Figure 16.

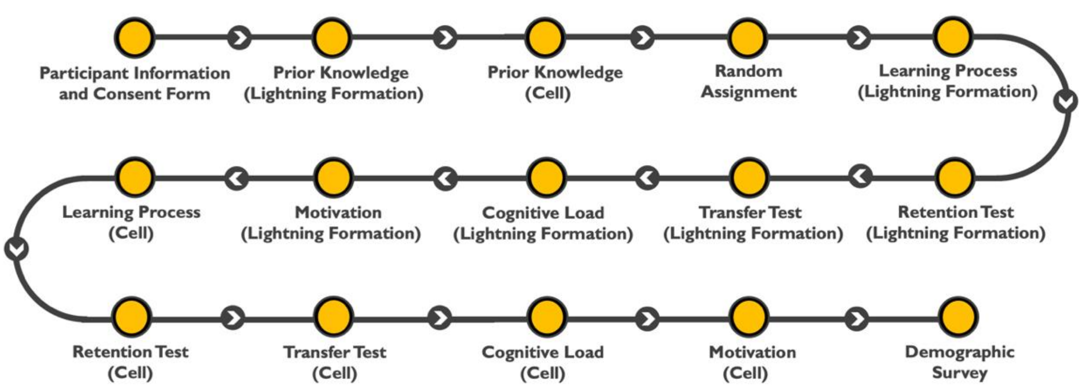

Figure 16. Data collection process

After giving information about the research and process, the participation information and consent form was delivered to the students. In this step, the purpose of the study, information about the study in the ethical context and information about the researchers were included. Students who agreed to partake in the research signed the consent form and proceeded to the subsequent stage. Since the questionnaires and tests were conducted through Survey Monkey, a program designed as an online research tool, a link was shared with the participants. The first two sections of the online survey research tool included prior knowledge test questions about lightning formation and cell structure. Once the participants finished the prior knowledge assessments, they were randomly allocated to one of the twelve experimental groups. Participants assigned to one of the virtual reality groups used a VR headset, whereas those assigned to one of the augmented reality groups used an Android mobile phone. They chose the lightning formation subject matter from the menu in the application and their version based on the assigned treatment group. There was no menu for traditional learning environment groups since there were different links for each version and subject matter. The participants watch the assigned learning materials. There were slight differences in the time of the learning process. For instance, the N version is system paced and takes two minutes 46 seconds, whereas there is no exact time in the NS version because the NS version is user-paced. The student clicks the “Next” button to continue the next part of the learning material. In addition, the students examine the extra screen without time restriction at the beginning of the learning material in the NP version to increase their prior knowledge. When they click the “Next” button on the extra screen in the NP version, the rest of the learning material is the same as in the N version. According to the researcher's observation, the duration of using learning materials for lightning formation took between three and five minutes. Upon finishing the learning phase, participants responded to the retention test, transfer test, cognitive load questionnaire, and motivation questionnaire sequentially through the online survey platform. There was no time restriction for them. After answering the questions, the participants used learning material about the cell structure subject matter. There were slight differences in the time of the learning process, as in the lightning formation. For instance, the N version is system paced and takes five minutes and eight seconds, whereas there is no exact time in the NS version because of user pace. Then, they finished the retention test, transfer test, cognitive load questionnaire, and motivation questionnaire. As the last step, they complete the demographic survey. Based on observation by the researcher, the data collection process for a person took at least 35 minutes and at most 60 minutes. Figure 17 and Figure 18 depict images of participants engaging in activities during the data collection process within the VR learning environments. The data collection process took approximately four months.

Figure 17. Participants in the data collection process - 1

Figure 18. Participants in the data collection process - 2

3.5 Data scoring

Three raters evaluated the participant’s retention and transfer test scores based on the rubrics (Appendix C). One of the raters is the researcher himself. The second rater is a Ph.D. candidate at Boğazici University. The third is an educational technologist with an M.A. as a science teacher. There are 24 sub-criteria of lightning formation subject matter for the retention test in the rubric, whereas the number is 38 for the cell structure. The raters gave points based on each criterion (full or partial). In the transfer tests, four sub-criterions correspond to each question. For each question, there were two acceptable answers. Each acceptable answer was graded one point. After the rating process was done, interrater reliability was measured. The values of agreement between raters are given in Table 4. An agreement between 81% and 100% is very good based on the classification of Cohen's kappa. Furthermore, the mean score of the three raters constitutes the participant's final score for each criterion or question. The total of these average scores signifies the participant's final score for each respective test.

Table 4. Interrater Reliability for Retention and Transfer Test

	Retention - Lightning Formation			Retention - Cell		
	R1	R2	R3	R1	R2	R3
R1	-	-	-	-	-	-
R2	87%	-	-	91%	-	-
R3	89%	89%	-	92%	92%	-
	Transfer - Lightning Formation			Transfer - Cell		
	R1	R2	R3	R1	R2	R3
R1	-	-	-	-	-	-
R2	82%	-	-	89%	-	-
R3	86%	85%	-	88%	90%	-

3.6 Data analysis

The collected data within the research scope were analyzed using IBM SPSS Statistics 27 software. Descriptive statistics were computed, including mean, median, standard deviation, frequency, skewness, and kurtosis values. The following were used to assess whether data are normally distributed: histogram, Q-Q probability plot (graphical), Shapiro-Wilk test, Kolmogorov-Smirnov test, and Skewness & Kurtosis values (analytical).

One-way ANOVA was carried out for the first and second research questions once the following three conditions were satisfied: (a) identifying outliers through boxplots, (b) evaluating data set normality for each cell in the design, and (c) assessing homogeneity of variances using Levene's test for equality of variances. The Bonferroni test was conducted as a post hoc analysis for ANOVA because it is commonly described as stricter than the Tukey test, tolerating type I errors and less conservative than Scheffé's method (Lee & Lee, 2018). Furthermore, it is applicable even with unequal sample sizes. If the ANOVA test assumptions were not met, the Kruskal-Wallis H test was employed as a non-parametric alternative. Dunn's (1964) procedure was performed as a post hoc test. The Paired-samples t-test was utilized for the third research question, as the data exhibited a normal distribution.

The significance level was determined as $p = .05$ for each analysis performed. Then, the Bonferroni correction was applied for multiple comparisons to minimize the possibility of Type 1 error that may arise in the case of multiple comparisons. While making the Bonferroni correction, the significance level ($p = .05$) was divided by the number of comparisons in multiple comparisons, and evaluations were made on the level of significance obtained (Huck, 2012). In addition, the omega square, eta square (for

ANOVA) and Cohen’s d (Paired-sample t-test) effect sizes were calculated. The partial eta-square and omega-square values represent a small effect in the range of .01 - .06, a medium effect in the range of .06-.14, and a high effect in the value range of .14 and above (Huck, 2012, p.27). Cohen's d effect size thresholds were established as d = 0.2 (small effect size), d = 0.5 (medium effect size), and d = 0.8 (large effect size) (Cohen, 1992).

CHAPTER 4

RESULTS

The normality assumption was checked before examining each research question using one of the parametric metric tests (e.g., ANOVA and t-test). Different methods were used to test the normality assumption. Firstly, the Shapiro–Wilk test (for $n < 50$) or Kolmogorov-Smirnov test (for $n \geq 50$) was checked. Skewness and kurtosis values are examined if the data did not meet the normality conditions according to Shapiro-Wilk or Kolmogorov-Smirnov values. Tabachnick, Fidell, and Ullman (2013) suggest that data is deemed normally distributed if skewness and kurtosis values fall between -1.5 and +1.5. Alongside analytical tests, graphical methods such as histograms, Q-Q plots, and box plots were visually inspected for validation. The subsequent sections detail findings corresponding to each research question and subject matter (lightning formation and Cell).

4.1 Research question 1

This section will show the results of the lightning formation and cell subject matters for the following question: How does the learning environment (AR, VR, and Traditional) affect university students' cognitive load, motivation, and learning outcomes (Retention and Transfer)?

4.1.1 Research question 1.1

This section will show the results of the lightning formation and cell subject matters for the following question: How does the learning environment affect university students' retention test scores?

4.1.1.1 Subject matter: Lightning formation

Upon examining a boxplot, no outliers were detected in the data. However, both normality ($p < .05$) and homogeneity of variance assumptions ($p < .05$) were not met. As a result, a Kruskal-Wallis H test was performed to investigate the impact of the learning environment on students' retention scores for the lightning formation topic. Table 5 shows the medians and numbers for each treatment group.

Table 5. Medians for Retention Scores of Each Learning Environment

Groups	N	Mean Rank
Virtual Reality	128	8.00
Augmented Reality	127	9.00
Traditional	128	7.17
Total	383	8.33

The retention score distributions appeared similar for all groups, as observed through a visual inspection of a boxplot (see Appendix I, Figure I1). A Kruskal-Wallis H test (χ^2 (2) = 4.773, p = .092) indicated no statistically significant differences in the medians of retention scores across the various learning environments. This suggests that the different learning environments did not significantly impact the students' abilities to retain the learned information.

4.1.1.2 Subject matter: Cell

Upon inspecting the boxplot, no outliers were found in the data. The assumptions of normality ($p < .05$) and homogeneity of variance were violated ($p < .05$). As a result, a Kruskal-Wallis H test was utilized to investigate the impact of the learning environment on the retention scores of university students for cell subject matter. Table 6 displays the medians and sample sizes for each treatment group.

Table 6. Medians for Retention Scores of Each Learning Environment

Groups	N	Mean Rank
Virtual Reality	128	4.33
Augmented Reality	127	6.67
Traditional	128	4.00
Total	383	5.00

Upon visual inspection using a boxplot, the distribution of retention scores seemed comparable for all groups. (see Appendix I, Figure I2). A Kruskal-Wallis H test (χ^2 (2) = 11.209, $p = .004$) demonstrated that the medians of retention scores were significantly different among the various learning environments. Additionally, pairwise comparisons were conducted using Dunn's (1964) procedure, incorporating a Bonferroni correction to account for multiple comparisons. The adjusted p-values are provided in Table 7. This post hoc analysis revealed that the Augmented Reality group (Mdn = 6.67) had significantly higher retention scores than the Traditional group (Mdn = 4.00) ($p = .003$). There is no statistically significant difference in any other group combination.

Table 7. Pairwise Comparison for Retention Scores of Students Using Different Learning Environments of the Cell Subject Matter

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Traditional – Virtual Reality	15.836	13.826	1.145	.252	.756
Traditional – Augmented Reality	45.686	13.853	3.298	.001	.003
Virtual Reality – Augmented Reality	-29.853	13.853	-2.155	.031	.094

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.1.2 Research question 1.2

This section will show the results of the lightning formation and cell subject matters for the following question: How does the learning environment affect university students' transfer test scores?

4.1.2.1 Subject matter: Lightning formation

Upon inspection of the boxplot, two outliers were observed in the data. These outliers were situated at the top in the virtual and augmented reality groups. The outliers were removed from the data to verify the other one-way ANOVA analysis assumptions, as Weisberg (2005) suggested. After this, Levene's test for equality of variances was conducted, which indicated homogeneity of variances ($p = .395$) (see Appendix J, Table J1). However, the data was not normally distributed for each learning environment based on the Kolmogorov-Smirnov values ($p < .05$), as shown in Appendix J, Table J2. The Skewness and Kurtosis values were also examined to validate the normality test. Skewness values ranged from .062 to .296, and Kurtosis values ranged from -.544 to -.176 (see Appendix J, Table J3). Based on the skewness & kurtosis and visual inspection, the data were normally distributed (see Appendix I, Figure I3). Table 8 shows descriptive statistics of each treatment group. The mean scores in the table reflect the transfer scores ranging from zero to eight.

Table 8. Descriptive Statistics for Students' Transfer Scores of Each Learning Environment

Groups	Mean	Std. Dev.	n
Virtual Reality	1.919	1.036	127
Augmented Reality	1.820	1.064	126
Traditional	2.021	1.182	128
Total	1.920	1.096	381

According to the one-way ANOVA test ($F(2, 378) = 1.065$, $p = .346$), there was no significant difference in the means of transfer scores among different learning environments.

4.1.2.2 Subject matter: Cell

Upon inspection of the boxplot, two outliers were observed in the data. These outliers were situated at the top in the augmented reality and traditional groups. The outliers were removed from the data to verify the other assumptions of the one-way ANOVA analysis (Weisberg, 2005) (see Appendix I, Figure I4). After this, Levene's test for equality of variances was conducted, which indicated homogeneity of variances ($p = .050$), as shown in Appendix J, Table J4. However, the data was not normally distributed for each learning environment based on the Kolmogorov-Smirnov values ($p < .05$), as shown in Appendix J, Table J5. Skewness values ranged from .106 to .364, and Kurtosis values ranged from -.588 to -.175 (see Appendix J, Figure J6). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 9 shows descriptive statistics of each treatment group. The one-way ANOVA test ($F(2, 378) =$

.921, p = .399) revealed that means of transfer scores were not significantly different between different learning environments.

Table 9. Descriptive Statistics for Students' Transfer Scores of Each Learning Environment

Groups	Mean	Std. Dev.	n
Virtual Reality	2.323	1.300	128
Augmented Reality	2.511	1.251	126
Traditional	2.299	1.505	127
Total	2.377	1.357	381

4.1.3 Research question 1.3

This section will show the results of the lightning formation and cell subject matters for the following question: How does the learning environment affect university students' cognitive load?

4.1.3.1 Subject matter: Lightning formation

4.1.3.1.1 Intrinsic Cognitive Load

Upon inspection of the boxplot, no outlier was observed in the data (Appendix I, Figure I5). Levene's test for equality of variances was conducted, which indicated homogeneity of variances (p = .050), as shown in Appendix J, Table J7. However, the data was found to be normally distributed only for the virtual reality group (p > .05), while it was not normally distributed for the augmented reality group (p = .005) and traditional group (p = .034), based on the Kolmogorov-Smirnov values (see Appendix J, Table J8). Skewness values ranged from .152 to .372, and Kurtosis values ranged from -.662 to -.436 (see Appendix J, Table J9). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 10 shows descriptive statistics of each treatment group. The mean scores in the table reflect the intrinsic cognitive scores ranging from zero to 10.

Table 10. Descriptive Statistics for Students' Intrinsic Cognitive Load Scores of Each Learning Environment

Groups	Mean	Std. Dev.	n
Virtual Reality	3.471	2.053	128
Augmented Reality	3.604	2.034	127
Traditional	3.359	2.414	128
Total	3.478	2.171	383

The one-way ANOVA test ($F(2, 380) = .403$, $p = .668$) revealed that means of intrinsic cognitive load scores of lightning formation subject matter were not significantly different between different learning environments.

4.1.3.1.2 Extraneous Cognitive Load

The assumptions of normality ($p < .05$) and homogeneity of variance ($p < .05$) were violated, so a Kruskal-Wallis H test was performed instead. Table 11 displays the medians and sample sizes for each treatment group.

Table 11. Medians for Extraneous Cognitive Load Scores of Each Learning Environment

Groups	N	Mean Rank
Virtual Reality	128	1.33
Augmented Reality	127	1.89
Traditional	128	1.22
Total	383	1.44

After visually inspecting the boxplot in Appendix I, Figure I6, it was observed that the distributions of extraneous cognitive load scores were similar for all groups. A

Kruskal-Wallis H test indicated that the medians of extraneous cognitive load scores were significantly different among different learning environments (χ^2 (2) = 7.451, p = .024). Adjusted p-values are presented in Table 12. The post hoc analysis revealed that the Augmented Reality group (Mdn = 1.89) had significantly higher extraneous cognitive load scores than the Traditional group (Mdn = 1.44) (p = .030). However, there was no statistically significant difference in any other group combination.

Table 12. Pairwise Comparison for Extraneous Cognitive Load Scores of Students Using Different Learning Environments of the Lightning Formation Subject Matter

Group 1 - Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Traditional – Virtual Reality	6.988	13.830	.505	.613	1.000
Traditional – Augmented Reality	35.707	13.857	2.577	.010	.030
Virtual Reality – Augmented Reality	-28.719	13.857	-2.073	.038	.115

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.1.3.1.3 Germane Cognitive Load

Upon inspection of the boxplot, no outlier was observed in the data (see Appendix I, Figure I7). Levene's test for equality of variances was conducted, which indicated homogeneity of variances (p = .386), as shown in Appendix J, Table J10. According to Kolmogorov-Smirnov values, the data was normally distributed for the virtual reality and augmented reality groups (p > .05) but not for the traditional group (p = .001) (see Appendix J, Table J11). Skewness values ranged from -.622 to -.205, and Kurtosis values ranged from -1.026 to -.346 (see Appendix J, Table J12). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 13 shows descriptive statistics of each treatment group.

Table 13. Descriptive Statistics for Students' Germane Cognitive Load Scores of Each Learning Environment

Groups	Mean	Std. Dev.	n
Virtual Reality	5.982	2.611	128
Augmented Reality	6.051	2.506	127
Traditional	6.565	2.665	128
Total	6.200	2.601	383

The one-way ANOVA test ($F(2, 380) = 1.921$, $p = .148$) revealed that means of germane cognitive load scores of lightning formation subject matter were not significantly different between different learning environments.

4.1.3.2 Subject matter: Cell

4.1.3.2.1 Intrinsic Cognitive Load

Upon inspection of the boxplot, no outlier was observed in the data (see Appendix I, Figure I8). Levene's test for equality of variances was conducted, which indicated homogeneity of variances ($p = .261$), as shown in Appendix J, Table J14. According to Kolmogorov-Smirnov values, the data was not normally distributed for all groups ($p < .05$) (see Appendix J, Table J15). Skewness values ranged from .052 to .176, and Kurtosis values ranged from -1.298 to -1.091 (see Appendix J, Table J16). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 14 shows descriptive statistics of each treatment group.

Table 14. Descriptive Statistics for Students' Intrinsic Cognitive Load Scores of Each Learning Environment

Groups	Mean	Std. Dev.	n
Virtual Reality	4.453	2.961	128
Augmented Reality	4.438	3.046	127
Traditional	4.844	3.290	128
Total	4.645	3.099	383

The one-way ANOVA test ($F(2, 380) = .508$, $p = .602$) revealed that means of intrinsic cognitive load scores of cell subject matter were not significantly different between different learning environments.

4.1.3.2.2 Extraneous Cognitive Load

The normality ($p < .05$) and homogeneity of variance assumptions ($p < .05$) were both violated. Therefore, a Kruskal-Wallis H test was conducted to examine the effect of the learning environment on university students' extraneous cognitive load scores for cell subject matter. Table 15 shows the medians and numbers for each treatment group.

Table 15. Medians for Extraneous Cognitive Load Scores of Each Learning Environment

Groups	N	Mean Rank
Virtual Reality	128	1.00
Augmented Reality	127	2.00
Traditional	126	2.00
Total	381	2.00

Upon examining the boxplot in Appendix I, Figure I9), the extraneous cognitive load scores distribution appeared similar across all groups. A Kruskal-Wallis H test was performed, revealing that the medians of extraneous cognitive load scores were statistically significantly different among the various learning environments ($\chi2$ (2) =

14.035, $p < .001$). Adjusted p-values can be found in Table 16. The post hoc analysis showed that the Augmented Reality group (Mdn = 2.00) had notably higher extraneous cognitive load scores compared to the Virtual Reality group (Mdn = 1.00) ($p < .001$). However, no statistically significant differences were observed in any other group comparisons.

Table 16. Pairwise Comparison for Extraneous Cognitive Load Scores of Students Using Different Learning Environments of the Cell Subject Matter

Group 1 - Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Virtual Reality – Traditional	-21.583	13.583	-1.614	.106	.319
Traditional – Augmented Reality	28.718	13.610	2.110	.035	.105
Virtual Reality – Augmented Reality	-50.645	13.556	-3.736	.000	.001

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.1.3.2.3 Germane Cognitive Load

Upon inspection of the boxplot, no outlier was observed in the data (see Appendix I, Figure I10). Levene's test for equality of variances was conducted, which indicated homogeneity of variances ($p = .058$), as shown in Appendix J, Table J17. According to Kolmogorov-Smirnov values, the data was not normally distributed for all groups ($p < .05$) (see Appendix J, Table J18). Skewness values ranged from -.575 to .014, and Kurtosis values ranged from -1.020 to -.351 (see Appendix J, Table J19). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 17 shows descriptive statistics of each treatment group.

Table 17. Descriptive Statistics for Students' Germane Cognitive Load Scores of Each Learning Environment

Groups	Mean	Std. Dev.	n
Virtual Reality	6.469	2.488	128
Augmented Reality	6.323	2.535	127
Traditional	5.297	2.885	128
Total	6.029	2.686	383

The one-way ANOVA test ($F(2, 380) = 7.473$, $p < .001$) revealed that means of germane cognitive load scores of cell subject matter were statistically significantly different between different learning environments. The effect size, determined using omega squared, was 0.005, signifying less than a small effect, while partial eta squared suggested a small effect (partial $\eta 2 = .038$). In essence, effect size estimates demonstrate that students' germane cognitive load scores in various learning environments exhibited a small difference. The Bonferroni post hoc analysis revealed that the Virtual Reality group ($M = 6.47$, $SD = 2.49$) ($p = .001$) and Augmented Reality group ($M = 6.32$, $SD = 2.54$) ($p = .006$) had statistically significantly higher mean germane cognitive load scores than the Traditional group ($M = 5.30$, $SD = 2.89$) based on after the Bonferroni correction ($p < .0166$) (see Table 18).

Table 18. Multiple Comparisons of Germane Cognitive Load Scores of Students Using Different Learning Environments

(I) Version	(J) Version	(I-J) Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Virtual Reality	Augmented Reality	-.146	.331	1.000	-.650	.942
	Traditional	1.172	.330	.001	.378	1.966
Traditional	Augmented Reality	-1.026	.331	.006	-1.822	-.230

*. The mean difference is significant at the .0166 level (Bonferroni Adjustment).

4.1.4 Research question 1.4

This section will show the results of the lightning formation and cell subject matters for the following question: How does the learning environment affect university students' motivation?

4.1.4.1 Subject matter: Lightning formation

4.1.4.1.1 Overall motivation

Upon inspection of the boxplot, no outlier was observed in the data. However, the normality ($p < .05$) and homogeneity of variance assumptions ($p < .05$) were both violated. As a result, a Kruskal-Wallis H test was conducted to examine the effect of the variables on the outcome. Table 19 shows the medians and numbers for each treatment group.

Table 19. Medians for Overall Motivation Scores of Each Learning Environment

Groups	N	Mean Rank
Virtual Reality	128	4.00
Augmented Reality	127	3.83
Traditional	128	4.00
Total	383	3.92

After visually inspecting the boxplot in Appendix I, Figure I11, it was observed that all groups' overall motivation scores were similar. A Kruskal-Wallis H test indicated that the medians of overall motivation scores were not significantly different among different learning environments (χ^2 (2) = .531, p = .767).

4.1.4.1.2 Attention

Upon inspection of the boxplot, no outlier was observed in the data (see Appendix I, Figure I12). Levene's test for equality of variances was conducted, which indicated homogeneity of variances (p = .536), as shown in Appendix J, Table J20. According to Kolmogorov-Smirnov values, the data was not normally distributed for all groups (see Appendix J, Table J21). Skewness values ranged from -.662 to -.381, and Kurtosis values ranged from -.610 to -.311 (see Appendix J, Table J22). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 20 shows descriptive statistics of each treatment group. The mean scores in the table reflect the attention scores ranging from zero to five.

Table 20. Descriptive Statistics for Students' Attention Scores in Each Learning Environment

Groups	Mean	Std. Dev.	n
Virtual Reality	3.630	1.027	128
Augmented Reality	3.643	.928	127
Traditional	3.758	1.056	128
Total	3.677	1.005	383

The one-way ANOVA test (F (2, 380) = .624, p = .536) revealed that means of attention scores of lightning formation subject matter were not statistically significantly different between different learning environments.

4.1.4.1.3 Relevance

Upon inspection of the boxplot, no outlier was observed in the data (see Appendix I, Figure I13). Levene's test for equality of variances was conducted, which indicated homogeneity of variances (p = .801), as shown in Appendix J, Table J23. According to

Kolmogorov-Smirnov values, the data was not normally distributed for all groups (p < .05) (see Appendix J, Table J24). Skewness values ranged from -.594 to -.338, and Kurtosis values ranged from -.795 to -.224 (see Appendix J, Table J25). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 21 shows descriptive statistics of each treatment group. The mean scores in the table reflect the relevance scores ranging from zero to five.

Table 21. Descriptive Statistics for Students' Relevance Scores of Each Learning Environment

Groups	Mean	Std. Dev.	n
Virtual Reality	3.698	.937	128
Augmented Reality	3.642	.986	127
Traditional	3.768	.939	128
Total	3.703	.953	383

The one-way ANOVA test (F (2, 380) = .512, p = .570) revealed that means of relevance scores of lightning formation subject matter were not statistically significantly different between different learning environments.

4.1.4.1.4 Confidence

Upon inspection of the boxplot, no outlier was observed in the data (see Appendix I, Figure I14). Levene's test for equality of variances was conducted, which indicated homogeneity of variances (p = .152), as shown in Appendix J, Table J26. According to Kolmogorov-Smirnov values, the data was not normally distributed for all groups (p < .05) (see Appendix J, Table J27). Skewness values ranged from -.985 to -.672, and Kurtosis values ranged from -.389 to .397 (see Appendix J, Table J28). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 22 shows descriptive statistics of each treatment group. The mean scores in the table reflect the confidence scores ranging from zero to five.

Table 22. Descriptive Statistics for Students' Confidence Scores in Each Learning Environment

Groups	Mean	Std. Dev.	n
Virtual Reality	3.888	.894	128
Augmented Reality	3.887	.885	127
Traditional	3.909	1.050	128
Total	3.895	.944	383

The one-way ANOVA test ($F(2, 380) = .022$, $p = .979$) revealed that means of confidence scores of lightning formation subject matter were not statistically significantly different between different learning environments.

4.1.4.1.5 Satisfaction

Upon inspection of the boxplot, no outlier was observed in the data. However, the normality ($p < .05$) and homogeneity of variance assumptions ($p < .05$) were both violated. As a result, a Kruskal-Wallis H test was conducted to examine the effect of the variables on the outcome. Table 23 displays the medians and sample sizes for each treatment group.

Table 23. Medians for Satisfaction Scores of Each Learning Environment

Groups	N	Mean Rank
Virtual Reality	128	4.33
Augmented Reality	127	4.00
Traditional	128	4.00
Total	383	4.00

Upon examining the boxplot in Appendix I, Figure I15, the distribution of satisfaction scores appeared to be similar across all groups. A Kruskal-Wallis H test was performed, revealing that the medians of satisfaction scores were statistically significantly different among the various learning environments (χ^2 (2) = 6.259, p = .044). Adjusted p-values can be found in Table 24. The post hoc analysis showed that the Virtual Reality group (Mdn = 4.33) had notably higher satisfaction scores compared to the Traditional group (Mdn = 4.00) (p = .041). Nevertheless, no significant differences were observed in any other group comparisons.

Table 24. Pairwise Comparison for Satisfaction Scores of Students Using Different Learning Environments of the Lightning Formation Subject Matter

Group 1-Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Traditional – Virtual Reality	33.637	13.625	2.469	.014	.041
Traditional – Augmented Reality	12.018	13.652	.880	.379	1.000
Augmented Reality – Virtual Reality	21.619	13.652	1.584	.113	.340

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.1.4.2 Subject matter: Cell

4.1.4.2.1 Overall motivation

Upon inspection of the boxplot, no outlier was observed in the data. However, the normality (p < .05) and homogeneity of variance assumptions (p < .05) were both violated. Therefore, a Kruskal-Wallis H test was conducted. Table 25 shows the medians and numbers for each treatment group.

Table 25. Medians for Overall Motivation Scores of Each Learning Environment

Groups	N	Mean Rank
Virtual Reality	128	4.25
Augmented Reality	127	3.83
Traditional	128	3.46
Total	383	3.92

After visually inspecting the boxplot in Appendix I, Figure I16, it was observed that all groups' overall motivation scores were similar. A Kruskal-Wallis H test was performed, revealing that the medians of overall motivation scores were statistically significantly different across various learning environments (χ^2 (2) = 18.044, p < .001). Adjusted p-values can be found in Table 26. The post hoc analysis showed that the Virtual Reality group (Mdn = 4.25) had notably higher overall motivation scores than the Traditional group (Mdn = 3.46) (p < .001). However, no statistically significant differences were observed in any other group comparisons.

Table 26. Pairwise Comparison for Overall Motivation Scores of Students Using Different Learning Environments of the Cell Subject Matter

Group 1-Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Traditional – Augmented Reality	27.318	13.857	1.971	.049	.146
Traditional – Virtual Reality	58.699	13.830	4.244	.000	.000
Augmented Reality – Virtual Reality	31.382	13.857	2.265	.024	.071

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.1.4.2.2 Attention

Upon examining the boxplot, there were no outliers in the data. The normality (p < .05) and the homogeneity of variance assumptions were violated (p < .05). Therefore, a Kruskal-Wallis H test was conducted. Table 27 shows the medians and numbers for each treatment group.

Table 27. Medians for Attention Scores of Each Learning Environment

Groups	N	Mean Rank
Virtual Reality	128	4.00
Augmented Reality	127	3.67
Traditional	128	3.33
Total	383	4.00

After visually inspecting the boxplot in Appendix I, Figure I17, it was observed that all groups' attention scores were similar. A Kruskal-Wallis H test indicated that the medians of attention scores differed significantly among different learning environments (χ^2 (2) = 12.288, p = .002). Adjusted p-values are presented in Table 28. The post hoc analysis revealed that the Virtual Reality group (Mdn = 4.00) had significantly higher attention scores than the Traditional group (Mdn = 3.33) (p = .002). Nonetheless, no significant differences were observed in any other group comparisons.

Table 28. Pairwise Comparison for Attention Scores of Students Using Different Learning Environments of the Cell Subject Matter

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Traditional – Augmented Reality	14.928	13.767	1.084	.278	.835
Traditional – Virtual Reality	47.125	13.740	3.430	.001	.002
Augmented Reality – Virtual Reality	32.197	13.767	2.339	.019	.058

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.1.4.2.3 Relevance

Upon examining the boxplot, there were no outliers in the data. The normality ($p < .05$) and the homogeneity of variance assumptions were violated ($p < .05$). Therefore, a Kruskal-Wallis H test was conducted. Table 29 shows the medians and numbers for each treatment group.

Table 29. Medians for Relevance Scores of Each Learning Environment

Groups	N	Mean Rank
Virtual Reality	128	4.00
Augmented Reality	127	4.00
Traditional	128	3.33
Total	383	4.00

Upon examining the boxplot in Appendix I, Figure I18, it can be observed that the relevance scores for all groups appeared to be similar. A Kruskal-Wallis H test (χ^2 (2) = 8.803, p = .012) revealed that the medians of relevance scores were statistically significantly different between different learning environments. Adjusted p-values are presented (see Table 30). This post hoc analysis revealed that the Virtual Reality group (Mdn = 4.00) had significantly higher attention scores than the Traditional group (Mdn = 3.33) (p = .010). There is no significant difference in any other group combination.

Table 30. Pairwise Comparison with a Bonferroni Adjustment for Relevance Scores of Students Using Different Learning Environments of the Cell Subject Matter

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Traditional – Augmented Reality	24.683	13.766	1.793	.073	.219
Traditional – Virtual Reality	40.438	13.739	2.943	.003	.010
Augmented Reality – Virtual Reality	15.754	13.766	1.144	.252	.757

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.1.4.2.4 Confidence

The normality (p < .05) and the homogeneity of variance assumptions were violated (p < .05). Therefore, a Kruskal-Wallis H test was conducted. Table 31 shows the medians and numbers for each treatment group.

Table 31. Medians for Confidence Scores of Each Learning Environment

Groups	N	Mean Rank
Virtual Reality	128	4.00
Augmented Reality	127	4.00
Traditional	128	3.67
Total	383	4.00

After visually inspecting the boxplot in Appendix I, Figure I19, it was observed that all groups' confidence scores were similar. A Kruskal-Wallis H test was performed, revealing that the medians of confidence scores were statistically significantly different across various learning environments (χ^2 (2) = 12.345, p = .002). Adjusted p-values can be found in Table 32. The post hoc analysis showed that the Virtual Reality group (Mdn = 4.00) had notably higher confidence scores compared to the Traditional group (Mdn = 3.67) (p = .001). However, no statistically significant differences were observed in any other group comparisons.

Table 32. Pairwise Comparison for Confidence Scores of Students Using Different Learning Environments of the Cell Subject Matter

Group 1 - Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Traditional – Augmented Reality	28.139	13.704	2.053	.040	.120
Traditional – Virtual Reality	47.809	13.677	3.495	.000	.001
Augmented Reality – Virtual Reality	19.669	13.704	1.435	.151	.454

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.1.4.2.5 Satisfaction

Upon examining the boxplot, there were no outliers in the data. The normality ($p < .05$) and the homogeneity of variance assumptions were violated ($p < .05$). Therefore, a Kruskal-Wallis H test was conducted. Table 33 shows the medians and numbers for each treatment group.

Table 33. Medians for Satisfaction Scores of Each Learning Environment

Groups	N	Mean Rank
Virtual Reality	128	4.67
Augmented Reality	127	4.00
Traditional	128	3.33
Total	383	4.00

Upon examining the boxplot in Appendix I, Figure 20, it can be observed that the distributions of satisfaction scores appeared to be similar across all groups. A Kruskal-Wallis H test (χ^2 (2) = 31.373, $p < .001$) revealed that the medians of satisfaction scores were statistically significantly different between different learning environments. Adjusted p-values are presented (see Table 34). This post hoc analysis revealed that the Virtual Reality group (Mdn = 4.67) had significantly higher satisfaction scores than the Traditional group (Mdn = 3.33) ($p < .001$) and the Augmented Reality group (Mdn = 4.00) ($p = .005$). Additionally, the Augmented Reality group (Mdn = 4.00) had significantly higher satisfaction scores than the Traditional group (Mdn = 3.33) ($p < .047$).

Table 34. Pairwise Comparison for Satisfaction Scores of Students Using Different Learning Environments of the Lightning Formation Subject Matter

Group 1 - Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Traditional – Augmented Reality	33.071	13.693	2.415	.016	.047
Traditional – Virtual Reality	76.320	13.666	5.585	.000	.000
Augmented Reality – Virtual Reality	43.249	13.693	3.159	.002	.005

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.1.5 Summary of results of research question 1

The study investigated how the learning environment (AR, VR, and Traditional) affects university students' cognitive load, motivation, and learning outcomes. For retention scores, there was no significant difference between learning environments for lightning formation, but AR had higher scores for cell structure than the Traditional group. There was no significant difference in transfer scores for either subject matter.

Regarding cognitive load, the extraneous cognitive load scores for lightning formation were notably higher in the traditional group than in the AR group. For cell structure subject matter, extraneous cognitive load scores were significantly higher in the VR group than in other groups. Additionally, the VR and AR groups had significantly higher germane cognitive load scores than the traditional group.

There were no significant differences in overall motivation scores between different learning environments for the lightning formation subject. However, the VR group had significantly higher overall motivation scores for cell structure subject matter than the traditional group. The VR group also had significantly greater attention, relevance, confidence, and satisfaction scores than the traditional group for cell structure subject matter. The AR group had significantly higher satisfaction scores for lightning formation subject matter than the traditional group. Detailed information is provided in Table 35.

Table 35. Summary Results for Research Question 1

Variables			Lightning Formation		Cell Structure	
Ind. Var.	Dep. Var.	Sub	p	Post hoc	p	Post hoc
Learning Environment	Retention Score	NA	.092	No difference	.004	AR > Traditional (p = .003)
Learning Environment	Transfer Score	NA	.346	No difference	.399	No difference
Learning Environment	Cognitive Load	Intrinsic CL	.668	No difference	.602	No difference
		Extraneous CL	.024	AR > Traditional (p = .030)	< .001	AR > VR (p < .001)
		Germane CL	.148	No difference	< .001	AR > Traditional (p = .001) VR > Traditional (p = .006)
Learning Environment	Motivation	Overall	.767	No difference	< .001	VR > Traditional (p < .001)
		Attention	.536	No difference	.002	VR > Traditional (p = .002)
		Relevance	.570	No difference	.012	VR > Traditional (p = .010)
		Confidence	.979	No difference	.002	VR > Traditional (p = .001)
		Satisfaction	.041	VR > Traditional (p = .041)	< .001	VR > AR (p < .001) VR > Traditional (p = .005) AR > Traditional (p = .047)

4.2 Research question 2

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence cognitive load and learning outcomes for university students across different learning environments?

4.2.1 Research question 2.1

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence retention test scores for university students across different learning environments?

4.2.1.1 Subject matter: Lightning formation

Upon examining the boxplot, no outliers were found in the data. However, the normality assumption ($p < .05$) and the homogeneity of variance assumption ($p < .05$) were both violated. Hence, a Kruskal-Wallis H test was performed to investigate the impact of segmenting, pre-training, and modality on undergraduate students' retention scores. Table 36 shows the medians and numbers for each treatment group.

Table 36. Medians for Retention Scores of Each Version

Groups	N	Mean Rank
N	96	6.33
NS	95	11.67
T	96	6.33
NP	96	10.50
Total	383	8.33

Upon examining the boxplot in Appendix I, Figure I21, it can be observed that the distributions of retention scores appeared to be similar across all groups. A Kruskal-Wallis H test (χ^2 (3) = 33.993, $p < .001$) was performed, revealing that the medians of retention scores were significantly different among the various version groups. Adjusted p-values can be found in Table 37. The post hoc analysis showed that the NP group (Mdn = 10.50) had notably higher retention scores than the N group (Mdn = 6.33) ($p < .001$) and the T group (Mdn = 6.33) ($p < .001$). Furthermore, the NS group (Mdn = 11.67) had significantly higher retention scores than the N group (Mdn = 6.33) ($p < .001$) and the T group (Mdn = 6.33) ($p < .001$). No statistically significant differences were observed in any other group comparisons.

Table 37. Pairwise Comparison for Retention Scores of Students Using Different Versions of the Lightning Formation Subject Matter

Group 1-Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
N-T	-5.052	15.975	-.316	.752	1.000
N-NP	-65.979	15.975	-4.130	.000	.000
N-NS	-70.628	16.017	-4.410	.000	.000
T-NP	-60.927	15.975	-3.814	.000	.001
T-NS	65.576	16.017	4.094	.000	.000
NP-NS	4.649	16.017	.290	.772	1.000

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.2.1.2 Subject matter: Cell

Six data points were found to be outliers upon examining the boxplot. The outlier in the N group and T group were at the top. The outliers were removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005). However, the normality ($p < .05$) and the homogeneity of variance assumptions were violated ($p < .05$). Therefore, a Kruskal-Wallis H test was conducted to examine the segmenting, pre-training, and modality effect on undergraduate students' retention scores. Table 38 shows the medians and numbers for each treatment group.

Table 38. Medians for Retention Scores of Each Version

Groups	N	Mean Rank
N	95	4.00
NS	95	6.00
T	96	3.00
NP	96	8.17
Total	382	5.00

Upon visually inspecting the boxplot in Appendix I, Figure I22, it was found that the distributions of retention scores were similar for all groups. A Kruskal-Wallis H test (χ^2 (3) = 30.041, $p < .0001$) showed that the medians of retention scores were significantly different among the version groups. Adjusted p-values can be found in

Table 39. The post hoc analysis revealed that the NP group (Mdn = 8.17) had significantly higher retention scores than the N group (Mdn = 4.00) ($p < .001$) and the T group (Mdn = 3.00) ($p < .001$). Furthermore, the NS group (Mdn = 6.00) had significantly higher retention scores than the N group (Mdn = 4.00) ($p = .049$) and the T group (Mdn = 3.00) ($p = .004$). No other group combinations showed statistically significant differences.

Table 39. Pairwise Comparison for Retention Scores of Students Using Different Versions of the Cell Subject Matter

Group 1-Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
T-N	12.130	15.965	.760	.447	1.000
T-NS	54.388	15.965	3.407	.001	.004
T-NP	-76.182	15.924	-4.784	.000	.000
N-NS	-42.258	16.007	-2.640	.008	.049
N-NP	-64.053	15.965	-4.012	.000	.000
NS-NP	-21.795	15.965	-1.365	.172	1.000

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.2.2 Research question 2.1.1

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence retention test scores for university students in virtual learning environments?

4.2.2.1 Subject matter: Lightning formation

Upon inspecting the boxplot, it was found that there were two outliers in the data, with one outlier each in the T and NP groups at the upper end of the distribution. Furthermore, the normality assumption ($p < .05$) and the homogeneity of variance assumption ($p < .05$) were both violated. As a result, a Kruskal-Wallis H test was conducted to investigate the impact of segmenting, pre-training, and modality on undergraduate students' retention scores in the virtual learning environment. Table 40 shows the medians and numbers for each treatment group.

Table 40. Medians for Retention Scores of Each Version in the Virtual Learning Environment

Groups	N	Mean Rank
N	32	6.67
NS	32	9.50
T	32	6.33
NP	32	11.50
Total	128	13.00

Upon visually inspecting the boxplot in Appendix I, Figure I23, it was found that the distributions of retention scores in the virtual learning environment were similar for all groups. A Kruskal-Wallis H test (χ^2 (3) = 13.020, $p < .005$) indicated that the medians of retention scores in the virtual reality learning environment were significantly different among the version groups. Adjusted p-values can be found in Table 41. The post hoc analysis revealed that the NP group (Mdn = 11.50) had significantly higher retention scores in the virtual reality learning environment than the T group (Mdn = 6.33) (p = .022) and the N group (Mdn = 6.67) (p = .028). No other group combinations showed significant differences.

Table 41. Pairwise Comparison for Retention Scores of Students Using Different Versions of the Material in a Virtual Learning Environment

Group 1 - Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
T-N	.719	9.269	.078	.938	1.000
T-NS	20.141	9.269	2.173	.030	.179
T-NP	-26.891	9.269	-2.901	.004	.022
N-NS	-19.422	9.269	-2.095	.036	.217
N-NP	-26.172	9.269	-2.824	.005	.028
NS-NP	-6.750	9.269	-.728	.466	1.000

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.2.2.2 Subject matter: Cell

Upon inspecting the boxplot, it was found that one outlier in the data was located at the upper end of the distribution in the N group. The outlier was removed from the dataset to test the other assumptions of the one-way ANOVA analysis (Weisberg, 2005). However, it was found that the normality assumption ($p < .05$) and the homogeneity of variance assumption ($p < .05$) were both violated. Consequently, a Kruskal-Wallis H test was conducted to investigate the impact of segmenting, pre-training, and modality on undergraduate students' retention scores in the virtual learning environment. The medians and sample sizes for each treatment group are presented in Table 42.

Table 42. Medians for Retention Scores of Each Version in the Virtual Learning Environment

Groups	N	Mean Rank
N	32	4.17
NS	32	5.00
T	32	3.50
NP	32	7.00
Total	128	4.33

Distributions of retention scores in the virtual learning environment were similar for all groups, as assessed by visual inspection of a boxplot (see Appendix I, Figure I24). A Kruskal-Wallis H test (χ^2 (3) = 9.397, $p = .024$) revealed that the medians of retention scores in the virtual reality learning environment were statistically significantly different between the different version groups. Adjusted p-values are presented (see Table 43). This post hoc analysis revealed that the NP group (Mdn = 7.00) had significantly higher retention scores in the virtual reality learning environment than the T group (Mdn =

3.50) ($p = .027$). There is no statistically significant difference in any other group combination.

4.2.3 Research question 2.1.2

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence retention test scores for university students in augmented learning environments?

4.2.3.1 Subject matter: Lightning formation

Upon inspecting the boxplot, it was found that there were no outliers in the data. However, the normality assumption ($p < .05$) and the homogeneity of variance assumption ($p < .05$) were both violated. Consequently, a Kruskal-Wallis H test was conducted to investigate the impact of segmenting, pre-training, and modality on undergraduate students' retention scores in the augmented learning environment. Table 43 shows the medians and numbers for each treatment group.

Table 43. Medians for Retention Scores of Each Version in the Augmented Learning Environment

Groups	N	Mean Rank
N	32	6.00
NS	31	13.00
T	32	8.17
NP	32	14.00
Total	127	9.00

All groups had similar retention score distributions in the augmented learning environment, as confirmed by examining a boxplot in Appendix I, Figure I25. However, a Kruskal-Wallis H test (χ^2 (3) = 18.752, p < .001) indicated that the medians of retention scores in the augmented learning environment were significantly different between the various version groups. Table 44 presents adjusted p-values. A posthoc analysis revealed that the NP group (Mdn = 14.00) (p < .001) and NS group (Mdn = 13.00) (p = .003) had significantly higher retention scores in the augmented reality learning environment than the N group (Mdn = 6.00). No other group combinations had a statistically significant difference.

Table 44. Pairwise Comparison for Retention Scores of Students Using Different Versions of the Material in Augmented Learning Environment

Group 1-Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
T-NS	19.959	9.272	2.153	.031	.188
T-NP	-21.656	9.198	-2.355	.019	.111
N-NS	-32.115	9.272	-3.464	.001	.003
N-NP	-33.812	9.198	-3.676	.000	.001
NS-NP	-1.697	9.272	-.183	.855	1.000
N-T	-12.156	9.198	-1.322	.186	1.000

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.2.3.2 Subject matter: Cell

Upon examining a boxplot, it was observed that there were two data points. The outliers in the N group were at the top. The outliers were removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005). The normality (p < .05) and the homogeneity of variance assumptions were violated (p < .05). Therefore, a Kruskal-Wallis H test was conducted to examine the segmenting, pre-training, and modality effect on undergraduate students' retention scores in the augmented learning environment. Table 45 shows the medians and numbers for each treatment group.

Table 45. Medians for Retention Scores of Each Version in the Augmented Learning Environment

Groups	N	Mean Rank
N	32	4.17
NS	31	8.67
T	32	5.67
NP	32	11.67
Total	127	6.67

After examining a boxplot (see Appendix I, Figure I26), it was determined that the distribution of retention scores in the augmented learning environment was similar for all groups. A Kruskal-Wallis H test (χ^2 (3) = 13.644, p = .003) revealed that the medians of retention scores in the augmented learning environment significantly differed between the different version groups. Adjusted p-values are presented (see Table 46). This post hoc analysis revealed that the NP group (Mdn = 11.67) had significantly higher retention scores in the augmented reality learning environment than the N group (Mdn = 4.17) (p = .011) and the T group (Mdn = 5.67) (p = .023). There is no significant difference in any other group combination.

Table 46. Pairwise Comparison for Retention Scores of Students Using Different Versions of the Material in the Augmented Learning Environment

Group 1 - Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
T-NS	17.549	9.267	1.894	.058	.350
T-NP	-26.625	9.193	-2.896	.004	.023
N-NS	-19.628	9.267	-2.118	.034	.205
N-NP	-28.703	9.193	-3.122	.002	.011
NS-NP	-9.076	9.267	-.979	.327	1.000
N-T	-2.078	9.193	-.226	.821	1.000

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.2.4 Research question 2.1.3

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence retention test scores for university students in traditional learning environments?

4.2.4.1 Subject matter: Lightning formation

Upon inspecting a boxplot, it was observed that one data point was considered an outlier. The outlier in the NP group was at the top. The outlier was removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005). Variances were homogeneous, as assessed by Levene's test for equality of variances ($p = .296$) (see Appendix J, Table J29). According to Shapiro-Wilk values, the data was normally distributed for NP, NS, and N groups ($p > .05$) but not for the T group ($p = .003$) (see Appendix J, Table J30). Skewness values ranged from .338 to .971, and Kurtosis values ranged from -1.028 to .007 (see Appendix J, Table J31). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 47 shows descriptive statistics of each treatment group. The mean scores in the table reflect the retention scores ranging from zero to 50.

Table 47. Descriptive Statistics for Students' Retention Scores of Each Version in the Traditional Learning Environment

Groups	Mean	Std. Dev.	n
N	7.917	5.858	32
NS	11.802	7.482	32
T	7.178	5.586	32
NP	8.731	5.735	31
Total	8.908	6.393	127

The one-way ANOVA test (F (3, 123) = 3.419, p = .020) indicated that the means of retention scores varied significantly between the different version groups. The Bonferroni correction was applied to account for multiple comparisons, and a p-value less than .0083 was considered statistically significant. The effect size was calculated using omega squared and was found to be 0.018, indicating a small effect. On the other hand, partial eta squared suggested a moderate effect (partial $\eta 2$ = .077). In other words, effect size estimates indicate that different versions of the educational material designed for the traditional learning environment moderately affected student retention scores. As in Table 48, the Bonferroni post hoc analysis revealed that the NS group (M = 11.80, SD = 7.48) had statistically significantly higher mean retention scores than the T group (M = 7.18, SD = 5.59) (p = .021) based on without Bonferroni correction. However, after the Bonferroni correction, the group means were not significantly different (p > .0083).

Table 48. Multiple Comparisons of Retention Scores of Students Using Different Versions of the Material

(I) Version	(J) Version	(I-J) Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
N	NS	-3.885	1.554	.082	-8.053	.282
	T	.740	1.554	1.000	-3.428	4.907
	NP	-.815	1.567	1.000	-5.015	3.386
NS	T	4.625*	1.554	.021	.458	8.792
	NP	3.071	1.567	.313	-1.130	7.272
T	NP	-1.554	1.567	1.000	-5.755	2.647

*. The mean difference is significant at the .0083 level (Bonferroni Adjustment).

4.2.4.2 Subject matter: Cell

Upon examining a boxplot, it was observed that there were four data points considered outliers. The outliers in the T, N, and NS groups were at the top. The outliers were removed from the data to check the other assumptions of the one-way ANOVA analysis

(Weisberg, 2005). The normality ($p < .05$) and the homogeneity of variance assumptions were violated ($p < .05$). Therefore, a Kruskal-Wallis H test was conducted to examine the segmenting, pre-training, and modality effect on undergraduate students' retention scores in the traditional learning environment. Table 49 shows the medians and numbers for each treatment group.

Table 49. Medians for Retention Scores of Each Version in the Traditional Learning Environment

Groups	N	Mean Rank
N	32	3.83
NS	32	4.83
T	32	2.33
NP	32	6.83
Total	128	4.00

After examining a boxplot (see Appendix I, Figure I27), it was determined that the distribution of retention scores in the traditional learning environment was similar for all groups. A Kruskal-Wallis H test (χ^2 (3) = 9.042, p = .029) revealed that the medians of retention scores in the traditional learning environment significantly differed between the different version groups. Adjusted p-values are presented (see Table 50). This post hoc analysis revealed that the NP group (Mdn = 6.83) had significantly higher retention scores in the traditional learning environment than the T group (Mdn = 2.33) (p = .034). There is no significant difference in any other group combination.

Table 50. Pairwise Comparison Using for Retention Scores of Students Using Different Versions of the Material in the Traditional Learning Environment

Group 1-Group 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
T-N	7.563	9.261	.817	.414	1.000
T-NS	18.531	9.261	2.001	.045	.272
T-NP	-25.594	9.261	-2.764	.006	.034
N-NS	-10.969	9.261	-1.184	.236	1.000
N-NP	-18.031	9.261	-1.947	.052	.309
NS-NP	-7.062	9.261	-.763	.446	1.000

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

4.2.5 Summary of results of research question 2.1

The study examined the effect of segmenting, pre-training, and modality principles on university students' retention scores in different learning environments. Regardless of the learning environment, the NP and NS groups had significantly higher retention scores for the lightning formation and cell structure subject matters than the T and N groups. In the virtual reality learning environment, the NP group had significantly higher retention scores for the lightning formation and cell structure subjects than the T and N groups. No other significant differences were found among other group combinations. In the augmented learning environment, the NP group had significantly higher retention scores for the lightning formation and cell structure subjects than the N and T groups. The NS group also had significantly higher lightning-formation retention scores than the N group. No other significant differences were observed among other group combinations. In the traditional learning environment, the NS group had significantly higher mean retention scores for lightning formation than the T group. However, this difference was not significant after the Bonferroni correction. The NP group had significantly higher retention scores for cell structure than the T group, with no other significant differences found among other group combinations. In conclusion, the segmenting, pre-training, and modality effects are generally held across various learning environments. The NP group often performed better than the other groups regarding retention test scores. Detailed information is provided in Table 51.

Table 51. Summary Results for Research Question 2.1

Subject	IV	DV	p	Post hoc	Effect of Principles
Lightning Formation	Versions (MML) (Environment: All)	Retention Score	< .0001	NP > N (p < .001) NP > T (p < .001) NS > N (p < .001) NS > T (p < .001)	Modality: Not observed Pre-training: Observed Segmenting: Observed
	Versions (MML) (Environment: VR)	Retention Score	< .005	NP > N (p = .028) NP > T (p = .022)	Modality: Not observed Pre-training: Observed Segmenting: Not observed
	Versions (MML) (Environment: AR)	Retention Score	< .001	NP > N (p < .001) NS > N (p = .003)	Modality: Not observed Pre-training: Observed Segmenting: Observed
	Versions (MML) (Environment: Traditional)	Retention Score	.020	NS > T (p = .021)	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
Cell	Versions (MML) (Environment: All)	Retention Score	< .001	NP > N (p < .001) NP > T (p < .001) NS > N (p = .049) NS > T (p = .004)	Modality: Not observed Pre-training: Observed Segmenting: Observed
	Versions (MML) (Environment: VR)	Retention Score	.024	NP > T (p = .027)	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
	Versions (MML) (Environment: AR)	Retention Score	.003	NP > N (p = .011) NP > T (p = .023)	Modality: Not observed Pre-training: Observed Segmenting: Not observed
	Versions (MML) (Environment: Traditional)	Retention Score	.029	NP > T (p = .034)	Modality: Not observed Pre-training: Not observed Segmenting: Not observed

4.2.6 Research question 2.2

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence transfer test scores for university students across different learning environments?

4.2.6.1 Subject matter: Lightning formation

Upon visual inspection of a boxplot, it was determined that the data contained four outliers (see Appendix I, Figure I28). The outliers in the T, NP, and NS groups were at the top. The outliers were removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005). Then, the homogeneity of variances in the data sets ($p > .05$) was observed (see Appendix J, Table J32). According to Kolmogorov-Smirnov values, the data was normally distributed for NP and NS groups ($p > .05$) but not for the T group ($p = .009$) and N group ($p = .005$) (see Appendix J, Table J33). Skewness values ranged from -.057 to .253, and Kurtosis values ranged from -.805 to -.493 (see Appendix J, Table J34). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 52 shows descriptive statistics of each treatment group.

Table 52. Descriptive Statistics for Students' Transfer Scores of Each Version

Groups	Mean	Std. Dev.	n
N	1.944	1.056	96
NS	1.996	1.063	93
T	1.656	1.099	95
NP	2.014	1.041	95
Total	1.902	1.070	379

The one-way ANOVA test ($F(3, 375) = 2.334$, $p = .074$) revealed that the means of transfer scores were not statistically different among the different version groups.

4.2.6.2 Subject matter: Cell

After examining a boxplot, it was observed that there was one data point considered an outlier (see Appendix I, Figure I29). The outlier in the NS group was at the top. The outlier was removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005). Then, the homogeneity of variances in the data sets ($p > .05$) was observed (see Appendix J, Table J35). According to Kolmogorov-Smirnov values, the data was not normally distributed for all groups ($p > .05$) (see Appendix J, Table J36). Skewness values ranged from .130 to .274, and Kurtosis values ranged from -.639 to -.017 (see Appendix J, Table J37). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 53 shows descriptive statistics of each treatment group.

Table 53. Descriptive Statistics for Students' Transfer Scores of Each Version

Groups	Mean	Std. Dev.	n
N	2.177	1.326	96
NS	2.206	1.366	94
T	2.267	1.300	96
NP	2.889	1.361	96
Total	2.389	1.365	382

The one-way ANOVA test ($F(3, 378) = 10.959$, $p < .001$) indicated that the means of transfer scores differed significantly between the different version groups. The Bonferroni correction was applied to account for multiple comparisons, and a p-value less than .0083 was considered statistically significant. The effect size calculated using omega squared was 0.038, indicating a small effect. Similarly, partial eta squared also suggested a small effect (partial $\eta^2 = .046$). As in Table 54, the Bonferroni post hoc analysis revealed that the NP group ($M = 2.89$, $SD = 1.36$) had statistically significantly higher mean transfer scores than the N group ($M = 2.18$, $SD = 1.33$) ($p = .002$), the NS group ($M = 2.21$, $SD = 1.37$) ($p = .003$), and the T group ($M = 2.27$, $SD = 1.30$) ($p = .008$). There was no statistically significant difference in any other group combination.

Table 54. Multiple Comparisons of Transfer Scores of Students Using Different Versions of the Material

(I) Version	(J) Version	(I-J) Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
N	NS	-.029	.194	1.000	-.544	.487
	T	-.090	.193	1.000	-.603	.422
	NP	-.712*	.193	.002	-1.224	-.199
NS	T	-.062	.194	1.000	-.577	.453
	NP	-.683*	.194	.003	-1.198	-.168
T	NP	-.622*	.193	.008	-1.134	-.109

*. The mean difference is significant at the .0083 level (Bonferroni Adjustment).

4.2.7 Research question 2.2.1

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence transfer test scores for university students in a virtual reality learning environment?

4.2.7.1 Subject matter: Lightning formation

Upon inspecting a boxplot, it was observed that one data point was considered an outlier. The outlier in the NS group was at the top. The outlier was removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005) (see Appendix I, Figure I30). Levene's test for equality of variances ($p = .164$) indicated that variances were homogeneous (see Appendix J, Table J38). According to Shapiro-Wilk values, the data was normally distributed for NS, T, and NP groups ($p > .05$) but not for the N group ($p = .042$) (see Appendix J, Table J39). Skewness values ranged from -.373 to .178, and Kurtosis values ranged from -.817 to -.432 (see Appendix J, Table J40). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 55 shows descriptive statistics of each treatment group.

Table 55. Descriptive Statistics for Students' Transfer Scores of Each Version in the Virtual Learning Environment

Groups	Mean	Std. Dev.	n
N	1.740	.871	32
NS	2.043	1.246	31
T	1.823	.912	32
NP	2.073	1.090	32
Total	1.919	1.036	127

The one-way ANOVA test ($F(3, 123) = .853$, $p = .501$) indicated no significant difference in the means of transfer scores between the different version groups.

4.2.7.2 Subject matter: Cell

There was one outlier in the data assessed by inspection of a boxplot. The outlier in the N group was at the top. The outlier was removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005) (see Appendix I, Figure I31). Then, normally distributed data for all groups ($p > .05$) were obtained (see Appendix J, Table J41), and the homogeneity of variances in the data sets ($p > .05$) was observed (see Appendix J, Table J42). Table 56 shows descriptive statistics of each treatment group.

Table 56. Descriptive Statistics for Students' Transfer Scores of Each Version in the Virtual Learning Environment

Groups	Mean	Std. Dev.	n
N	1.699	1.212	31
NS	2.406	1.261	32
T	2.344	1.237	32
NP	2.708	1.182	32
Total	2.294	1.263	127

The one-way ANOVA test ($F(3, 123) = 3.776$, $p = .012$) revealed that means of transfer scores were significantly different among the different version groups. The effect size calculated using omega squared was 0.014, indicating a small effect size. However, partial eta squared suggested a moderate effect (partial $\eta 2 = .084$). This indicates that the different versions of the educational material designed for the virtual learning environment moderately affected students' transfer scores. As in Table 57, the Bonferroni post hoc analysis revealed that the NP group ($M = 2.71$, $SD = 1.18$) had significantly higher mean transfer scores ($p = .0082$) than the N group ($M = 1.70$, $SD = 1.21$) but no other group differences were statistically significant.

Table 57. Multiple Comparisons of Transfer Scores of Students Using Different Versions of the Material in the Virtual Learning Environment

(I) Version	(J) Version	(I-J) Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
N	NS	-.707	.308	.141	-1.534	.119
	T	-.645	.308	.231	-1.472	.182
	NP	-1.009*	.308	.0082	-1.836	-.183
NS	T	.063	.306	1.000	-.758	.883
	NP	-.302	.305	1.000	-1.122	.518
T	NP	-.365	.306	1.000	-1.185	.455

*. The mean difference is significant at the .0083 level (Bonferroni Adjustment).

4.2.8 Research question 2.2.2

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence transfer test scores for university students in an augmented reality learning environment?

4.2.8.1 Subject matter: Lightning formation

Upon visual inspection of a boxplot, it was determined that the data contained one outlier. The outlier in the NS group was at the top. The outlier was removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005) (see Appendix I, Figure I32). Then, normally distributed data for all groups ($p > .05$) were obtained (see Appendix J, Table J43), and the homogeneity of variances in the data sets ($p > .05$) was observed (see Appendix J, Table J44). Table 58 shows descriptive statistics of each treatment group.

Table 58. Descriptive Statistics for Students' Transfer Scores of Each Version in the Augmented Learning Environment

Groups	Mean	Std. Dev.	n
N	2.371	1.007	32
NS	1.744	.904	30
T	1.396	1.086	32
NP	1.865	1.097	32
Total	2.820	1.064	126

The one-way ANOVA test ($F (3, 122) = 3.940$, $p = .010$) revealed that means of transfer scores were significantly different among the different version groups. The effect size calculated using omega squared was 0.017, indicating a small effect, whereas partial eta squared indicated a moderate effect (partial $\eta 2 = .088$). In other words, effect size estimates indicate that different versions of the educational material designed for the augmented learning environment moderately affected students' transfer scores. As in Table 59, the Bonferroni post hoc analysis revealed that the N group ($M = 2.37$, $SD = 1.00$) had significantly higher mean transfer scores than the T group ($M = 1.40$, $SD = 1.09$) ($p = .005$). However, no other group differences were statistically significant.

Table 59. Multiple Comparisons of Transfer Scores of Students Using Different Versions of the Material in the Augmented Learning Environment

(I) Version	(J) Version	(I-J) Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
N	NS	.526	.261	.277	-.174	1.227
	T	.875*	.257	.005	.186	1.564
	NP	.406	.257	.700	-.283	1.096
NS	T	.349	.261	1.000	-.352	1.049
	NP	-.120	.261	1.000	-.821	.581
T	NP	-.469	.257	.424	-1.158	.221

*. The mean difference is significant at the .0083 level (Bonferroni Adjustment).

4.2.8.2 Subject matter: Cell

Upon visual inspection of a boxplot, it was determined that the data contained two outliers. The outliers in the N group were at the top and the bottom. The outlier was removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005) (see Appendix I, Figure I33). Then, normally distributed data for all groups ($p > .05$) were obtained (see Appendix J, Table J45), and the homogeneity of variances in the data sets ($p > .05$) was observed (see Appendix J, Table J46). Table 60 shows descriptive statistics of each treatment group.

Table 60. Descriptive Statistics for Students' Transfer Scores of Each Version in the Augmented Learning Environment

Groups	Mean	Std. Dev.	n
N	2.622	.857	30
NS	2.258	1.310	31
T	2.156	1.179	32
NP	3.104	1.387	32
Total	2.536	1.248	125

The one-way ANOVA test ($F(3, 121) = 4.036$, $p = .009$) revealed that means of transfer scores were significantly different among the different version groups. The effect size calculated using omega squared was 0.013, indicating a small effect, whereas partial eta squared indicated a moderate effect (partial $\eta2$ = .091). In other words, effect size estimates indicate that different versions of the educational material designed for the augmented learning environment moderately affected students' transfer scores. As in Table 61, the Bonferroni post hoc analysis revealed that the NP group (M = 3.10, SD = 1.39) had statistically significantly higher mean transfer scores than the NS group (M = 2.26, SD = 1.31) (p = .037) and T group (M = 2.16, SD = 1.18) (p = .012) based on without Bonferroni correction. However, after the Bonferroni correction, the group means were not significantly different ($p > .0083$).

Table 61. Multiple Comparisons of Transfer Scores of Students Using Different Versions of the Material in the Augmented Learning Environment

(I) Version	(J) Version	(I-J) Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
N	NS	.364	.309	1.000	-.463	1.192
	T	.466	.306	.732	-.355	1.287
	NP	-.482	.306	.708	-1.303	.339
NS	T	.102	.304	1.000	-.712	.916
	NP	-.846	.304	.037	-1.660	-.032
T	NP	-.948	.301	.012	-1.756	-.140

*. The mean difference is significant at the .0083 level (Bonferroni Adjustment).

4.2.9 Research question 2.2.3

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence transfer test scores for university students in a traditional learning environment?

4.2.9.1 Subject matter: Lightning formation

Upon visual inspection of a boxplot, it was determined that the data contained one outlier. The outlier in the NP group was at the top. The outlier was removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005) (see Appendix I, Figure I34). Variances were homogeneous, as assessed by Levene's test for equality of variances (p = .050) (see Appendix J, Table J47). According to Shapiro-Wilk values, the data was normally distributed for NS, T, and NP groups (p > .05) but not for the N group (p = .014) (see Appendix J, Table J48). Skewness values ranged from -.210 to .294, and Kurtosis values ranged from -1.212 to -.124 (see Appendix J, Table J49). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 62 shows descriptive statistics of each treatment group.

Table 62. Descriptive Statistics for Students' Transfer Scores of Each Version in the Traditional Learning Environment

Groups	Mean	Std. Dev.	n
N	1.823	1.215	32
NS	2.188	.991	32
T	1.854	1.368	32
NP	2.108	.940	31
Total	1.92	1.140	127

The one-way ANOVA test (F (3, 123) =.806, p = .496) revealed that means of transfer scores were not significantly different among the different version groups.

4.2.9.2 Subject matter: Cell

There were three outliers in the data assessed by inspection of a boxplot. The outlier in the N group was at the top, whereas the outliers in the NS group were at the top and the bottom. The outliers were removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005) (see Appendix I, Figure I35). Then, the homogeneity of variances in the data sets ($p > .05$) was observed (see Appendix J, Table J50). According to Shapiro-Wilk values, the data was normally distributed for T, N, and NP groups ($p > .05$) but not for the NS group ($p = .018$) (see Appendix J, Table J51). Skewness values ranged from .120 to .275, and Kurtosis values ranged from -.894 to .094 (see Appendix J, Table J52). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 63 shows descriptive statistics of each treatment group.

Table 63. Descriptive Statistics for Students' Transfer Scores of Each Version in the Traditional Learning Environment

Groups	Mean	Std. Dev.	n
N	1.979	1.294	31
NS	1.822	1.378	30
T	2.302	1.496	32
NP	2.854	1.507	32
Total	2.248	1.461	125

The one-way ANOVA test ($F(3, 121) = 3.216$, $p = .025$) revealed that means of transfer scores were significantly different among the different version groups. The effect size calculated using omega squared was 0.050, indicating a small effect, whereas partial eta squared indicated a moderate effect (partial $\eta 2 = .074$). In other words, effect size estimates indicate that different versions of the educational material designed for the traditional learning environment moderately affected students' transfer scores. As in Table 64, the Bonferroni post hoc analysis revealed that the NP group ($M = 2.85$, $SD = 1.51$) had statistically significantly higher mean transfer scores ($p = .031$) than the NS group (M = 1.82, SD = 1.38) based on before Bonferroni correction. However, after the Bonferroni correction, the group means were not significantly different ($p > .0083$).

Table 64. Multiple Comparisons of Transfer Scores of Students Using Different Versions of the Material in the Traditional Learning Environment

(I) Version	(J) Version	(I-J) Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
N	NS	.156	.365	1.000	-.822	1.134
	T	-.324	.359	1.000	-1.286	.639
	NP	-.876	.359	.096	-1.838	.086
NS	T	-.480	.362	1.000	-1.450	.490
	NP	-1.032*	.362	.031	-2.002	-.062
T	NP	-.552	.356	.740	-1.507	.402

*. The mean difference is significant at the .0083 level (Bonferroni Adjustment).

4.2.10 Summary of results of research question 2.2

The study examined the effect of segmenting, pre-training, and modality principles in different learning environments on university students' transfer scores. For the lightning formation subject matter, the transfer scores were not significantly different among version groups, except in the augmented learning environment, where the N group had significantly higher mean transfer scores than the T group. The transfer scores significantly differed among version groups for the cell structure subject matter. The NP group had higher mean transfer scores than the N, NS, and T groups in the overall cell structure subject matter analysis. In the virtual learning environment, the NP group had significantly higher mean transfer scores than the N group. In the augmented learning environment, the NP group had significantly higher mean transfer scores than the NS and T groups, but these differences were insignificant after the Bonferroni correction. In the traditional learning environment, the NP group had significantly higher mean transfer scores than the NS group, but this difference was insignificant after the

Bonferroni correction. In conclusion, the segmenting, pre-training, and modality effects on transfer test scores were generally less consistent across learning environments than retention test scores. The NP group often performed better than other groups but not always with statistical significance after adjusting for multiple comparisons. Detailed information is provided in Table 65.

Table 65. Summary Results for Research Question 2.2

Subject	IV	DV	p	Post hoc	Effect of Principles
Lightning Formation	Versions (MML) (Environment: All)	Transfer Score	.074	No difference	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
	Versions (MML) (Environment: VR)	Transfer Score	.501	No difference	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
	Versions (MML) (Environment: AR)	Transfer Score	.010	N > T (p = .005)	Modality: Observed Pre-training: Not observed Segmenting: Not observed
	Versions (MML) (Environment: Traditional)	Transfer Score	.496	No difference	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
Cell	Versions (MML) (Environment: All)	Transfer Score	< .001	NP > N (p = .002) NP > NS (p = .003) NP > T (p = .008)	Modality: Not observed Pre-training: Observed Segmenting: Not observed
	Versions (MML) (Environment: VR)	Transfer Score	.012	NP > N (p = .008)	Modality: Not observed Pre-training: Observed Segmenting: Not observed
	Versions (MML) (Environment: AR)	Transfer Score	.009	NP > NS (p = .037) NP > T (p = .012)	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
	Versions (MML) (Environment: Traditional)	Transfer Score	.025	NP > NS (p = .031)	Modality: Not observed Pre-training: Not observed Segmenting: Not observed

4.2.10 Research question 2.3

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence cognitive load scores for university students across different learning environments?

4.2.10.1 Subject matter: Lightning formation

4.2.10.1.1 Intrinsic cognitive load

There were three outliers in the data assessed by inspection of a boxplot. The outliers in the NS group and NP group were at the top. The outliers were removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005) (see Appendix I, Figure I36). Levene's test for equality of variances ($p = .134$) indicated that variances were homogeneous (see Appendix J, Table J53). According to Kolmogorov-Smirnov values, the data was normally distributed for N, NS, and NP groups ($p > .05$) but not for the T group ($p = .046$) (see Appendix J, Table J54). Skewness values ranged from -.243 to .528, and Kurtosis values ranged from -.832 to -.016. Based on the skewness & kurtosis and visual inspection, the data were normally distributed (see Appendix J, Table J55). Table 66 shows descriptive statistics of each treatment group. The mean scores in the table reflect the intrinsic cognitive scores ranging from zero to 10.

Table 66. Descriptive Statistics for Students' Intrinsic Cognitive Load Scores of Each Version

Groups	Mean	Std. Dev.	n
N	3.556	2.195	96
NS	3.072	2.099	93
T	3.816	2.457	96
NP	3.361	1.867	95
Total	3.454	2.174	380

The one-way ANOVA test ($F(3, 376) = 1.988$, $p = .115$) revealed that the means of intrinsic cognitive load scores were not statistically significantly different among the different version groups.

4.2.10.1.2 Extraneous cognitive load

Upon examining a boxplot, it was determined that the data had eight outliers. The outliers in the N, T, and NP groups were located at the upper end of the distribution. Levene's test for equality of variances ($p = .257$) indicated that there was homogeneity of variances, suggesting that the variances of the different groups were not significantly different from each other. However, Kolmogorov-Smirnov and Skewness-Kurtosis values indicated that the data for each group was not normally distributed ($p > .05$). Therefore, a Kruskal-Wallis H test was performed to investigate the segmenting, pre-training, and modality effect on undergraduate students' extraneous cognitive load scores. Table 67 shows the medians and numbers for each treatment group.

Table 67. Medians for Extraneous Cognitive Load Scores of Each Version

Groups	N	Mean Rank
N	96	1.39
NS	95	1.00
T	96	1.33
NP	96	1.94
Total	383	1.44

After examining a boxplot (see Appendix I, Figure I37), it was determined that the distribution of extraneous cognitive load scores was similar for all groups. A Kruskal-Wallis H test ($\chi2$ (3) = 4.019, $p = .259$) revealed that the medians of extraneous cognitive load scores were not statistically significantly different between the different version groups.

4.2.10.1.3 Germane cognitive load

Upon inspecting a boxplot, it was observed that one outlier in the N group was located at the lower end of the distribution. Additionally, normality ($p < .05$) and homogeneity of variance assumptions were violated ($p < .05$). Therefore, a Kruskal-Wallis H test was conducted to investigate the segmenting, pre-training, and modality effect on undergraduate students' germane cognitive load scores. Table 68 shows the medians and numbers for each treatment group.

Table 68. Medians for Germane Cognitive Load Scores of Each Version

Groups	N	Mean Rank
N	95	6.25
NS	95	7.00
T	96	6.00
NP	96	6.38
Total	382	6.50

Distributions of germane cognitive load scores were similar for all groups, as assessed by visual inspection of a boxplot (see Appendix I, Figure I38). A Kruskal-Wallis H test ($\chi 2$ (3) = 2.767, $p = .429$) revealed that the medians of germane cognitive load scores were not statistically significantly different between the different version groups.

4.2.10.2 Subject matter: Cell

4.2.10.2.1 Intrinsic cognitive load

After examining a boxplot, it was determined that the data had no outliers (see Appendix I, Figure I39). Levene's test for equality of variances ($p = .153$) indicated that there was homogeneity of variances, suggesting that the variances of the different groups were not significantly different from each other (see Appendix J, Table J56). According to Kolmogorov-Smirnov values, the data was not normally distributed for all groups ($p < .05$) (see Appendix J, Table J57). Skewness values ranged from -.055 to .354, and Kurtosis values ranged from -1.452 to -.992 (see Appendix J, Table J58). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 69 shows descriptive statistics of each treatment group.

Table 69. Descriptive Statistics for Students' Intrinsic Cognitive Load Scores of Each Version

Groups	Mean	Std. Dev.	n
N	4.885	2.927	96
NS	4.747	3.294	95
T	4.479	3.132	96
NP	4.469	3.061	96
Total	4.645	3.099	383

The one-way ANOVA test ($F(3, 379) = .420$, $p = .738$) revealed that means of intrinsic cognitive load scores were not statistically significantly different among the different version groups.

4.2.10.2.2 Extraneous cognitive load

There was no outlier in the data assessed by inspection of a boxplot (see Appendix I, Figure I40). Variances were homogeneous, as assessed by Levene's test for equality of variances ($p = .192$) (see Appendix J, Table J59). According to Kolmogorov-Smirnov values, the data was not normally distributed for all groups ($p < .05$) (see Appendix J, Table J60). Skewness values ranged from .740 to .866, and Kurtosis values ranged from -.473 to -.352 (see Appendix J, Table J61). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 70 shows descriptive statistics of each treatment group.

Table 70. Descriptive Statistics for Students' Extraneous Cognitive Load Scores of Each Version

Groups	Mean	Std. Dev.	n
N	2.031	1.861	96
NS	2.126	1.858	95
T	2.250	2.205	96
NP	2.406	2.034	96
Total	2.204	1.992	383

The one-way ANOVA test ($F(3, 379) = .634$, $p = .594$) revealed that means of extraneous cognitive load scores were not statistically significantly different among the different version groups.

4.2.10.2.3 Germane cognitive load

After examining a boxplot, it was determined that the data had no outliers (see Appendix I, Figure I41). Levene's test for equality of variances ($p = .649$) indicated homogeneity, suggesting that the variances of the different groups were not significantly different (see

Appendix J, Table J62). According to Kolmogorov-Smirnov values, the data was not normally distributed for all groups ($p < .05$) (see Appendix J, Table J63). Skewness values ranged from -.692 to -.207, and Kurtosis values ranged from -.962 to -.242 (see Appendix J, Table J64). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 71 shows descriptive statistics of each treatment group.

Table 71. Descriptive Statistics for Students' Germane Cognitive Load Scores of Each Version

Groups	Mean	Std. Dev.	n
N	5.760	2.682	96
NS	6.326	2.773	95
T	5.854	2.644	96
NP	6.094	2.821	96
Total	6.008	2.729	383

The one-way ANOVA test ($F(3, 379) = .826$, $p = .480$) revealed that means of germane cognitive load scores were not significantly different among the different version groups.

4.2.11 Research question 2.3.1

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence cognitive load scores for university students in a virtual reality learning environment?

4.2.11.1 Subject matter: Lightning formation

4.2.11.1.1 Intrinsic cognitive load

Upon visual inspection of a boxplot, it was determined that one outlier was in the data. The outlier in the NS group and NP group was at the top. The outlier was removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005) (see Appendix I, Figure I42). Variances were homogeneous, as assessed by Levene's test for equality of variances ($p = .685$) (see Appendix J, Table J65). According to Shapiro-Wilk values, the data was normally distributed for N, NS, and T groups ($p > .05$) but not for the NP group ($p = .039$) (see Appendix J, Table J66). Skewness values ranged from -.089 to .103, and Kurtosis values ranged from -1.331 to -.184 (see Appendix J, Table J67). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 72 shows descriptive statistics of each treatment group.

Table 72. Descriptive Statistics for Students' Intrinsic Cognitive Load Scores of Each Version in the Virtual Reality Environment

Groups	Mean	Std. Dev.	n
N	3.240	1.907	32
NS	3.366	1.910	31
T	3.813	2.154	32
NP	3.292	2.061	32
Total	3.428	2.000	127

The one-way ANOVA test ($F(3, 123) = .542$, $p = .654$) revealed that means of intrinsic cognitive load scores in the virtual reality learning environment were not significantly different among the different version groups.

4.2.11.1.2 Extraneous cognitive load

Upon examining a boxplot, it was observed that one data point was considered an outlier. The outlier in the N group was at the top. The normality ($p < .05$) and the homogeneity of variance assumptions were violated ($p < .05$). Thus, a Kruskal-Wallis H test was conducted to examine the segmenting, pre-training, and modality effect on undergraduate students' extraneous cognitive load scores. Table 73 shows the medians and numbers for each treatment group.

Table 73. Medians for Extraneous Cognitive Load Scores of Each Version in the Virtual Learning Environment

Groups	N	Mean Rank
N	31	1.33
NS	32	1.39
T	32	1.22
NP	32	1.50
Total	127	1.33

Distributions of extraneous cognitive load scores were similar for all groups, as assessed by visual inspection of a boxplot (see Appendix I, Figure I43). A Kruskal-Wallis H test ($\chi 2$ (3) = 1.174, $p = .759$) revealed that the medians of extraneous cognitive load scores were not statistically significantly different between the different version groups.

4.2.11.1.3 Germane cognitive load

After examining a boxplot, it was determined that the data had no outliers (see Appendix I, Figure I44). Variances were homogeneous, as assessed by Levene's test for equality of variances ($p = .121$) (see Appendix J, Table J68). According to Shapiro-Wilk values, the data was normally distributed for N, T, and NP groups ($p > .05$) but not for the NS group ($p = .015$) (see Appendix J, Table J69). Skewness values ranged from -.546 to .067, and Kurtosis values ranged from -1.266 to -.859 (see Appendix J, Table J70). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 74 shows descriptive statistics of each treatment group.

Table 74. Descriptive Statistics for Students' Germane Cognitive Load Scores of Each Version in the Virtual Reality Environment

Groups	Mean	Std. Dev.	n
N	6.117	2.080	32
NS	6.266	2.859	32
T	5.867	2.744	32
NP	5.680	2.775	32
Total	5.598	2.611	128

The one-way ANOVA test ($F(3, 124) = .313$, $p = .816$) revealed that means of germane cognitive load scores in the virtual reality learning environment were not significantly different among the different version groups.

4.2.11.2 Subject matter: Cell

4.2.11.2.1 Intrinsic cognitive load

After examining a boxplot, it was determined that the data had no outliers (see Appendix I, Figure I45). Variances were homogeneous, as assessed by Levene's test for equality of variances ($p = .738$) (see Appendix J, Table J68). According to Shapiro-Wilk values, the data was normally distributed for the N group ($p > .05$) but not for the NS group ($p = .003$), T group ($p = .009$), and NP group ($p = .016$) (see Appendix J, Table J69). Skewness values ranged from -.450 to .500, and Kurtosis values ranged from -1.149 to -

.868 (see Appendix J, Table J70). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 75 shows descriptive statistics of each treatment group.

Table 75. Descriptive Statistics for Students' Intrinsic Cognitive Load Scores of Each Version in the Virtual Reality Environment

Groups	Mean	Std. Dev.	n
N	4.969	2.857	32
NS	4.875	2.745	32
T	4.188	3.177	32
NP	3.781	3.024	32
Total	4.435	2.961	128

The one-way ANOVA test ($F (3, 124) = 1.180$, $p = .320$) revealed that means of intrinsic cognitive load scores in the virtual reality learning environment were not significantly different among the different version groups.

4.2.11.2.2 Extraneous cognitive load

After examining a boxplot, it was determined that the data had no outliers (see Appendix I, Figure I46). The normality ($p < .05$) and the homogeneity of variance assumptions were violated ($p < .05$) (see Appendix J, Table J71-73). Thus, a Kruskal-Wallis H test was conducted to examine the effects of segmenting, pre-training, and modality on undergraduate students' extraneous cognitive load scores. Table 76 shows the medians and numbers for each treatment group.

Table 76. Medians for Extraneous Cognitive Load Scores of Each Version in the Virtual Learning Environment

Groups	N	Mean Rank
N	32	1.00
NS	32	2.00
T	32	1.00
NP	32	1.00
Total	128	1.50

Distributions of extraneous cognitive load scores were similar for all groups, as assessed by visual inspection of a boxplot. A Kruskal-Wallis H test ($\chi 2$ (3) = 2.462, p = .482) revealed that the medians of extraneous cognitive load scores were not statistically significantly different between the different version groups.

4.2.11.2.3 Germane cognitive load

After examining a boxplot, it was determined that the data had no outliers. The normality ($p < .05$) and the homogeneity of variance assumptions were violated ($p < .05$). Thus, a Kruskal-Wallis H test was conducted to examine the segmenting, pre-training, and modality effect on undergraduate students' germane cognitive load scores. Table 77 shows the medians and numbers for each treatment group.

Table 77. Medians for Germane Cognitive Load Scores of Each Version in the Virtual Learning Environment

Groups	N	Mean Rank
N	32	7.00
NS	32	7.00
T	32	7.00
NP	32	7.00
Total	128	7.00

Distributions of germane cognitive load scores were similar for all groups, as assessed by visual inspection of a boxplot (see Appendix I, Figure I47). A Kruskal-Wallis H test ($\chi2$ (3) = 1.953, p = .582) revealed that the medians of germane cognitive load scores were not statistically significantly different between the different version groups.

4.2.12 Research question 2.3.2

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence cognitive load scores for university students in augmented reality learning environments?

4.2.12.1 Subject matter: Lightning formation

4.2.12.1.1 Intrinsic cognitive load

There was no outlier in the data assessed by inspection of a boxplot (see Appendix I, Figure I48). Variances were homogeneous, as assessed by Levene's test for equality of variances (p = .150) (see Appendix J, Table J74). According to Shapiro-Wilk values, the data was normally distributed for N, NS, and NP groups ($p > .05$) but not for the T group (p = .035) (see Appendix J, Table J75). Skewness values ranged from .049 to .648, and Kurtosis values ranged from -.625 to -.029. The data were normally distributed based on the skewness & kurtosis and visual inspection (see Appendix J, Table J76). Table 78 shows descriptive statistics of each treatment group.

Table 78. Descriptive Statistics for Students' Intrinsic Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	Mean	Std. Dev.	n
N	3.260	1.983	32
NS	3.602	1.997	31
T	4.458	2.799	32
NP	3.344	1.738	32
Total	3.667	2.195	127

The one-way ANOVA test ($F(3, 123) = 2.043$, $p = .111$) revealed that the means of intrinsic cognitive load scores in the augmented reality learning environment were not significantly different among the different version groups.

4.2.12.1.2 Extraneous cognitive load

There was no outlier in the data assessed by inspection of a boxplot (see Appendix I, Figure I49). Variances were homogeneous, as assessed by Levene's test for equality of variances ($p = .108$) (see Appendix J, Table J77). According to Shapiro-Wilk values, the data was not normally distributed for N, NS, T and NP groups ($p < .05$) (see Appendix J, Table J78). Skewness values ranged from .307 to .876, and Kurtosis values ranged from -1.482 to -.286. The data were normally distributed based on the skewness & kurtosis and visual inspection (see Appendix J, Table J79). Table 79 shows descriptive statistics of each treatment group.

Table 79. Descriptive Statistics for Students' Extraneous Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	Mean	Std. Dev.	n
N	2.181	1.886	32
NS	2.093	1.003	31
T	3.285	2.839	32
NP	2.542	1.972	32
Total	2.528	2.234	127

The one-way ANOVA test ($F(3, 123) = 1.914$, $p = .131$) revealed that means of extraneous cognitive load scores in the augmented reality learning environment were not significantly different among the different version groups.

4.2.12.1.3 Germane cognitive load

After examining a boxplot, it was determined that the data had no outliers. Then, the normality ($p < .05$) and the homogeneity of variance assumptions were violated ($p < .05$) (see Appendix J, Table J80-82). Therefore, a Kruskal-Wallis H test was conducted to examine the effects of segmenting, pre-training, and modality on undergraduate students' germane cognitive load scores in the augmented reality learning environment. Table 80 shows the medians and numbers for each treatment group.

Table 80. Medians for Germane Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	N	Mean Rank
N	32	5.63
NS	31	6.75
T	32	6.00
NP	32	6.86
Total	127	6.00

Distributions of germane cognitive load scores were similar for all groups, as assessed by visual inspection of a boxplot (see Appendix I, Figure 80). A Kruskal-Wallis H test ($\chi 2$ (3) = 4.098, $p = .251$) revealed that the medians of germane cognitive load scores in the augmented reality learning environment were not significantly different between the different versions.

4.2.12.2 Subject matter: Cell

4.2.12.2.1 Intrinsic cognitive load

There was no outlier in the data assessed by inspection of a boxplot (see Appendix I, Figure I51). Variances were homogeneous, as assessed by Levene's test for equality of variances (p = .302) (see Appendix J, Table J83). According to Shapiro-Wilk values, the data was normally distributed for N and NP groups ($p > .05$) but not for the T group ($p = .021$) and NS group ($p = .003$) (see Appendix J, Table J84). Skewness values ranged from -.148 to .469, and Kurtosis values ranged from -1.463 to -.618 (see Appendix J, Table J85). Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 81 shows descriptive statistics of each treatment group.

Table 81. Descriptive Statistics for Students' Intrinsic Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	Mean	Std. Dev.	n
N	4.625	2.721	32
NS	5.000	3.474	31
T	4.563	3.037	32
NP	4.375	3.035	32
Total	4.638	3.046	127

The one-way ANOVA test ($F(3, 123) = .228$, $p = .877$) revealed that means of intrinsic cognitive load scores in the augmented reality learning environment were not significantly different among the different version groups.

4.2.12.2.2 Extraneous cognitive load

There was no outlier in the data assessed by inspection of a boxplot (see Appendix I, Figure I52). Variances were homogeneous, as assessed by Levene's test for equality of variances (p = .075). According to Kolmogorov-Smirnov, the data was not normally distributed for each group (p > .05). Skewness values ranged from -.005 to 1.178, and Kurtosis values ranged from -1.250 to.434. Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 82 shows descriptive statistics of each treatment group.

Table 82. Descriptive Statistics for Students' Extraneous Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	Mean	Std. Dev.	n
N	2.406	2.408	32
NS	2.452	1.823	31
T	3.156	3.017	32
NP	3.219	2.196	32
Total	2.811	2.406	127

The one-way ANOVA test (F (3, 123) = 1.060, p = .369) revealed that means of extraneous cognitive load scores in the augmented reality learning environment were not significantly different among the different version groups.

4.2.12.2.3 Germane cognitive load

There was no outlier in the data assessed by inspection of a boxplot (see Appendix I, Figure I53). Variances were homogeneous, as assessed by Levene's test for equality of variances (p = .057) (see Appendix J, Table J80). According to Shapiro-Wilk values, the data was normally distributed for NP groups (p > .05) but not for the N group (p = .030), T group (p = .027), and NS group (p = .019). Skewness values ranged from -.417 to -.105, and Kurtosis values ranged from -1.019 to -.494. Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 83 shows descriptive statistics of each treatment group.

Table 83. Descriptive Statistics for Students' Germane Cognitive Load Scores of Each Version in the Augmented Learning Environment

Groups	Mean	Std. Dev.	n
N	6.625	2.060	32
NS	7.129	2.247	32
T	5.219	3.098	32
NP	6.313	2.442	31
Total	6.315	2.563	127

The one-way ANOVA test ($F(3, 123) = 3.325$, $p = .022$) revealed that means of germane cognitive load scores were statistically significantly different among the different version groups. The effect size calculated using omega squared was 0.007, indicating a lower than small effect, whereas partial eta squared indicated a moderate effect (partial $\eta^2 = .075$). In other words, effect size estimates indicate that students' germane cognitive load scores were moderately affected by different versions of the educational material designed for the augmented learning environment. As in Table 84, the Bonferroni post hoc analysis revealed that the NS group ($M = 7.13$, $SD = 2.25$) had statistically significantly higher mean germane cognitive load scores ($p = .017$) than the T group ($M = 5.22$, $SD = 3.10$) based on before Bonferroni correction. However, after the Bonferroni correction, the group means were not significantly different ($p > .0083$).

Table 84. Multiple Comparisons of Germane Cognitive Load Scores of Students Using Different Versions of the Material in the Augmented Learning Environment

(I) Version	(J) Version	(I-J) Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
N	NS	-.504	.629	1.000	-2.190	1.182
	T	1.406	.624	.155	-.266	3.079
	NP	.313	.624	1.000	-1.360	1.985
NS	T	1.910	.629	.017	.225	3.596
	NP	.817	.629	1.000	-.869	2.502
T	NP	-1.094	.624	.492	-2.766	.579

*. The mean difference is significant at the .0083 level (Bonferroni Adjustment).

4.2.13 Research question 2.3.3

This section will show the results of the lightning formation and cell subject matters for the following question: To what extent do segmenting, pre-training, and modality effects influence cognitive load scores for university students in traditional learning environment?

4.2.13.1 Subject matter: Lightning formation

4.2.13.1.1 Intrinsic cognitive load

There was one outlier in the data assessed by inspection of a boxplot. The outlier in the NP group was at the top. The outlier was removed from the data to check the other assumptions of the one-way ANOVA analysis (Weisberg, 2005) (see Appendix I, Figure I54). Variances were homogeneous, as assessed by Levene's test for equality of variances ($p = .160$). According to Shapiro-Wilk values, the data was normally distributed for N and NP groups ($p > .05$) but not for the T group ($p = .031$) and NS group ($p = .003$). Skewness values ranged from -.295 to .842, and Kurtosis values ranged from -1.077 to -.214. Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 85 shows descriptive statistics of each treatment group.

Table 85. Descriptive Statistics for Students' Intrinsic Cognitive Load Scores of Each Version in the Traditional Learning Environment

Groups	Mean	Std. Dev.	n
N	4.167	2.571	32
NS	2.438	2.403	32
T	3.177	2.275	32
NP	3.452	1.841	31
Total	3.307	2.350	127

The one-way ANOVA test ($F(3, 123) = 3.109$, $p = .029$) revealed that means of intrinsic cognitive load scores were statistically significantly different among the different version groups. The effect size calculated using omega squared was 0.016, indicating a small effect, whereas partial eta squared indicated a moderate effect (partial $\eta 2 = .070$). In other words, effect size estimates indicate that students' intrinsic cognitive load scores were moderately affected by different versions of the educational material designed for the traditional learning environment. As in Table 86, the Bonferroni post hoc analysis revealed that the N group ($M = 4.17$, $SD = 2.57$) had statistically significantly higher mean intrinsic cognitive load scores ($p = .019$) than the NS group ($M = 2.44$, $SD = 2.40$) based on before Bonferroni correction. However, the group means were not statistically significantly different ($p > .0083$) after the Bonferroni correction.

Table 86. Multiple Comparisons of Intrinsic Cognitive Load Scores of Students Using Different Versions of the Material in the Traditional Learning Environment

(I) Version	(J) Version	(I-J) Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
N	NS	1.729	.573	.019	.192	3.266
	T	.990	.573	.521	-.548	2.528
	NP	.715	.578	1.000	-.835	2.265
NS	T	-.740	.573	1.000	-2.277	.798
	NP	-1.014	.578	.490	-2.564	.535
T	NP	-.275	.578	1.000	-1.824	1.275

*. The mean difference is significant at the .0083 level (Bonferroni Adjustment).

4.2.13.1.2 Extraneous cognitive load

There were two outliers in the data assessed by inspection of a boxplot. The outliers in the NS and NP groups were at the top. Then, the normality ($p < .05$) and the homogeneity of variance assumptions were violated ($p < .05$). Therefore, a Kruskal-Wallis H test was conducted to examine the segmenting, pre-training, and modality effect on undergraduate students' extraneous cognitive load scores in the traditional reality learning environment. Table 87 shows the medians and numbers for each treatment group.

Table 87. Medians for Extraneous Cognitive Load Scores of Each Version in the Traditional Reality Learning Environment

Groups	N	Mean Rank
N	32	1.17
NS	31	.89
T	32	1.22
NP	31	2.11
Total	126	1.22

Distributions of extraneous cognitive load scores were similar for all groups, as assessed by visual inspection of a boxplot (see Appendix I, Figure I55). A Kruskal-Wallis H test ($\chi 2$ (3) = 6.263, $p = .099$) revealed that the medians of extraneous cognitive load scores in the traditional learning environment were not significantly different between groups.

4.2.13.1.3 Germane cognitive load

There was one outlier in the data assessed by inspection of a boxplot. The outlier in the N group was at the bottom (see Appendix I, Figure I56). Variances were homogeneous, as assessed by Levene's test for equality of variances (p = .146). According to Shapiro-Wilk values, the data was normally distributed for the NP group (p > .05) but not for the N group (p = .041), T group (p = .009), and NS group (p = .028). Skewness values ranged from -.830 to -.339, and Kurtosis values ranged from -1.016 to .203. Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 88 shows descriptive statistics of each treatment group.

Table 88. Descriptive Statistics for Students' Germane Cognitive Load Scores of Each Version in the Traditional Learning Environment

Groups	Mean	Std. Dev.	n
N	7.032	2.113	31
NS	6.844	2.623	32
T	6.563	2.832	32
NP	6.117	2.700	31
Total	6.636	2.577	127

The one-way ANOVA test (F (3, 123) = .750, p = .524) revealed that means of germane cognitive load scores were not significantly different among the different version groups in the traditional learning environments.

4.2.13.2 Subject matter: Cell

4.2.13.2.1 Intrinsic cognitive load

There was no outlier in the data assessed by inspection of a boxplot (see Appendix I, Figure I57). Variances were homogeneous, as assessed by Levene's test for equality of variances ($p = .262$). According to Shapiro-Wilk values, the data was normally distributed for the N group ($p > .05$) but not for the NS group ($p = .001$), T group ($p = .039$), and NP group ($p = .029$). Skewness values ranged from -.188 to .332, and Kurtosis values ranged from -1.452 to -1.012. Based on the skewness & kurtosis and visual inspection, the data were normally distributed. Table 89 shows descriptive statistics of each treatment group.

Table 89. Descriptive Statistics for Students' Intrinsic Cognitive Load Scores of Each Version in the Traditional Learning Environment

Groups	Mean	Std. Dev.	n
N	5.063	3.252	32
NS	4.375	3.670	32
T	4.688	3.257	32
NP	5.250	3.037	32
Total	4.844	3.290	128

The one-way ANOVA test ($F(3, 124) = .444$, $p = .722$) revealed that means of intrinsic cognitive load scores in the traditional learning environment were not significantly different among the different version groups.

4.2.13.2.2 Extraneous cognitive load

The normality ($p < .05$) and the homogeneity of variance assumptions were violated ($p < .05$). Therefore, a Kruskal-Wallis H test was conducted to examine the segmenting, pre-training, and modality effect on undergraduate students' extraneous cognitive load scores. Table 90 shows the medians and numbers for each treatment group.

Table 90. Medians for Extraneous Cognitive Load Scores of Each Version in the Traditional Learning Environment

Groups	N	Mean Rank
N	32	2.00
NS	32	1.00
T	32	2.00
NP	32	2.00
Total	128	2.00

Distributions of extraneous cognitive load scores were similar for all groups, as assessed by visual inspection of a boxplot (see Appendix I, Figure I58). A Kruskal-Wallis H test ($\chi2$ (3) = 1.192, $p = .755$) revealed that the medians of extraneous cognitive load scores were not significantly different between the different version groups in the traditional learning environment.

4.2.13.2.3 Germane cognitive load

The normality ($p < .05$) and the homogeneity of variance assumptions were violated ($p < .05$). Therefore, a Kruskal-Wallis H test was conducted to examine the segmenting, pre-training, and modality effect on undergraduate students' germane cognitive load scores. Table 91 shows the medians and numbers for each treatment group.

Table 91. Medians for Germane Cognitive Load Scores of Each Version in the Traditional Learning Environment

Groups	N	Mean Rank
N	32	5.00
NS	32	6.00
T	32	6.00
NP	32	5.00
Total	128	5.00

Distributions of germane cognitive load scores were similar for all groups, as assessed by visual inspection of a boxplot (see Appendix I, Figure I59). A Kruskal-Wallis H test ($\chi 2$ (3) = 2.662, p = .447) revealed that the medians of germane cognitive load scores were not significantly different between the different version groups in the traditional learning environment.

4.2.14 Summary of results of research question 2.3

The study investigated the effect of segmenting, pre-training, and modality principles in different learning environments on university students' cognitive load scores. Overall, the intrinsic, extraneous, and germane cognitive load scores for the lightning formation and cell structure subject matters were not significantly different among the different version groups. In the virtual reality learning environment, no significant differences were found among the different version groups for either lightning formation or cell structure subject matters regarding cognitive load. In the augmented reality learning environment, no significant differences existed among the different version groups for the lightning formation subject matter. The NS group had significantly higher mean germane cognitive load scores for the cell structure subject matter than the T group before the Bonferroni correction. However, the difference was not significant after the correction. In the traditional learning environment, the N group had significantly higher mean intrinsic cognitive load scores for the lightning formation subject matter than the NS group before the Bonferroni correction. However, the difference was not significant after the correction. No significant differences were found in the extraneous and germane cognitive load scores for lightning formation or in any cognitive load scores for cell structure subject matter. In summary, the segmenting, pre-training, and modality effects on cognitive load were not consistently observed across different learning environments, and significant differences were limited and often not significant after adjusting for multiple comparisons. Detailed information is provided in Table 92.

Table 92. Summary Results for Research Question 2.3

	IV	DV	p	Post hoc	Effect of Principles
Lightning Formation	Versions (MML) Environment: All	Intrinsic CL Extraneous CL Germane CL	.115 .259 .429	No difference	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
	Versions (MML) Environment: VR	Intrinsic CL Extraneous CL Germane CL	.654 .759 .816	No difference	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
	Versions (MML) Environment: AR	Intrinsic CL Extraneous CL Germane CL	.111 .131 .251	No difference	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
	Versions (MML) Environment: Traditional	Intrinsic CL Extraneous CL Germane CL	.029 .099 .524	No difference	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
Cell	Versions (MML) Environment: All	Intrinsic CL Extraneous CL Germane CL	.738 .594 .480	No difference	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
	Versions (MML) Environment: VR	Intrinsic CL Extraneous CL Germane CL	.320 .482 .582	No difference	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
	Versions (MML) Environment: AR	Intrinsic CL Extraneous CL Germane CL	.877 .369 .022	No difference	Modality: Not observed Pre-training: Not observed Segmenting: Not observed
	Versions (MML) Environment: Traditional	Intrinsic CL Extraneous CL Germane CL	.722 .755 .447	No difference	Modality: Not observed Pre-training: Not observed Segmenting: Not observed

4.3 Research question 3

This section will show the results of the lightning formation and cell subject matters for the following question: Is there a significant difference between the cognitive load scores of the students using the lightning formation learning material and those using the cell learning material?

4.3.1 Research question 3.1

This section will show the results of the lightning formation and cell subject matters for the following question: Is there a significant difference between the intrinsic cognitive load scores of the students using the lightning formation learning material and those using the cell learning material?

There was no outlier in the data assessed by inspection of a boxplot (see Appendix I, Figure I60). The normality ($p < .05$) was violated based on Shapiro-Wilk's test values (see Appendix J, Table J86). The Skewness value was .351, and the Kurtosis value was -.375 (see Appendix J, Table J87). Based on the skewness & kurtosis and visual inspection, the data were normally distributed (see Appendix I, Figure I61-62). Table 93 shows descriptive statistics of intrinsic cognitive load scores.

Table 93. Descriptive Statistics for Students' Intrinsic Cognitive Load Scores

	Groups	Mean	Std. Dev.	n
Intrinsic Cognitive	Cell	4.645	3.099	383
Load	Lightning Formation	3.499	2.223	383

A Paired-samples t-test revealed that the intrinsic cognitive load score was measured higher when using the educational material for cell subject matter ($M = 4.645$, SD = 3.099) as opposed to using the educational material for lightning formation subject matter (M = 3.499, SD = 2.223), a statistically significant mean increase of 1.146, 95% Cl [0.808, 1.485], t(382) = 6.656, p < .001, d = 0.34.

4.3.2 Research question 3.2

This section will show the results of the lightning formation and cell subject matters for the following question: Is there a significant difference between the extraneous cognitive load scores of the students using the lightning formation learning material and those using the cell learning material?

Three extreme outliers were in the data assessed by inspection of a boxplot (see Appendix I, Figure I63). They were removed from the data. The normality (p < .05) was violated based on Shapiro-Wilk's test values (see Appendix J, Table J88). The Skewness value was .193, and the Kurtosis value was -.180 (see Appendix J, Table J89). Based on the skewness & kurtosis and visual inspection, the data were normally distributed (see Appendix I, Figure I64-65). Table 94 shows descriptive statistics of extraneous cognitive load scores.

Table 94. Descriptive Statistics for Students' Extraneous Cognitive Load Scores

	Groups	Mean	Std. Dev.	n
Extraneous	Cell	2.216	2.117	380
Cognitive Load	Lightning Formation	2.078	1.926	380

A Paired-samples t-test showed that the extraneous cognitive load score was almost the same when using the educational material for cell subject matter (M = 2.216, SD = 2.117) as opposed to using the educational material for lightning formation subject matter (M = 2.078, SD = 1.926), not a statistically significant mean increase of .138, 95% Cl [-0.005, .281], t(379) = 1.893, p = .059, d = 0.09.

4.3.3 Research question 3.3

This section will show the results of the lightning formation and cell subject matters for the following question: Is there a significant difference between the germane cognitive load scores of the students using the lightning formation learning material and those using the cell learning material?

One extreme outlier was in the data assessed by inspection of a boxplot (see Appendix I, Figure I66). They were removed from the data. Based on Shapiro-Wilk's test values, the normality ($p < .05$) was violated (see Appendix J, Table J90). The Skewness value was -.166, and the Kurtosis value was -.143 (see Appendix J, Table J91). Based on the skewness & kurtosis and visual inspection, the data were normally distributed (see Appendix I, Figure I67-68). Table 95 shows descriptive statistics of germane cognitive load scores.

Table 95. Descriptive Statistics for Students' Germane Cognitive Load Scores

	Groups	Mean	Std. Dev.	n
Germane Cognitive	Cell	6.024	2.715	382
Load	Lightning Formation	6.190	2.597	382

A Paired-samples t-test revealed that the germane cognitive load score was almost the same when using the educational material for cell subject matter (M = 6.024, SD = 2.715) as opposed to using the educational material for lightning formation subject matter (M = 6.190, SD = 2.597), not a statistically significant mean increase of .166, 95% Cl [-0.413, .099], t(381) = -1.234, p = .218, d = 0.06.

CHAPTER 5

DISCUSSION

In this study, the impact of different learning environments (AR, VR, and Traditional) on university students' cognitive load, motivation, and learning outcomes were investigated, as well as the efficacy of segmenting, pre-training, and modality principles in managing cognitive load and enhancing learning outcomes within these environments. The differences in cognitive load scores between students learning about lightning formation and cell structure were also examined. The findings provide valuable insights into the complex interplay between learning environments, principles of multimedia learning, and subject matter, contributing to the understanding of the cognitive theory of multimedia learning and its practical applications in educational settings. The discussion delved into the implications of our results concerning the research questions and explored their significance for theoretical development and practical implementation in instructional design and education.

5.1 Influence of learning environments on learning outcomes

The results offer valuable insights into the influence of different learning environments (AR, VR, and Traditional) on university students' retention and transfer test scores across two subject matters: lightning formation and cell structure. While significant differences were found in retention test scores for the cell structure subject matter, no significant differences were observed in retention scores for the lightning formation and transfer test scores for both subject cases.

Regarding the lightning formation subject matter, the findings showed no significant differences in retention and transfer test scores across learning environments. This aligns with previous research indicating that the selection of a learning environment may not have a considerable influence on students' learning outcomes (Allcoat & von Mühlenen, 2018; Babu, Krishna, Unnikrishnan, & Bhavani, 2018; Greenwald, Corning, Funk, & Maes, 2018; Harrington et al., 2018; Makransky et al., 2019; Meyer et al., 2019; Moreno & Mayer, 2002; Sung & Mayer, 2013). For instance, a study by Makransky et al. (2019) reported that students in a VR learning environment did not exhibit higher learning outcomes than those in a traditional setting when learning about laboratory safety. Similarly, Gopalan et al. (2015) found no significant differences in learning performance between AR and traditional learning environments for science education.

In contrast, for the cell structure subject matter, the AR learning environment significantly enhanced retention test scores compared to the traditional learning environment. This finding aligns with studies that have reported positive effects of AR-based learning on student outcomes in other subject areas, such as marine education (Lu & Liu, 2015), electromagnetism (Ibáñez et al., 2014), lab techniques (Akçayır et al., 2016) and solid geometry (Lin et al., 2015). However, no significant differences were observed in transfer test scores for cell structure across different learning environments, as shown in the research by Cai et al. (2017), Chen and Wang (2018), and Meyer et al. (2019). This discrepancy between retention and transfer performance for the cell structure subject matter underscores the complexity of the relationship between learning environments and learning outcomes (Meyer et al., 2019; Harrington et al., 2018).

One potential reason for the differing outcomes between the two subjects could be the varying complexity levels inherent in each topic since the cell structure subject matter is typically considered more intricate than lightning formation. The disparity in complexity might have played a role in the efficacy of the learning environments in fostering learning outcomes (Kalyuga, 2007; Moreno & Mayer, 2007). For instance, the complexity of cell structure could be better understood through augmented reality's engaging and dynamic aspects. On the other hand, grasping the concept of the lightning formation may be more straightforward for students, resulting in similar educational results regardless of the learning setting. Moreno and Mayer (2007) also highlighted that the effectiveness of interactive multimodal learning environments could be influenced by the complexity of the learning material and learners' prior knowledge. Furthermore, studies in AR (Wu et al., 2018; Ibáñez et al., 2014) and VR (Makransky et al., 2019; Parong & Mayer, 2018) highlight the significance of considering the interplay between subject complexity and learning environment design when assessing the impact of technology-enhanced learning.

Another possible explanation is that the instructional design of the learning materials and activities contributed to the observed differences in learning outcomes (Clark & Mayer, 2016; Sweller, Ayres, Kalyuga, & Chandler, 2003; Mayer, 2009). Effective instructional design is crucial in maximizing the potential benefits of immersive learning environments (Cai et al., 2017; Hwang et al., 2016; Mikropoulos & Natsis, 2011). If the instructional design for the cell structure subject matter was particularly well-suited to the AR environment, it might explain the improved retention scores observed in the study.

Students' familiarity with the technology used in different learning environments could also influence their learning outcomes. Those accustomed to traditional learning approaches might initially find adapting to new technologies like VR and AR challenging, potentially leading to varied outcomes across diverse subjects based on individual experiences and comfort levels (Wu et al., 2018; Tarng, Lin, Lin, & Ou, 2016). For instance, Ray and Deb (2016) discovered that students' performance was inferior to those taught through traditional techniques in the beginning stages of immersive virtual reality (I-VR) learning sessions. However, once they grew accustomed to the technology, their learning exceeded that of the control group. Likewise, research that provided extended access to I-VR (Akbulut, Catal, & Yıldız, 2018; Alhalabi, 2016; Molina-Carmona, Pertegal-Felices, Jimeno-Morenilla, & Mora-Mora, 2018) typically revealed its superiority over non-immersive or conventional methods. Concerns about comparing different media, initially brought up by Clark (1983) and later revisited by Parong and Mayer (2018), emphasize the need to consider the comfort level with I-VR technology when directly comparing it to alternative teaching methods. The novelty of head-mounted displays (HMDs) and I-VR might temporarily impede learning results and classroom implementation.

5.2 Influence of learning environments on cognitive load

The intrinsic and germane cognitive load scores did not show statistically significant differences between the different learning environments for the lightning formation subject matter. This indicates that neither the AR nor the VR environments significantly affected the inherent complexity of the subject matter or the students' more profound understanding and processing of the topic. However, the AR group had a higher extraneous cognitive load than the traditional group, suggesting that the AR environment might have introduced additional elements or distractions that increased cognitive load unrelated to the learning task. This could be due to the novelty of the AR technology, which might have caused students to allocate more cognitive resources to navigating the AR environment rather than focusing on the subject matter (Allcoat & von Mühlenen, 2018; Dunleavy, Dede, & Mitchell, 2009; Parong & Mayer, 2018). Additionally, the design of the specific learning materials in the AR environment might not have been optimized for minimizing extraneous cognitive load, which could have contributed to this group's higher extraneous cognitive load (Ibáñez & Delgado-Kloos, 2018; Kavanagh et al., 2017).

Regarding the cell structure subject matter, the intrinsic cognitive load was not significantly different between different learning environments, indicating that the complexity of the subject matter remained consistent across the different environments. In this case, the AR group had a higher extraneous cognitive load than the VR group, which might imply that the AR environment introduced additional cognitive demands compared to the VR environment. This could be due to the overlay of digital information on the natural world in the AR environment, which might have made it more challenging for students to focus on the subject matter (Ibáñez & Delgado-Kloos, 2018; Radu, 2014). Furthermore, integrating AR elements might have required students to switch attention between the natural world and the augmented content, potentially increasing extraneous cognitive load (Dunleavy & Dede, 2014; Kavanagh et al., 2017).

The germane cognitive load scores for the cell subject matter significantly differed between the learning environments. The VR and AR groups had significantly higher mean germane cognitive load scores than the traditional groups. Since higher germane cognitive load scores indicate a decrease in the overall cognitive load, the result suggests that both VR and AR environments were more effective in facilitating deeper understanding and processing of the cell subject matter than the traditional environment. The immersive nature of VR and the interactive elements of AR might have supported the students' engagement and allowed them to make better connections with the content, leading to improved germane cognitive load (Dunleavy et al., 2009; Ibáñez & Delgado-Kloos, 2018; Makransky et al., 2019). For instance, the 3D visualization in VR and AR environments could have provided students with a more intuitive understanding of complex cell structures, enhancing their comprehension of the subject matter (Akçayır & Akçayır, 2017; Küçük et al., 2016). Moreover, using interactive simulations in VR and AR environments might have encouraged active learning and fostered advanced cognitive abilities, leading to a rise in germane cognitive load (Dunleavy et al., 2009; Merchant et al., 2014).

5.3 Influence of learning environments on motivation

For lightning formation subject matter, the study found no significant differences in overall motivation, attention, relevance, and confidence scores across the VR, AR and traditional learning environments. However, satisfaction scores were significantly different, with the VR group demonstrating higher satisfaction scores than the traditional group. This result could be attributed to the immersive and interactive nature of VR, which has been shown to enhance learners' satisfaction in previous research (Chang, Sung, Guo, Chang, & Kuo, 2022; Shim & Lee, 2022). VR's ability to present complex phenomena visually engagingly could increase satisfaction (Shim & Lee, 2022).

In contrast, the study found significant differences in overall motivation scores between the different learning environments for cell structure subject matter. As in prior studies, the VR group had higher overall motivation scores than the traditional group (Chang et al., 2022; Liu, Wang, Koszalka, & Wan, 2022). Additionally, the VR group showed significantly higher scores in attention, relevance, and confidence, suggesting that the immersive and interactive features of VR may enhance students' engagement and interest in this subject matter. The VR environment may provide a more effective way of visualizing complex cell structures, which could contribute to students' heightened motivation (Chang et al., 2022; Liu et al., 2022).

The satisfaction scores for cell structure subject matter differed significantly between the learning environments. The VR group had higher satisfaction scores than the traditional and AR groups, while the AR group had higher satisfaction scores than the traditional group, as in the previous research (Akçayır et al., 2016; Cai et al., 2017; Cascales-Martínez, Martínez-Segura, Pérez-López, & Contero, 2016; Bursztyn, Shelton, Walker, & Pederson, 2017; Lu & Liu, 2015). This result implies that VR and AR environments may offer more engaging learning experiences for cell structure subject matter than traditional environments. However, VR is more effective in promoting satisfaction, possibly due to its higher immersion level.

In summary, the findings of this study suggest that the impact of learning environments on student motivation may vary depending on the subject matter. While VR and AR technologies may not significantly affect motivation in certain subjects, such as lightning formation, they can positively impact motivation for other subjects, like cell structure. The immersive and interactive nature of VR and AR may enhance students' attention, relevance, confidence, and satisfaction in some contexts, leading to higher overall motivation (Akçayır et al., 2016; Cai et al., 2017; Cascales-Martínez et al., 2016; Chang et al., 2022; Bursztyn et al., 2017; Shim & Lee, 2022; Liu et al., 2022; Lu & Liu, 2015).

5.4 Modality effect in different learning environments

The modality effect relates to how information is presented to learners through visual or auditory channels and how these presentations impact learning outcomes (Mayer & Moreno, 2003). The modality effect suggests that presenting information in a multimedia format, where visual and auditory information complements each other, can lead to better learning outcomes than single-mode presentations (Mayer, 2009). This study observed the modality effect for the lightning formation subject matter based on retention scores in the AR learning environment but not for the cell structure subject matter in the AR learning environment. This finding implies that the impact of the modality effect may be subject-specific.

The lack of modality effect for cell structure subject matter in the AR learning environment could be due to the inherent complexity of cell structures, making it more challenging for learners to process the information in multiple modalities effectively. Additionally, the nature of AR technology, which typically superimposes digital information in the real world, may need to provide a sufficient level of immersion for learners to benefit from the modality effect in this context fully.

The study also found no modality effect for the lightning formation subject matter based on retention scores in the VR and traditional learning environments as in prior studies (Chen & Yang, 2020; Liu et al., 2018; Oberfoell & Correia, 2016; Schweppe & Rummer, 2016; Wang et al., 2018). This result could be attributed to the fact that VR environments, while immersive, may not always provide an optimal combination of visual and auditory information for specific subject matters (Liberman & Dubovi, 2023; Moreno & Mayer, 2002). In traditional learning environments, the modality effect may be limited by the lack of interactivity and the static nature of the presented information. On the other hand, many studies also observe the modality effect in different learning environments and topics (Craig et al., 2002; Oliveira et al., 2015).

Moreover, the modality effect was not observed for the lightning formation and cell structure subject matters based on transfer and cognitive load scores in VR, AR, or traditional learning environments. This finding suggests that the modality effect may not consistently influence different learning outcomes and measures. The lack of modality effect on transfer scores could be because integrating multiple modalities may only sometimes facilitate the application of learned knowledge to new situations (Cheon et al., 2014; Chen & Yen, 2021; Mayer et al., 2019). Similarly, the absence of a modality effect on cognitive load scores indicates that using multiple modalities may not consistently reduce the cognitive burden associated with processing complex information (Chen & Yen, 2021; Fiorella et al., 2012; Greenberg et al., 2021).

The instructional design of the educational materials played a role in the absence of the modality effect. Previous studies suggest that the effectiveness of different modalities in learning environments can be influenced by how the materials are designed and presented (Mayer, 2005; Sweller et al., 2011). The instructional design in the study could have effectively leveraged the affordances of the different modalities, or else the modality effect might not have been observed. Moreover, the lack of the modality effect could be due to the individual differences among learners. Students' prior knowledge, academic achievement, learning preferences, and cognitive abilities can influence the effectiveness of different modalities in learning environments (Kalyuga, 2007; Lin et al., 2015; Mayer, 2009).

In conclusion, this study's findings reveal that the modality effect may not be universally applicable across different subject matters, learning environments, and learning outcomes. The presence or absence of the modality effect may depend on various factors, such as the complexity of the subject matter, the level of immersion provided by the learning environment, instructional design, individual differences, and the measures used to assess learning outcomes. Further research is needed to identify the specific conditions under which the modality effect can be most effectively leveraged to enhance learning.

5.5 Segmenting effect in different learning environments

The segmenting effect refers to the improved learning outcomes when instructional material is broken down into smaller, more manageable segments, allowing learners to process information at their own pace. In this study, the segmenting effect was observed for the lightning formation and cell structure subject matters based on retention test scores regardless of the learning environment as in the prior research (Ibrahim, Antonenko, Greenwood, & Wheeler, 2012; Kuehl, Eitel, Damnik, Koerndle, 2014). However, there are some contradictions in the literature, as seen in the current study, where the segmenting effect was not observed for the lightning formation and cell structure subjects, according to transfer test results and cognitive load scores, regardless of the learning environment (Chen & Yang, 2020; Chen & Yen, 2021; Soicher & Becker-Blease, 2020).

The segmenting effect for retention test scores in both subject matters may be attributed to the fact that breaking down complex information into smaller parts can help learners focus on individual elements and build their understanding incrementally (Mayer & Chandler, 2001; Mayer & Moreno, 2003). This process may facilitate better retention, as learners can dedicate more cognitive resources to understanding and remembering each segment (Mayer & Pilegard, 2014).

The lack of segmenting effect for transfer test scores suggests that dividing instructional material into smaller parts may only sometimes enhance the application of learned knowledge to new situations (Chen & Yang, 2020; Chen & Yen, 2021; Soicher & Becker-Blease, 2020). This result may be because segmenting might hinder the integration of different pieces of information, which is necessary to transfer knowledge effectively (Izmirli & Kurt, 2016). Additionally, learners might struggle to see the connections between segmented pieces of information, making it challenging to apply their understanding to novel contexts.

The absence of segmenting effect on cognitive load scores implies that breaking instructional material into smaller segments does not consistently reduce the cognitive burden associated with processing complex information. This finding could be explained by the fact that segmenting may sometimes increase extraneous cognitive load, as learners need to switch their attention between different segments and mentally organize the information (Chen & Yen, 2021; Ibrahim et al., 2012; Kuehl et al., 2014; Mayer et al., 2019).

In the virtual reality and traditional learning environments, the segmenting effect was not observed for the lightning formation and cell structure subject matters based on retention, transfer, and cognitive load scores. The absence of segmenting effect in these learning environments could be attributed to the specific characteristics of the instructional materials or the learning environments themselves, which may need to be more conducive to effective segmentation (Mayer & Pilegard, 2014).

In the augmented reality learning environment, the segmenting effect was observed for the lightning formation based on retention test scores but not observed for the cell structure subject matter. This result suggests that the effectiveness of segmenting may be subject-specific in the AR learning environment. The segmenting effect may be more beneficial for some topics, such as lightning formation, where smaller segments may help learners better understand and retain complex information. In contrast, segmenting may be less effective for more complex subject matters, such as cell structure, where integrating different pieces of information is crucial for understanding.

In conclusion, the presence or absence of the segmenting effect across different subject matters, learning environments, and learning outcomes may depend on various factors, such as the complexity of the subject matter, the characteristics of the instructional material, and the learning environment. Further research is needed to understand better the conditions under which segmenting can be most effectively applied to enhance learning outcomes.

5.6 Pre-training effect in different learning environments

The present study observed the pre-training effect for both the lightning formation and cell structure subject matters based on retention test scores when considering data from all learning environments (VR, AR, and traditional). In addition, the pre-training effect was observed for the cell structure subject matter based on transfer test scores when considering data from all learning environments but not for the lightning formation subject. However, the pre-training effect was not observed for the lightning formation and cell structure subject matters based on cognitive load scores regardless of the learning environments.

The study's findings are consistent with previous research demonstrating the benefits of pre-training for enhancing learning outcomes (Mayer et al., 2002; Clarke et al., 2005; Eitel et al., 2013). However, it is essential to note that the effectiveness of pre-training can depend on factors such as the type of pre-training, learners' prior knowledge level, and the media and method used (Mayer et al., 2002; Mayer & Pilegard, 2014; Meyer et al., 2019).

When analyzing the data for each learning environment separately, the pre-training effect was observed for the lightning formation based on retention test scores in both the VR and AR learning environments but not for the cell structure subject matter. Moreover, the pre-training effect was not observed for the lightning formation and cell structure subject matters based on retention scores in the traditional learning environment. This finding might be related to the specific affordances of VR and AR technologies, such as the potential for immersive, interactive, and engaging experiences, which could differentially affect the impact of pre-training on various subjects (Meyer et al., 2019; Petersen et al., 2020).

Regarding the transfer test scores, the pre-training effect was observed for the cell structure subject matter in the VR learning environment but not for the cell structure subject matter in the AR learning environment. Furthermore, based on transfer test scores in the AR and traditional learning environments, the pre-training effect was not observed for the lightning formation and cell structure. These results align with previous research suggesting that the effectiveness of pre-training may vary depending on the subject matter and the complexity of the learning tasks (Kester et al., 2006; Çeken & Taşkın, 2022).

Lastly, based on cognitive load scores in the VR, AR, and traditional learning environments, the pre-training effect was not observed for the subject of lightning formation and cell structure. This finding could be due to several factors, such as the instructional design, learning tasks, or learner characteristics not effectively promoting a reduction in cognitive load through pre-training (Kalyuga, 2007). The measures used to assess cognitive load were not sensitive enough to detect the pre-training effect in this study.

In conclusion, the present study's findings provide valuable insights into the pre-training effect in VR, AR, and traditional learning environments. The results highlight the importance of considering the specific subject matter, learning tasks, learner characteristics, and learning environment when designing and implementing pre-training interventions to maximize their effectiveness. Further research is needed to explore the pre-training effect's underlying mechanisms and boundary conditions in different contexts and learning.

5.7 Theoretical implications

This study contributes to the existing body of literature on the influence of different learning environments (AR, VR, and Traditional), principles of multimedia learning (segmenting, pre-training, and modality), and subject matter complexity (lightning formation and cell structure) on university students' cognitive load, motivation, and learning outcomes. The findings provide valuable insights into the complex interplay between these factors, contributing to understanding the cognitive theory of multimedia learning and its practical applications in educational settings.

Influence of learning environments on learning outcomes: The study's findings highlight that the impact of learning environments on learning outcomes may vary depending on the subject matter. While no significant differences were observed in learning outcomes for the lightning formation subject matter, the AR environment significantly enhanced retention test scores for the cell structure subject matter. This suggests that the complexity of the subject matter and instructional design could influence the efficacy of different learning environments. It underscores the importance of exploring the interplay between subject complexity, learning environment design, and instructional design when assessing the impact of technology-enhanced learning.

Influence of learning environments on cognitive load: The study reveals that the effect of learning environments on cognitive load may depend on the subject matter and the specific learning environment. The findings indicate that AR and VR environments can affect intrinsic, extraneous, and germane cognitive load differently across different subjects. This highlights the importance of considering the factors that contribute to these differences, such as the design of learning materials, the integration of digital information, and the degree of immersion and interaction provided by the learning environments.

Influence of learning environments on motivation: The study's findings emphasize that the impact of learning environments on student motivation may vary depending on the subject matter. While VR and AR technologies may not significantly affect motivation in certain subjects, such as lightning formation, they can positively impact motivation for other subjects, like cell structure. The immersive and interactive nature of VR and AR may enhance students' attention, relevance, confidence, and satisfaction in some contexts, leading to higher overall motivation. This finding emphasizes the need to examine the contributing factors behind these differences and to determine the most efficient methods for integrating these technologies into diverse learning scenarios.

Modality effect: The results of this study suggest that the modality effect may not be universally applicable across different subject matters, learning environments, and learning outcomes. The presence or absence of the modality effect may depend on various factors, such as the complexity of the subject matter, the level of immersion provided by the learning environment, instructional design, individual differences, and the measures used to assess learning outcomes. This finding accentuates the importance of pinpointing the conditions that optimize the modality effect for improved learning outcomes.

Segmenting effect: The study reveals that the segmenting effect may vary depending on the complexity of the subject matter, the characteristics of the instructional material, and the learning environment. The presence or absence of the segmenting effect across different subject matters, learning environments, and learning outcomes may depend on these factors. This finding emphasizes the significance of understanding the optimal conditions for applying segmentation to boost learning outcomes. Moreover, the study implies that the segmenting effect may benefit certain content or pedagogical approaches. This finding could encourage the development of theories that account for content-specific factors when explaining the segmenting effect and its impact on learning outcomes.

Pre-training effect: The study's findings provide valuable insights into the pre-training effect in VR, AR, and traditional learning environments. The results highlight the importance of considering the specific subject matter, learning tasks, learner characteristics, and learning environment when designing and implementing pre-training interventions to maximize their effectiveness. This underscores the need for personalized and context-specific approaches to enhance the effectiveness of pre-training strategies. The findings also suggest that the pre-training effect may not be universally effective in reducing cognitive load, which warrants further exploration of the pre-training effect's underlying mechanisms and boundary conditions in different contexts and learning settings. This observation suggests that the efficiency of pre-training interventions may depend on factors such as the intricacy of learning materials, learners' prior knowledge, cognitive abilities, and the specific requirements of learning tasks.

In summary, the study's findings have important theoretical implications for understanding the complex interplay between learning environments, principles of multimedia learning, and subject matter in shaping students' cognitive load, motivation, and learning outcomes. The results emphasize the need for a more nuanced approach when applying these principles in instructional design and educational research, considering the complexity of the subject matter, the characteristics of instructional materials, the learning environment, and individual differences among learners. Further research is needed to understand better the conditions under which these principles can be most effectively applied to enhance learning outcomes and to develop tailored instructional strategies for different contexts and learners.

5.8 Practical implications

The study's practical implications can be divided into several categories, including instructional design, technology integration, and professional development for educators. Key takeaways that may be useful for future educational practices and technology implementation include the followings:

The study's findings highlight the importance of carefully designing instructional materials and activities to maximize the potential benefits of immersive learning environments (AR, VR, and traditional). Different subject matters may require tailored instructional design approaches to capitalize on the unique strengths of each learning environment. The study also underscores the need to consider the complexity of the subject matter when designing learning experiences. More complex subjects may benefit from the enhanced visualization and interactivity provided by AR and VR environments. However, less complex subjects may achieve similar outcomes across various learning environments, indicating that careful consideration should be given to the cost-benefit analysis of implementing immersive technologies for specific topics.

Based on retention test scores, the segmenting effect observed for both subject matters suggests that breaking complex information into smaller segments can enhance retention. Educators and instructional designers should consider segmenting complex content to facilitate understanding and retention while examining the potential limitations of the segmenting effect on transfer and cognitive load. As the modality effect was observed in the AR learning environment for the lightning formation subject matter but not for the cell structure, educators and instructional designers should consider the specific subject matter and the characteristics of the learning environment when deciding whether to utilize different modalities to optimize learning outcomes.

The results suggest that the effectiveness of AR and VR environments may depend on the subject matter being taught. Therefore, educators and instructional designers should carefully evaluate the potential benefits of these technologies for specific topics and learning objectives. When integrating AR and VR technologies, optimizing the learning environment to minimize extraneous cognitive load is crucial. This may involve reducing distractions, simplifying interfaces, and providing clear instructions for navigation and interaction within the environment. It is also essential to consider the potential impact of students' familiarity and comfort level with the technology when integrating AR and VR into the learning process. Offering opportunities for students to familiarize themselves with the technology before engaging in learning activities may help improve their overall learning experience and outcomes.

The study highlights the importance of providing professional development opportunities for educators to understand the unique strengths and limitations of different learning environments, including AR, VR, and traditional settings. This knowledge will enable them to make informed decisions about incorporating these technologies into their instructional practices. Professional development programs should also guide best practices for designing learning experiences that leverage the unique capabilities of AR and VR environments, as well as strategies for minimizing extraneous cognitive load and maximizing germane cognitive load. Moreover, educators should be encouraged to collaborate and share their experiences with AR and VR integration, discussing successes and challenges and learning from one another's expertise to improve the overall effectiveness of these technologies in educational settings.

The pre-training effect observed for both subject matters based on retention and transfer test scores highlights the importance of familiarizing students with key concepts and terminology before diving into complex multimedia instruction. Educators should incorporate pre-training activities into their lesson plans to enhance learning outcomes while considering the potential limitations of pre-training on cognitive load management.

In conclusion, this study offers valuable insights for educators, instructional designers, and developers of learning materials to enhance the effectiveness of multimedia learning experiences. By considering the modality, segmenting, pre-training effects, and cognitive load experienced by learners across different subject matters and learning environments, practitioners can make informed decisions that optimize learning outcomes and promote more profound understanding.

Learning outcomes and assessment: The study focused on specific learning outcomes, such as retention, transfer, cognitive load, and motivation. Further research should consider other learning outcomes, such as problem-solving skills, collaboration, and critical thinking, to comprehensively understand different learning environments' potential benefits and drawbacks.

Mediating factors: The study might not have explored all potential mediating factors that could influence the effectiveness of the learning environments, such as the role of instructional support, feedback, and scaffolding. Future research should examine the impact of these factors on the effectiveness of VR, AR, and traditional learning environments to inform the development of more effective instructional designs.

By addressing these limitations and considering the suggested areas for further research, future studies can contribute to a more nuanced and comprehensive understanding of the relative strengths and weaknesses of VR, AR, and traditional learning environments in different educational contexts.

CHAPTER 6

CONCLUSION

This study explored the influence of various learning environments (AR, VR, and Traditional) on university students' cognitive load, motivation, and learning outcomes. In addition, it rigorously evaluated the effectiveness of segmenting, pre-training, and modality principles in managing cognitive load and enhancing learning outcomes. The research focused on two subject matters: lightning formation and cell structure, providing a rich comparative context.

One of the key findings of this research was the negligible difference in retention scores across the learning environments for lightning formation. However, AR outperformed the traditional learning environment for cell structure, underscoring the potential of AR in teaching complex biological concepts. The study demonstrated that the subject matter significantly influenced the effectiveness of the learning environment, suggesting a nuanced approach to choosing the appropriate learning environment based on the complexity and nature of the subject. In the context of transfer scores, no significant differences were found across different learning environments and subject matters. This highlights the complex interaction between learning environments and the nature of the learning task, with the effectiveness of AR and VR likely dependent on the specific requirements of the task.

An interesting dynamic concerning cognitive load score was also discovered. The data suggested that these scores can fluctuate significantly, depending on both the learning environment and subject matter. This underscores the necessity to consider cognitive load management when designing learning environments, particularly when dealing with complex subjects that might naturally impose a higher intrinsic cognitive load.

Motivation scores provided another layer of complexity. For the lightning formation subject matter, there were no significant differences in motivation scores across the learning environments. However, when the subject matter was switched to the cell structure, the VR group had significantly higher motivation scores. This suggests that the immersive and interactive nature of VR can have a positive impact on student motivation, especially in complex and abstract subjects where visualization and interaction can enhance understanding.

The application of segmenting, pre-training, and modality principles across various learning environments yielded interesting findings. The pre-training group frequently outperformed the others in retention test scores, suggesting that pre-training may provide a solid foundation upon which further learning can be built. This could be due to pre-training enhancing learners' prior knowledge, enabling them to integrate new information more effectively, thereby reducing cognitive load and improving retention. This finding underscores the potential value of incorporating pre-training phases into learning designs, particularly in AR and VR environments, which can often introduce additional cognitive load due to their immersive and interactive nature.

However, the impact of these principles on transfer test scores and cognitive load scores was often insignificant after adjusting for multiple comparisons. This suggests that while these principles have some effectiveness in managing cognitive load, they may not always translate into significant improvements in learning outcomes across different learning environments. For instance, while segmenting or breaking down information into manageable parts might ease the cognitive load, it doesn't necessarily guarantee that learners can transfer this knowledge to different contexts or tasks effectively.

In terms of modality principles, the mixed results observed in this study suggest a complex interplay between the type of learning environment and the modality of information presentation. While these principles propose that using different sensory modes for presenting information can reduce cognitive load, their effectiveness might depend on a variety of factors, including the nature of the subject matter and the learners' individual preferences and abilities. The variation in the effectiveness of modality principles across the learning environments in this study invites further investigation into how these principles can be optimized for different learning contexts.

These findings hint at the need for a more nuanced understanding of how segmenting, pre-training, and modality principles operate within different learning environments. Recognizing that these principles may interact differently with various factors, such as the nature of the subject matter, the type of learning environment, and individual learner characteristics, could lead to the development of more tailored and effective instructional designs.

Of note is the role of familiarity with technology, as it can significantly influence the effectiveness of AR and VR environments. Students' comfort level with these advanced technologies can impact their cognitive load and motivation, thereby affecting learning outcomes. As technological literacy varies among individuals, understanding the role of technological familiarity in learning environments remains a pertinent area for future investigation. Additionally, integrating new technology into educational settings should consider the learners' familiarity with it to enhance its effectiveness.

The study also acknowledges the significant role of emotions, stress, and uncertainty as contributors to cognitive load and learning outcomes (Choi, van Merriënboer, & Paas, 2014). Traditionally seen as impediments to learning, these elements are now recognized as integral parts of vocational and professional learning contexts, contributing to intrinsic cognitive load. This perspective suggests that managing these states is a crucial part of students' competency development (Plass & Kaplan, 2016). Future research could explore how these elements interact with the learning environments and principles examined in this study.

When comparing the two subject matters, cell structure presented a higher intrinsic cognitive load score than lightning formation, with extraneous and germane cognitive load scores not showing significant differences between the subjects. This aligns with the perspective that certain subjects might contribute more to the intrinsic cognitive load, especially when they involve elements of stress, emotion, and uncertainty that are intrinsic to the subject matter.

In conclusion, this study significantly contributes to the field by providing a comprehensive understanding of the complex relationship between learning environments, multimedia learning principles, and subject matters. It enhances our understanding of the cognitive theory of multimedia learning and its potential applications in various educational settings. The insights gained from this study provide a solid foundation for future research in this area, and they carry significant implications for instructional design and the integration of emerging technologies in education.

While the current study offers valuable insights, it is essential to acknowledge its limitations. These limitations provide opportunities for refining the study design in future research, ensuring a more comprehensive understanding of the relationship between learning environments and educational outcomes. The following limitations should be considered when interpreting the results of this study:

Sample size and generalizability: The study may have a limited sample size, which could affect the generalizability of the results. Future research should involve more extensive and diverse samples and different educational contexts to enhance the external validity and applicability of the findings.

Subject matters: The study focused on two specific subject matters, lightning formation and cell structure. Beyond the scope of the current study, future research could also delve into the investigation of the effectiveness of AR, VR, and traditional learning environments in other academic domains. For instance, exploring the impact of these learning environments in subjects with a high degree of abstract concepts, such as theoretical physics or philosophy, might yield different results. Similarly, subjects that inherently involve a lot of spatial reasoning, such as geology or architecture, might particularly benefit from the immersive nature of AR and VR technologies. Furthermore, it could be valuable to examine the role of these learning environments in practical, hands-on disciplines, such as medical studies or engineering. In these fields, AR and VR could potentially offer unique benefits by providing safe and controlled environments for simulation-based learning, thereby potentially enhancing the students' practical skills and preparedness for real-world situations.

Duration and long-term effects: The study may have been limited in its duration, which could impact the assessment of long-term learning outcomes. Further research should investigate the long-term retention, transfer, and motivational effects of using VR, AR, and traditional learning environments over extended periods.

Design and implementation: The design and implementation of the VR, AR, and traditional learning environments in the study might only capture some possible variations and affordances of these technologies. Future research should examine different design and implementation strategies and explore the potential interactions between learning environment features and instructional strategies, such as segmenting and modality effects.

Learner characteristics: The study might not have considered individual differences in learners' cognitive and motivational profiles, which could influence the effectiveness of the learning environments. Further research should investigate how prior knowledge, learning preferences, cognitive abilities, and motivation might affect the effectiveness of VR, AR, and traditional learning environments.

APPENDIX A

LIGHTNING FORMATION PRIOR KNOWLEDGE QUESTIONS

Please place a checkmark next to the items that apply to you.

1. I regularly read the weather maps in the newspaper. [ ]

2. I know what a cold front is. [ ]

3. I can distinguish between cumulous and nimbus clouds. [ ]

4. I know what a low-pressure system is. [ ]

5. I can explain what makes the wind blow. [ ]

6. I know what this symbol means:  [ ]

7. I know what this symbol means:  [ ]

Please grade your knowledge of meteorology (weather) on a 5-point scale ranging from *very little* (1) to *very much* (5):

APPENDIX B

CELL STRUCTURE PRIOR KNOWLEDGE QUESTIONS

Please place a checkmark next to the items that apply to you.

1. I know what microfilaments are. [ ]

2. I know what cytoplasm is. [ ]

3. I know what ribosomes are. [ ]

4. I know what ATP regeneration is. [ ]

5. I know what the following picture is. [ ]

6. I know what the following picture is. [ ]

7. I know what the following picture is. [ ]

8. I know what the following picture is. [ ]

Please grade your knowledge of cell structure on a 7-point scale ranging from *very little* (1) to *very much* (7).

APPENDIX C

RUBRICS FOR RETENTION TEST

Table C1. Rubric for Retention Test About Lightning Formation

Cümleler (*Statements*)	Tam Puan (*Full Point*)	Yarım puan (*Partial Point*)
Soğuk nemli hava, daha sıcak bir yüzey üzerinde hareket eder ve ısınır. (*Cool moist air moves over a warmer surface and heats up*)	2	1
Dünya yüzeyine yakın ısınmış nemli hava hızla yükselir. (*Warm moist air near the Earth's surface rises rapidly*)	2	1
Bu yukarı doğru hareket ile hava soğudukça, su buharı su damlacıklarına yoğunlaşır ve bir bulut oluşturur. (*As the air cools with this upward movement, the water vapor condenses into water droplets and forms a cloud*)	2	1
Bulutun tepesi donma seviyesinin üzerine uzanır, bu nedenle bulutun üst kısmı küçük buz kristallerinden oluşur. (*The top of the cloud lies above the freezing level, so the top of the cloud is made up of small ice crystals*)	2	1
Sonunda, su damlacıkları ve buz kristalleri, yükselmelerle askıya alınamayacak kadar büyük hale gelir. (*Eventually, the water droplets and ice crystals become too large to be suspended by uplifts*)	2	1
Yağmur damlaları ve buz kristalleri bulutun içine düştükçe, buluttaki havanın bir kısmını aşağı doğru sürükleyerek aşağı hava akımları oluştururlar. (*As raindrops and ice crystals fall into the cloud, they create downdrafts, dragging some of the air in the cloud downwards*)	2	1
Düşüşler yere çarptığında, her yöne yayılırlar ve insanların yağmur başlamadan hemen önce hissettiği soğuk rüzgarların esintilerini üretirler. (*When the drops hit the ground, they spread out in all directions and produce gusts of cold winds that people feel just before the rain starts*)	2	1
Bulutun içinde yükselen ve düşen hava akımları elektrik yüklerinin oluşmasına neden olur. (*Air currents rising and falling within the cloud cause electric charges to be created*)	2	1
Elektrik yükü, bulutun içindeki yükselen su damlacıklarının, daha ağır olan ve düşen buz parçalarına çarpmasından kaynaklanır. (*The electric charge is caused by the rising water droplets in the cloud hitting heavier and falling pieces of ice*)	2	1
Negatif yüklü parçacıklar bulutun alt kısmında toplanır ve pozitif yüklü parçacıkların çoğu tepe noktasına doğru yükselir. (*Negatively charged particles collect in the lower part of the cloud, and most positively charged particles rise towards the apex*)	2	1
Yıldırımın temelini oluşturan ilk yıldırım filizi "Adım lider" olarak adlandırılır. (*The first lightning shoot, which forms the basis of lightning, is called "Step leader"*)	2	1

Negatif yüklerin adım lideri, bir dizi adımda aşağı doğru hareket eder ve neredeyse yere yakındır. (The step leader of negative charges moves down in a series of steps and is almost close to the ground)	2	1
Pozitif yüklü bir adım lider, ağaçlar ve binalar gibi nesnelerden yukarı yönde hareket eder. (*A positively charged step leader moves upward from objects such as trees and buildings*)	2	1
İki lider genellikle yerden yaklaşık 50 metre yüksekte buluşur. (*The two leaders usually meet at about 50 meters above the ground*)	2	1
Negatif yüklü parçacıklar daha sonra liderler tarafından oluşturulan yol boyunca buluttan yere doğru ilerler. Çok parlak değildir ve genellikle çok sayıda dalı vardır. (*Negatively charged particles then travel from the cloud to the ground along the path created by the leaders. It is not very bright and usually has many branches*)	2	1
Adım lider yere yaklaştıkça, zıt bir yük içerir, bu nedenle yerden pozitif yüklü parçacıklar aynı yol boyunca yukarı doğru ilerler. (*As the step leader approaches the ground, it contains an opposite charge, so positively charged particles from the ground travel upward along the same path*)	2	1
Akımın bu yukarı doğru hareketi, geri dönüş darbesidir. (*This upward movement of current is the return pulse*)	2	1
Geridönüş darbasi buluta yaklaşık 50 mikrosaniye içinde ulaşır. (*The return pulse reaches the cloud in about 50 microseconds*)	2	1
Geri dönüş darbesi, insanların bir şimşek çakmasıyla fark ettiği parlak ışığı üretir. (*The return pulse produces the bright light that people notice with a flash of lightning*)	2	1
Akım o kadar hızlı hareket eder ki yukarı doğru hareketi algılanamaz. (*The current moves so fast that its upward movement cannot be detected*)	2	1
Şimşek çakması genellikle yüz milyonlarca voltluk bir elektrik potansiyelinden oluşur. (*A lightning flash usually consists of an electrical potential of hundreds of millions of volts*)	2	1
Yıldırım kanalı boyunca hava kısa bir süre çok yüksek bir sıcaklığa kadar ısıtılır. (*The air along the lightning channel is briefly heated to a very high temperature*)	2	1
Böylesine yoğun bir ısınma, havanın patlayarak genişlemesine ve gök gürültüsü dediğimiz bir ses dalgası üretmesine neden olur. (*Such intense heating causes the air to explode and expand, producing a sound wave we call thunder*)	2	1
Pozitif yük, negatif yük, adım lider, buz kristali, su damlacıkları, donma noktası, geri dönüş darbesi, gök gürültüsü. (*Positive charge, negative charge, step leader, ice crystal, water droplets, freezing point, return pulse, thunder*)	4	from three to one

Table C2. Rubric for Retention Test About Cell

Cümleler (*Statements*)	Tam Puan (*Full Point*)	Yarım puan (*Partial Point*)
Hücre, bir organizmanın en küçük canlı birimidir. (*The cell is the smallest living unit of an organism*)	2	1
Hücreler üç ortak özelliğe sahiptir. (*Cells have three common features*)	2	1
Bütün hücreler hücre içini çevreden izole eden hücre zarına, jel benzeri stoplazmaya ve hücrenin kalıtsal materyali olan DNA'ya sahiptir. (*All cells have a cell membrane that insulates the inside of the cell from the environment, a gel-like cytoplasm, and DNA, the hereditary material of the cell*)	2	1
Genel olarak hücreler iki sınıfa ayrılır (Ökaryotik & Prokaryot) (*In general, cells are divided into two classes (Eukaryotic & Prokaryotic)*)	2	1
Ökaryotik hücreler, çekirdek ve diğer özel kısımları içeren organellere sahiptir. (*Eukaryotic cells have organelles that contain the nucleus and other specialized parts*)	2	1
Ökaryotik hücreler, bitki ve hayvanlarda bulunan ve çok daha gelişmiş hücrelerdir. (*Eukaryotic cells are much more advanced cells found in plants and animals*)	2	1
Prokaryot hücrelerde zarla çevrili organel ve çekirdekleri yoktur. (*Prokaryotic cells do not have membrane-enclosed organelles and nuclei*)	2	1
Prokaryot hücreler genetik materyale sahiptirler ancak bu materyaller çekirdek gibi zarlı bir yapı ile oluşmamıştır. (*Prokaryotic cells have genetic material, but these materials were not formed with a membranous structure such as a nucleus*)	2	1
Prokaryotik hücreler tek hücreli organizmalardır. (Tıpkı bakteriler gibi) (*Prokaryotic cells are unicellular organisms. (Just like bacteria)*)	2	1
Organeller, küçük organ anlamına gelir ve belirli işleri yerine getirmek için özelleşmiş hücre kısımlarıdır. (*Organelles means small organs and are specialized cell parts to perform certain jobs*)	2	1
Çekirdek, DNA gibi kalıtsal materyalleri içerir. (*The nucleus contains hereditary materials such as DNA*)	2	1
DNA, hücrenin neyi ve nasıl yapacağını söyler. (*DNA tells the cell what to do and how*)	2	1
Kromotin, çekirdek zarının altında DNA nın depolanmış formudur. (*Chromatin is the stored form of DNA under the nuclear membrane*)	2	1
DNA bölünmeye hazır olduğu zaman DNA, kromozom olarak bilinen yapıları oluşturur. (*When DNA is ready to divide, the DNA forms structures known as chromosomes*)	2	1
Çekirdek ayrıca ribozomların ve RNA'ların üretildiği yer olan nükleolus'ları, yani çekirdekçikleri içerir. (*The nucleus also contains the nucleoluses, or nucleoli, where ribosomes and RNAs are produced*)	2	1
Ribozomlar, çekirdekten ayrıldıktan sonra protein yapımı ve sentezlenmesi gibi önemli işlere dahil olacaktır. (*After leaving the nucleus, ribosomes will be involved in important jobs such as protein making and synthesis*)	2	1

Çekirdek hariç ribozom ve geriye kalan organeller, jel benzeri madde olan sitoplazma içerisinde yüzer. (*Except for the nucleus, the ribosome and the remaining organelles float in the cytoplasm, which is a gel-like substance*)	2	1
Ribozomlar sitoplazma içerisinde serbestçe dolaşabilir. Aynı zamanda Endoplazmik Retikulum'a tutunabilirler. (*Ribosomes can circulate freely in the cytoplasm. They can also attach to the Endoplasmic Reticulum*)	2	1
Endoplazmik Retikulum'un iki çeşidi vardır. Ribozomun bağlı olduğu "Granürlü Endoplazmik Retikulum" (rER) ve Ribozom'un bağlı olmadığı "Granürsüz Endoplazmik Retikulum" (sER). (*There are two types of Endoplasmic Reticulum. "Granular Endoplasmic Reticulum" (rER) to which the Ribosome is attached, and "Non-grained Endoplasmic Reticulum" (sER) to which the Ribosome is not attached*)	2	1
Endoplazmik Retikulum, Ribozom'lar tarafından sentezlenen proteinler için zarla kaplı bir geçit yapısıdır. (*The Endoplasmic Reticulum is a membrane-enclosed gateway structure for proteins synthesized by Ribosomes*)	2	1
Protein ve diğer materyaller Endoplazmik Retikulum'den küçük kesecikler içinde Golgi Aygıtı'na taşınır. (*Protein and other materials are transported from the Endoplasmic Reticulum to the Golgi Apparatus in small vesicles*)	2	1
Golgi Aygıtı bu proteinleri uygun şekillerde bükerek veya onlara lipid ya da karbonhidrat gibi materyallerin eklenmesi ile dönüşüm devam eder. (*The transformation continues by the Golgi Apparatus bending these proteins into appropriate shapes or adding materials such as lipids or carbohydrates to them*)	2	1
Lizozom'lar zarar görmüş ya da yıpranmış hücre parçalarını içine alan öğütücülerdir. (*Lysosomes are grinders that engulf damaged or worn-out cell parts*)	2	1
Lizozomlar hücresel atıkları parçalayan enzimler ile doludur. (*Lysosomes are packed with enzymes that break down cellular waste*)	2	1
Kofullar ise tek katlı zarla çevrili içi sıvı dolu keselerdir. (*Vacuoles are fluid-filled sacs surrounded by a single-layered membrane*)	2	1
Kofullar bitki hücrelerinde büyük ve sayısı az, hayvan hücrelerinde ise küçük sayıca fazladır. (*Vacuoles are large and few in number in plant cells, and small in number in animal cells*)	2	1
Mitokondri, hem bitki hem hayvan hücresinde bulunan enerjiden sorumlu organeldir. (*Mitochondria are the organelles responsible for energy found in both plant and animal cells*)	2	1
Mitokondri bütün hücresel aktiviteler için enerji sağlayan ATP moleküllerini üretir. (*Mitochondria produce ATP molecules that provide energy for all cellular activities*)	2	1
Çok fazla enerjiye ihtiyaç duyan hücrelerde Mitokondri sayısı oldukça fazladır. (*The number of mitochondria is quite high in cells that need a lot of energy*)	2	1
Ökaryot hücrelere şeklini veren ve hücre içi organizasyonu sağlayan yapıların tümü hücre iskeleti olarak adlandırılır. (*All the structures that give shape to eukaryotic cells and provide intracellular organization are called cytoskeleton*)	2	1

Hücre iskeleti, proteinlerden yapılmış iplik benzeri mikrofilament'lerin ve ince, içleri boş olan mikrotübül'leri içerir. (*The cytoskeleton consists of thread-like microfilaments made of proteins and thin, hollow microtubules*)	2	1
Bitki hücreleri, enerji için güneş ışığını emen Kloroplast adı verilen organele sahip hücrelerdir. (*Plant cells are cells with organelles called Chloroplasts that absorb sunlight for energy*)	2	1
Kloroplast, fotosentezin gerçekleştiği yerdir. (*The chloroplast is where photosynthesis takes place*)	2	1
Bitkiler klorofil taşıdıkları için yeşil renklidirler. (*Plants are green in color because they contain chlorophyll*)	2	1
Bitki hücrelerinde, hücre zarının dışında onları koruyan, şekil ve destek veren hücre duvarına sahiptir. (*Plant cells have a cell wall outside the cell membrane that protects them and gives them shape and support*)	2	1
Hücre duvarı, cansız, hücre zarına göre daha kalın ve dayanıklı bir yapıya sahip olduğundan bitki hücrelerini dışarıdan gelebilecek mekanik etkilere karşı korur. (*Since the cell wall has a thicker and more durable structure than the non-living cell membrane, it protects the plant cells against mechanical effects that may come from outside*)	2	1
Hücre zarı seçici geçirgen özelliğe, hücre çeperi ise cansız olduğundan tam geçirgen özelliğe sahiptir. (*The cell membrane is selectively permeable, and the cell wall is fully permeable because it is non-living*)	2	1
Ökaryotik, Prokaryot, hücre zarı, stoplazma, çekirdek, DNA, ribozom, endoplazmik retikulum, golgi aygıtı, lizozom, koful, mitokondri, hücre iskeleti, kloroplast, hücre duvarı. (*Eukaryotic, Prokaryotic, cell membrane, cytoplasm, nucleus, DNA, ribosome, endoplasmic reticulum, golgi apparatus, lysosome, vacuole, mitochondria, cytoskeleton, chloroplast, cell wall*)	6	from 5 to 1

APPENDIX D

TRANSFER TEST QUESTIONS

Lightning Formation

1. What could you do to reduce the intensity of lightning?
2. Suppose you see clouds in the sky but no lightning. Why not?
3. What does air temperature have to do with lightning?
4. What causes lightning?

Cell

1. Why does the cell not get flooded with all the various elements surrounding the cell?
2. You are in a lab, looking at a blood sample where an infection has spread more than it should have. What could be the cause of this?
3. Why is the membrane that surrounds the nucleus important?
4. How does the cell keep functioning?

APPENDIX E

RUBRICS FOR TRANSFER TEST

Table E1. Rubric for Transfer Test About Lightning Formation

Question	Acceptable Answers	Points
Yıldırımın şiddetini azaltmak için ne yapabilirsiniz? (*What could you do to reduce the intensity of lightning?*)	1. Dünyanın yüzeyinden pozitif parçacıkları çıkarın. (*Remove positive particles from the earth's surface*) 2. Pozitif parçacıkları bulutun yakınına yerleştirmek. (*Placing positive particles near the cloud*)	2
Diyelim ki; gökyüzünde bulut var ama yıldırım yok. Nedeni ne olabilir? (*Suppose you see clouds in the sky but no lightning. Why not?*)	1. Bulutun tepesi donma seviyesinin üzerinde olmayabilir. (*Top cloud may not be above freezing level*) 2. Hiçbir buz kristali formu yoktur. (*No ice crystals form*)	2
Hava sıcaklığının yıldırım oluşmasıyla alakası nedir? (*What does air temperature have to do with lightning?*)	1. Dünyanın yüzeyi sıcak ve yaklaşan hava soğuk. (*Earth's surface is warm and oncoming air cool*) 2. Bulutun üstü donma seviyesinin üzerinde ve bulutun altı donma seviyesinin altında. (*The top of the cloud is above freezing level and the bottom cloud is below freezing level*)	2
Yıldırımın oluşmasına ne sebep olur? (*What causes lightning?*)	1. Bulut içindeki elektrik yükündeki farklılıklar. (*Differences in electrical charge within the cloud*) 2. Bulut içindeki hava sıcaklığındaki fark. (*Difference in air temperature within the cloud*)	2

Table E2. Rubric for Transfer Test About Cell

Question	Acceptable Answers	Points
Hücre neden kendini çevreleyen tüm çeşitli unsurlarla dolup taşmaz? (*Why does the cell not get flooded with all the various elements surrounding the cell?*)	1. Hücre, hücre zarı denilen bir zarla çevrilidir. (*The cell is surrounded by a membrane called the cell membrane.*) 2. Hücre zarı seçici geçirgendir. (*The cell membrane is selectively permeable*)	2
Diyelim ki laboratuvardasınız, incelediğin hücrenin içinde mitokondri sayısının normalden daha fazla olduğunu gördünüz. Bunun nedeni ne olabilir? (*You are in the lab, and you find that the number of mitochondria is higher than normal. What could be the reason for this?*)	1. Hücre içindeki miktarları o hücrenin enerji ihtiyacı ile doğru orantılıdır. (*The amount of mitochondria in the cell is directly proportional to the energy need of that cell*) 2. Kas ve sinir hücreleri gibi enerji ihtiyacı fazla olan hücredir. (*It is a cell that needs more energy, such as muscle and nerve cells*)	2
Çekirdeği çevreleyen zar neden önemlidir? (*Why is the membrane that surrounds the nucleus important?*)	1.Çekirdeği (DNA'yı) sitoplazmadan ayırır. (korur). (*Separates the nucleus (DNA) from the cytoplasm*) 2. Çekirdek zarında, sitoplazmaya madde alışverişini sağlayan porlar (delikler) vardır. (*There are pores (holes) in the nuclear membrane that allow the exchange of substances to the cytoplasm*)	2
Hücre işleyişini nasıl sürdürür? (*How does the cell keep functioning?*)	1. DNA, hücresel faaliyetlerin yönetimini gerçekleştirir. (*DNA carries out the management of cellular activities*) 2. Organeller (ökaryotik hücrelerde) yaşamsal faaliyetlerin yerine getirilmesine yardımcı olur. (*Organelles (in eukaryotic cells) help to perform vital activities*)	2

APPENDIX F

COGNITIVE LOAD QUESTIONNAIRE

Answer the following questions on a scale of 1-10, where 10 indicates highly agreed, and one is highly disagreed.

	Highly disagreed (1)	Highly agreed (10)
The topic/topics covered in the activity was/were very complex.		
The activity covered formulas that I perceived as very complex.		
The activity covered concepts and definitions that I perceived as very complex.		
The instructions and/or explanations during the activity were very unclear.		
The instructions and/or explanations were full of unclear language.		
The activity really enhanced my understanding of the topic(s) covered.		
The activity really enhanced my knowledge and understanding of statistics.		
The activity really enhanced my understanding of the formulas covered.		
The activity really enhanced my understanding of concepts and definitions.		
The instructions and/or explanations were full of unclear language.		

APPENDIX G

MOTIVATION QUESTIONNAIRE

Answer the following questions on a scale of 1-5, where five indicates highly agreed, and one highly disagrees.

	Highly disagreed (1)	Highly agreed (5)
The quality of the text helped to hold my attention.		
The way the information is arranged on the pages helped keep my attention.		
The variety of reading passages, exercises, illustrations, etc., helped keep my attention on the user instructions.		
It is clear to me how the content of these user instructions is related to things I already know.		
The content and style of writing in these user instructions convey the impression that being able to work with the telephone is worth it.		
The content of these user instructions will be useful to me.		
As I worked with these user instructions, I was confident that I could learn how to work well with the telephone.		
After working with these user instructions for a while, I was confident that I would be able to complete exercises with the telephone.		
The good organization of the content helped me be confident that I would learn to work with the telephone.		
I enjoyed working with these user instructions so much that I was stimulated to keep on working.		
I really enjoyed working with these user instructions.		
It was a pleasure to work with such well-designed user instructions.		

APPENDIX H

ETHICS COMMITTEE APPROVAL

Evrak Tarih ve Sayısı: 15.04.2021-11776

T.C.
BOĞAZİÇİ ÜNİVERSİTESİ
SOSYAL VE BEŞERİ BİLİMLER YÜKSEK LİSANS VE DOKTORA TEZLERİ ETİK İNCELEME KOMİSYONU
TOPLANTI TUTANAĞI

Toplantı Sayısı	:	15
Toplantı Tarihi	:	15.04.2021
Toplantı Saati	:	13:00
Toplantı Yeri	:	Zoom Sanal Toplantı
Bulunanlar	:	Dr. Öğr. Üyesi Yasemin Sohtorik İlkmen, Prof. Dr. Ebru Kaya, Prof. Dr. Fatma Nevra Seggie
Bulunmayanlar	:	

Burç Çeken

Yönetim Bilişim Sistemleri

Sayın Araştırmacı,

"Examination of multimedia learning principles in augmented reality and virtual reality learning environments" başlıklı projeniz ile ilgili olarak yaptığınız SBB-EAK 2021/6 sayılı başvuru komisyonumuz tarafındann 15 Nisan 2021 tarihli toplantıda incelenmiş ve uygun bulunmuştur.

Bu karar tüm üyelerin toplantıya çevrimiçi olarak katılımı ve oybirliği ile alınmıştır. COVID-19 önlemleri kapsamında kurul üyelerinden ıslak imza alınamadığı için bu onam mektubu üye ve raportör olarak Ebru Kaya tarafından bütün üyeler adına e-imzalanmıştır.

Saygılarımızla, bilgilerinizi rica ederiz.

Prof. Dr. Ebru KAYA
ÜYE

e-imzalıdır
Prof. Dr.Ebru KAYA
Raportör

APPENDIX I

VISUAL INSPECTIONS OF THE DATA

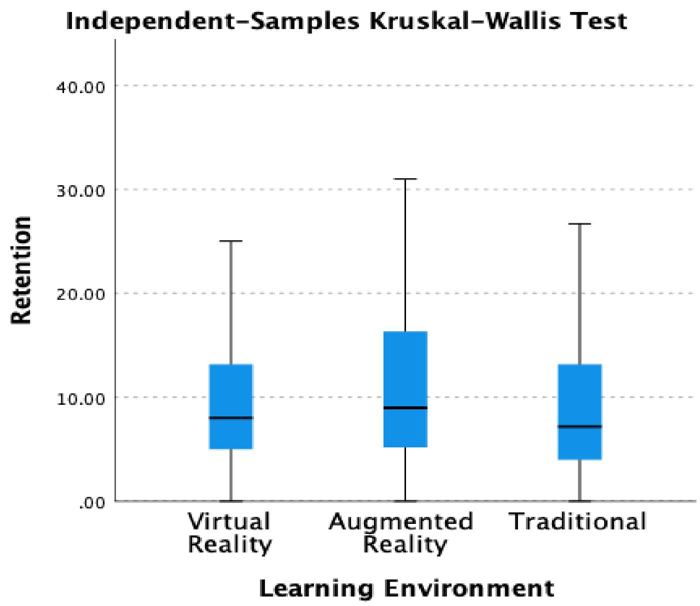

Figure I1. Boxplot of the retention scores of each learning environment about lightning formation subject matter

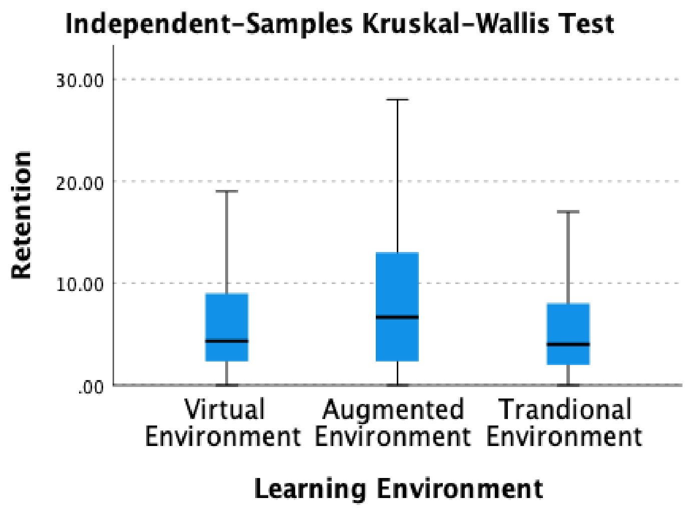

Figure I2. Boxplot of the retention scores of each learning environment about cell structure subject matter

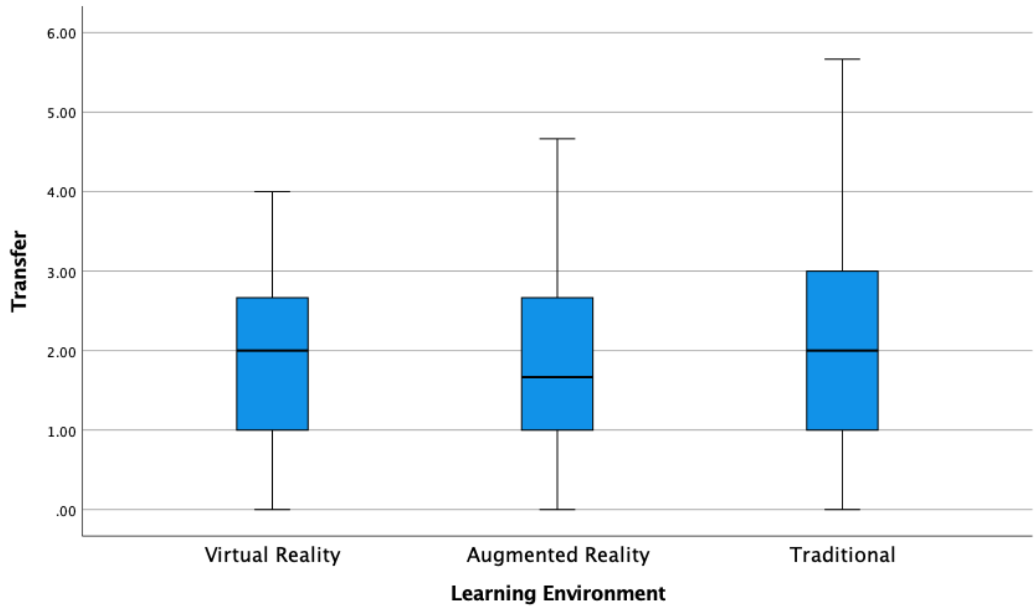

Figure I3. Boxplot of the transfer scores of each learning environment about lightning formation subject matter

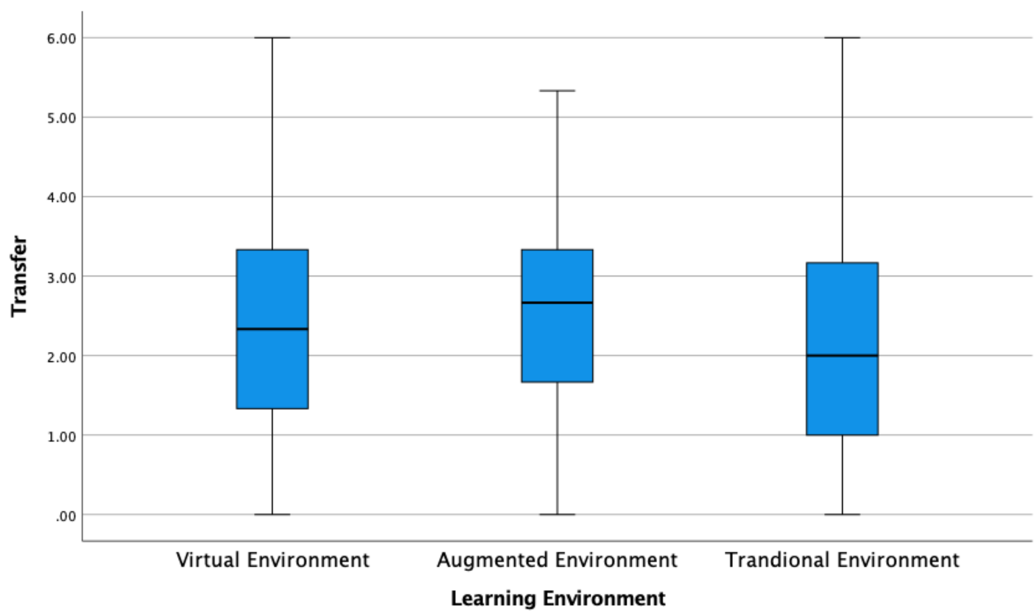

Figure I4. Boxplot of the transfer scores of each learning environment about cell structure subject matter

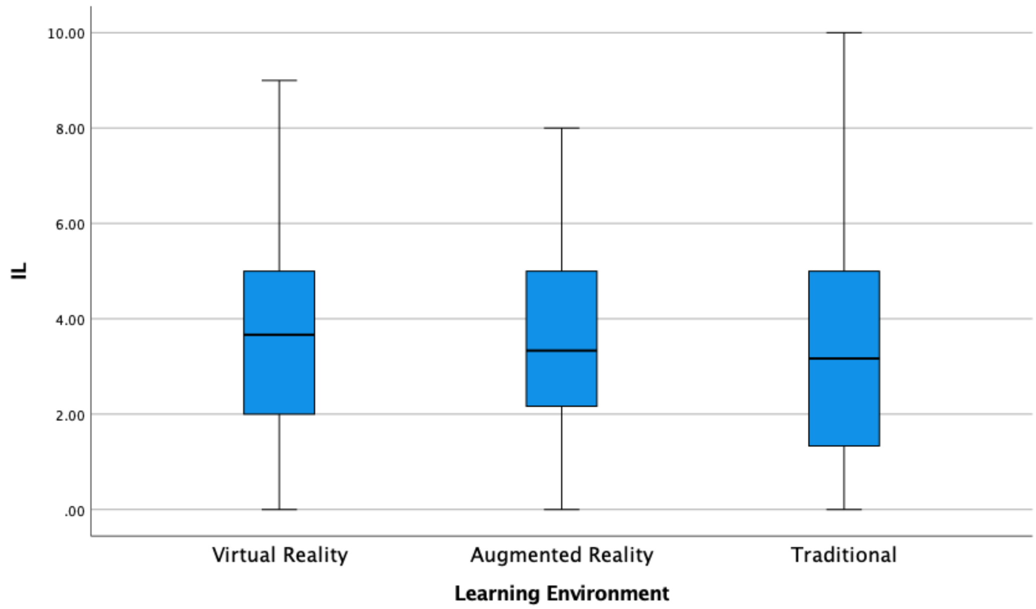

Figure I5. Boxplot of the intrinsic cognitive load scores of each learning environment about lightning formation subject matter

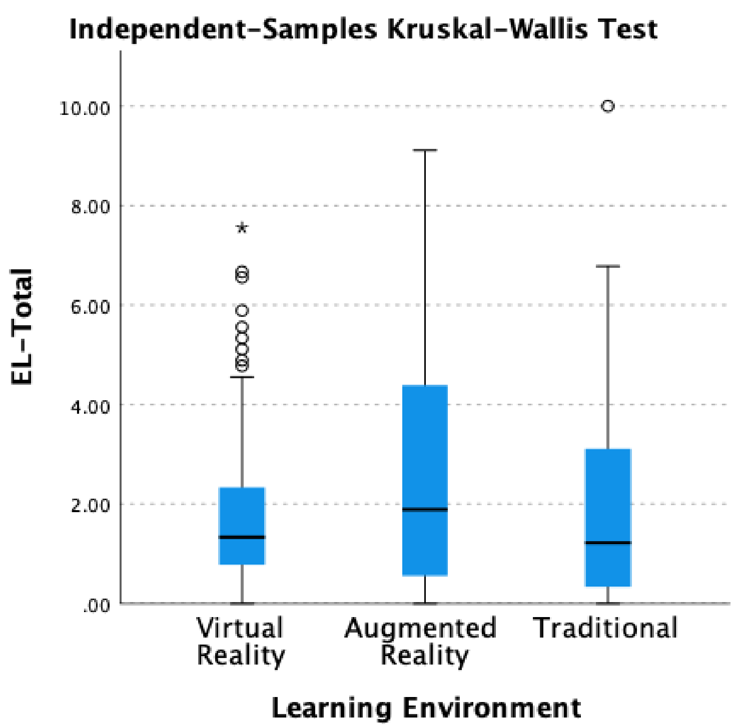

Figure I6. Boxplot of the extraneous cognitive load scores of each learning environment about lightning formation subject matter

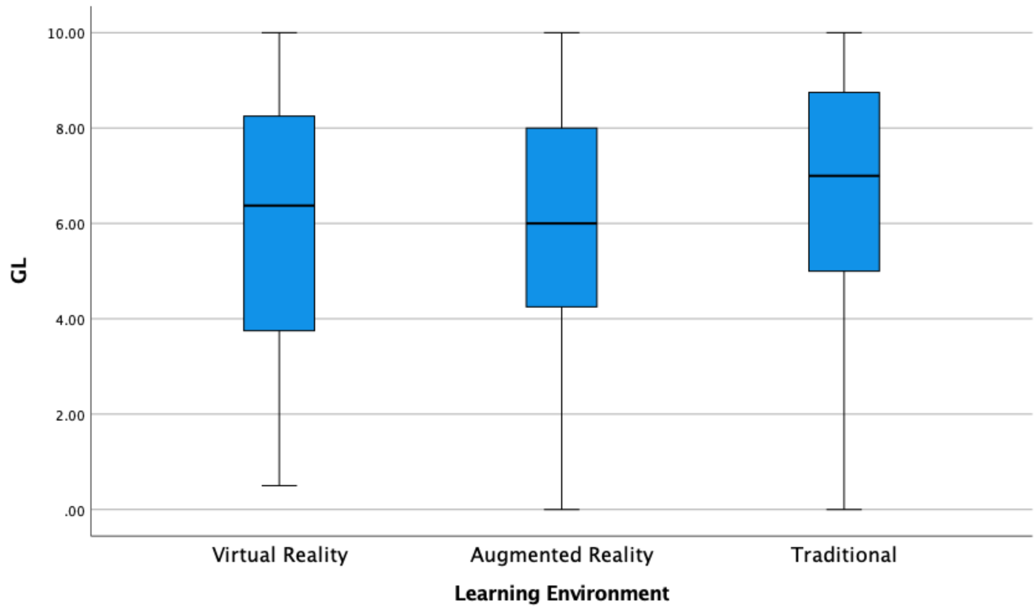

Figure I7. Boxplot of the germane cognitive load scores of each learning environment about lightning formation subject matter

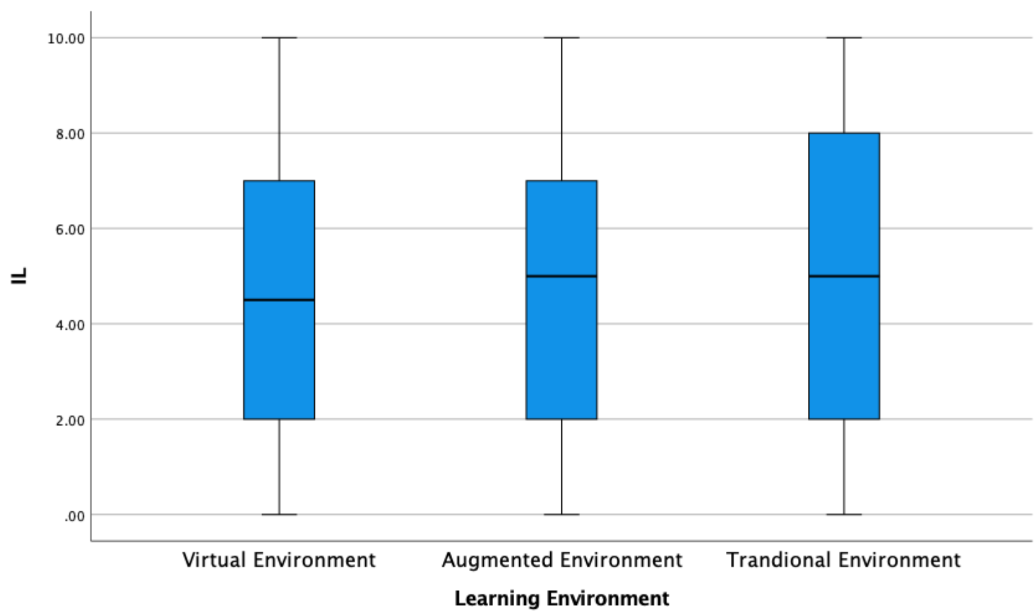

Figure I8. Boxplot of the intrinsic cognitive load scores of each learning environment about cell structure subject matter

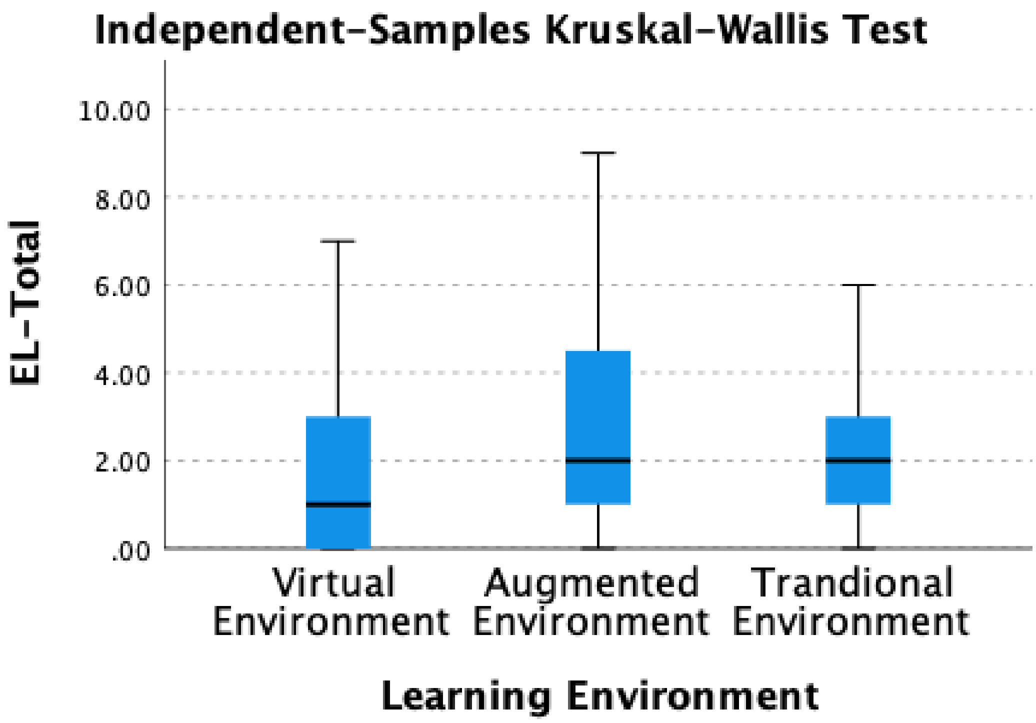

Figure I9. Boxplot of the extraneous cognitive load scores of each learning environment about cell structure subject matter

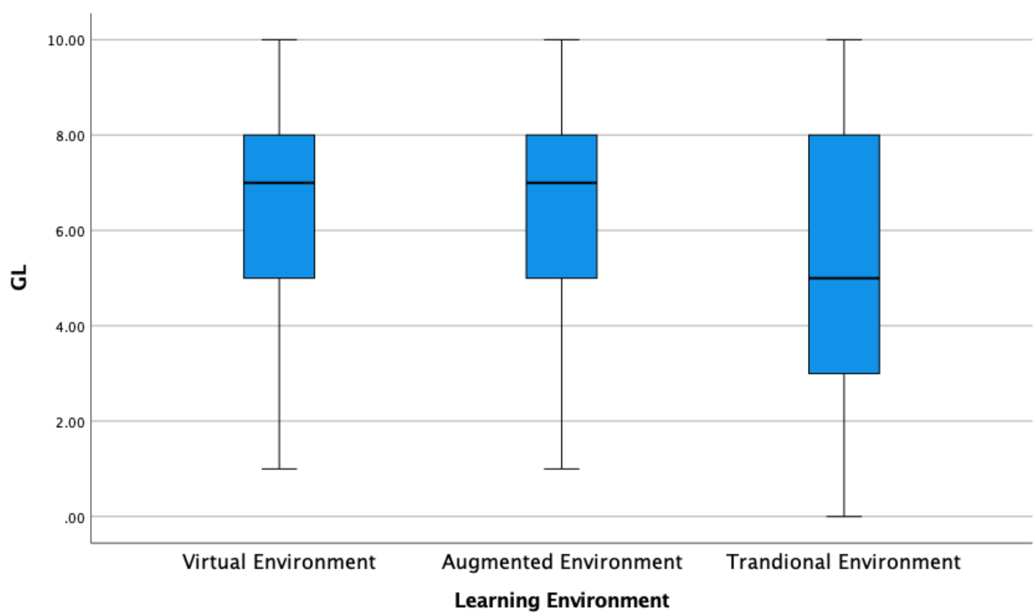

Figure I10. Boxplot of the germane cognitive load scores of each learning environment about cell structure subject matter

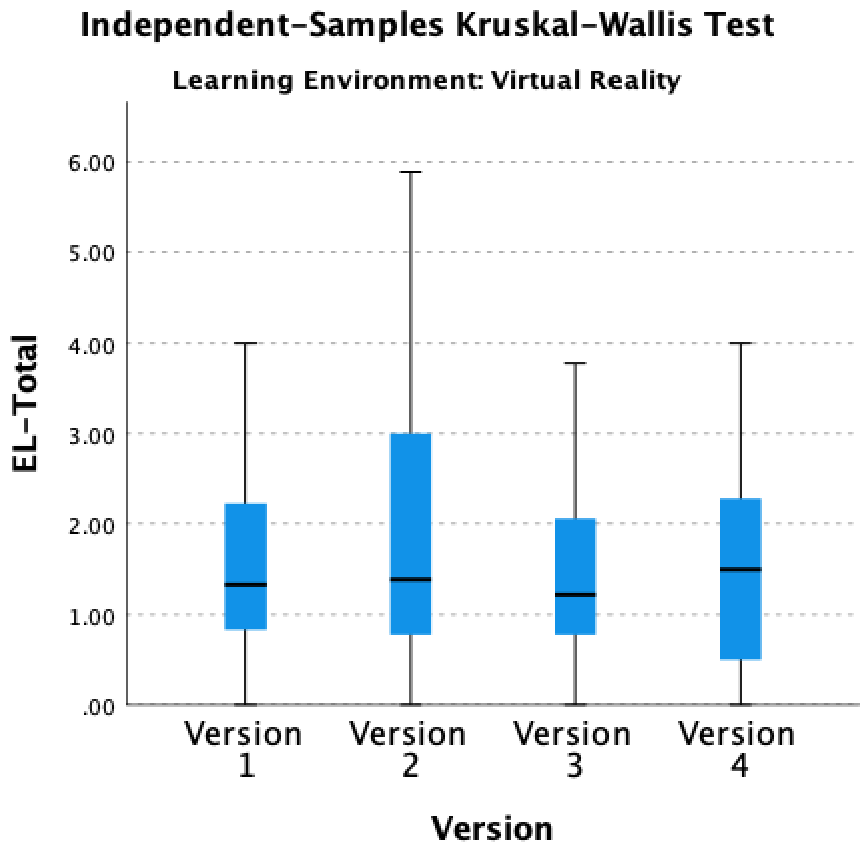

Figure I11. Boxplot of the overall motivation scores of each learning environment about lightning formation subject matter

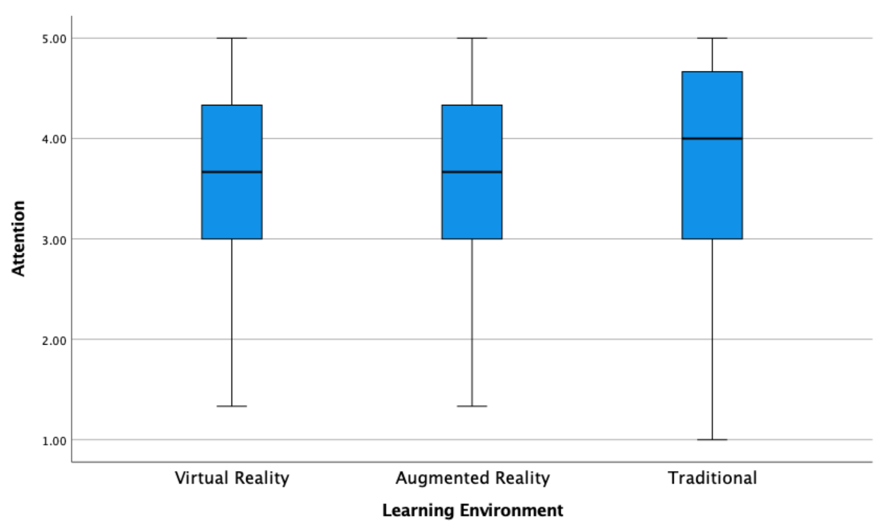

Figure I12. Boxplot of the attention scores of each learning environment about lightning formation subject matter

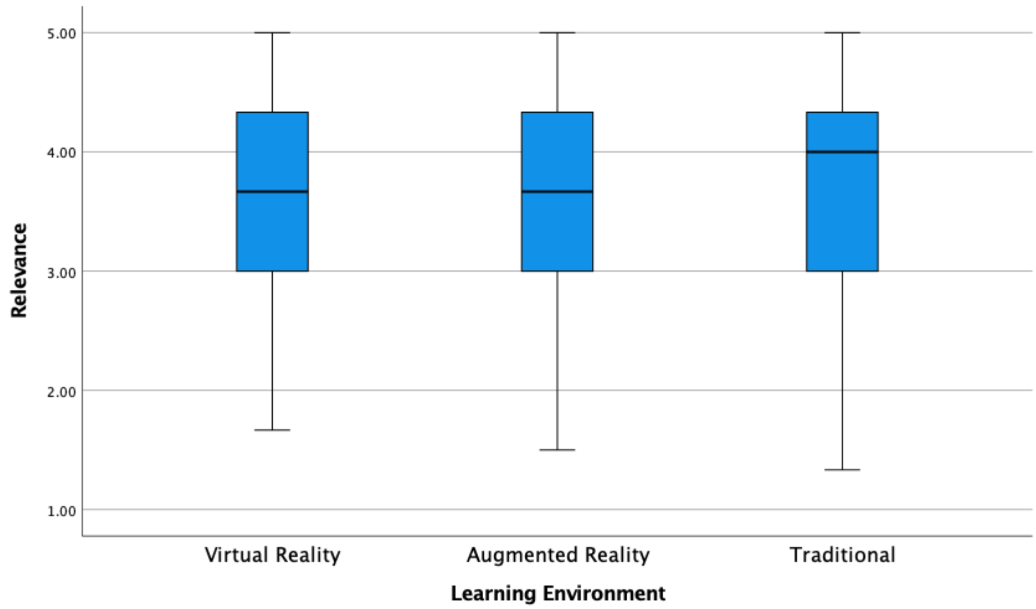

Figure I13. Boxplot of the relevance scores of each learning environment about lightning formation subject matter

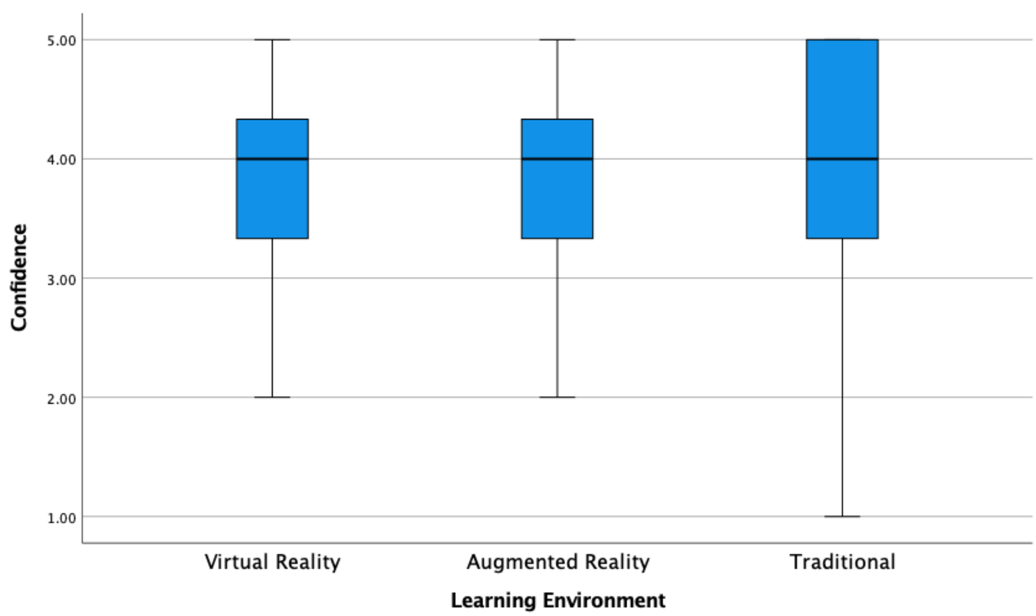

Figure I14. Boxplot of the confidence scores of each learning environment about lightning formation subject matter

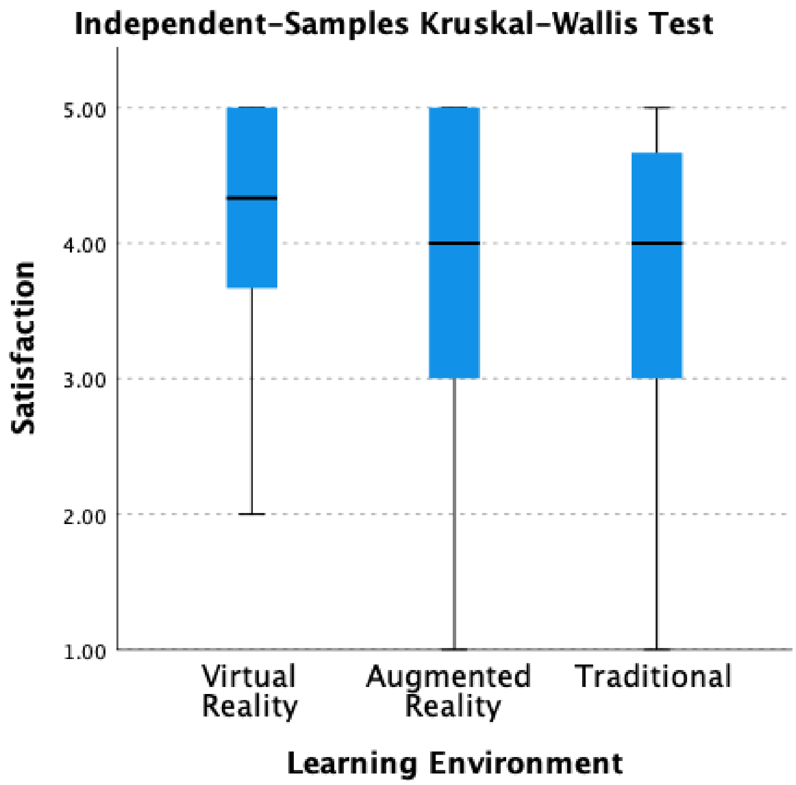

Figure I15. Boxplot of the satisfaction scores of each learning environment about lightning formation subject matter

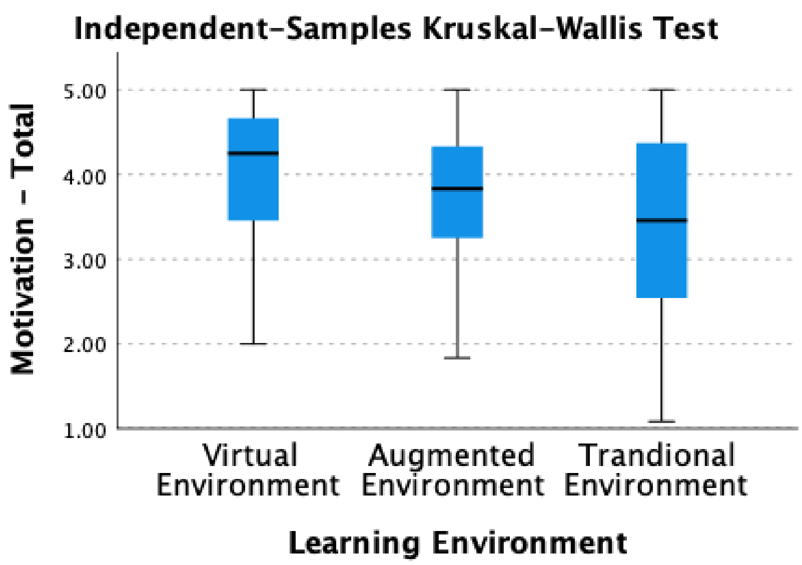

Figure I16. Boxplot of the overall motivation scores of each learning environment about cell structure subject matter

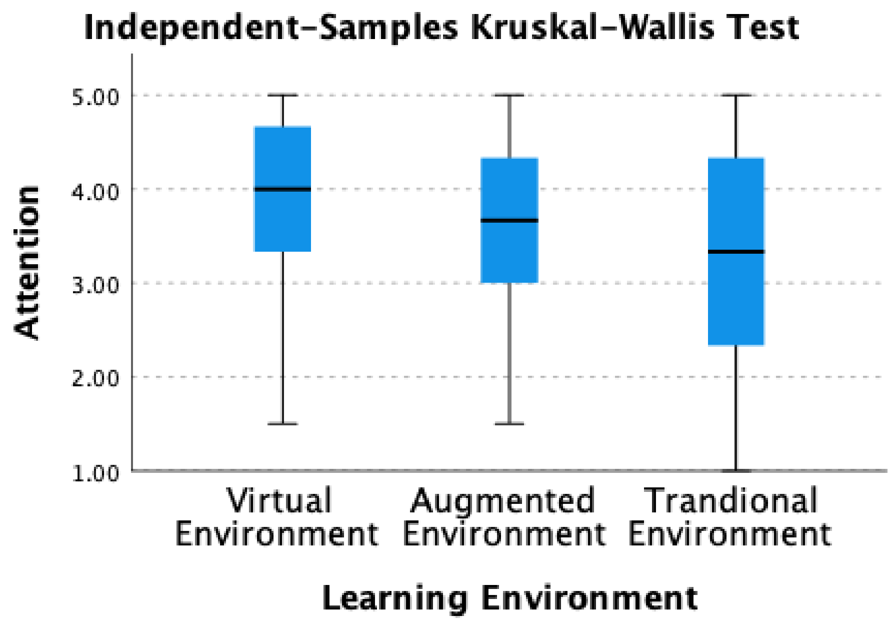

Figure I17. Boxplot of the attention scores of each learning environment about cell structure subject matter

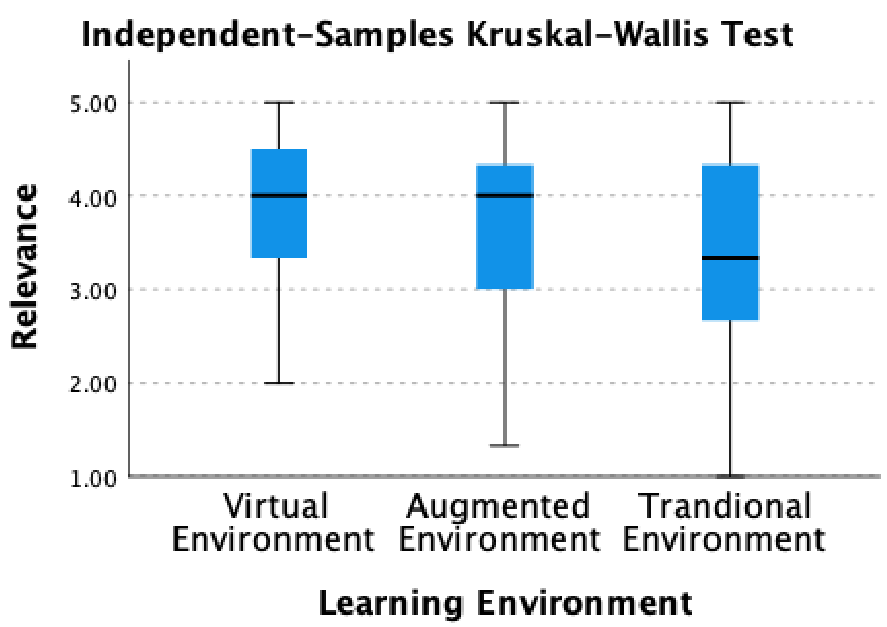

Figure I18. Boxplot of the relevance scores of each learning environment about cell structure subject matter

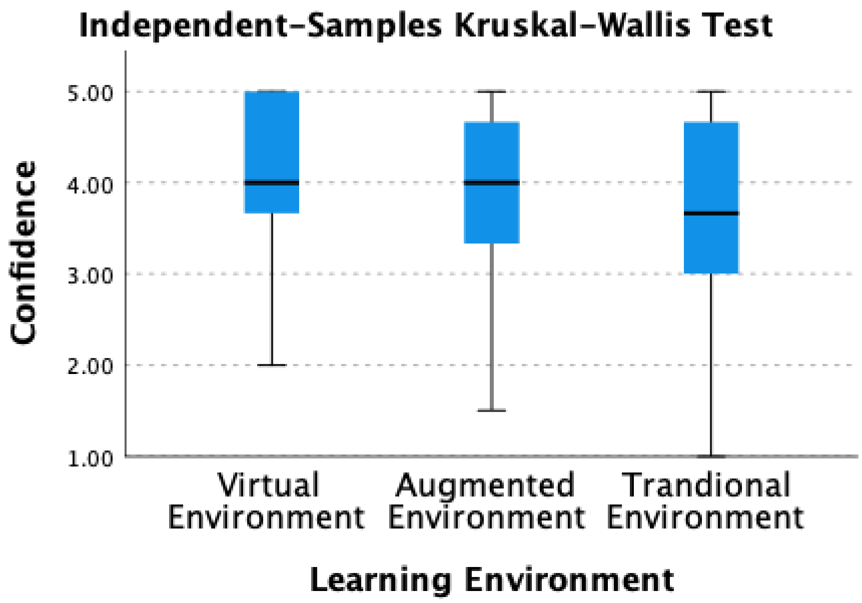

Figure I19. Boxplot of the confidence scores of each learning environment about cell structure subject matter

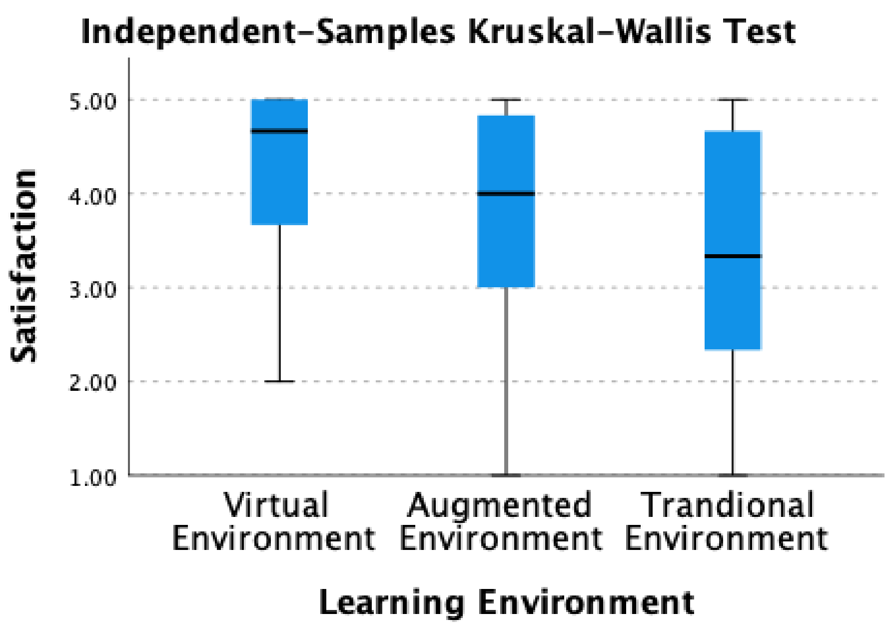

Figure I20. Boxplot of the satisfaction scores of each learning environment about cell structure subject matter

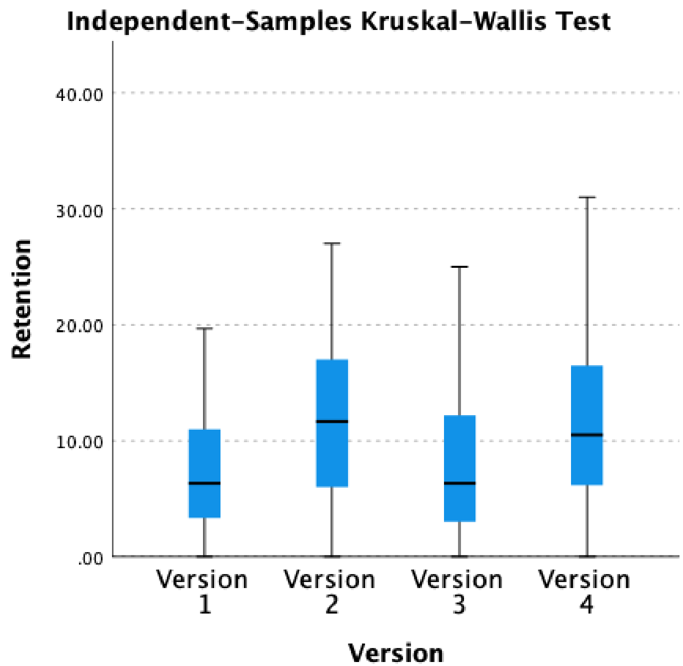

Figure I21. Boxplot of the retention scores of each version about lightning formation subject matter

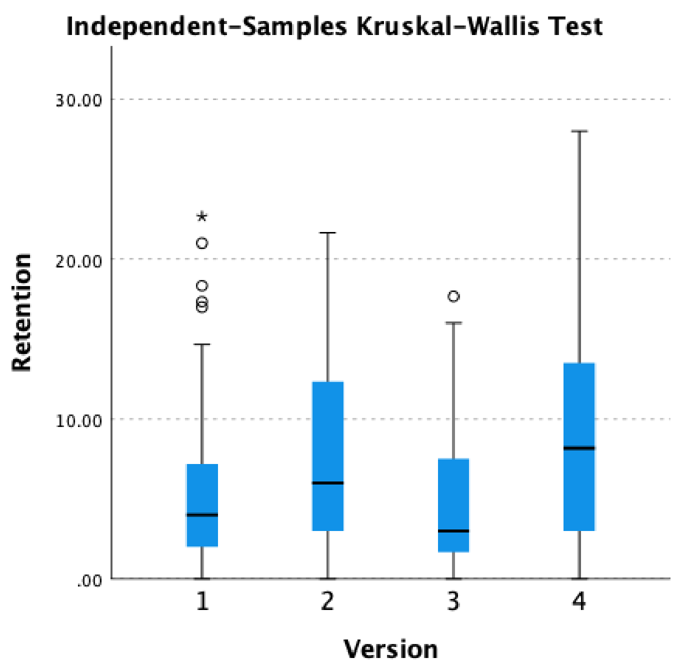

Figure I22. Boxplot of the retention scores of each version about cell structure subject matter

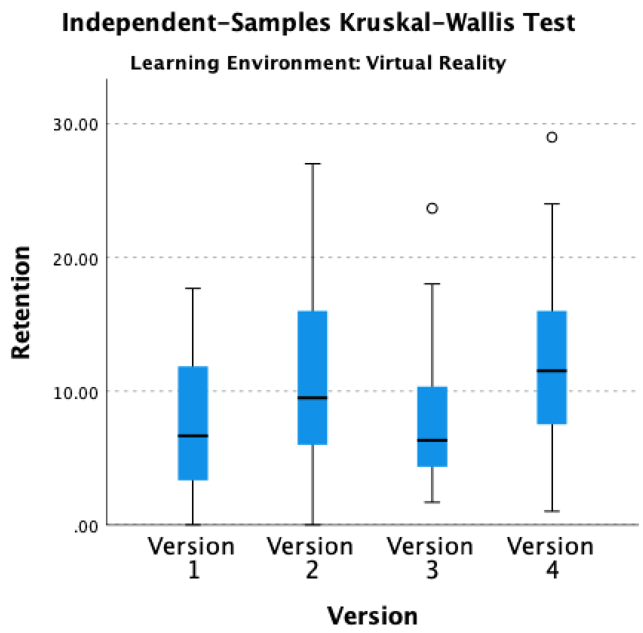

Figure I23. Boxplot of the retention scores of each version in the virtual reality learning environment about lightning formation subject matter

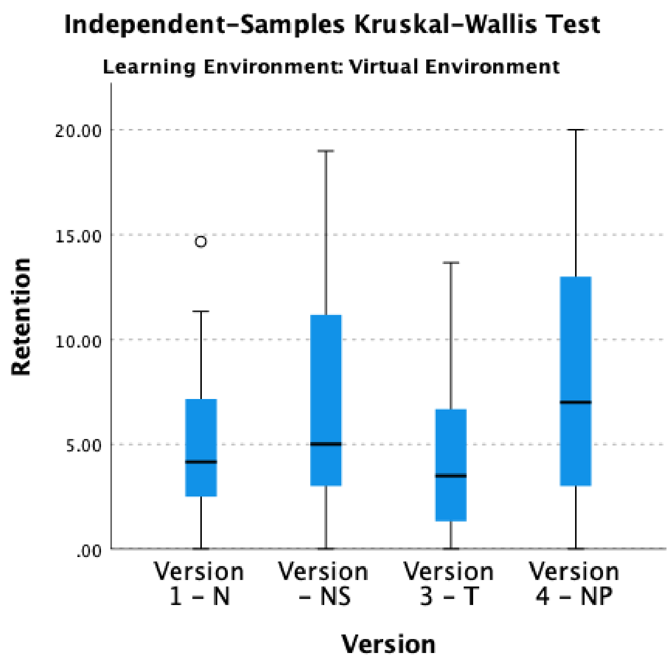

Figure I24. Boxplot of the retention scores of each version in the virtual reality learning environment about cell structure subject matter

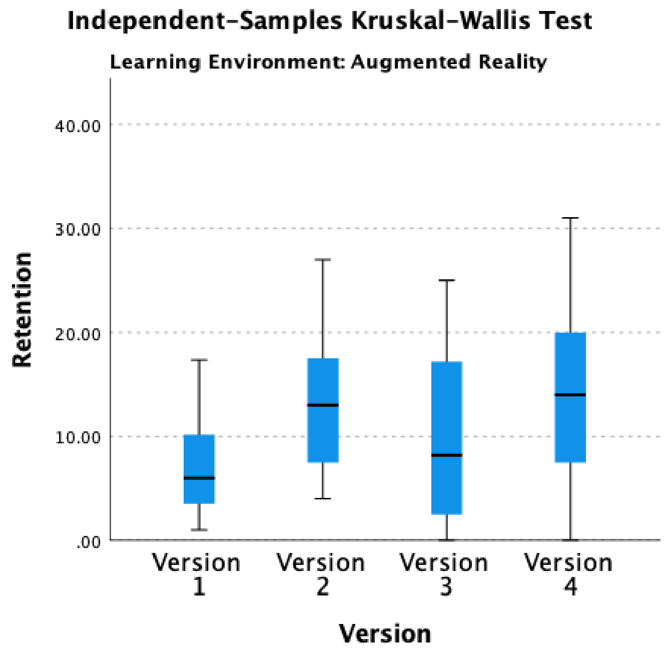

Figure I25. Boxplot of the retention scores of each version in the augmented reality learning environment about lightning formation subject matter

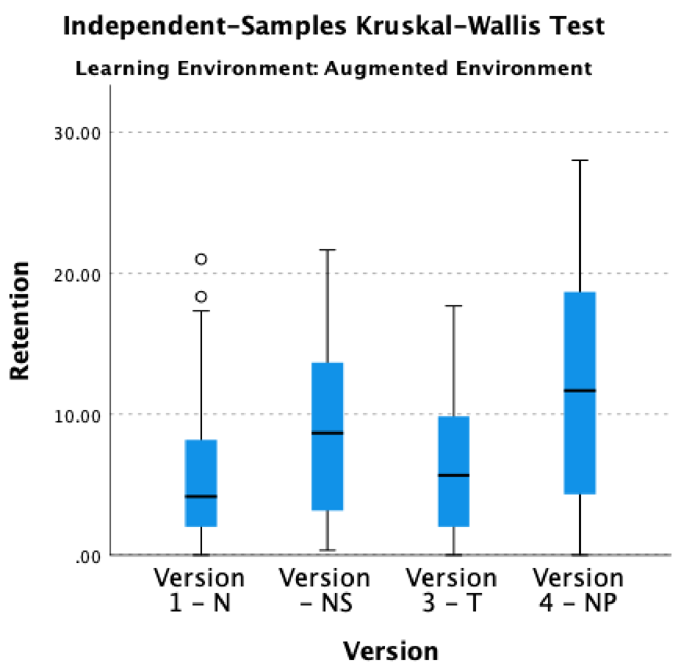

Figure I26. Boxplot of the retention scores of each version in the augmented reality learning environment about cell structure subject matter

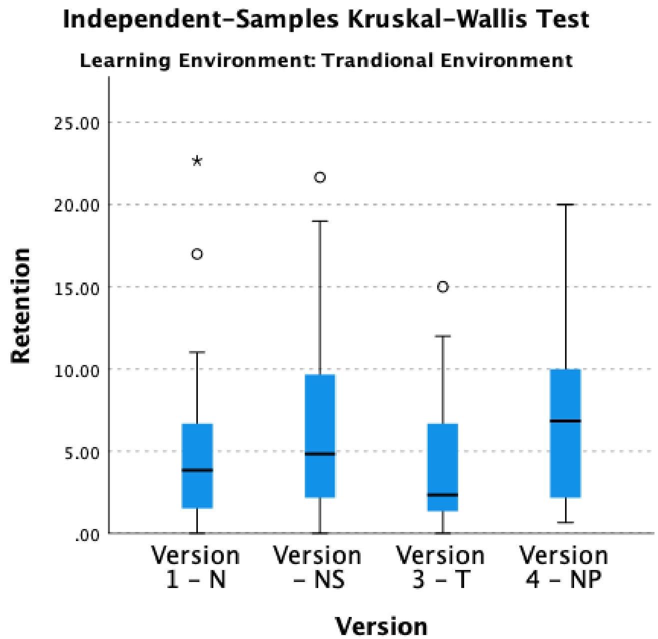

Figure I27. Boxplot of the retention scores of each version in the traditional reality learning environment about cell structure subject matter

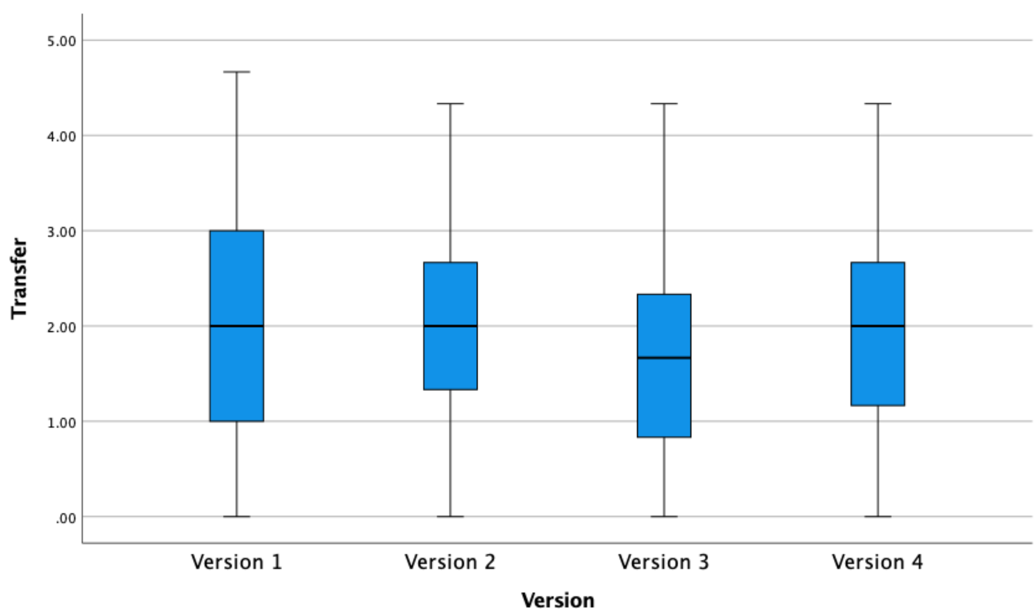

Figure I28. Boxplot of the transfer scores of each version about lightning formation subject matter

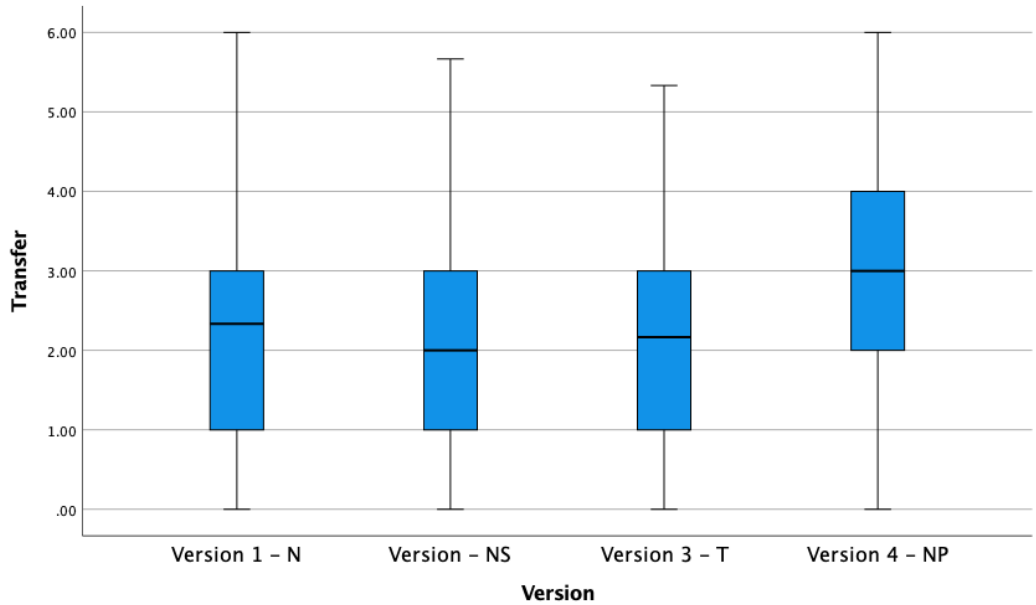

Figure I29. Boxplot of the transfer scores of each version about cell structure subject matter

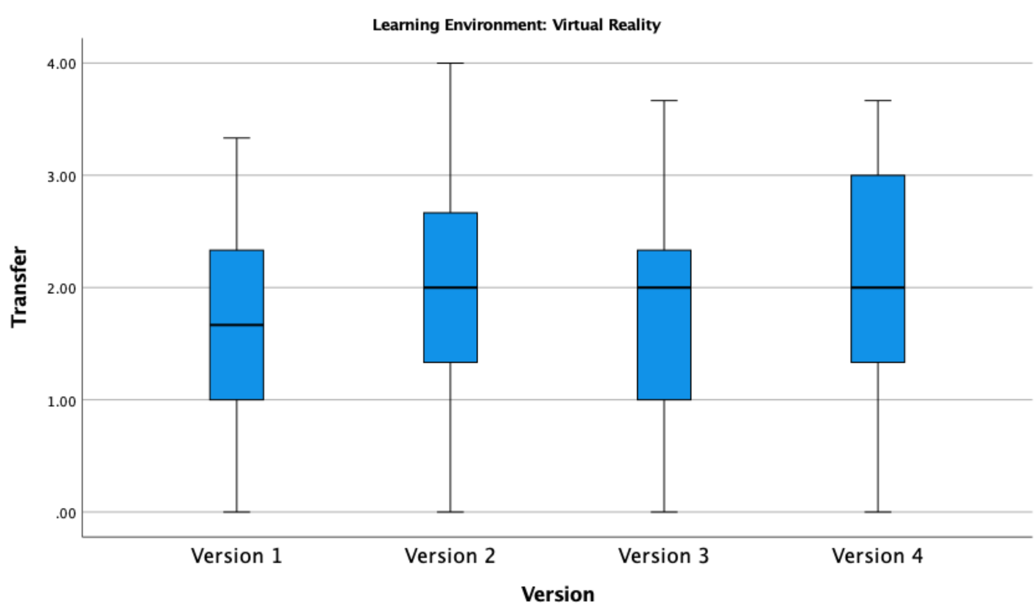

Figure I30. Boxplot of the transfer scores of each version in the virtual reality learning environment about lightning formation subject matter

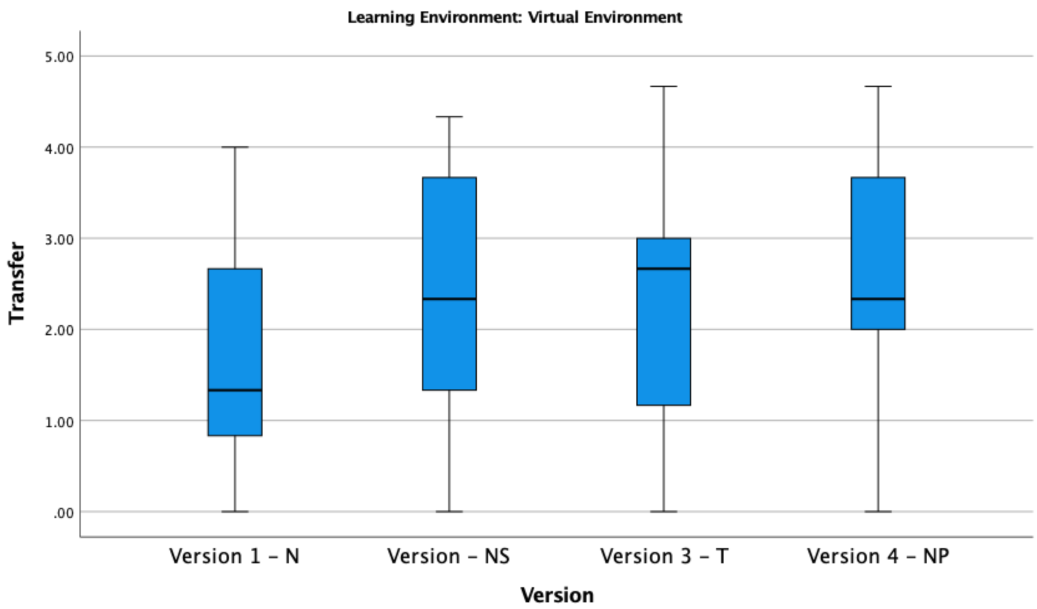

Figure I31. Boxplot of the transfer scores of each version in the virtual reality learning environment about cell structure subject matter

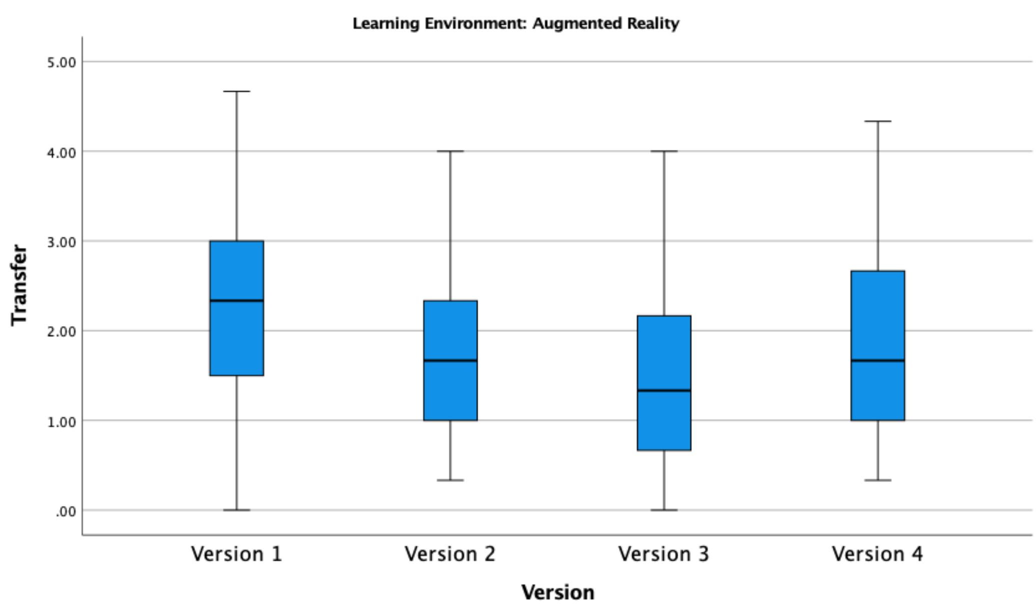

Figure I32. Boxplot of the transfer scores of each version in the augmented reality learning environment about lightning formation subject matter

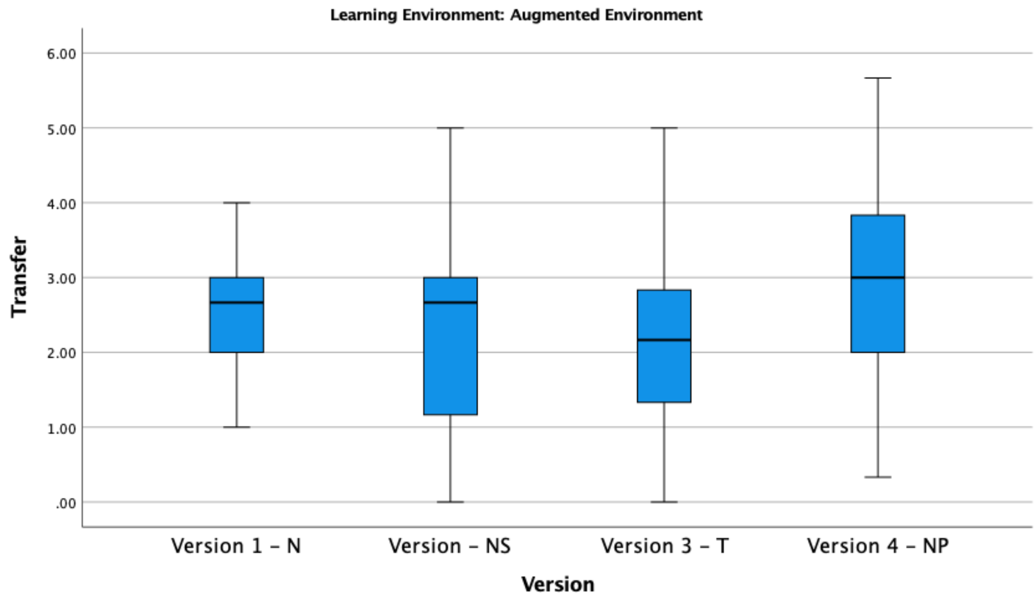

Figure I33. Boxplot of the transfer scores of each version in the augmented reality learning environment about cell structure subject matter

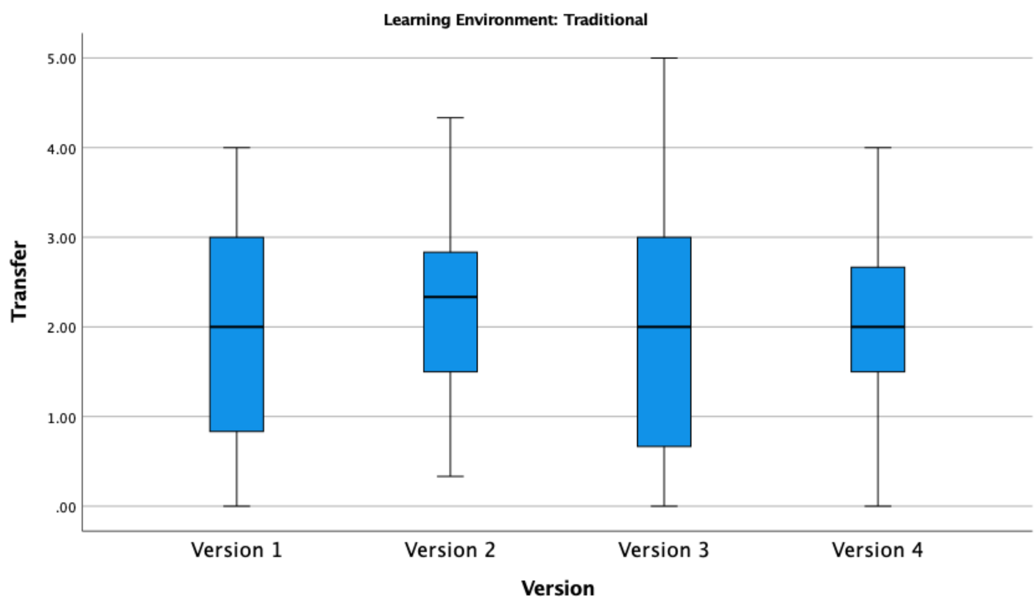

Figure I34. Boxplot of the transfer scores of each version in the traditional learning environment about lightning formation subject matter

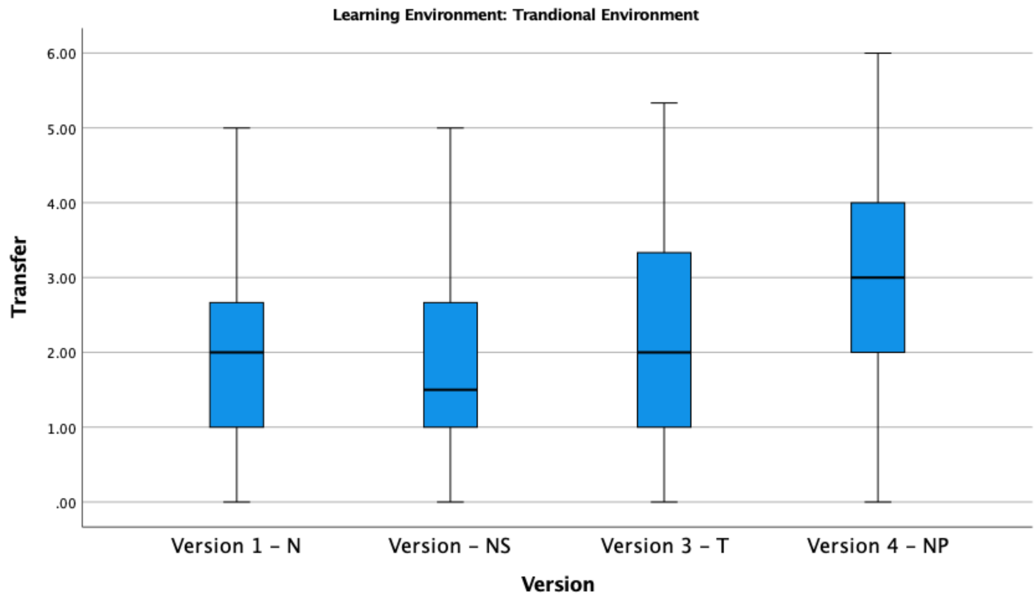

Figure I35. Boxplot of the transfer scores of each version in the traditional learning environment about cell structure subject matter

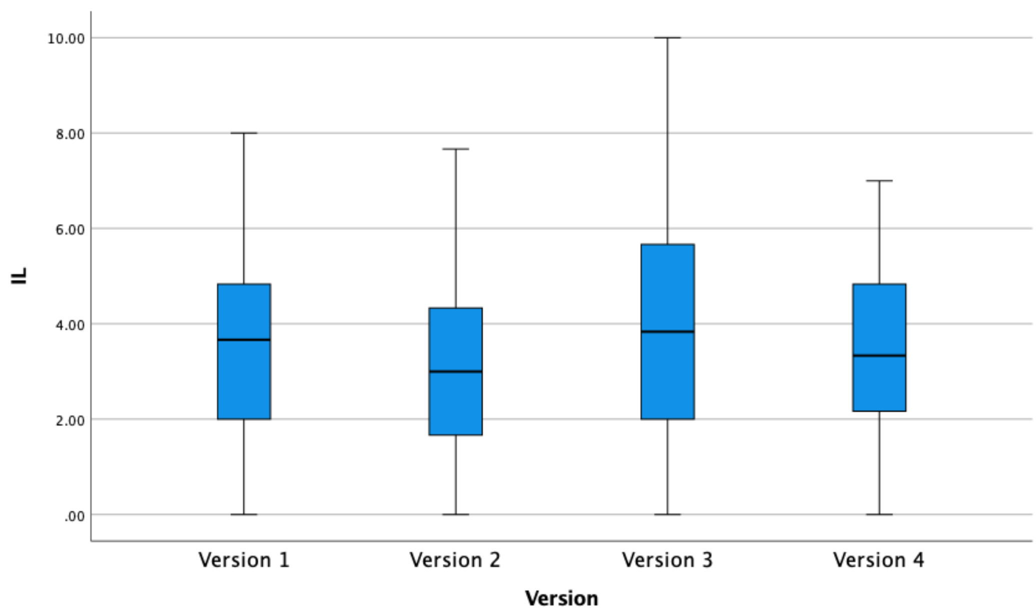

Figure I36. Boxplot of the intrinsic cognitive load scores of each version about lightning formation subject matter

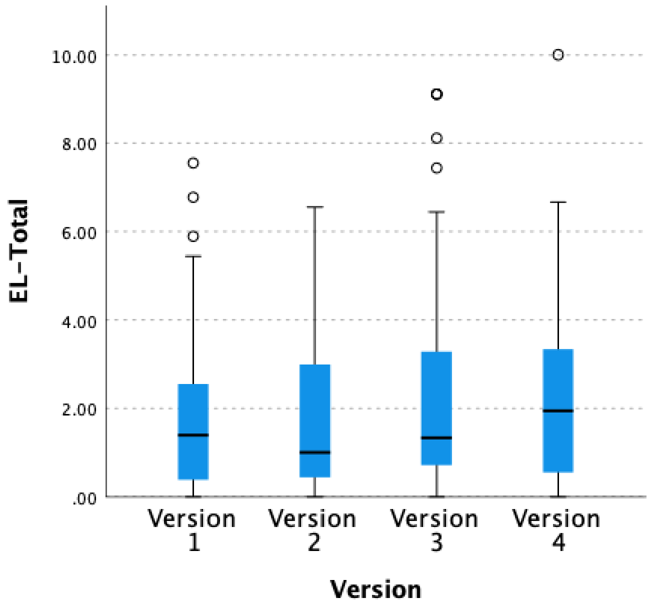

Figure I37. Boxplot of the extraneous cognitive load scores of each version about lightning formation subject matter

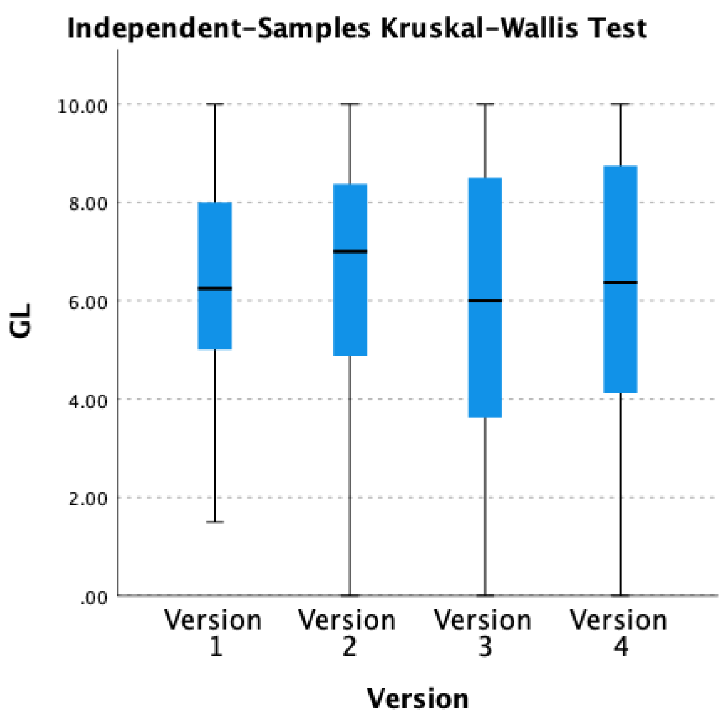

Figure I38. Boxplot of the germane cognitive load scores of each version about lightning formation subject matter

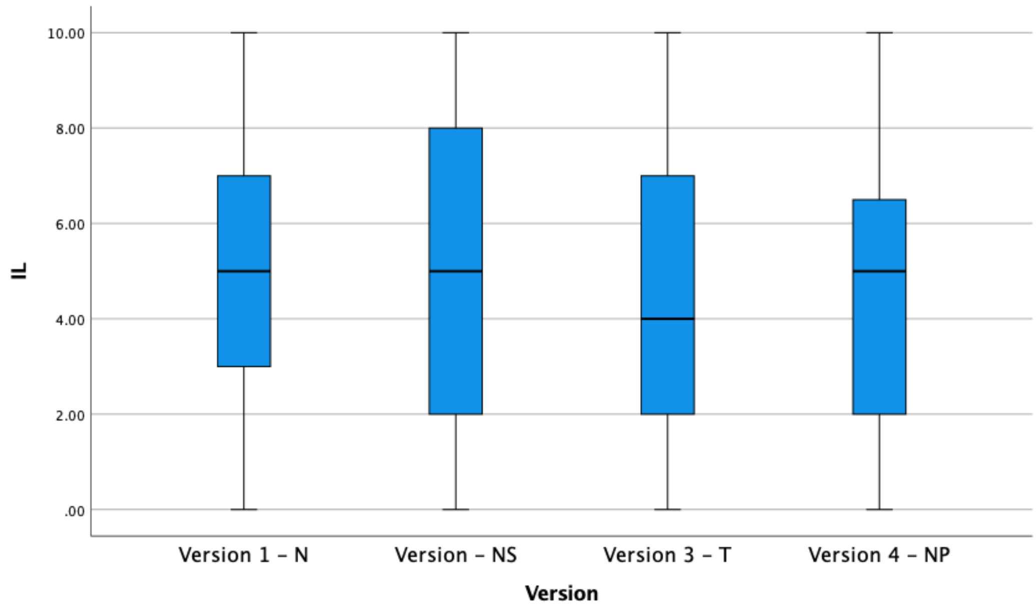

Figure I39. Boxplot of the intrinsic cognitive load scores of each version about cell structure subject matter

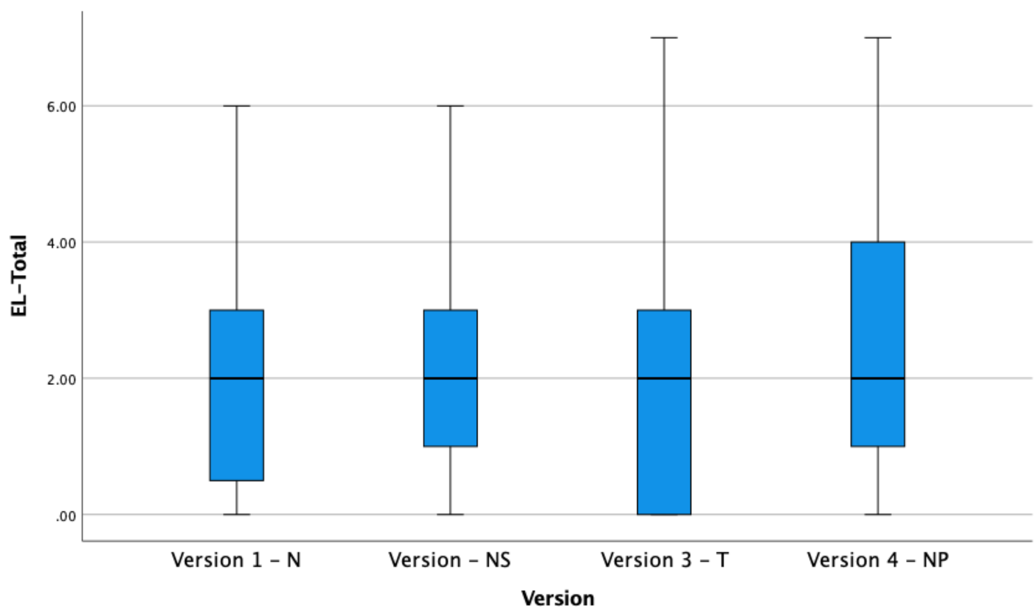

Figure I40. Boxplot of the extraneous cognitive load scores of each version about cell structure subject matter

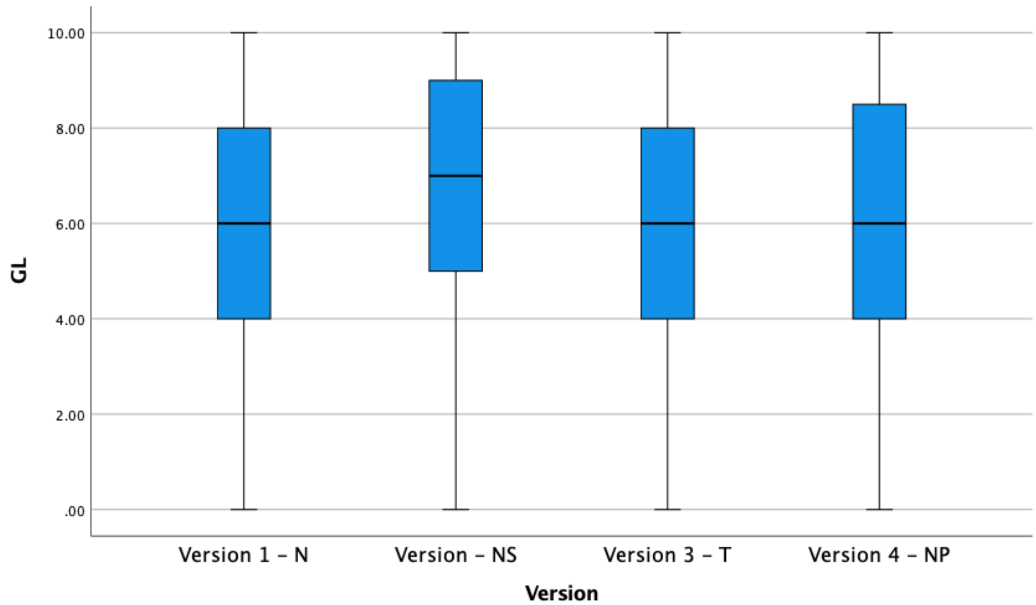

Figure I41. Boxplot of the germane cognitive load scores of each version about cell structure subject matter

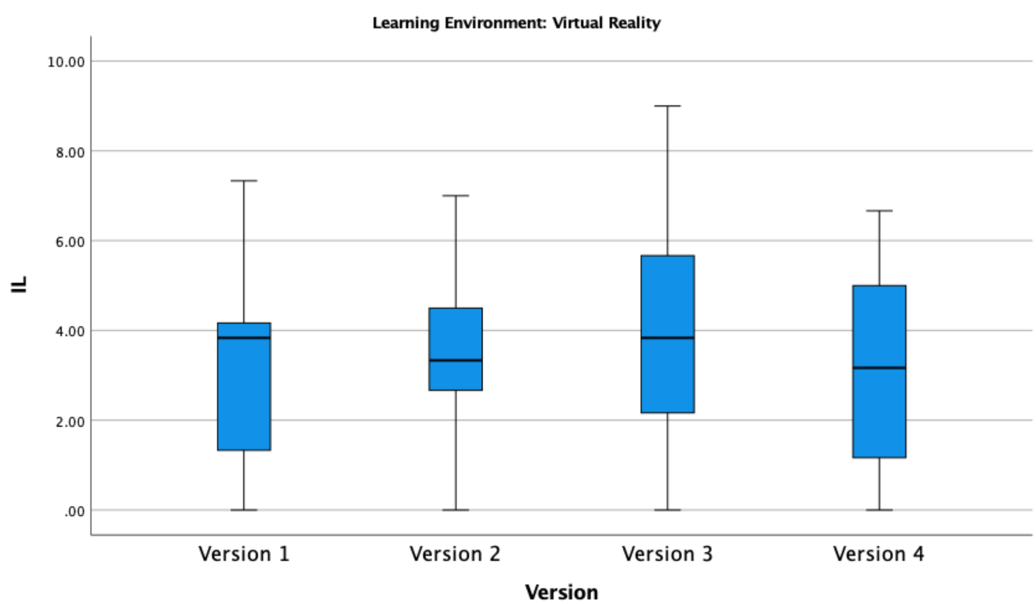

Figure I42. Boxplot of the intrinsic cognitive load scores of each version in the virtual reality learning environment about lightning formation subject matter

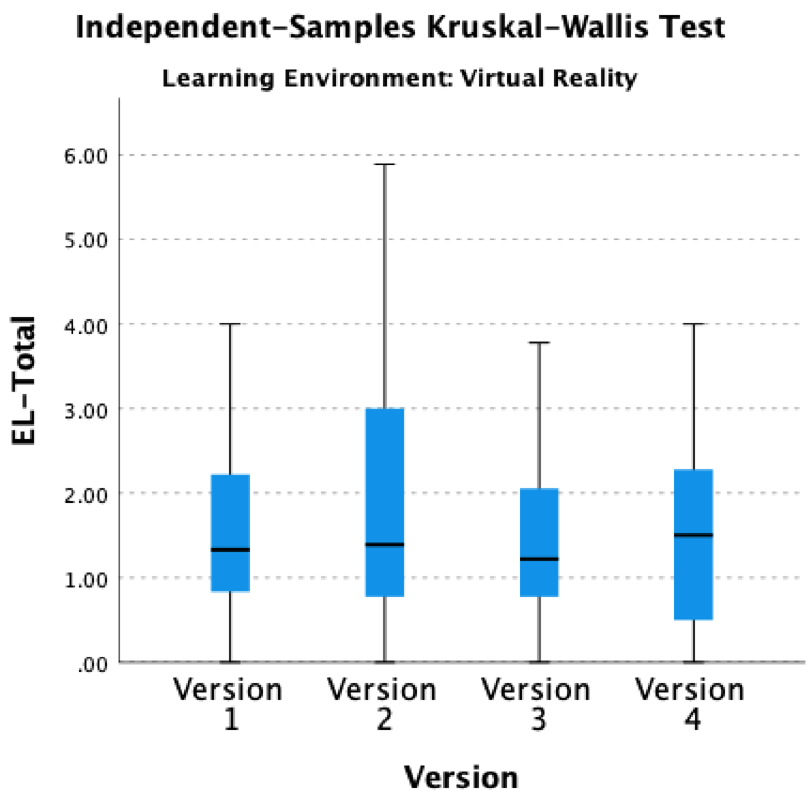

Figure I43. Boxplot of the extraneous cognitive load scores of each version in the virtual reality learning environment about lightning formation subject matter

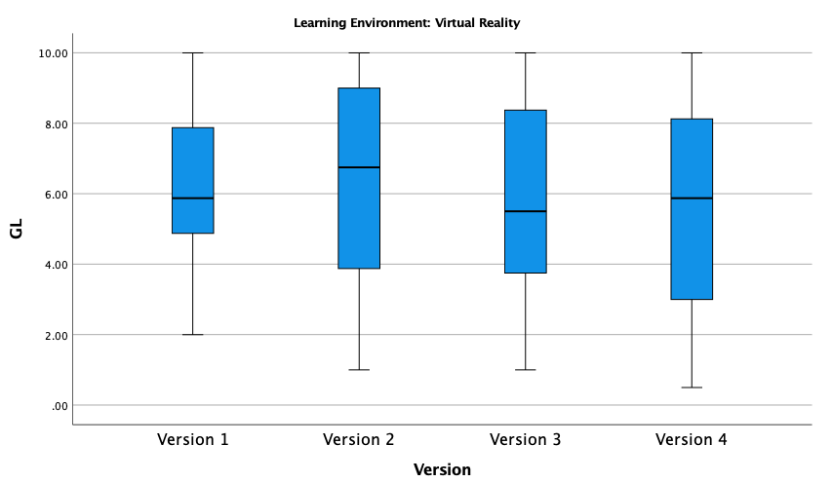

Figure I44. Boxplot of the germane cognitive load scores of each version in the virtual reality learning environment about lightning formation subject matter

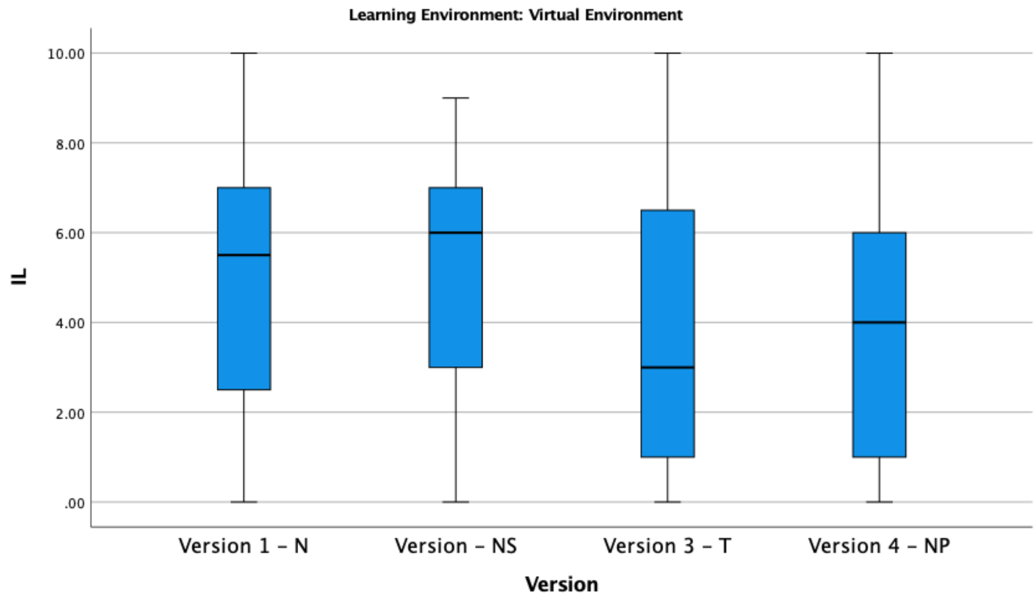

Figure I45. Boxplot of the intrinsic cognitive load scores of each version in the augmented reality learning environment about cell structure subject matter

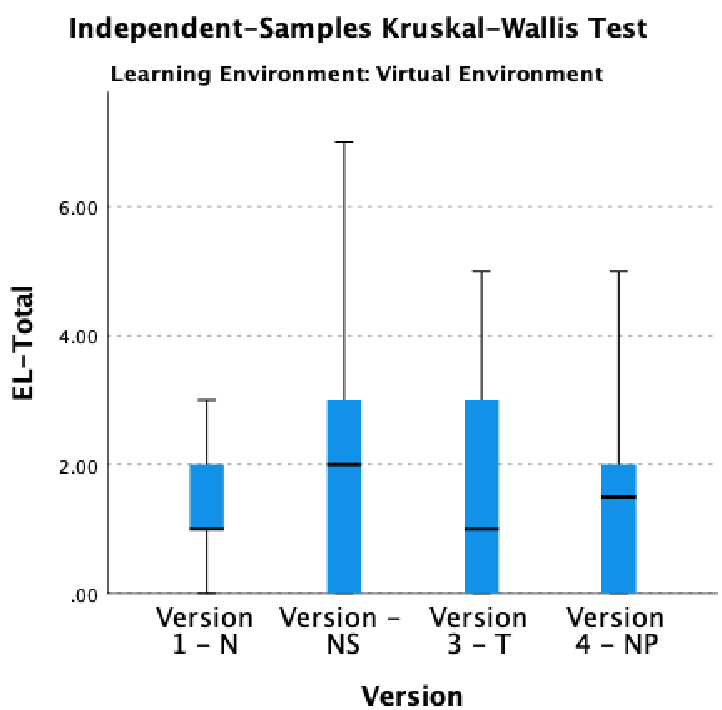

Figure I46. Boxplot of the extraneous cognitive load scores of each version in the augmented reality learning environment about cell structure subject matter

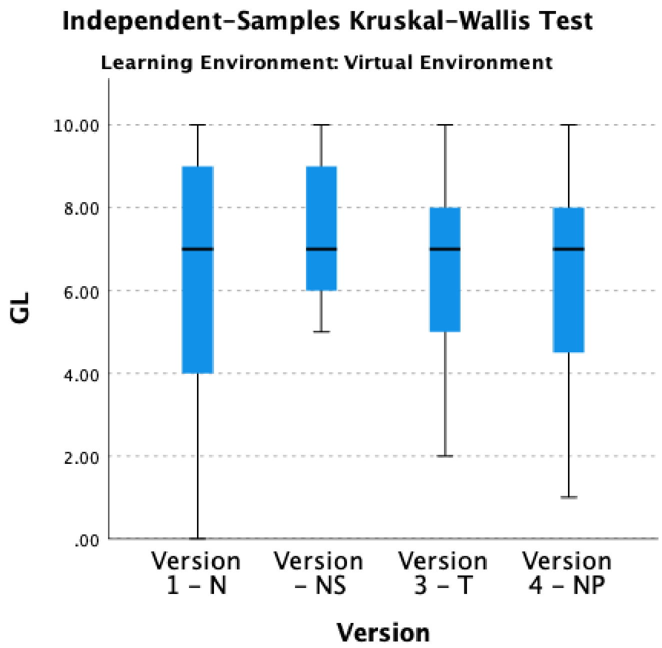

Figure I47. Boxplot of the germane cognitive load scores of each version in the virtual reality learning environment about cell structure subject matter

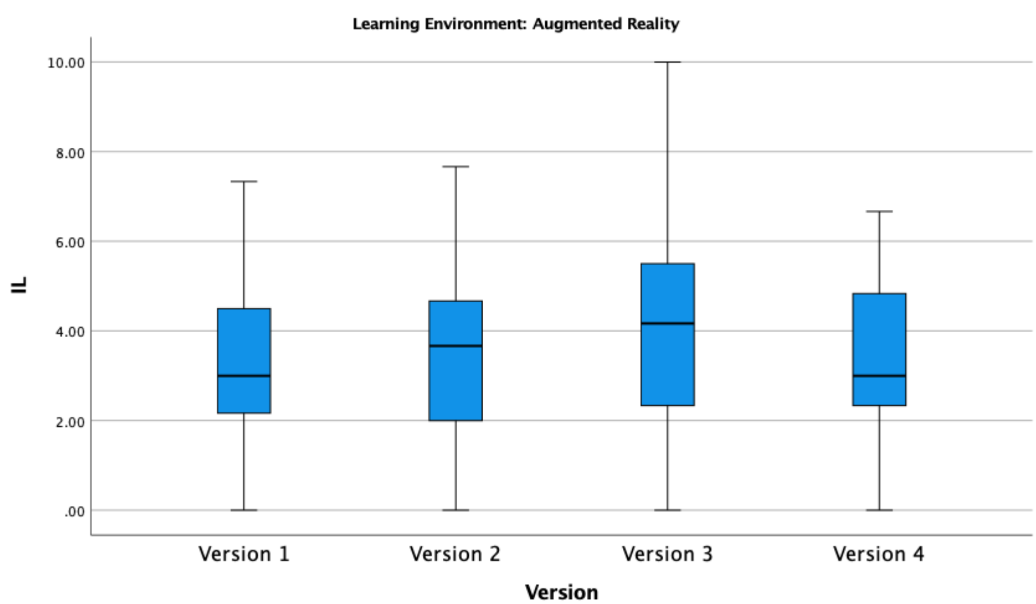

Figure I48. Boxplot of the intrinsic cognitive load scores of each version in augmented reality learning environment about lightning formation subject matter

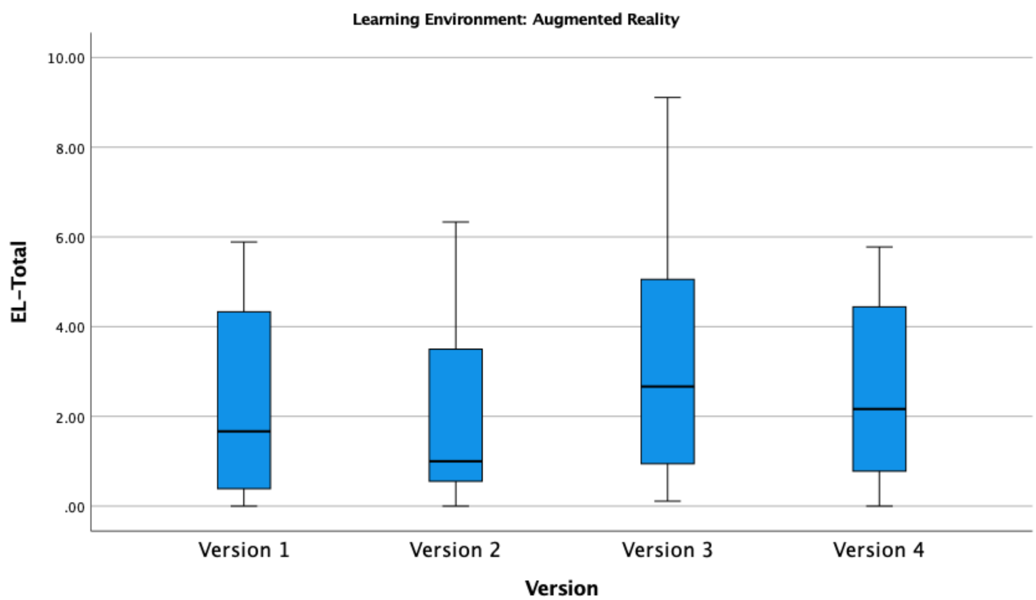

Figure I49. Boxplot of the extraneous cognitive load scores of each version in augmented reality learning environment about lightning formation subject

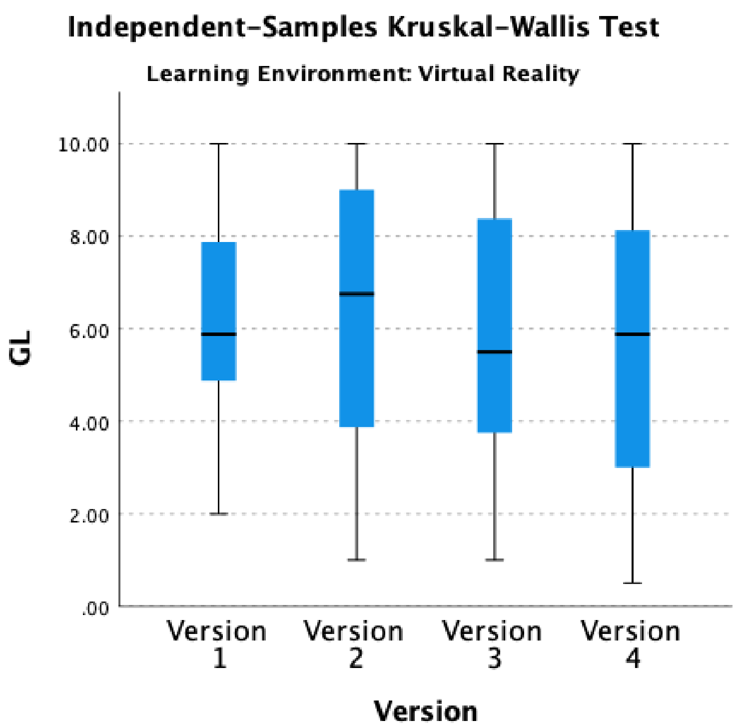

Figure I50. Boxplot of the germane cognitive load scores of each version in augmented reality learning environment about lightning formation subject matter

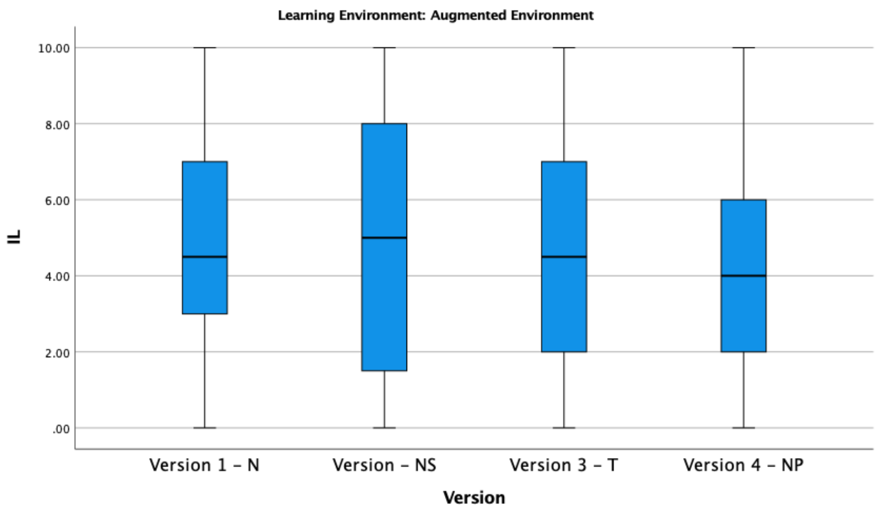

Figure I51. Boxplot of the intrinsic cognitive load scores of each version in the augmented reality learning environment about cell structure subject matter

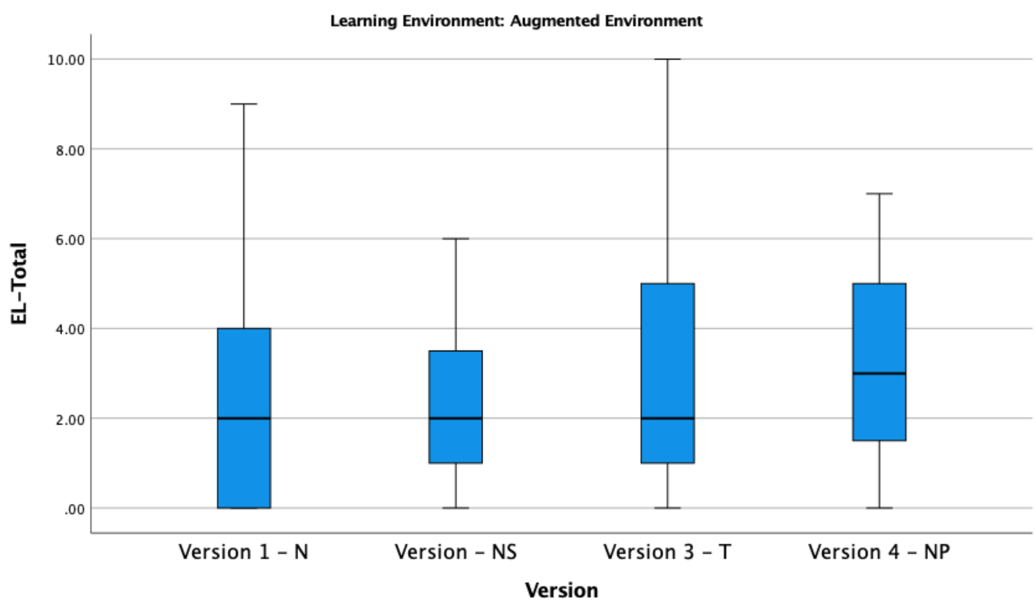

Figure I52. Boxplot of the extraneous cognitive load scores of each version in the augmented reality learning environment about cell structure subject matter

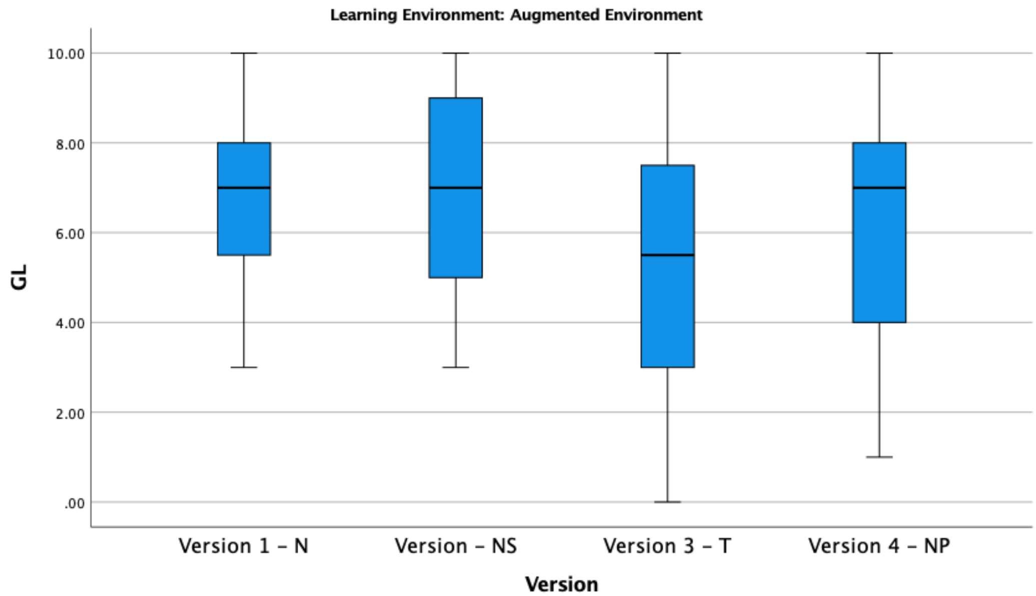

Figure I53. Boxplot of the germane cognitive load scores of each version in the augmented reality learning environment about cell structure subject matter

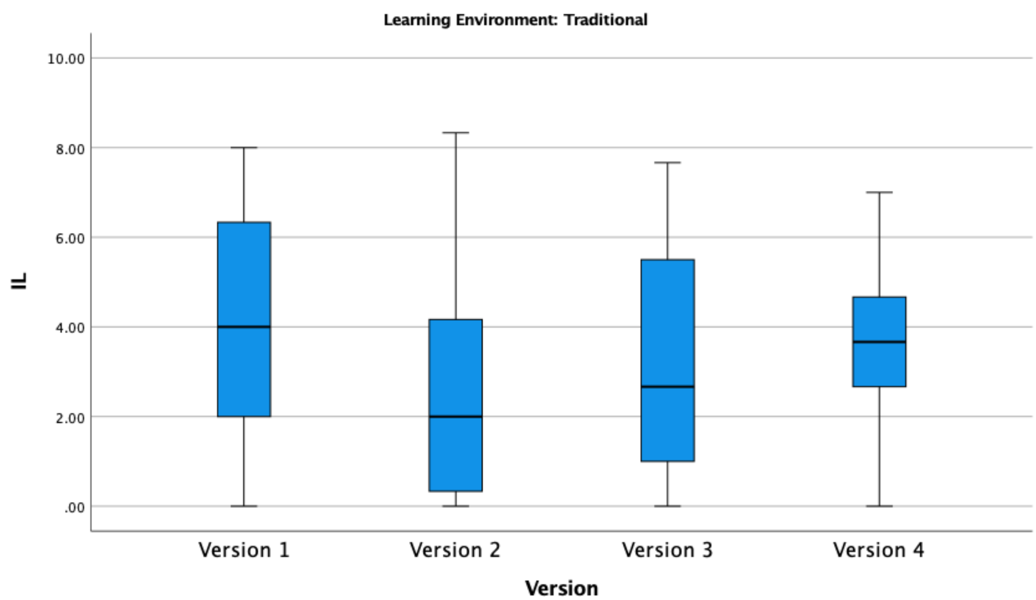

Figure I54. Boxplot of the intrinsic cognitive load scores of each version in the traditional learning environment about lightning formation subject matter

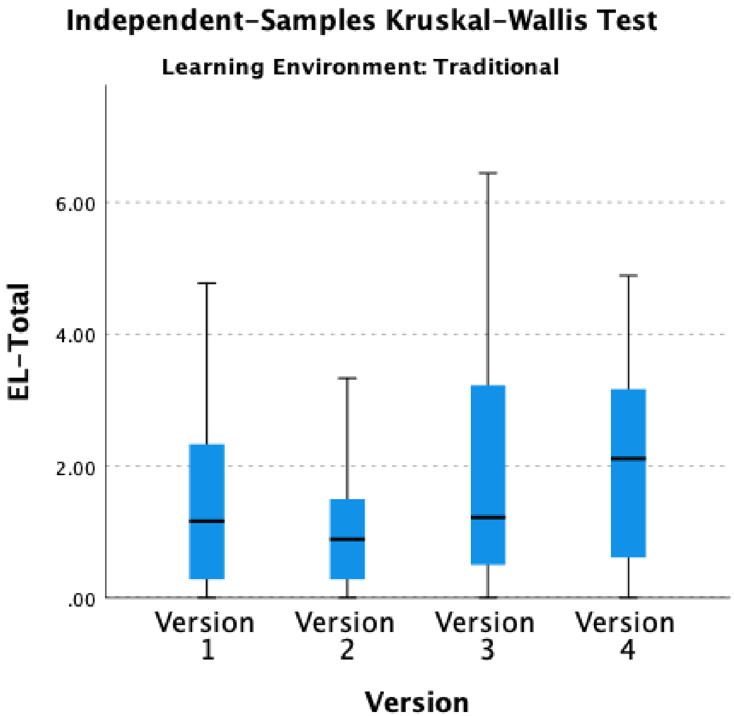

Figure I55. Boxplot of the extraneous cognitive load scores of each version in the traditional learning environment about lightning formation subject matter

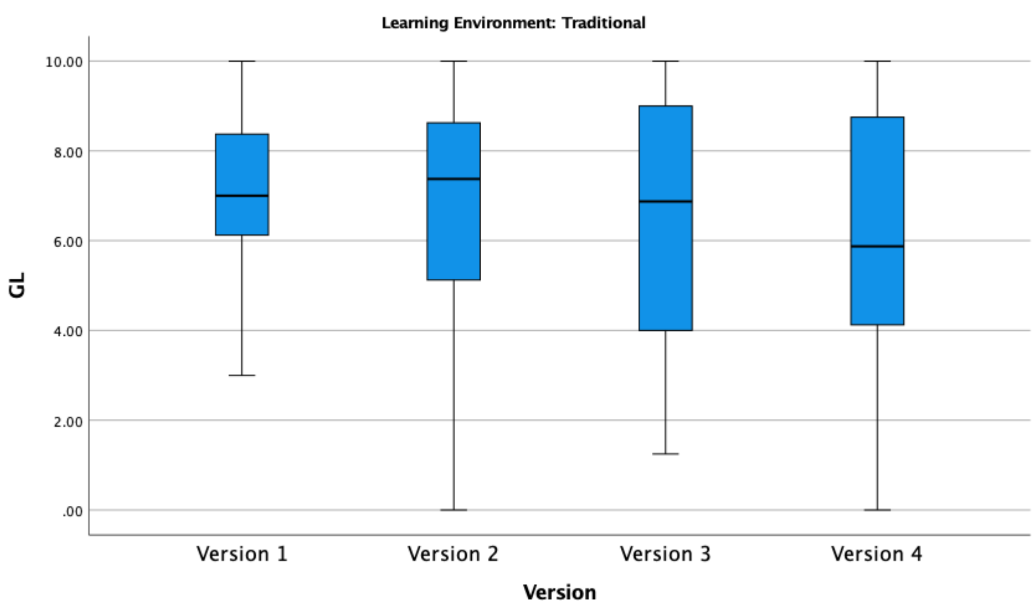

Figure I56. Boxplot of the germane cognitive load scores of each version in the traditional learning environment about lightning formation subject matter

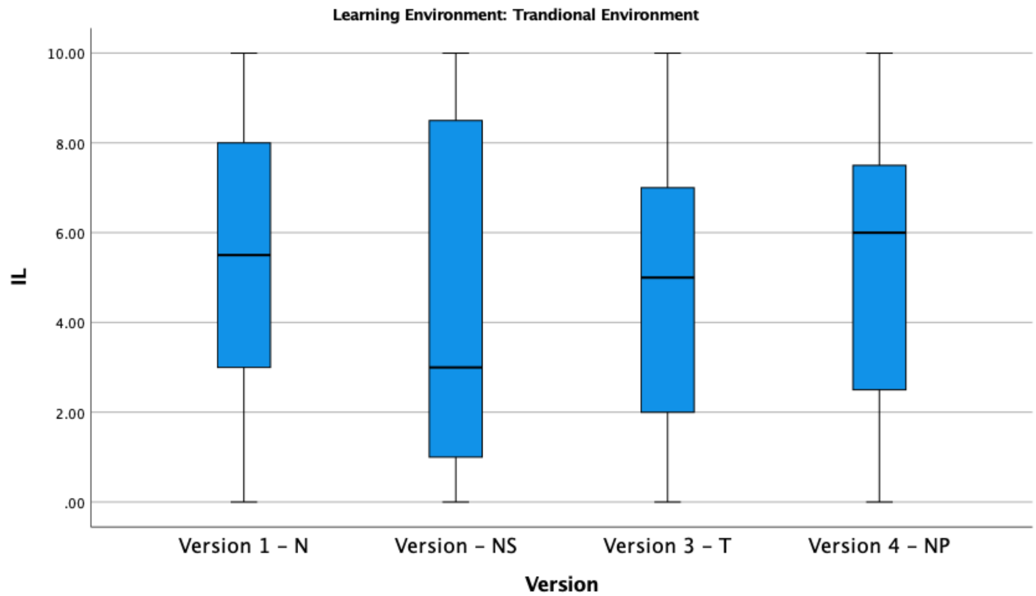

Figure I57. Boxplot of the intrinsic cognitive load scores of each version in the traditional learning environment about cell structure subject matter

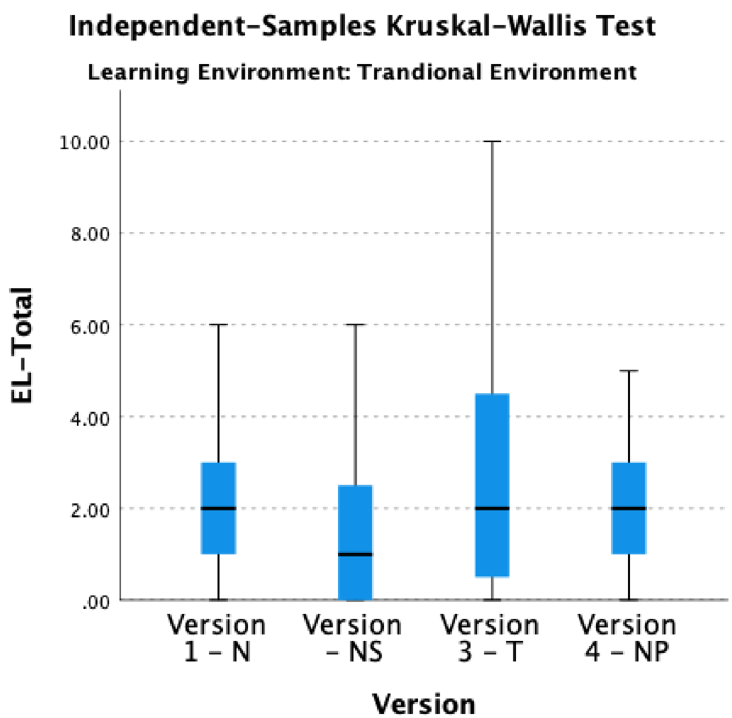

Figure I58. Boxplot of the extraneous cognitive load scores of each version in the traditional learning environment about cell structure subject matter

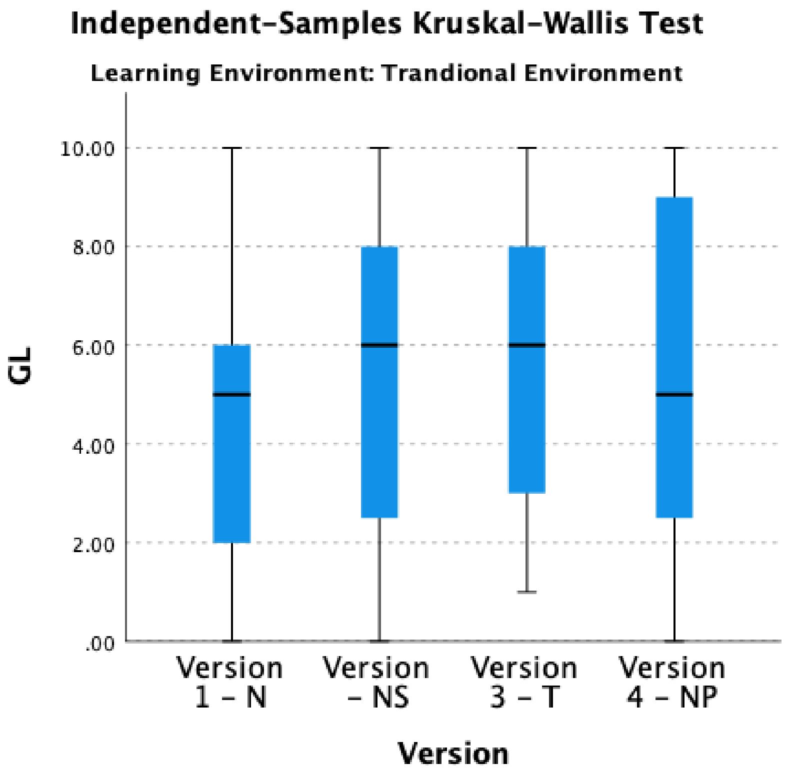

Figure I59. Boxplot of the germane cognitive load scores of each version in the traditional learning environment about cell structure subject matter

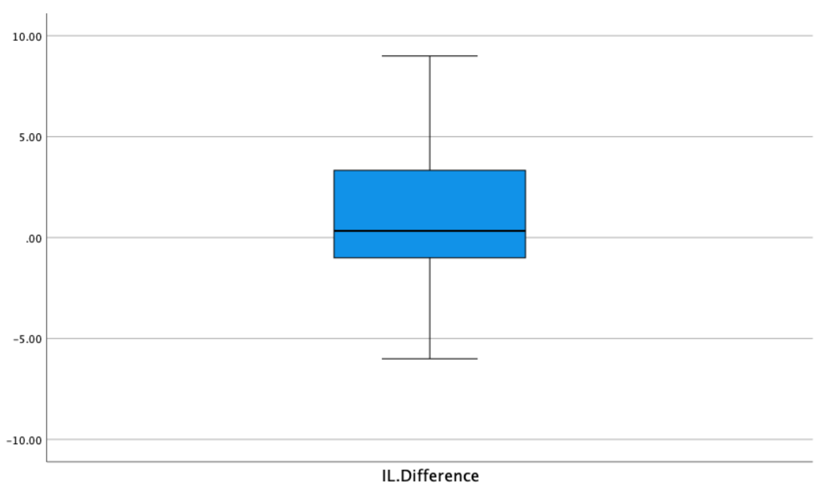

Figure I60. Boxplot of the intrinsic cognitive load scores

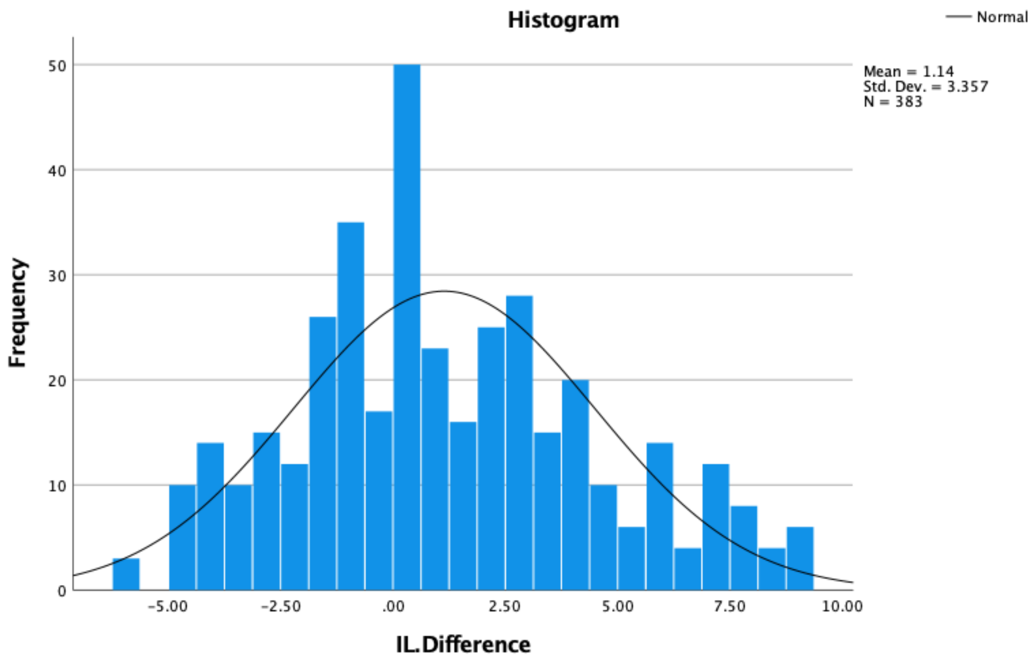

Figure I61. Histogram of the intrinsic cognitive load scores

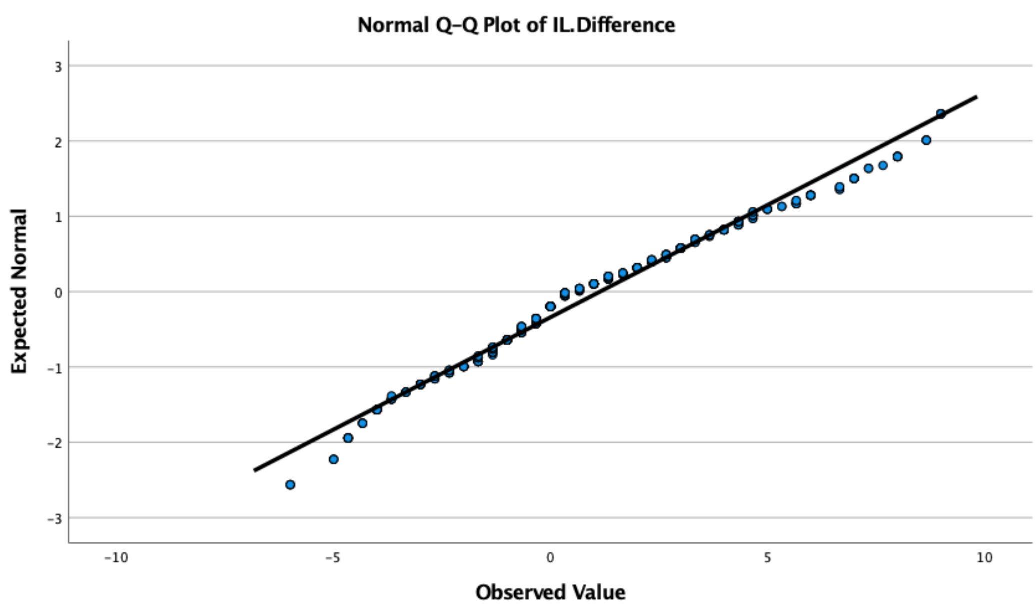

Figure I62. Q-Q plot of the intrinsic cognitive load scores

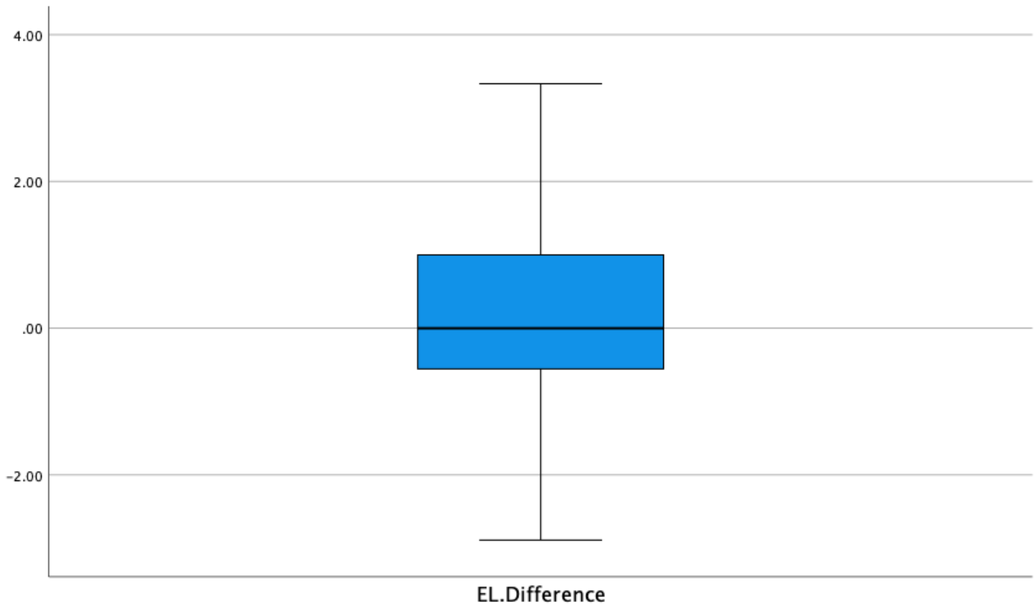

Figure I63. Boxplot of the extraneous cognitive load scores

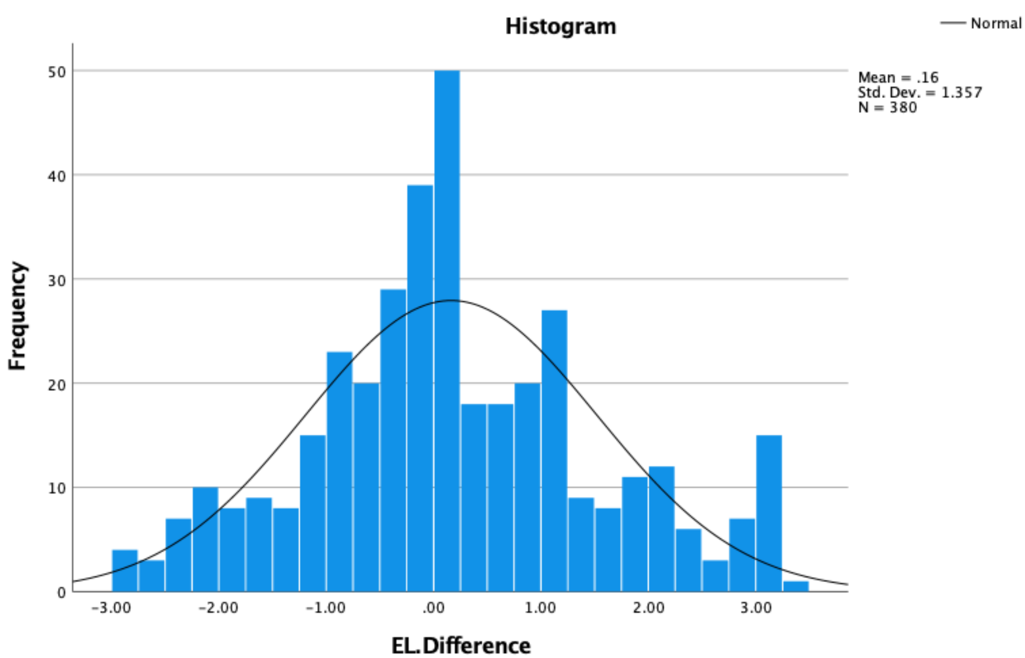

Figure I64. Histogram of the extraneous cognitive load scores

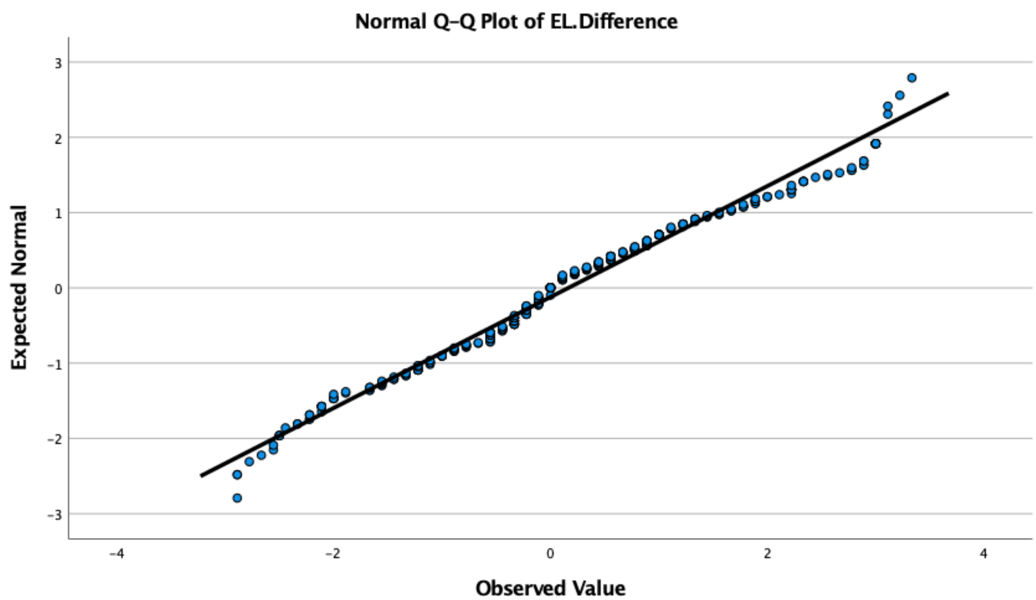

Figure I65. Q-Q plot of the extraneous cognitive load scores

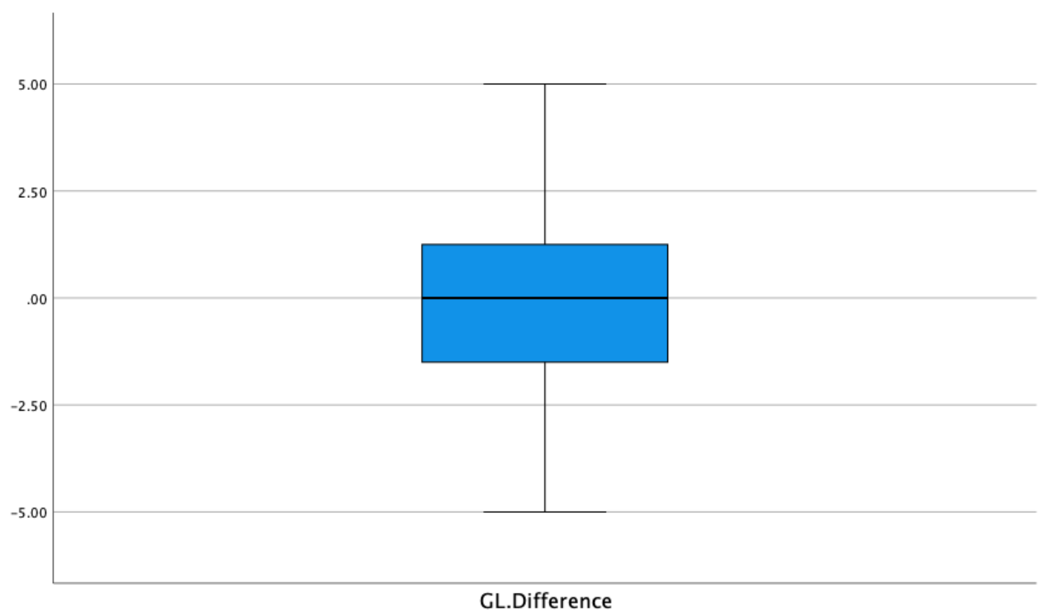

Figure I66. Boxplot of the germane cognitive load scores

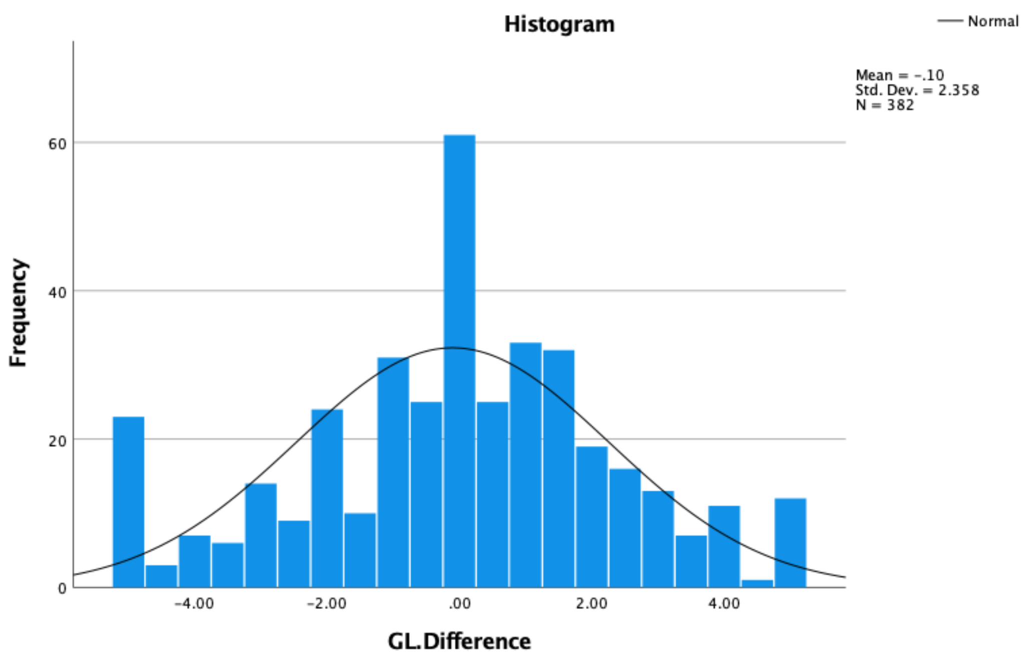

Figure I67. Histogram of the germane cognitive load scores

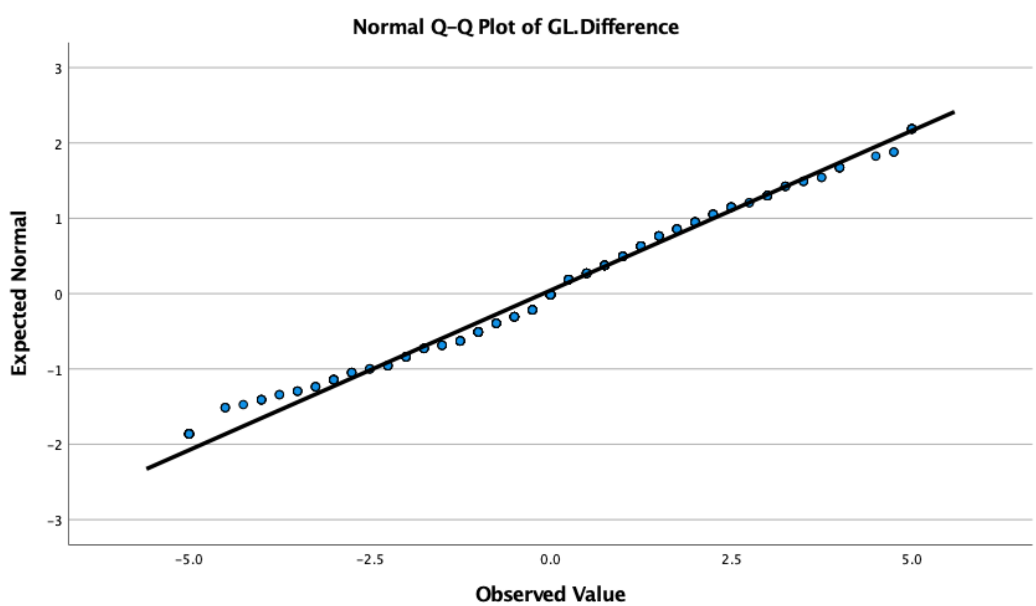

Figure I68. Q-Q plot of the germane cognitive load scores

APPENDIX J

RESULTS OF ASSUMPTION TESTS

Table J1. Homogeneity of Variance Tests for the Transfer Scores of Each Learning Environment

		Levene Statistic	df1	df2	Sig.
Transfer Scores	Based on Mean	.930	2	378	.395

Table J2. Normality Tests for the Transfer Scores of Each Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Virtual Reality	.083	127	.032	.972	127	.010
Augmented Reality	.105	126	.002	.972	126	.011
Traditional	.102	128	.002	.970	128	.006

Table J3. Skewness and Kurtosis Test for the Transfer Scores of Each Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Virtual Reality	127	.062	.215	-.544	.427
Augmented Reality	126	.296	.216	-.522	.428
Traditional	128	.120	.214	-.176	.425

Table J4. Homogeneity of Variance Tests for the Transfer Scores of Each Learning Environment

		Levene Statistic	df1	df2	Sig.
Transfer Scores	Based on Mean	3.033	2	378	.050

Table J5. Normality Tests for the Transfer Scores of Each Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Virtual Reality	.089	128	.014	.974	128	.015
Augmented Reality	.094	126	.008	.977	126	.027
Traditional	.113	127	.000	.959	127	.001

Table J6. Skewness and Kurtosis Test for the Transfer Scores of Each Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Virtual Reality	128	.106	.214	-.588	.425
Augmented Reality	126	.110	.216	-.175	.428
Traditional	127	.364	.215	-.531	.427

Table J7. Homogeneity of Variance Tests for the Intrinsic Cognitive Load Scores of Each Learning Environment

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Load	Based on Mean	3.072	2	380	.050

Table J8. Normality Tests for the Intrinsic Cognitive Load Scores of Each Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Virtual Reality	.073	128	.089	.976	128	.020
Augmented Reality	.097	127	.005	.977	127	.026
Traditional	.082	128	.034	.957	128	.000

Table J9. Skewness and Kurtosis Test for the Intrinsic Cognitive Load Scores of Each Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Virtual Reality	128	.214	.214	-.436	.425
Augmented Reality	127	.224	.215	-.506	.427
Traditional	128	.372	.214	-.662	.425

Table J10. Homogeneity of Variance Tests for the Germane Cognitive Load Scores of Each Learning Environment

		Levene Statistic	df1	df2	Sig.
Germane Cognitive Scores	Based on Mean	.954	2	380	.386

Table J11. Normality Tests for the Germane Cognitive Load Scores of Each Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Virtual Reality	.079	128	.051	.958	128	.001
Augmented Reality	.075	127	.080	.968	127	.004
Traditional	.111	128	.001	.937	128	.000

Table J12. Skewness and Kurtosis Test for the Germane Cognitive Load Scores of Each Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Virtual Reality	128	-.205	.214	-1.026	.425
Augmented Reality	127	-.411	.215	-.346	.427
Traditional	128	-.622	.214	-.464	.425

Table J13. Homogeneity of Variance Tests for the Intrinsic Cognitive Load Scores of Each Learning Environment

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Scores	Based on Mean	1.349	2	380	.261

Table J14. Normality Tests for the Intrinsic Cognitive Load Scores of Each Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Virtual Reality	.129	128	.000	.941	128	.000
Augmented Reality	.138	127	.000	.943	127	.000
Traditional	.123	128	.000	.924	128	.000

Table J15. Skewness and Kurtosis Test for the Intrinsic Cognitive Load Scores of Each Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Virtual Reality	128	.052	.214	-1.151	.425
Augmented Reality	127	.176	.215	-1.091	.427
Traditional	128	.057	.214	-1.298	.425

Table J16. Homogeneity of Variance Tests for the Germane Cognitive Load Scores of Each Learning Environment

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Scores	Based on Mean	2.876	2	380	.058

Table J17. Normality Tests for the Germane Cognitive Load Scores of Each Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Virtual Reality	.147	128	.000	.936	128	.000
Augmented Reality	.149	127	.000	.942	127	.000
Traditional	.107	128	.001	.952	128	.000

Table J18. Skewness and Kurtosis Test for the Germane Cognitive Load Scores of Each Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Virtual Reality	128	-.575	.214	-.351	.425
Augmented Reality	127	-.426	.215	-.563	.427
Traditional	128	.014	.214	-1.020	.425

Table J19. Homogeneity of Variance Tests for the Attention Scores of Each Learning Environment

		Levene Statistic	df1	df2	Sig.
Attention Scores	Based on Mean	1.262	2	380	.284

Table J20. Normality Tests for the Attention Scores of Each Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Virtual Reality	.133	128	.000	.935	128	.000
Augmented Reality	.106	127	.001	.956	127	.000
Traditional	.145	128	.000	.915	128	.000

Table J21. Skewness and Kurtosis Test for the Attention Scores of Each Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Virtual Reality	128	-.528	.214	-.610	.425
Augmented Reality	127	-.382	.215	-.405	.427
Traditional	128	-.662	.214	-.311	.311

Table J22. Homogeneity of Variance Tests for the Relevance Scores of Each Learning Environment

		Levene Statistic	df1	df2	Sig.
Relevance Score	Based on Mean	.222	2	380	.801

Table J23. Normality Tests for the Relevance Scores of Each Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Virtual Reality	.118	128	.000	.942	128	.000
Augmented Reality	.095	127	.007	.946	127	.000
Traditional	.129	128	.000	.938	128	.000

Table J24. Skewness and Kurtosis Test for the Relevance Scores of Each Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Virtual Reality	128	-.407	.214	-.795	.425
Augmented Reality	127	-.338	.215	-.672	.427
Traditional	128	-.594	.214	-.224	.425

Table J25. Homogeneity of Variance Tests for the Confidence Scores of Each Learning Environment

		Levene Statistic	df1	df2	Sig.
Confidence Scores	Based on Mean	1.895	2	380	.152

Table J26. Normality Tests for Confidence Scores of Each Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Virtual Reality	.183	128	.000	.902	128	.000
Augmented Reality	.173	127	.000	.910	127	.000
Traditional	.167	128	.000	.881	128	.000

Table J27. Skewness and Kurtosis Test for the Confidence Scores of Each Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Virtual Reality	128	-.683	.214	-.389	.425
Augmented Reality	127	-.672	.215	-.348	.427
Traditional	128	-.985	.214	-.397	.425

Table J28. Homogeneity of Variance Tests for the Retention Scores of Each Version in the Traditional Learning Environment

		Levene Statistic	df1	df2	Sig.
Retention Scores	Based on Mean	1.247	3	123	.296

Table J29. Normality Tests for the Retention Scores of Each Version in the Traditional Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.169	32	.021	.940	32	.076
NS	.115	32	.200	.947	32	.116
T	.161	32	.033	.887	32	.003
NP	.129	31	.200	.944	31	.109

Table J30. Skewness and Kurtosis Test for the Retention Scores of Each Version in the Traditional Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	.338	.414	-1.028	.809
NS	32	.487	.414	-.698	.809
T	32	.971	.414	.007	.809
NP	31	.449	.421	-.727	.821

Table J31. Homogeneity of Variance Tests for the Transfer Scores of Each Version

		Levene Statistic	df1	df2	Sig.
Transfer Scores	Based on Mean	.349	3	375	.790

Table J32. Normality Tests for the Transfer Scores of Each Version

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.112	96	.005	.968	96	.020
NS	.090	93	.060	.970	93	.029
T	.107	95	.009	.956	95	.003
NP	.088	95	.068	.975	95	.068

Table J33. Skewness and Kurtosis Test for the Transfer Scores of Each Version

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	96	-.057	.246	-.590	.488
NS	93	.253	.250	-.493	.495
T	95	.129	.247	-.805	.490
NP	95	.005	.247	-.775	.490

Table J34. Homogeneity of Variance Tests for the Transfer Scores of Each Version

		Levene Statistic	df1	df2	Sig.
Transfer Scores	Based on Mean	.330	3	378	.804

Table J35. Normality Tests for the Transfer Scores of Each Version

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.114	96	.004	.965	96	.012
NS	.109	94	.008	.966	94	.014
T	.099	96	.022	.968	96	.018
NP	.118	96	.002	.979	96	.131

Table J36. Skewness and Kurtosis Test for the Transfer Scores of Each Version

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	96	.27**4**	.246	-.017	.488
NS	94	.278	.249	-.639	.493
T	96	.140	.246	-.554	.488
NP	96	.130	.246	-.424	.488

Table J37. Homogeneity of Variance Tests for the Retention Scores of Each Version in the Virtual Learning Environment

		Levene Statistic	df1	df2	Sig.
Transfer Scores	Based on Mean	1.414	3	123	.242

Table J38. Normality Tests for the Transfer Scores of Each Version in the Virtual Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.115	32	.200	.966	32	.402
NS	.135	31	.156	.929	31	.042
T	.139	32	.116	.960	32	.280
NP	.113	31	.200	.945	32	.105

Table J39. Skewness and Kurtosis Test for the Transfer Scores of Each Version in the Virtual Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	.174	.414	-.495	.809
NS	31	.178	.421	-.680	.821
T	32	-.216	.414	-.432	.809
NP	32	-.373	.414	-.817	.809

Table J40. Homogeneity of Variance Tests for the Transfer Scores of Each Version in the Virtual Learning Environment

		Levene Statistic	df1	df2	Sig.
Transfer Scores	Based on Mean	.093	3	123	.964

Table J41. Normality Tests for the Transfer Scores of Each Version in the Virtual Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.137	31	.143	.942	31	.091
NS	.144	32	.090	.936	32	.056
T	.140	32	.115	.952	32	.168
NP	.156	32	.047	.956	32	.211

Table J42. Homogeneity of Variance Tests for the Retention Scores of Each Version in the Augmented Learning Environment

		Levene Statistic	df1	df2	Sig.
Transfer Scores	Based on Mean	.604	3	122	.614

Table J43. Normality Tests for the Transfer Scores of Each Version in the Augmented Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.122	32	.200	.973	32	.577
NS	.128	30	.200	.959	30	.298
T	.148	32	.073	.936	32	.059
NP	.165	32	.026	.935	32	.055

Table J44. Homogeneity of Variance Tests for the Transfer Scores of Each Version in the Augmented Learning Environment

		Levene Statistic	df1	df2	Sig.
Transfer Scores	Based on Mean	2.067	3	121	.108

Table J45. Normality Tests for the Transfer Scores of Each Version in the Augmented Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.137	30	.157	.931	30	.052
NS	.139	31	.135	.967	31	.428
T	.097	32	.200*	.979	32	.760
NP	.102	32	.200*	.967	32	.415

Table J46. Homogeneity of Variance Tests for the Retention Scores of Each Version in the Traditional Learning Environment

		Levene Statistic	df1	df2	Sig.
Transfer Scores	Based on Mean	2.681	3	123	.050

Table J47. Normality Tests for the Transfer Scores of Each Version in the Traditional Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.177	32	.012	.914	32	.014
NS	.090	32	.200	.979	32	.779
T	.120	32	.200	.946	32	.108
NP	.132	31	.182	.980	31	.806

Table J48. Skewness and Kurtosis Test for the Transfer Scores of Each Version in the Traditional Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	-.143	.414	-1.212	.809
NS	32	.039	.414	-.470	.809
T	32	.294	.414	-.615	.809
NP	31	-.210	.421	-.124	.821

Table J49. Homogeneity of Variance Tests for the Transfer Scores of Each Version in the Traditional Learning Environment

		Levene Statistic	df1	df2	Sig.
Transfer Scores	Based on Mean	.979	3	121	.405

Table J50. Normality Tests for the Transfer Scores of Each Version in the Traditional Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.152	31	.067	.936	31	.079
NS	.191	30	.007	.914	30	.018
T	.148	32	.071	.952	32	.159
NP	.152	32	.058	.958	32	.235

Table J51. Skewness and Kurtosis Test for the Transfer Scores of Each Version in the Traditional Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	31	.275	.421	.094	.821
NS	30	.779	.427	.015	.833
T	32	.125	.414	-.894	.809
NP	32	.120	.414	-.403	.809

Table J52. Homogeneity of Variance Tests for the Intrinsic Cognitive Load Scores of Each Version

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Load	Based on Mean	1.869	3	376	.134

Table J53. Normality Tests for the Intrinsic Cognitive Load Scores of Each Version

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.076	96	.200	.963	96	.008
NS	.087	93	.081	.955	93	.003
T	.092	96	.046	.958	96	.004
NP	.082	95	.117	.969	95	.025

Table J54. Skewness and Kurtosis Test for the Intrinsic Cognitive Load Scores of Each Version

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	96	-.243	.246	-.645	.488
NS	93	.161	.250	-.863	.495
T	96	.528	.246	-.016	.488
NP	95	-.039	.247	-.832	.490

Table J55. Homogeneity of Variance Tests for the Intrinsic Cognitive Load Scores of Each Version

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Scores	Based on Mean	1.769	3	379	.153

Table J56. Normality Tests for the Intrinsic Cognitive Load Scores of Each Version

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.117	96	.002	.953	96	.002
NS	.174	95	.000	.909	95	.000
T	.140	96	.000	.928	96	.000
NP	.111	96	.005	.940	96	.000

Table J57. Skewness and Kurtosis Test for the Intrinsic Cognitive Load Scores of Each Version

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	96	-.055	.246	-.992	.488
NS	95	-.022	.247	-1.452	.490
T	96	.354	.246	-1.038	.488
NP	96	.156	.246	-1.035	.488

Table J58. Homogeneity of Variance Tests for the Extraneous Cognitive Load Scores of Each Version

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Scores	Based on Mean	1.586	3	379	.192

Table J59. Normality Tests for Extraneous Cognitive Load Scores of Each Version

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.194	96	.000	.875	96	.000
NS	.201	95	.000	.877	95	.000
T	.183	96	.000	.859	96	.000
NP	.204	96	.000	.893	96	.000

Table J60. Skewness and Kurtosis Test for the Extraneous Cognitive Load Scores of Each Version

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	96	.765	.246	-.454	.488
NS	95	.788	.247	-.369	.490
T	96	.246	.246	-.352	.488
NP	96	.246	.246	-.473	.488

Table J61. Homogeneity of Variance Tests for the Germane Cognitive Load Scores of Each Version

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Scores	Based on Mean	.549	3	379	.649

Table J62. Normality Tests for the Germane Cognitive Load Scores of Each Version

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.136	96	.000	.955	96	.002
NS	.143	95	.000	.924	95	.000
T	.126	96	.001	.952	96	.001
NP	.116	96	.003	.940	96	.000

Table J63. Skewness and Kurtosis Test for the Germane Cognitive Load Scores of Each Version

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	96	-.347	.246	-.659	.488
NS	95	-.692	.247	-.242	.490
T	96	-.436	.246	-.588	.488
NP	96	-.207	.246	-.962	.488

Table J64. Homogeneity of Variance Tests for the Intrinsic Cognitive Load Scores of Each Version in the Virtual Reality Learning Environment

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Load	Based on Mean	.497	3	123	.685

Table J65. Normality Tests for the Intrinsic Cognitive Load Scores of Each Version in the Virtual Reality Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.155	32	.049	.939	32	.069
NS	.131	31	.185	.961	31	.317
T	.118	32	.200	.968	32	.453
NP	.141	32	.103	.930	32	.039

Table J66. Skewness and Kurtosis Test for the Intrinsic Cognitive Load Scores of Each Version in the Virtual Reality Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	.003	.414	-.481	.809
NS	31	-.089	.421	-.403	.821
T	32	.079	.414	-.184	.809
NP	32	.103	.414	-1.331	.809

Table J67. Homogeneity of Variance Tests for the Germane Cognitive Load Scores of Each Version in the Virtual Reality Learning Environment

		Levene Statistic	df1	df2	Sig.
Germane Cognitive Load	Based on Mean	1.973	3	124	.121

Table J68. Normality Tests for the Germane Cognitive Load Scores of Each Version in the Virtual Reality Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.117	32	.200	.975	32	.654
NS	.131	32	.174	.914	32	.015
T	.128	32	.196	.939	32	.071
NP	.120	32	.200	.944	32	.098

Table J69. Skewness and Kurtosis Test for the Germane Cognitive Load Scores of Each Version in the Virtual Reality Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	.032	.414	-.751	.809
NS	32	-.546	.414	-.859	.809
T	32	.067	.414	-1.266	.809
NP	32	-.172	.414	-1.220	.809

Table J70. Homogeneity of Variance Tests for the Intrinsic Cognitive Load Scores of Each Version in the Virtual Reality Learning Environment

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Scores	Based on Mean	.421	3	124	.738

Table J71. Normality Tests for the Intrinsic Cognitive Load Scores of Each Version in the Virtual Reality Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.141	32	.106	.950	32	.148
NS	.249	32	.000	.888	32	.003
T	.177	32	.012	.905	32	.009
NP	.165	32	.027	.916	32	.016

Table J72. Skewness and Kurtosis Test for the Intrinsic Cognitive Load Scores of Each Version in the Virtual Reality Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	-.206	.414	-.925	.809
NS	32	-.450	.414	-1.149	.809
T	32	.500	.414	-1.074	.809
NP	32	.313	.414	-.868	.809

Table J73. Homogeneity of Variance Tests for the Intrinsic Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Load	Based on Mean	1.803	3	123	.150

Table J74. Normality Tests for the Intrinsic Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.115	32	.200	.969	32	.474
NS	.111	31	.200	.972	31	.577
T	.127	32	.200	.928	32	.035
NP	.141	32	.107	.966	32	.405

Table J75. Skewness and Kurtosis Test for the Intrinsic Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	.291	.414	-.563	.809
NS	31	.059	.421	-.625	.821
T	32	.648	.414	-.029	.809
NP	32	.049	.414	-.450	.809

Table J76. Homogeneity of Variance Tests for the Extraneous Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

		Levene Statistic	df1	df2	Sig.
Extraneous Cognitive Load	Based on Mean	2.067	3	123	.108

Table J77. Normality Tests for the Extraneous Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.173	32	.016	.877	32	.002
NS	.223	31	.000	.864	31	.001
T	.149	32	.067	.882	32	.002
NP	.158	32	.041	.892	32	.004

Table J78. Skewness and Kurtosis Test for the Extraneous Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	.430	.414	-1.350	.809
NS	31	.751	.421	-.855	.821
T	32	.876	.414	-.286	.809
NP	32	.307	.414	-1.482	.809

Table J79. Homogeneity of Variance Tests for the Intrinsic Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Scores	Based on Mean	1.230	3	123	.302

Table J80. Normality Tests for the Intrinsic Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.162	32	.032	.949	32	.111
NS	.193	31	.005	.885	31	.003
T	.165	32	.026	.920	32	.021
NP	.112	32	.200	.939	32	.069

Table J81. Skewness and Kurtosis Test for the Intrinsic Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	.013	.414	-.894	.809
NS	31	-.148	.421	-1.463	.821
T	32	.469	.414	-.896	.809
NP	32	.378	.414	-.618	.809

Table J82. Homogeneity of Variance Tests for the Extraneous Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

		Levene Statistic	df1	df2	Sig.
Extraneous Cognitive Scores	Based on Mean	2.359	3	123	.075

Table J83. Normality Tests for the Extraneous Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.161	32	.035	.881	32	.002
NS	.239	31	.000	.830	31	.000
T	.243	32	.000	.834	32	.000
NP	.135	32	.145	.924	32	.027

Table J84. Skewness and Kurtosis Test for the Extraneous Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	.892	.414	.244	.809
NS	31	.895	.421	-.418	.821
T	32	1.178	.414	.434	.809
NP	32	-.005	.414	-1.250	.809

Table J85. Homogeneity of Variance Tests for the Germane Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

		Levene Statistic	df1	df2	Sig.
Extraneous Cognitive Scores	Based on Mean	2.573	3	123	.057

Table J86. Normality Tests for the Germane Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.197	32	.003	.926	32	.030
NS	.152	31	.065	.917	31	.019
T	.132	32	.167	.924	32	.027
NP	.142	32	.099	.937	32	.061

Table J87. Skewness and Kurtosis Test for the Germane Cognitive Load Scores of Each Version in the Augmented Reality Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	-.285	.414	-.494	.809
NS	31	-.417	.421	-.946	.821
T	32	-.315	.414	-1.019	.809
NP	32	-.105	.414	-.682	.809

Table J88. Homogeneity of Variance Tests for the Intrinsic Cognitive Load Scores of Each Version in the Traditional Learning Environment

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Load	Based on Mean	1.749	3	123	.160

Table J89. Normality Tests for the Intrinsic Cognitive Load Scores of Each Version in the Traditional Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.084	32	.200	.943	32	.090
NS	.156	32	.047	.886	32	.003
T	.151	32	.060	.926	32	.031
NP	.109	31	.200	.964	31	.367

Table J90. Skewness and Kurtosis Test for the Intrinsic Cognitive Load Scores of Each Version in the Traditional Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	.013	.414	-1.077	.809
NS	32	.842	.414	-.214	.809
T	32	.392	.414	-1.090	.809
NP	31	-.295	.421	-.272	.821

Table J91. Homogeneity of Variance Tests for the Germane Cognitive Load Scores of Each Version in the Traditional Learning Environment

		Levene Statistic	df1	df2	Sig.
Germane Cognitive Load	Based on Mean	1.825	3	123	.146

Table J92. Normality Tests for the Germane Cognitive Load Scores of Each Version in the Traditional Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.119	31	.200	.929	31	.041
NS	.124	32	.200	.925	32	.028
T	.163	32	.030	.906	32	.009
NP	.165	32	.027	.936	32	.059

Table J93. Skewness and Kurtosis Test for the Germane Cognitive Load Scores of Each Version in the Traditional Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	31	-.339	.421	-.382	.821
NS	32	-.830	.414	.203	.809
T	32	-.511	.414	-1.016	.809
NP	32	-.443	.421	-.748	.809

Table J94. Homogeneity of Variance Tests for the Intrinsic Cognitive Load Scores of Each Version in the Traditional Learning Environment

		Levene Statistic	df1	df2	Sig.
Intrinsic Cognitive Scores	Based on Mean	1.348	3	124	.262

Table J95. Normality Tests for the Intrinsic Cognitive Load Scores of Each Version in the Traditional Learning Environment

Groups	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
N	.113	32	.200	.934	32	.052
NS	.210	32	.001	.863	32	.001
T	.126	32	.200	.930	32	.039
NP	.129	32	.193	.925	32	.029

Table J96. Skewness and Kurtosis Test for the Intrinsic Cognitive Load Scores of Each Version in the Traditional Learning Environment

Groups	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
N	32	-.054	.414	-1.126	.809
NS	32	.332	.414	-1.552	.809
T	32	.161	.414	-1.012	.809
NP	32	-.188	.414	-1.256	.809

Table J97. Normality Tests for the Intrinsic Cognitive Load Scores

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Intrinsic Cognitive Load Scores	.106	383	.000	.978	383	.000

Table J98. Skewness and Kurtosis Test for the Intrinsic Cognitive Load Scores

	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Intrinsic Cognitive Load Scores	383	.351	.125	-.375	.249

Table J99. Normality Tests for the Extraneous Cognitive Load Scores

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Extraneous Cognitive Load Scores	.088	380	.000	.981	380	.000

Table J100. Skewness and Kurtosis Test for the Extraneous Cognitive Load Scores

	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Extraneous Cognitive Load Scores	380	.193	.125	-.180	.250

Table J101. Normality Tests for the Germane Cognitive Load Scores

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	*df*	Sig.	Statistic	*df*	Sig.
Germane Cognitive Load Scores	.086	382	.000	.977	382	.000

Table J102. Skewness and Kurtosis Test for the Germane Cognitive Load Scores

	N	Skewness		Kurtosis	
		Statistic	Std. Error	Statistic	Std. Error
Germane Cognitive Load Scores	382	-.166	.125	-.143	.249

REFERENCES

Akbulut, A., Catal, C., & Yıldız, B. (2018). On the effectiveness of virtual reality in the education of software engineering. *Computer Applications in Engineering Education, 26*(4), 918-927. https://doi.org/10.1002/cae.21935

Akbulut, Y. (2011). Bilişsel yük kuramı ve çoklu ortam tasarımı [Cognitive load theory and multimedia design]. In H.F. Odabaşı & Ö.Ö. Dursun (Eds.), *Çoklu ortam tasarımı [Multimedia design]* (pp. 37-55). Ankara: Pegem.

Akkoyunlu, B., & Yılmaz, M. (2005). Türetimci çoklu ortam öğrenme kuramı. *Hacettepe Üniversitesi Eğitim Fakültesi,* (28), 9-18.

Akman, E. (2019). *İlkokul matematik dersi kesirler konusunda geliştirilen sanal gerçeklik uygulamasının farklı değişkenler açısından etkisinin incelenmesi* (PhD thesis). Available from YOKSIS database (Order No. 587141). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/giris.jsp

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, 20, 1-11. https://doi.org/10.1016/j.edurev.2016.11.002

Akçayır, M., Akçayır, G., Pektaş, H. M., & Ocak, M. A. (2016). Augmented reality in science laboratories: The effects of augmented reality on university students' laboratory skills and attitudes toward science laboratories. *Computers in Human Behavior, 57*, 334-342. https://doi.org/10.1016/j.chb.2015.12.054

Alhalabi, W. (2016). Virtual reality systems enhance students' achievements in engineering education. *Behaviour & Information Technology*, 35(11), 919–925.

Allcoat, D., & von Mühlenen, A. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology*, 26.

Almasseri, M., & AlHojailan, M. I. (2019). How flipped learning based on the cognitive theory of multimedia learning affects students' academic achievements. *Journal of Computer Assisted Learning, 35*(6), 769-781.

Altinpulluk, H., Kilinc, H., Firat, M., & Yumurtaci, O. (2020). The influence of segmented and complete educational videos on the cognitive load, satisfaction, engagement, and academic achievement levels of learners. *Journal of Computers in Education*, *7*(2), 155-182.

Altmeyer, K., Kapp, S., Thees, M., Malone, S., Kuhn, J., & Brünken, R. (2020). The use of augmented reality to foster conceptual knowledge acquisition in STEM laboratory courses—Theoretical background and empirical results. *British Journal of Educational Technology*, *51*(3), 611-628.

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching* (1st ed.). Jossey-Bass.

Andersen, M. S., & Makransky, G. (2021). The validation and further development of a multidimensional cognitive load scale for virtual environments. *Journal of Computer Assisted Learning*, *37*(1), 183-196.

Asthana, A. (2008). Multimedia in Education. In: Furht, B. (eds) *Encyclopedia of multimedia*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-78414-4_140

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94*(2), 416–427.

Avcı, Ş. K., Coklar, A. N., & İstanbullu, A. (2019). The effect of three dimensional virtual environments and augmented reality applications on the learning achievement: A meta-analysis study. *Egitim ve Bilim-Education and Science, 44*(198), 149–182.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications, 21*(6), 34–47.

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environment, 6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Babu, S. K., Krishna, S., Unnikrishnan, R., & Bhavani, R. R. (2018). Virtual reality learning environments for vocational education: A comparison study with conventional instructional media on knowledge retention. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)* (pp. 385–389). IEEE.

Baceviciute, S., Lucas, G., Terkildsen, T., & Makransky, G. (2022). Investigating the redundancy principle in immersive virtual reality environments: An eye-tracking and EEG study. *Journal of Computer Assisted Learning, 38*(1), 120-136.

Baceviciute, S., Mottelson, A., Terkildsen, T., & Makransky, G. (2020). Investigating representation of text and audio in educational VR using learning outcomes and EEG. *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-13).

Baddeley, A. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review*, *5*(3), 243-260.

Bakhshialiabad, H., Bakhshi, M., & Hassanshahi, G. (2015). Students' perceptions of the academic learning environment in seven medical sciences courses based on DREEM. *Advances in Medical Education and Practice, 6*, 195–203.

Batdi, V., & Talan, T. (2019). Augmented reality applications: A Meta-analysis and thematic analysis. *Turkish Journal of Education*, *8*(4), 276-297.

Billinghurst, M., Kato, H., & Poupyrev, I. (2001). The MagicBook: a transitional AR interface. *Computers & Graphics, 25*(5), 745–753.

Boucheix, J.-M., & Schneider, E. (2009). Static and animated presentations in learning dynamic mechanical systems. *Learning and Instruction*, *19*(2), 112-127.

Bratfisch, O., Borg, G., & Dornic, S. (1972). *Perceived Item-difficulty in Three Tests of Intellectual Performance Capacity.* Report No. 29. Stockholm: Institute of Applied Psychology.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage (Orlando, Fla.), 5*(1), 49-62.

Bressler, D. M., & Bodzin, A. M. (2013). A mixed methods assessment of students' flow experiences during a mobile augmented reality science game. *Journal of Computer Assisted Learning, 29*(6), 505-517.

Brown, A., & Green, T. (2016). Virtual reality: Low-cost tools and resources for the classroom. *Techtrends, 60*(5), 517-519.

Bruder, E. A. (2018). *Manipulation of cognitive load in simulation-based medical education* (PhD thesis). Queen's University, Canada

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 53-61.

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science, 32*(1/2), 115-132.

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49*(2), 109-119.

Burdea, G. C., & Coiffet, P. (2003). *Virtual reality technology*. John Wiley & Sons.

Bursztyn, N., Shelton, B., Walker, A., & Pederson, J. (2017). Increasing undergraduate interest to learn geoscience with GPS-based augmented reality field trips on students’ own smartphones. *GSA Today, 27*(6), 4-10.

Buttussi, F., & Chittaro, L. (2017). Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE Transactions on Visualization and Computer Graphics, 24*(2), 1063-1076.

Cai, S., Chiang, F., Sun, Y., Lin, C., & Lee, J. J. (2017). Applications of augmented reality-based natural interactive learning in magnetic field instruction. *Interactive Learning Environments, 25*(6), 778–791.

Calle-Bustos, A. M., Juan, M. C., García-García, I., & Abad, F. (2017). An augmented reality game to support therapeutic education for children with diabetes. *PloS One, 12*(9), e0184645-e0184645.

Cascales-Martínez, A., Martínez-Segura, M. J., Pérez-López, D., & Contero, M. (2016). Using an augmented reality enhanced tabletop system to promote learning of mathematics: A case study with students with special educational needs. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(2), 355-380.

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

Chang, C, & Hwang, G. (2018). Trends of mobile technology-enhanced medical education: a review of journal publications from 1998 to 2016. *International Journal of Mobile Learning and Organisation, 12*(4), 373–393.

Chang, C.-Y., Sung, H.-Y., Guo, J.-L., Chang, B.-Y., & Kuo, F.-R. (2022). Effects of spherical video-based virtual reality on nursing students' learning performance in childbirth education training. *Interactive Learning Environments, 30*(3), 400-416.

Chang, K.-E., Chang, C.-T., Hou, H.-T., Sung, Y.-T., Chao, H.-L., & Lee, C.-M. (2014). Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. *Computers and Education, 71*, 185-197.

Chang, M., Evans, M. A., Kim, S., Norton, A., Deater-Deckard, K., & Samur, Y. (2016). The effects of an educational video game on mathematical engagement. *Education and Information Technologies, 21*(5), 1283–1297.

Chang, S., Hsu, T., Kuo, W., & Jong, M. S. (2020). Effects of applying a VR-based two-tier test strategy to promote elementary students' learning performance in a geology class. *British Journal of Educational Technology*, *51*(1), 148-165.

Chen, C. (2020). Impacts of augmented reality and a digital game on students' science learning with reflection prompts in multimedia learning. *Educational Technology Research and Development, 68*(6), 3057–3076.

Chen, C.-H., Chou, Y.-Y., & Huang, C.-Y. (2016). An augmented-reality-based concept map to support mobile learning for science. *The Asia-Pacific Education Researcher, 25*(4), 567-578.

Chen, C., & Wang, C. (2015). Employing augmented-reality-embedded instruction to disperse the imparities of individual differences in earth science learning. *Journal of Science Education and Technology, 24*(6), 835-847.

Chen, C., & Yang, Y. (2020). Investigation of the effectiveness of common representational formats in online learner-paced software training materials. *Innovations in Education and Teaching International, 57*(1), 97-108.

Chen, C., & Yen, P. (2021). Learner control, segmenting, and modality effects in animated demonstrations used as the before-class instructions in the flipped classroom. *Interactive Learning Environments, 29*(1), 44-58.

Chen, Y., & Wang, C. (2018). Learner presence, perception, and learning achievements in augmented-reality-mediated learning environments. *Interactive Learning Environments, 26*(5), 695-708.

Cheng, K. (2017). Reading an augmented reality book: An exploration of learners' cognitive load, motivation, and attitudes. *Australasian Journal of Educational Technology, 33*(4), 53–69.

Cheng, K., & Tsai, C. (2013). Affordances of augmented reality in science learning: Suggestions for future research. *Journal of Science Education and Technology, 22*(4), 449-462.

Cheng, K., & Tsai, C. (2020). Students' motivational beliefs and strategies, perceived immersion and attitudes towards science learning with immersive virtual reality: A partial least squares analysis. *British Journal of Educational Technology, 51*(6), 2139-2158.

Cheon, J., Crooks, S., & Chung, S. (2014). Does segmenting principle counteract the modality principle in instructional animation? *British Journal of Educational Technology, 45*(1), 56–64.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others?. *Psychological Bulletin*, *143*(1), 1.

Chiang, T. H. C., Yang, S. J. H., & Hwang, G. (2014). Students' online interactive patterns in augmented reality-based inquiry activities. *Computers and Education, 78*, 97–108.

Chittaro, L., Corbett, C. L., McLean, G. A., & Zangrando, N. (2018). Safety knowledge transfer through mobile virtual reality: A study of aviation life preserver donning. *Safety Science*, *102*, 159-168.

Choi, H. H., van Merriënboer, J. J. G., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. *Educational Psychology Review*, *26*, 225–244.

Christou, C. (2010). Virtual reality in education. In *Affective, interactive and cognitive methods for e-learning design: creating an optimal education experience* (pp. 228–243). IGI Global.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445.

Clark, R. C., & Mayer, R. E. (2010). Applying the segmenting and pretraining principles: managing complexity by breaking a lesson into parts. *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia*, *207*, 218.

Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (Fourth;4; ed.). Wiley.

Clark, R. C., Nguyen, F., & Sweller, J. (2011). *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. John Wiley & Sons.

Clarke, T., Ayres, P., & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications: Research on cognitive load theory and its design implications for E-learning. *Educational Technology Research and Development, 53*(3), 15–24.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features and redundancy. *Journal of Educational Psychology, 94*(2), 428.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Crooks, S. M., Cheon, J., Inan, F., Ari, F., & Flores, R. (2012). Modality and cueing in multimedia learning: Examining cognitive and perceptual explanations for the modality effect. *Computers in Human Behavior, 28*(3), 1063–1071.

Cubillo, J., Martín, S., Castro, M., Diaz, G., Colmenar, A., & Botički, I. (2014). A learning environment for augmented reality mobile learning. In *2014 IEEE frontiers in education conference (FIE) proceedings* (pp. 1-8). IEEE.

Curcio, I. D., Dipace, A., & Norlund, A. (2016). Virtual realities and education. *Research on Education and Media, 8*(2), 60-68.

Çeken, B., & Taşkın, N. (2022). Multimedia learning principles in different learning environments: a systematic review. *Smart Learning Environments, 9*(1), 1-22.

Çoban, M. (2020). Artırılmış gerçeklikle desteklenmiş videolarla öğretimin akademik başarı, bilişsel yük ve motivasyona etkisi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 20*(2), 1079-1098.

Çoban, M., Bolat, Y. I., & Göksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review, 36*, 100452.

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105–134.

de Oliveira Neto, J. D., Huang, W. D., & de Azevedo Melli, N. C. (2015). Online learning: Audio or text? *Educational Technology Research and Development, 63*(4), 555-573.

Di Natale, A. F., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology, 51*(6), 2006-2033.

Di Serio, Á., Ibáñez, M. B., & Kloos, C. D. (2013). Impact of an augmented reality system on students' motivation for a visual art course. *Computers and Education, 68*, 586-596.

Dousay, T. A., & Trujillo, N. P. (2019). An examination of gender and situational interest in multimedia learning environments. *British Journal of Educational Technology, 50*(2), 876-887.

Dubovi, I., Levy, S. T., & Dagan, E. (2017). Now I know how! The learning process of medication administration among nursing students with non-immersive desktop virtual reality simulation. *Computers and Education, 113*, 16-27.

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. *Handbook of research on educational communications and technology* (pp. 735–745). Springer, New York.

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology, 18(1),* 7–22.

Dunn, O. (1964). Multiple comparisons using rank sums. *Technometrics*, *6*(3), 241–252.

Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction, 28*, 48-63.

Erden, M., & Altun, S. (2006). *Learning styles*. Istanbul: Morpa Culture Publications.

Estapa, A., & Nadolny, L. (2015). The effect of an augmented reality enhanced mathematics lesson on student achievement and motivation. *Journal of STEM Education, 16*(3), 40.

Ferdig, R., Gandolfi, E., & Immel, Z. (2018). Educational opportunities for immersive virtual reality. In J. Voogt, G. Knezek, R. Christensen, & K. W. Lai (Eds.), *Second handbook of information technology in primary and secondary education*, (pp. 955–966). Springer.

Fiorella, L., Vogel-Walcutt, J. J., & Schatz, S. (2012). Applying the modality principle to real-time feedback and the acquisition of higher-order cognitive skills. *Educational Technology Research and Development, 60*(2), 223-238.

Fong, S. F., Lily, L. P. L., & Por, F. P. (2012). Reducing cognitive overload among students of different anxiety levels using segmented animation. *Procedia, Social and Behavioral Sciences, 47*, 1448–1456.

Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: state of the art and perspectives. In *The International Scientific Conference eLearning and Software for Education* (Vol. 1, No. 133, pp. 10–1007).

Garzón, J., & Acevedo, J. (2019). Meta-analysis of the impact of Augmented Reality on students' learning gains. *Educational Research Review, 27*, 244-260.

Garzón, J., Pavón, J. & Baldiris, S. (2019). Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality* 23, 447–459. https://doi.org/10.1007/s10055-019-00379-9

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*(4), 313–331.

Godwin-Jones, R. (2016). Augmented reality and language learning: From annotated vocabulary to place-based mobile games. *Language Learning & Technology, 20*(3), 9–19.

Gonzato, J. C., Arcila, T., & Crespin, B. (2008). Virtual objects on real oceans. In *GRAPHICON'2008* (pp. 49-54).

Gopalan, V., Zulkifli, A. N., Mohamed, N. F. F., Alwi, A., Che Mat, R., Abu Bakar, J. A., & Saidin, A. Z. (2015). Evaluation of E-star: An enhanced science textbook using augmented reality among lower secondary school students. *Jurnal Teknologi*, *77*(29), 55-61.

Graham, M., Zook, M., & Boulton, A. (2022). Augmented reality in urban places: contested content and the duplicity of code. *Machine Learning and the City: Applications in Architecture and Urban Design*, 341–366.

Greenberg, K., Zheng, R., Gardner, M., & Orr, M. (2021). Individual differences in visuospatial working memory capacity influence the modality effect. *Journal of Computer Assisted Learning*, *37*(3), 735–744.

Greenwald, S. W., Corning, W., Funk, M., & Maes, P. (2018). Comparing learning in virtual reality with learning on a 2D screen using electrostatics activities. *J.UCS (Annual Print and CD-ROM Archive Ed.), 24*(2), 220–245.

Gun, E., & Atasoy, B. (2017). The effects of augmented reality on elementary school students' spatial ability and academic achievement. *Egitim Ve Bilim-Education and Science, 42*(191), 31-51.

Haji, F. A., Rojas, D., Childs, R., de Ribaupierre, S., & Dubrowski, A. (2015). Measuring cognitive load: Performance, mental effort and simulation task complexity. *Medical Education, 49*(8), 815-827.

Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education, 8*(1), 1-32.

Harrington, C. M., Kavanagh, D. O., Wright Ballester, G., Wright Ballester, A., Dicker, P., Traynor, O., Hill, A., & Tierney, S. (2018). 360° operative videos: A randomised cross-over study evaluating attentiveness and information retention. *Journal of Surgical Education, 75*(4), 993–1000.

Harskamp, E. G., Mayer, R. E., & Suhre, C. (2007). Does the modality principle for multimedia learning apply to science classrooms? *Learning and Instruction, 17*(5), 465-477.

Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology, 21*(6), 713-729.

Hassanabadi, H., Robatjazi, E. S., & Savoji, A. P. (2011). Cognitive consequences of segmentation and modality methods in learning from instructional animations. *Procedia-Social and Behavioral Sciences, 30*, 1481-1487.

Henderson, S., & Feiner, S. (2010). Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE Transactions on Visualization and Computer Graphics, 17*(10), 1355-1368.

Hsiao, K., Chen, N., & Huang, S.-Y. (2012). Learning while exercising for science education in augmented reality among adolescents. *Interactive Learning Environments, 20*(4), 331–349.

Huang, X., Zou, D., Cheng, G., & Xie, H. (2021). A systematic review of AR and VR enhanced language learning. *Sustainability, 13*(9), 4639.

Huck, S. (2012). Multivariate tests on means. *Reading Statistics and Research. 6th ed. Boston (MA): Pearson*, 458–478.

Hwang, G., Wu, P., Chen, C., & Tu, N. (2016). Effects of an augmented reality-based educational game on students' learning achievements and attitudes in real-world observations. *Interactive Learning Environments, 24*(8), 1895-1906.

Hwang, W., & Hu, S. (2013). Analysis of peer learning behaviors using multiple representations in virtual reality and their impacts on geometry problem solving. *Computers and Education, 62*, 308-319.

Ibrahim, M., Antonenko, P. D., Greenwood, C. M., & Wheeler, D. (2012). Effects of segmenting, signalling, and weeding on learning from educational video. *Learning, Media and Technology, 37*(3), 220–235.

Ibáñez, M., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers and Education, 123*, 109-123.

Ibáñez, M. B., Di Serio, Á., Villarán, D., & Delgado Kloos, C. (2014). Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers and Education, 71*, 1-13.

Inan, F. A., Crooks, S. M., Cheon, J., Ari, F., Flores, R., Kurucay, M., & Paniukov, D. (2015). The reverse modality effect: Examining student learning from interactive computer-based instruction. *British Journal of Educational Technology, 46*(1), 123–130.

Izmirli, S., & Kurt, A. A. (2016). Effects of modality and pace on achievement, mental effort, and positive affect in multimedia learning environments. *Journal of Educational Computing Research, 54*(3), 299-325.

İzmirli, S. (2012). Öğrenen ve sistem hızında ilerleyen farklı çoklu ortam sunum türlerinin çeşitli değişkenler açısından incelenmesi (PhD thesis). Available from YOKSIS database (Order No. 315449). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/giris.jsp

Jamali, S. S., Shiratuddin, M. F., Wong, K. W., & Oskam, C. L. (2015). Utilising mobile-augmented reality for learning human anatomy. *Procedia, Social and Behavioral Sciences, 197*, 659-668.

Jeung, H. J., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17*(3), 329-345.

Jimenez, Y. A., Cumming, S., Wang, W., Stuart, K., Thwaites, D. I., & Lewis, S. J. (2018). Patient education using virtual reality increases knowledge and positive experience for breast cancer patients undergoing radiation therapy. Patient education using virtual reality increases knowledge and positive experience for breast cancer patients undergoing radiation therapy. *Supportive Care in Cancer, 26*(8), 2879–2888.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*(4), 509–539.

Kalyuga, S. (2009). *Cognitive load factors in instructional design for advanced learners*. Nova Science Publishers, Incorporated.

Kamarainen, A. M., Metcalf, S., Grotzer, T., Browne, A., Mazzuca, D., Tutwiler, M. S., & Dede, C. (2013). EcoMOBILE: Integrating augmented reality and probeware with environmental education field trips. *Computers and Education, 68*, 545-556.

Kaufmann, H., & Dünser, A. (2007). Summary of usability evaluations of an educational augmented reality application. In *Virtual Reality: Second International Conference, ICVR 2007, Held as part of HCI International 2007, Beijing, China, July 22-27, 2007. Proceedings 2* (pp. 660-669). Springer Berlin Heidelberg.

Kavanagh, S., Luxton-Reilly, A., Wuensche, B., & Plimmer, B. (2017). A systematic review of virtual reality in education. *Themes in Science and Technology Education, 10*(2), 85–119.

Keller, J. (2009). *Motivational design for learning and performance: The ARCS model approach* (1. Aufl. ed.). Springer-Verlag.

Keller, J. M. (1979). Motivation and instructional design: A theoretical perspective. *Journal of Instructional Development, 2*(4), 26-34.

Kester, L., Kirschner, P. A., & Van Merriënboer, J. J. (2004). Timing of information presentation in learning statistics. *Instructional Science, 32*(3), 233-252.

Kester, L., Kirschner, P. A., & van Merriënboer, J. J. (2006). Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology, 31*(2), 167-185.

Kester, L., Lehnen, C., Van Gerven, P. W., & Kirschner, P. A. (2006). Just-in-time, schematic supportive information presentation during cognitive skill acquisition. *Computers in Human Behavior, 22*(1), 93-112.

Klingenberg, S., Fischer, R., Zettler, I., & Makransky, G. (2023). Facilitating learning in immersive virtual reality: Segmentation, summarizing, both or none? *Journal of Computer Assisted Learning, 39*(1), 218-230.

Kuehl, T., Eitel, A., Damnik, G., & Koerndle, H. (2014). The impact of disfluency, pacing, and students' need for cognition on learning with multimedia. *Computers in Human Behavior, 35*, 189-198.

Kutbay, E., & Akpinar, Y. (2020). Investigating modality, redundancy and signaling principles with abstract and concrete representation. *International Journal of Education in Mathematics, Science and Technology, 8*(1), 131-145.

Küçük, S., Kapakin, S., & Göktaş, Y. (2016). Learning anatomy via mobile augmented reality: Effects on achievement and cognitive load: Learning anatomy. *Anatomical Sciences Education, 9*(5), 411-421.

Küçük, S., Yılmaz, R. M., & Göktaş, Y. (2014). Augmented reality for learning english: Achievement, attitude and cognitive load levels of students. *Egitim Ve Bilim-Education and Science, 39*(176), 393-404.

Lai, A. F., Chen, C. H., & Lee, G. Y. (2019). An augmented reality-based learning approach to enhancing students' science reading performances from the perspective of the cognitive load theory. *British Journal of Educational Technology, 50*(1), 232–247.

Laine, T. H., Nygren, E., Dirin, A., & Suk, H.-J. (2016). Science spots AR: A platform for science learning games with augmented reality. *Educational Technology Research and Development, 64*(3), 507-531.

Law, N., Niederhauser, D. S., Christensen, R., & Shear, L. (2016). A multilevel system of quality technology-enhanced learning and teaching indicators. *Educational Technology & Society, 19*(3), 72-83.

Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology*, *17*(4), 401–418.

Lee, E. A., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers and Education, 79*, 49-58.

Lee, S., & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology, 71*(5), 353-360.

Lee, S. H., Sergueeva, K., Catangui, M., & Kandaurova, M. (2017). Assessing google cardboard virtual reality as a content delivery system in business classrooms. *Journal of Education for Business, 92*(4), 153-160.

Leppink, J., Gog van, T., Paas, F., & Sweller, J. (2015). Cognitive load theory: researching and planning teaching to maximise learning. *Researching Medical Education*, 207-218.

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058-1072.

Li, J., Antonenko, P. D., & Wang, J. (2019). Trends and issues in multimedia learning research in 1996–2016: A bibliometric analysis. *Educational Research Review, 28*, 100282.

Liberman, L., & Dubovi, I. (2023). The effect of the modality principle to support learning with virtual reality: An eye-tracking and electrodermal activity study. *Journal of Computer Assisted Learning, 39*(2), 547-557.

Lin, H. K., Chen, M., & Chang, C. (2015). Assessing the effectiveness of learning solid geometry by using an augmented reality-assisted learning system. *Interactive Learning Environments, 23*(6), 799–810.

Lin, H. K., Hsieh, M., Wang, C., Sie, Z., & Chang, S. (2011). Establishment and Usability Evaluation of an Interactive AR Learning System on Conservation of Fish. *Turkish Online Journal of Educational Technology-TOJET*, *10*(4), 181-187.

Liou, H., Yang, S. J. H., Chen, S. Y., & Tarng, W. (2017). The influences of the 2D image-based augmented reality and virtual reality on student learning. *Educational Technology & Society, 20*(3), 110–121.

Liu, R., Wang, L., Koszalka, T. A., & Wan, K. (2022). Effects of immersive virtual reality classrooms on students' academic achievement, motivation and cognitive load in science lessons. *Journal of Computer Assisted Learning, 38*(5), 1422-1433.

Liu, Y., Jang, B. G., & Roy-Campbell, Z. (2018). Optimum input mode in the modality and redundancy principles for university ESL students' multimedia learning. *Computers and Education*, *127*, 190–200.

Loorbach, N., Peters, O., Karreman, J., & Steehouder, M. (2015). Validation of the instructional materials motivation survey (IMMS) in a self-directed instructional setting aimed at working with technology. *British Journal of Educational Technology*, *46*(1), 204-218.

Lu, S., & Liu, Y. (2015). Integrating augmented reality technology to enhance children's learning in marine education. *Environmental Education Research, 21*(4), 525-541.

Ludwig, C., & Reimann, C. (2005). *Augmented reality: Information at focus.* Cooperative Computing & Communication Laboratory (Volume 4. No. 1). Universität Paderborn.

Lumby, J. (2011). Enjoyment and learning: Policy and secondary school learners' experience in England. *British Educational Research Journal*, *37*(2), 247–264.

Luo, H., Li, G., Feng, Q., Yang, Y., & Zuo, M. (2021). Virtual reality in K-12 and higher education: A systematic review of the literature from 2000 to 2019. *Journal of Computer Assisted Learning*, *37*(3), 887-901.

Lusk, D. L., Evans, A. D., Jeffrey, T. R., Palmer, K. R., Wikstrom, C. S., & Doolittle, P. E. (2009). Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. *British Journal of Educational Technology*, *40*(4), 636–651.

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, *66*(5), 1141-1164.

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, *60*, 225-236.

Mantovani, F., Castelnuovo, G., Gaggioli, A., & Riva, G. (2003). Virtual reality training for health-care professionals. *Cyberpsychology & Behavior*, *6*(4), 389-395.

Manzoor, A. (2016). Technology-Enabled Learning Environments. In *Handbook of research on applied learning theory and design in modern education*. IGI Global, 2016. p. 545–559.

Martin-Gonzalez, A., Chi-Poot, A., & Uc-Cetina, V. (2016). Usability evaluation of an augmented reality system for teaching Euclidean vectors. *Innovations in Education and Teaching International*, *53*(6), 627-636.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1–19.

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press.

Mayer, R. (2009). *Multimedia learning* (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511811678

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning, 33*(5), 403–423. https://doi.org/https://doi.org/10.1111/jcal.12197.

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88*(1), 64–73. https://doi.org/10.1037/0022-0663.88.1.64

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*(2), 390-397.

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 96*(2), 389–395.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. *The Cambridge handbook of multimedia learning* (pp. 279-315). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.015.

Mayer, R. E., Howarth, J. T., Kaplan, M., & Hanna, S. (2018). Applying the segmenting principle to online geography slideshow lessons. *Educational Technology Research and Development, 66*(3), 563–577.

Mayer, R. E., & Jackson, J. (2005). The case for coherence in scientific explanations: Quantitative details can hurt qualitative understanding. *Journal of Experimental Psychology. Applied, 11*(1), 13–18.

Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages through pre-training: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology. Applied, 8*(3), 147-154.

Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94*(1), 171-185.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312-320.

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

Mayer, R. E., & Pilegard, C. (2014). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 316–344). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.016.

Mayer, R. E., Wells, A., Parong, J., & Howarth, J. T. (2019). Learner control of the pacing of an online slideshow lesson: Does segmenting help? *Applied Cognitive Psychology, 33*(5), 930–935.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers and Education, 70,* 29-40.

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers and Education,140*, 103603.

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers and Education*, *56*(3), 769-780.

Milgram, P., & Kishino, F. (1994). Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems, 77*(12), 1321–1329.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 101*(2), 343-352. https://doi.org/10.1037/h0043158.

Molina-Carmona, R., Pertegal-Felices, M. L., Jimeno-Morenilla, A., & Mora-Mora, H. (2018). Virtual reality learning activities for multimedia students to enhance spatial ability. *Sustainability (Basel, Switzerland), 10*(4), 1074.

Moos, D. C., & Azevedo, R. (2009). Learning with computer-based learning environments: A literature review of computer self-efficacy. *Review of Educational Research, 79*(2), 576-600.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22*(3), 149–158.

Moreno, R. (2007). Optimising learning from animations by minimising cognitive load: Cognitive and affective consequences of signalling and segmentation methods. *Applied Cognitive Psychology, 21*(6), 765–781.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments: Special issue on interactive learning environments: Contemporary issues and trends. *Educational Psychology Review*, *19*(3), 309-326.

Moreno, R., Mayer, R., & Lester, J. (2000). Life-like pedagogical agents in constructivist multimedia environments: Cognitive consequences of their interaction. In *EdMedia+ Innovate Learning* (pp. 776–781). Association for the Advancement of Computing in Education (AACE).

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358-368. https://doi.org/10.1037/0022-0663.91.2.358.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(1), 117–125. https://doi.org/10.1037/0022-0663.92.1.117.

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology, 94*(3), 598-610.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, *19*(2), 177-213.

Moro, C., Štromberga, Z., Raikos, A., & Stirling, A. (2017). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education*, *10*(6), 549-559.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319–334. https://doi.org/10.1037/0022-0663.87.2.319.

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. *Computers and Education*, *141*, 103618.

Mystakidis, S., Christopoulos, A., & Pellas, N. (2022). A systematic mapping review of augmented reality applications to support STEM learning in higher education. *Education and Information Technologies*, *27*(2), 1883-1927.

Oberfoell, A., & Correia, A. (2016). Understanding the role of the modality principle in multimedia learning environments. *Journal of Computer Assisted Learning, 32*(6), 607-617.

Onyesolu, M. O., & Eze, F. U. (2011). Understanding virtual reality technology: advances and applications. *Adv. Comput. Sci. Eng*, 53-70.

Ozerem, A., & Akkoyunlu, B. (2015). Learning environments designed according to learning styles and its effects on mathematics achievement. *Eurasian Journal of Educational Research, 15*(61), 61-80.

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. *The Cambridge handbook of multimedia learning* (pp. 27-42). Cambridge University Press.

Paas, F. G., & Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737-743.

Paas, F. G., & Van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255–287.

Pan, Z., Cheok, A. D., Yang, H., Zhu, J., & Shi, J. (2006). Virtual reality and mixed reality for virtual learning environments. *Computers & Graphics, 30*(1), 20-28.

Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior, 44*, 267-278.

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

Parong, J., & Mayer, R. E. (2021). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning, 37*(1), 226–241.

Petersen, G. B., Klingenberg, S., Mayer, R. E., & Makransky, G. (2020). The virtual field trip: Investigating how to optimize immersive virtual learning in climate change education. *British Journal of Educational Technology, 51*(6), 2098-2114.

Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In S. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 131–162). Oxford: Elsevier.

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

Potkonjak, V., Gardner, M., Callaghan, V., Mattila, P., Guetl, C., Petrović, V. M., & Jovanović, K. (2016). Virtual laboratories for education in science, technology, and engineering: A review. *Computers and Education, 95*, 309-327.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers and Education, 147*, 103778.

Radu, I. (2014). Augmented reality in education: A meta-review and cross-media analysis. *Personal and Ubiquitous Computing, 18*(6), 1533–1543.

Raja, R., & Nagasubramani, P. (2018). Impact of modern technology in education. *Journal of Applied and Advanced Research, 3*(1), 33-35.

Ray, A. B., & Deb, S. (2016). Smartphone based virtual reality systems in classroom teaching—a study on the effects of learning outcome. In *2016 IEEE eighth international conference on technology for education (T4E)* (pp. 68-71). IEEE.

Reinwein, J. (2012). Does the modality effect exist? And if so, which modality effect? *Journal of Psycholinguistic Research, 41*(1), 1–32.

Rey, G. D., Beege, M., Nebel, S., Wirzberger, M., Schmitt, T. H., & Schneider, S. (2019). A meta-analysis of the segmenting effect. *Educational Psychology Review, 31*(2), 389-419.

Richards, D., & Taylor, M. (2015). A comparison of learning gains when using a 2D simulation tool versus a 3D virtual world: An experiment to find the right representation involving the marginal value theorem. *Computers and Education, 86*, 157–171.

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review, 17,* 19-36.

Riva, G. (2003). Applications of virtual environments in medicine. *Methods of Information in Medicine, 42*(5), 524–534.

Sally Wu, Y.-H., & Alan Hung, S.-T. (2022). The effects of virtual reality infused instruction on elementary school students' english-speaking performance, willingness to communicate, and learning autonomy. *Journal of Educational Computing Research, 60*(6), 1558-1587.

Schmidt, J. T. (2007). Preparing students for success in blended learning environments: Future oriented motivation & self-regulation. *Educational Research*, 301–302.

Schnotz, W. (2005). An integrated model of text and picture comprehension. *The Cambridge handbook of multimedia learning* (pp. 49–70). Cambridge University Press.

Schroeder, R. (1996). *Possible worlds: The social dynamic of virtual reality technology*. Westview Press, Inc.

Schweppe, J., & Rummer, R. (2016). Integrating written text and graphics as a desirable difficulty in long-term multimedia learning. *Computers in Human Behavior, 60*, 131-137.

Schüler, A., Scheiter, K., & Gerjets, P. (2013). Is spoken text always better? Investigating the modality and redundancy effect with longer text presentation. *Computers in Human Behavior, 29*(4), 1590-1601.

Shim, H., & Lee, H. (2022). The effect of design education using virtual reality-based coding on student competence and educational satisfaction. *Education and Information Technologies, 27*(4), 4577–4597.

Singh, G., Mantri, A., Sharma, O., Dutta, R., & Kaur, R. (2019). Evaluating the impact of the augmented reality learning environment on electronics laboratory skills of engineering students. *Computer Applications in Engineering Education, 27*(6), 1361–1375.

Sırakaya, M., & Alsancak Sırakaya, D. (2022). Augmented reality in STEM education: A systematic review. *Interactive Learning Environments, 30*(8), 1556-1569.

Soicher, R. N., & Becker-Blease, K. A. (2020). Testing the segmentation effect of multimedia learning in a biological system. *Journal of Computer Assisted Learning, 36*(6), 825-837.

Sommerauer, P., & Müller, O. (2014). Augmented reality in informal learning environments: A field experiment in a mathematics exhibition. *Computers and Education, 79,* 59-68.

Squire, K. D., & Jan, M. (2007). Mad city mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology, 16*(1), 5–29.

Stiller, K., & Zinnbauer, P. (2011). Does Segmentation of Complex Instructional Videos in Big Steps Foster Learning? The Segmentation Principle Applied to a Classroom Video. In *Proceedings of the Global Learn Asia Pacific 2011-Global Conference on Learning and Technology* (pp. 2044-2053).

Sung, E., & Mayer, R. E. (2012). When graphics improve liking but not learning from online lessons. *Computers in Human Behavior, 28*(5), 1618-1625.

Sung, E., & Mayer, R. E. (2013). Online multimedia learning with mobile devices and desktop computers: An experimental test of Clark's methods-not-media hypothesis. *Computers in Human Behavior, 29*(3), 639-647.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Intrinsic and extraneous cognitive load. *Cognitive load theory* (pp. 57–69). Springer, New York.

Sweller, J., Ayres, P. L., Kalyuga, S., & Chandler, P. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using multivariate statistics* (Vol. 6, pp. 497-516). Boston, MA: Pearson.

Tabbers, H. (2002). *The modality of text in multimedia instruction: Refining the design guidelines* (Unpublished PhD thesis). Open University of the Netherlands, Heerlen.

Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, *74*(1), 71-81.

Tarng, W., Lin, Y., Lin, C., & Ou, K. (2016). Development of a lunar-phase observation system based on augmented reality and mobile learning technologies. *Mobile Information Systems, 2016*.

Thees, M., Kapp, S., Strzys, M. P., Beil, F., Lukowicz, P., & Kuhn, J. (2020). Effects of augmented reality on learning and cognitive load in university physics laboratory courses. *Computers in Human Behavior*, *108*, 106316.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology. Applied*, *3*(4), 257–287.

Timar, Z. Ş. n. (2020). *Çoklu ortamla değerlendirme materyalinde duygusal tasarımın işe koşulmasının bilişsel yüke ve motivasyona etkileri* (Unpublished PhD thesis). Anadolu University.

Tugtekin, U., & Odabasi, H. F. (2022). Do interactive learning environments have an effect on learning outcomes, cognitive load and metacognitive judgments? *Education and Information Technologies, 27*(5), 7019-7058.

Turan, Z., Meral, E., & Sahin, I. F. (2018). The impact of mobile augmented reality in geography education: Achievements, cognitive loads and views of university students. *Journal of Geography in Higher Education, 42*(3), 427-441.

UK Authority. (2019). VR and AR attract education sector interest. Retrieved from https://www.ukauthority.com/articles/vr-and-ar-attract-education-sector-interest/

van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 263–278). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.014

Van Merrienboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147-177.

Vinales, J. J. (2015). The learning environment and learning styles: A guide for mentors. *British Journal of Nursing, 24*(8), 454-457.

Wang, C. X., & Kinuthia, W. (2004). Defining technology enhanced learning environment for pre-service teachers. In *Society for information technology & teacher education international conference* (pp. 2724-2727). Association for the Advancement of Computing in Education (AACE).

Wang, F., Li, W., Mayer, R. E., & Liu, H. (2018). Animated pedagogical agents as aids in multimedia learning: Effects on eye-fixations during learning and learning outcomes. *Journal of Educational Psychology, 110*(2), 250-268.

Wang, Z., Sundararajan, N., Adesope, O. O., & Ardasheva, Y. (2017). Moderating the seductive details effect in multimedia learning with note-taking. *British Journal of Educational Technology, 48*(6), 1380-1389.

Weisberg, S. (2005). *Applied linear regression* (Vol. 528). John Wiley & Sons.

Weng, C., Otanga, S., Weng, A., & Cox, J. (2018). Effects of interactivity in E-textbooks on 7th graders science learning and cognitive load. *Computers & Education*, *120*, 172-184.

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review, 2*(1), 1–12.

Wu, H., Lee, S. W., Chang, H., & Liang, J. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers and Education, 62*, 41-49.

Wu, P. H., Hwang, G. J., Yang, M. L., & Chen, C. H. (2018). Impacts of integrating the repertory grid into an augmented reality-based learning design on students' learning achievements, cognitive load and degree of satisfaction. *Interactive Learning Environments*, *26*(2), 221-234.

Yang, S., Mei, B., & Yue, X. (2018). Mobile augmented reality assisted chemical education: Insights from elements 4D. *Journal of Chemical Education, 95*(6), 1060-1062.

Yen, J.-C., Tsai, C.-H., & Wu, M. (2013). Augmented reality in the higher education: Students' science concept learning and academic achievement in astronomy. *Procedia, Social and Behavioral Sciences, 103*, 165-173.

Yoon, S., Anderson, E., Lin, J., & Elinich, K. (2017). How augmented reality enables conceptual understanding of challenging science content.*Educational Technology & Society, 20*(1), 156-168.

Yoon, S. A., Elinich, K., Wang, J., Steinmeier, C., & Tucker, S. (2012). Using augmented reality and knowledge-building scaffolds to improve learning in a science museum. *International Journal of Computer-Supported Collaborative Learning, 7*(4), 519–541.

Yu, J., Denham, A. R., & Searight, E. (2022). A systematic review of augmented reality game-based learning in STEM education. *Educational Technology Research and Development, 70*(4), 1169–1194.

Yu, Z., & Xu, W. (2022). A meta-analysis and systematic review of the effect of virtual reality technology on users' learning outcomes. *Computer Applications in Engineering Education, 30*(5), 1470–1484.

Zhao, J., Lin, L., Sun, J., & Liao, Y. (2020). Using the summarizing strategy to engage learners: Empirical evidence in an immersive virtual reality environment. *The Asia-Pacific Education Researcher, 29*(5), 473–482.