

THESIS

INNOVATIVE EDUCATIONAL TECHNOLOGY: SUPPORTING COGNITIVE LOAD
MANAGEMENT AND SELF-REGULATED LEARNING IN ANATOMY EDUCATION

Submitted by

Brandon Lowry

Department of Biomedical Sciences

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2026

Master's Committee:

Advisor: Tod Clapp

Quint Winger

Seonil Kim

David Most

Copyright by Brandon Lee Lowry 2026

All Rights Reserved

ABSTRACT

INNOVATIVE EDUCATIONAL TECHNOLOGY: SUPPORTING COGNITIVE LOAD MANAGEMENT AND SELF-REGULATED LEARNING IN ANATOMY EDUCATION

Admin and faculty across the educational landscape are charged with understanding and deciding on how, if at all, to integrate technological advancements, such as virtual reality (VR) and artificial intelligence (AI), into their curriculum. Currently, the educational value of many of these technologies is unclear, with research studies contradicting their efficacy and value, ultimately suggesting that these tools, while innovative, must be used correctly to be effective. Therefore, educators across disciplines need evidence-based recommendations to successfully integrate the emergent, innovative technologies of the modern era and to understand how these technologies may impact student experience and learning. This thesis utilizes two complementary lenses—Cognitive Load and Self-Directed Learning—to examine technology-enhanced approaches for understanding student learning and experience within STEM courses.

The following chapters recount two studies that explored how technology can be used to measure and support student learning and regulation. Chapter 1 affords a broad examination of emerging technologies within STEM education, while underscoring the value of understanding the learner's experience using those technologies. Research design and methodology utilized in each study is briefly presented alongside the rationale for each research effort.

Chapter 2 is a reproduced version of the published manuscript, *Cognitive Load in Virtual Reality Anatomy Education: Comparing 2D and 3D Learning Experiences*. The chapter is composed of an experiment that measured the cognitive load of students learning anatomy

concepts in virtual reality. Our research team hypothesized that student cognitive load would be lower when viewing 3D content than when viewing 2D content. Quantitative data were collected via the NASA Task Load Index and the HP OMNICEPT REVERB G2 head-mounted display (HMD) for viewing virtual reality content. The HMD collects and analyzes biometric data to provide a measure for cognitive load. The NASA Task Load Index is a standardized measure for subjective reports of cognitive load experienced during an activity. Cognitive Load is defined as the total amount of mental effort required to complete a learning task. Additionally, qualitative data were collected via focus group and surveys to assess student experience, as well as the impact of various characteristics and prior experiences on cognitive load. The study highlights the value of VR technology in not only measuring student learning but helping to manage the required mental resources to complete the task.

Chapter 3 is a reproduced version of the published manuscript in the Journal of Microbiology and Biology Education. The study conducted—*Leveraging Generative AI to Foster Metacognition and Self-Directed Learners*—utilized generative AI to foster metacognitive and self-directed learning (SDL) behaviors among graduate biomedical science students. The rapidly expanding field of AI has left many educators perplexed with respect to how, when, and why to use AI in their classroom. Our research team conducted an 8-week observational study to examine how AI could support the development of metacognitive awareness among participants, and further evaluating how the metacognitive awareness impacted the student’s readiness for SDL. Eight participants engaged with a trained AI model that facilitated reflective conversations related to the student’s learning efforts and experience in the context of their graduate neuroanatomy course. Quantitative data were collected via the Metacognitive Awareness Inventory (MAI) and the Self-Directed Learning Instrument (SDLI) to determine the impact of

engaging with the model on students' use of metacognitive practices and readiness of SDL, respectively. Qualitative data were collected via focus group interviews. The model provides a working example of how AI may be used to support the development of students' self-regulation, rather than simply an answer generator.

Chapter 4 is a broad summary of the two studies, how they are connected, and how they contribute to our understanding of emergent educational technologies. Recommendations for future research are briefly discussed alongside recommendations for immediate integration of the technologies into STEM education.

The research studies described here help us understand the impact of innovation in education through technology. Both studies provide evidence that technology can support student learning by offering deeper and nuanced insights into the student experience. Virtual reality has the innate potential to reduce overall effort via the way it displays anatomical data to students. Generative AI may be utilized to support students in developing the requisite skills for success through AI-guided learning. Together, these emerging technologies provide a mechanism for innovative change within STEM education by offering educators effective tools for measuring and supporting student learning experiences beyond the typical approaches many have come to rely upon.

TABLE OF CONTENTS

ABSTRACT	ii
CHAPTER 1: INTRODUCTION	1
LITERATURE REVIEW	1
Innovation with Technology in STEM Education	1
Technology to Measure Student Learning and Experience	3
Technology to Support Student Regulation of Learning	4
REFERENCES	8
CHAPTER 2: Cognitive Load in Virtual Reality Anatomy Education: Comparing 2D and 3D Learning Experiences.....	12
ABSTRACT.....	12
INTRODUCTION	13
Changes in Anatomy Education.....	13
Adoption of VR Technology.....	14
Challenges with VR	14
MATERIALS AND METHODS	16
Participants.....	16
Study Design.....	17
Modules.....	18
Statistical Analysis	23
RESULTS.....	23
Participant Information	23
DISCUSSION	28
Anatomy Education and VR Integration.....	28
Cognitive Load.....	29
Experiences Between Modules	30
Implications for VR-Based Instructional Design.....	31
Limitations	33
Future Directions	33
Conclusion	34
SUPPLEMENTARY	35
REFERENCES	38
CHAPTER 3: Leveraging Generative AI to Foster Metacognition and Self-Directed Learners..	41
ABSTRACT.....	41

INTRODUCTION	42
SDL, Metacognition, and Motivation	43
Research Overview	44
Research Questions and Hypotheses	45
METHODS	45
Research Design.....	45
Participants.....	46
Materials	46
Procedure	48
Data Analysis	49
RESULTS.....	49
Self-directed learning instrument.....	49
Metacognition Awareness Inventory.....	50
Focus group and survey analysis	53
DISCUSSION	54
SUPPLEMENTARY	57
REFERENCES	59
CHAPTER 4: SUMMARY AND CONCLUSION.....	61
PURPOSE	61
SUMMARY OF FINDINGS	61
FUTURE RESEARCH	62
CONCLUSION.....	63

CHAPTER 1: INTRODUCTION

LITERATURE REVIEW

Innovation with Technology in STEM Education

The following section focuses on how educators utilize technology to measure aspects of student learning and experience (e.g., cognitive load, affect), support students' regulation of their unique learning process, and create innovative and effective technology-enabled learning opportunities for students across educational domains.

On a day-to-day basis, a university student will interact with various technologies, ranging from their school's learning management system (LMS), like Canvas or Blackboard, to various video or audio communication tools, such as Microsoft Teams and Zoom, and virtual environments such as augmented and virtual reality [1, 2, 3]. These technologies allow students to interact with each other and digital content in collaborative and engaging ways, improving student retention and persistence within education [1]. Interestingly, these technological innovations have expanded the idea of what a learning environment can be, expanding beyond physical spaces to include digital and virtual classrooms [4] that seamlessly blend the learning environment [5]. However, concerns exist regarding the integration of some of the more modern technologies (e.g., virtual reality, AI, large language models (LLMs)), warranting further research to better understand how these technologies can be intentionally integrated into curricula to support a more standardized and effective technology integration process [6].

Immersive technologies like virtual and augmented reality have changed how educators teach complex STEM concepts and tasks. These innovative approaches often support student engagement, allow for better comprehension, and lead to lower attrition in many disciplines [7,

8]. While still a new field, AI-supported approaches to facilitating teaching and learning have also shown positive impacts on learning outcomes [9]. However, across the spectrum of educational levels, these technologies present unique challenges to their successful integration into permanent curricula [10].

Research demonstrates a strong need for instructor support and preparation in using these technologies, in addition to concerns over a lack of inclusive practices or more globally centric approaches in their development [11]. For example, there are observed difficulties in students' use of VR hardware, like cybersickness, or reports of user distraction and discomfort that render the results of the immersive technology less effective [8]. There are also concerns over AI-supported chatbots, how they collect, manage, and utilize student data [12].

Particular attention has been given to STEM disciplines due to the inherent complexity of many STEM topics as they are often invisible, abstract, or difficult to spatially visualize. For example, STEM students are often tasked with interpreting maps, models, and 3D structures—requiring them to oscillate between conceptualizing a problem in 2D and 3D [13, 14]. Technologies like VR manage the difficulties students face in the abstract conceptualization required of them by allowing them to view the information or data being visualized in a manner more consistent with how they perceive objects in the natural world (i.e., three dimensions) [15, 16]. The combination of AI and VR shows promise in supporting students as they navigate spatially complex or abstract topics. Moreover, the addition of AI to AR has increased the ability for, and practicality of, more inclusive, student-centered practices [17]. Together, these opportunities and remaining areas of uncertainty warrant a concerted research effort among university stakeholders to ensure the appropriate integration of these technologies.

Technology to Measure Student Learning and Experience

Research suggests that many traditional measures for knowledge gains or learning outcomes are best suited for capturing established measures of factual knowledge within STEM education [18]. For example, traditional assessments have been studied and explored for impacts on learning, not just assessment, showcasing unique benefits to the learner dependent upon the nature, timing, and frequency of assessment [19]. However, little exists with respect to how technology-enabled assessments (e.g., AI-supported assessments, biometric data) differ in their approach and capability for assessing student learning and experience [20]. The research described in Chapter 2 addresses some of these issues by seeking to understand available mechanisms to objectively and quantitatively measure students' experiences in VR through assessments of cognitive load.

It is important to understand how students experience technology because it influences how they perceive and use the technology in the future. Student perception of a technology's efficacy or appropriateness for a learning goal, in addition to their comfort using the technology, leads to different levels of successful interaction [21]. Practically speaking, repeatedly surveying hundreds of students regarding education technology is simply too labor intensive for most instructors to accomplish at scale. While surveys provide valuable feedback, repeated surveying can lead to low-effort responses and low levels of survey completion [22]. Innovative technologies within the realm of VR, like the HP OMNICEPT, allow for a more passive collection of meaningful data by way of biometric data (i.e., heart rate variability (HRV), and pupillometry), affording a real-time assessment of student experience that can be collected more quickly and universally across a group of learners [23, 24].

The HP OMNICEPT, as discussed in Chapter 2, collects various biometric data before aggregating and computing those data into a measure of their cognitive load [23]. Cognitive Load (CL) is defined as the total amount of mental resources required to accomplish a learning task; cognitive load can become overwhelming when demands exceed working memory capacity, or insufficient when a challenge fails to pique a student's motivation to engage in a learning task [25]. Until the 21st century, subjective surveys were among the most widely used and accepted measures for determining a person's CL. However, these measures are post-experience and rely heavily on the user's ability to accurately quantify their CL during a particular event after it has already occurred [26, 27]. The real-time insight afforded by the HP OMNICEPT allows for a more targeted application of CL theory for learning purposes. Specifically, researchers can leverage the customizability of VR learning spaces to assess the impact of various instructional designs and methodologies on student CL. This adaptability provides educators with a more nuanced understanding of the student experience, permitting the evaluation of instructional approaches through complementary lenses: what students learn and how they experience the learning process.

Technology to Support Student Regulation of Learning

Educational technology has the potential to go beyond simply providing a novel way to present content; it carries the potential to influence and positively change the way a student views and engages with their own learning processes both within and beyond a learning environment [28]. STEM education and careers require a great deal of adaptability and self-regulation to be successful, as the complex and intricate nature of many STEM subjects is often unfamiliar, requiring students to be able to identify and mitigate learning gaps and needs [29]. Unfortunately, not all educators are not equipped with appropriate tools for helping students

develop that skill set. Many students struggle to adapt to the unique challenges presented to them and fail to appropriately utilize available resources [30]. As such, STEM courses face high levels of attrition both in courses and across majors [29, 31]. Researchers have begun to look at how educational technology can be used not only to facilitate learning activities, but also to help students become more adaptive lifelong learners, which is paramount for successful careers across STEM domains and disciplines [32].

To be an adaptable learner requires a great deal of self-awareness and reflection to regulate one's learning processes (i.e., to monitor and adjust one's approach to learning to ensure success) [33]. Students who engage in reflective practice as it relates to their learning have reported more enjoyment and course satisfaction, contributing to lower levels of attrition [34]. However, many students come into undergraduate study unfamiliar with metacognition, or reflective practice, and how it can be a valuable tool in learning new subjects [35]. Typical courseloads allow little time for educators to spend developing students' reflective skill set. Recent developments in the technological world have spilled over into the educational landscape, affording unique opportunities to support personal development alongside subject-specific instruction [32]. The research described in Chapter 3 explores the potential use case of generative AI in developing students' self-regulatory skills. Educational technologies can serve both as tools for understanding learner experiences and as mechanisms for supporting learner development.

Opportunity Through Educational Technology

The use of educational technology carries the potential to expand the capabilities of instructors as they provide unique resources and opportunities for their students [37]. In addition to improving teaching and instructional design, educational technology has the potential to greatly impact student learning and access by providing effective tools for feedback and support

[37, 38]. However, despite the promise of educational technology in providing effective support for students, its implementation faces several challenges [39]. Quite often the major barrier to implementation stems from limitations to infrastructure and access to the proper technologies [38]. In other cases, difficulties arise from faculty resistance and a lack of competence among student users [40]. Research that centers on examining strategies for implementation may mitigate some of these challenges. The research presented here offer solutions by exploring two use cases of educational technology that may

METHODOLOGY

The following chapters are peer-reviewed manuscripts from two separate IRB-approved research projects that examined the potential use cases of innovative technologies in anatomy education.

Chapter 2 methodology

The first research project used a comparative experimental design to assess the cognitive load of students viewing various types of content or data in VR, seeking to establish best practices for VR-based instructional design. In the study, students completed two learning activities and cognitive load was assessed during the activity using the HP OMNICEPT and afterward using the NASA Task Load Index (NASA-TLX). Additional demographic information was collected before their session, and student experience was assessed through a post-activity survey.

Students completed both learning activities in VR while wearing the HP OMNICEPT head-mounted display (HMD). This HMD was developed by HP Labs to be used in VR development, both in and out of academic contexts. The research team affiliated with its development used

machine learning (ML) and multi-sensor input (e.g., eye tracking, pulse plethymography, face camera, accelerometers, and gyroscopes) to establish a reliable cognitive load measurement tool that moved beyond a single physiological marker. The cognitive load score output from the HMD is a continuous value between 0.0 and 1.0. Findings from the technology development process met statistical thresholds for reliability and validity within the context of cognitive load measure [36]. Below is a brief description of the process.

Chapter 3 methodology

The second research project utilized a quasi-experimental design to explore how generative AI could be used to support student development of metacognitive skills and, subsequently, the ability to engage in self-regulatory behavior. Students in the experimental group interacted with a customized AI model that facilitated reflective conversations twice a week for 10 minutes per session for 8 weeks. Students in the comparative group were given suggestions for developmental resources. Impact on metacognitive awareness and self-regulatory behaviors was assessed using student scores on standardized assessments. Focus group data provide more personalized insight into the impact of educational technology on the student's learning.

The research team designed the AI model using OpenAI's custom GPT builder offered through professional level subscriptions. This feature allows for users to create a custom, fine-tuned GPT model that operates within pre-determined and user-selected parameters, limitations, and features. The custom GPT model is not the same as the consumer level ChatGPT with which most individuals are familiar. Users can limit or specify what knowledge resources the GPT will use in its generative process. In the case of this study, the research team limited the GPT to only focusing on metacognitive and reflective processes. It did not have any neuroanatomical

knowledge to use—this ensured the model did not give “wrong” answers, hallucinate, or steer students in any direction with respect to neuroanatomy content.

Throughout the interaction with the model, the students would have the opportunity to respond to 4 pre-determined questions. The model was instructed to probe for depth if the students’ responses lacked sufficient detail. During the development process, the research team fine-tuned the GPT prompt to ensure the model only probed for depth when appropriate.

Combined, these research studies utilized both quantitative and qualitative data to demonstrate how educational technology can shape student learning while also supporting their personal development.

REFERENCES

1. Nungu, L., Mukama, E. & Nsabayeze, E. Online collaborative learning and cognitive presence in mathematics and science education. Case study of university of Rwanda, college of education. *Educ Inf Technol* **28**, 10865–10884 (2023). <https://doi.org/10.1007/s10639-023-11607-w>
2. Le, B., Lawrie, G.A. & Wang, J.T.H. Student Self-perception on Digital Literacy in STEM Blended Learning Environments. *J Sci Educ Technol* **31**, 303–321 (2022). <https://doi.org/10.1007/s10956-022-09956-1>
3. Choi-Lundberg, D. L., Butler-Henderson, K., Harman, K., & Crawford, J. (2023). A systematic review of digital innovations in technology-enhanced learning designs in higher education. *Australasian Journal of Educational Technology*, 39(3), 133–162. <https://doi.org/10.14742/ajet.76154>
4. Papaioannou, G., Volakaki, M.-G., Kokolakis, S., & Vouyioukas, D. (2023). Learning Spaces in Higher Education: A State-of-the-Art Review. *Trends in Higher Education*. <https://doi.org/10.3390/higheredu2030032>
5. Lamb, J., Carvalho, L., Gallagher, M. *et al.* The Postdigital Learning Spaces of Higher Education. *Postdigit Sci Educ* **4**, 1–12 (2022). <https://doi.org/10.1007/s42438-021-00279-9>
6. Mena-Guacas, A. F., López-Catalán, L., Bernal-Bravo, C., & Ballesteros-Regaña, C. (2025). Educational Transformation Through Emerging Technologies: Critical Review of Scientific Impact on Learning. *Education Sciences*, 15(3), 368. <https://doi.org/10.3390/educsci15030368>

7. Tene T, Marcatoma Tixi JA, Palacios Robalino ML, Mendoza Salazar MJ, Vacacela Gomez C and Bellucci S (2024) Integrating immersive technologies with STEM education: a systematic review. *Front. Educ.* 9:1410163. <https://doi.org/10.3389/feduc.2024.1410163>
8. Jiang, H., Zhu, D., Chugh, R. *et al.* Virtual reality and augmented reality-supported K-12 STEM learning: trends, advantages and challenges. *Educ Inf Technol* **30**, 12827–12863 (2025). <https://doi.org/10.1007/s10639-024-13210-z>
9. Li, S., Zeng, C., Liu, H. *et al.* A meta-analysis of AI-enabled personalized STEM education in schools. *IJ STEM Ed* **12**, 58 (2025). <https://doi.org/10.1186/s40594-025-00566-y>
10. Saldívar-Almorejo, M. M., Flores-Herrera, L. A., Rivera-Blas, R., Niño-Suárez, P. A., Rivera-Blas, E. Z., & Rodríguez-Contreras, N. (2024). e-Learning Challenges in STEM Education. *Education Sciences*, *14*(12), 1370. <https://doi.org/10.3390/educsci14121370>
11. Nancy Mohd Al Hamad, Ololade Elizabeth Adewusi, Chika Chioma Unachukwu, Blessing Osawaru and Onyebuchi Nneamaka Chisom. A review on the innovative approaches to STEM education. *International Journal of Science and Research Archive*, 2024, 11(01), 244–252. Article DOI: <https://doi.org/10.30574/ijrsra.2024.11.1.0026>
12. Chiu, T.K.F., Li, Y. How Can Emerging Technologies Impact STEM Education?. *Journal for STEM Educ Res* **6**, 375–384 (2023). <https://doi.org/10.1007/s41979-023-00113-w>
13. Hayes, J.C., Kraemer, D.J.M. Grounded understanding of abstract concepts: The case of STEM learning. *Cogn. Research* **2**, 7 (2017). <https://doi.org/10.1186/s41235-016-0046-z>
14. Gagnier, K. M., Holochwost, S. J., & Fisher, K. R. (2021). Spatial thinking in science, technology, engineering, and mathematics: Elementary teachers' beliefs, perceptions, and self-efficacy. *Journal of Research in Science Teaching*, *59* (1), 95–126. <https://doi.org/10.1002/tea.21722>
15. Koç, A., Kanadlı, S. Effect of Interactive Learning Environments on Learning Outcomes in Science Education: A Network Meta-Analysis. *J Sci Educ Technol* **34**, 681–703 (2025). <https://doi.org/10.1007/s10956-025-10202-7>
16. Lowry et al., 2026
17. Garg, N., Kaur, A., Ahmad, F., & Dutta, R. (2025). Augmenting Education: The Transformative Power of AR, AI, and Emerging Technologies. *Human Behavior and Emerging Technologies*. <https://doi.org/10.1155/hbe2/5681184>
18. Gao, X., Li, P., Shen, J. et al. Reviewing assessment of student learning in interdisciplinary STEM education. *IJ STEM Ed* **7**, 24 (2020). <https://doi.org/10.1186/s40594-020-00225-4>
19. Murphy, D.H., Little, J.L. & Bjork, E.L. The Value of Using Tests in Education as Tools for Learning—Not Just for Assessment. *Educ Psychol Rev* **35**, 89 (2023). <https://doi.org/10.1007/s10648-023-09808-3>
20. Xu, W., Ouyang, F. The application of AI technologies in STEM education: a systematic review from 2011 to 2021. *IJ STEM Ed* **9**, 59 (2022). <https://doi.org/10.1186/s40594-022-00377-5>

21. Kemp, A., Palmer, E., Strelan, P., & Thompson, H. M. (2024). Testing a novel extended educational technology acceptance model using student attitudes towards virtual classrooms. *British Journal of Educational Technology*, 55, 2110–2131. <https://doi.org/10.1111/bjet.13440>
22. Fass-Holmes, B. (2022). Survey Fatigue—Literature Search and Analysis of Implications for Student Affairs Policies and Practices. *Journal of Interdisciplinary Studies in Education*, 11(1), 56-73. <https://ojed.org/jise/article/view/3262>
23. Siegel, E.H., Wei, J., Gomes, A., Oliviera, M., Sundaramoorthy, P., Smathers, K., Vankipuram, M., Ghosh, S., Horii, H., Bailenson, J., & Ballagas, R. “HP Omnicept Cognitive Load Database (HPO-CLD) – Developing a Multimodal Inference Engine for Detecting Real-time Mental Workload in VR” Technical Report, HP Labs, Palo Alto, CA
24. Wei, J., Siegel, E.H., Sundaramoorthy, P., Gomes, A., Zhang, S., Vankipuram, M., Smathers, K., Ghosh, S., Horii, H., Bailenson, J.N., & Ballagas, R. ‘. (2025). Cognitive Load Inference Using Physiological Markers in Virtual Reality. 2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR), 759-769.
25. Sweller, J. (2024). Cognitive load theory and individual differences. *Learning and Individual Differences*, 110, 102423.
26. Ouwehand K, Kroef A, Wong J and Paas F (2021) Measuring Cognitive Load: Are There More Valid Alternatives to Likert Rating Scales?. *Front. Educ.* 6:702616. <https://doi:10.3389/educ.2021.702616>
27. Leppink, J., Paas, F., Van der Vleuten, C.P.M. et al. Development of an instrument for measuring different types of cognitive load. *Behav Res* 45, 1058–1072 (2013).
28. Rashid, T., & Asghar, H. M. (2016). Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. *Computers in Human Behavior*, 63, 604–612. <https://doi.org/10.1016/j.chb.2016.05.084>
29. Blackmore, Conner & Vitali, Julian & Louise, Ainscough & Langfield, Tracey & Colthorpe, Kay. (2021). A Review of Self-Regulated Learning and Self-Efficacy: The Key to Tertiary Transition in Science, Technology, Engineering and Mathematics (STEM). *International Journal of Higher Education*.
30. Huh, Y., Lee, D., & Reigeluth, C. M. (2024). Teacher supports and technology needs to develop students’ self-regulated learning at U.S. K-12 online schools. *Contemporary Educational Technology*, 16(3), ep523. <https://doi.org/10.30935/cedtech/14956>
31. Graham, M. J., Frederick, J., Byars-Winston, A., Hunter, A. B., & Handelsman, J. (2013). Science education. Increasing persistence of college students in STEM. *Science* (New York, N.Y.), 341(6153), 1455–1456. <https://doi.org/10.1126/science.1240487>
32. Jin, SH., Im, K., Yoo, M. et al. Supporting students’ self-regulated learning in online learning using artificial intelligence applications. *Int J Educ Technol High Educ* 20, 37 (2023). <https://doi.org/10.1186/s41239-023-00406-5>
33. Zimmerman, Barry. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*. 41. 64-70. Doi:10.1207/s15430421tip4102_2.

34. Alt D, Raichel N and Naamati-Schneider L (2022) Higher Education Students' Reflective Journal Writing and Lifelong Learning Skills: Insights From an Exploratory Sequential Study. *Front. Psychol.* 12:707168. doi:10.3389/fpsyg.2021.707168
35. Stanton, J. D., Sebesta, A. J., & Dunlosky, J. (2021). Fostering metacognition to support student learning and performance. *CBE—Life Sciences Education*, 20(2)
36. O. Augereau, G. Brocheton and P. P. Do Prado Neto, "An Open Platform for Research about Cognitive Load in Virtual Reality," 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Christchurch, New Zealand, 2022, pp. 54-55, doi: 10.1109/VRW55335.2022.00020.
37. Chugh, R., Turnbull, D., Cowling, M.A. et al. Implementing educational technology in Higher Education Institutions: A review of technologies, stakeholder perceptions, frameworks and metrics. *Educ Inf Technol* 28, 16403–16429 (2023).
<https://doi.org/10.1007/s10639-023-11846-x>
38. Criollo-C, S., González-Rodríguez, M., Guerrero-Arias, A., Urquiza-Aguilar, L. F., & Luján-Mora, S. (2024). A Review of Emerging Technologies and Their Acceptance in Higher Education. *Education Sciences*, 14(1), 10. <https://doi.org/10.3390/educsci14010010>
39. Feng J, Yu B, Tan WH, Dai Z, Li Z (2025) Key factors influencing educational technology adoption in higher education: A systematic review. *PLOS Digit Health* 4(4): e0000764.
<https://doi.org/10.1371/journal.pdig.0000764>
40. Ahmad, S., Mohd Noor, A. S., Alwan, A. A., Gulzar, Y., Khan, W. Z., & Reegu, F. A. (2023). eLearning Acceptance and Adoption Challenges in Higher Education. *Sustainability*, 15(7), 6190. <https://doi.org/10.3390/su15076190>

CHAPTER 2: Cognitive Load in Virtual Reality Anatomy Education: Comparing 2D and 3D Learning Experiences

ABSTRACT

With virtual reality (VR) becoming increasingly integrated into educational settings, it remains crucial to understand how different formats of educational content impact cognitive processing. This study examined the impact of 2D versus 3D VR modules on cognitive load and learning outcomes in undergraduate anatomy education. Using real-time biometric measurements collected through VR head-mounted displays, cognitive load was assessed as students observed recorded laryngeal anatomy modules presented by virtual instruction utilizing 2D images or 3D models. Despite no significant impact on learning outcomes, participants in the 2D group consistently demonstrated a higher mean cognitive load ($M = 0.58$, $SD = 0.05$) compared to the 3D group ($M = 0.49$, $SD = 0.07$), indicating a significant reduction in cognitive effort with 3D visualization. Pre- and post-survey results revealed a mixed landscape in perceptions of VR technology, with some students reporting positive views and comfort, while others expressed significant reservations. Notably, however, no significant difference was found between individually reported attitudes and cognitive load, suggesting that students' cognitive load during the intervention was independent of these variables. Further, prior experience with VR, video game use, and general attitudes toward technology did not significantly influence cognitive load, suggesting that instructional design had a more substantial impact than user familiarity. These findings support the pedagogical value of incorporating 3D VR in anatomy curricula and highlighted the importance of evidence-based design to improve learning outcomes while managing cognitive demand.

-This chapter is a reproduction of, *Cognitive Load in Virtual Reality Education: Comparing 2D and 3D Learning Experiences*, as published in *Medical Sciences Educator*. List of Authors: Brandon Lowry, Samantha McGrath, Chad Eitel, Kenneth Ivie Jr., Carolyn Meyer, Becky Wiltgen, Heather Hall, Landon Williams, Samantha Scherner & Tod Clapp

INTRODUCTION

Changes in Anatomy Education

Historically, anatomy educators have situated time-intensive, cadaveric dissection alongside lecture-based instruction as the ideal approach for teaching. Recent medical and university curricula changes have reduced the time allocated to the foundational subject by incorporating less time-intensive and more integrated methods, such as problem-based learning [1] and team-based approaches [2].

Compounding the impact of these changes is the noticeable decrease in support for willed body programs, which are essential for traditional dissection-based teaching [3]. Despite the value and impact of these donors on students' education, growing ethical concerns regarding body donation and the increased demand for donated bodies from other disciplines have resulted in a notable negative shift in public opinion [4–6]. Educators have raised concerns over the reduced opportunity for hands-on dissection and observed shortages of donated bodies, with many questioning the level of anatomy knowledge students can reasonably obtain [7].

New technologies, such as virtual reality (VR), have supplemented traditional methods in many curricula. As a computer generated, headset delivered, immersive environment that allows for user interaction, VR offers a scalable and detailed anatomical model that students can manipulate. VR has supported many educators in preserving the quality of anatomy education despite the limitations of curricular changes, dwindling donor programs, and even the logistical challenges during the COVID-19 pandemic [8].

Adoption of VR Technology

Virtual reality (VR) presents complex anatomical structures in ways traditional methods cannot, as VR can explore three-dimensional models of human anatomy by facilitating intuitive understanding of spatial relationships [9]. Pottle [10] demonstrated VR's effectiveness in anatomy education, noting increased student engagement and improved learning outcomes [10]. Many educators observe positive results with VR, highlighting the capacity to boost engagement and motivation while achieving learning outcomes comparable to traditional methods [11, 12]. Despite this, the adoption of VR in anatomy education has yet to generate a consensus regarding its utility.

Some researchers have cited the higher initial investment and reoccurring costs associated with installing and maintaining VR technologies as notable considerations for both educators and researchers [13–15]. However, research suggests that long-term, maintaining VR technologies are more cost effective, enabling opportunities for large scale deployments that better support educational settings with limited resources [15, 16].

Further compounding the ambiguity associated with the impact of VR-based anatomy education, Tene et al. [17] conducted a systematic review of the technology, revealing a range of opinions on its effectiveness [17]. The ongoing debate highlights the need for further research to identify and understand optimal conditions for VR that most effectively contribute to anatomy education. More specifically, we need to understand how to overcome some of the challenges with individuals using the technology.

Challenges with VR

A challenge when incorporating VR-based education is the potential for cognitive overload. This can hinder learning outcomes by overburdening students' working memory [18].

Cognitive overload occurs when the demands on a student's cognitive system exceed their ability to process information, making it difficult for students to manage the complexity of a topic. Cognitive Load Theory (CLT) provides a useful framework for examining relationship between design and cognitive load. Sweller et al. [19] describe cognitive load as the interaction between a student's cognitive ability and the learning environment [19]. CLT identifies three distinct subtypes of cognitive load: intrinsic, extraneous, and germane. Intrinsic load relates to the complexity or interactivity of the subject being studied. It is specific to the content itself and cannot be altered by the instructional design (i.e., calculus is calculus; neuroanatomy is neuroanatomy) [19]. Teaching methods can reduce the intrinsic load by dividing complex information into smaller, more manageable segments. Extraneous load comes from poor instructional design and may detract from learning goals [19]. CLT theorists recommend that instructional designers reduce extraneous load by delivering content with clear and concise language and learning materials, avoiding unnecessary complexity and distractions. Lastly, germane load pertains to the processes that facilitate learning (e.g., schema construction, automation) [19]. Instructional design choices that enable active learning and engagement can enhance germane load, leading to better understanding and longer-lasting material retention [19].

In VR, factors such as the immersive environment, the volume of information presented, and the level of interactivity can contribute to cognitive overload, as shown by Makransky and Lilleholt [20]. By identifying elements of the VR experience that negatively impact a student's working memory, educators and VR designers can create more effective and manageable learning experiences that mitigate the risk of cognitive overload (i.e., best practices for instructional design in VR).

Addressing the obstacles of VR-based anatomy education is crucial for successful integration of the technology in educational settings. Educators and instructional designers must understand how VR-based education impacts cognitive processes and learning experiences. By identifying best practices for VR instructional design that support the optimal learning conditions, researchers can make design choices that mitigate the adverse effects of VR while maximizing utility and accessibility.

This current study addresses the gap in the literature by examining the interaction of 2D and 3D data, student experience, and student characteristics with cognitive load. The purpose of this study was to evaluate the effect of different variables on student cognitive load and experience. Specifically, researchers sought to: the

1. Explore differences in students' cognitive load when viewing 2D or 3D data in immersive virtual reality.
2. Examine how student characteristics impact cognitive load.
3. Determine the impact of visuospatial ability on student cognitive load.
4. Evaluate the impact of cognitive load on learning outcomes.

MATERIALS AND METHODS

Participants

Seventy undergraduate students in the spring of 2024 were recruited for the primary analyses (2D group $n = 33$, 3D group $n = 37$). Preliminary data were collected across the 2023 calendar year (i.e., spring, summer, and fall 2023 semesters) and modifications in procedure and assessment helped to inform the current procedures. The 2023 data are treated as preliminary due

to differences in procedure, issues with data contamination, and edits to assessment questions throughout 2023.

Participants were excluded if they were under 18 years old or not enrolled in an upper-level Human Gross Anatomy course. They received course credit for participation; extra credit and monetary incentives were not offered. After completing the assignments, students were given the opportunity to provide informed consent on whether their data could (or could not) be used in research analysis. Participants with data collection represented in the results are from the 2024 cohort. General demographics were collected, but specific demographics were not paired with experimental and observation data, as many marginalized identities (e.g., female, BIPOC [Black, Indigenous, People of Color]) were not adequately represented within the research team. Additionally, analysis related to cognitive load differences among demographic groups was not conducted on these groups to maintain anonymity of BIPOC students. The program has a small number of BIPOC students which would allow for identification. This study adhered to CSU Institutional Review Board policies and was approved under IRB #4132.

Study Design

The research team used the mental rotation test scores to randomize participants into two homogeneous groups based on visuospatial ability. Students were divided into a group that participated in either the two-dimensional module or the three-dimensional module (2D $n = 33$; 3D $n = 37$). Before the start of the intervention, all participants completed a pre-test and a pre-survey to measure existing understanding of laryngeal anatomy. These pre assessments also captured student attitude and perception of VR in anatomy education, their prior use of VR and video games, and their overall physical comfort with using VR. Participants then watched a module covering the anatomy of the larynx using either 2D or 3D presentation depending on

their experimental group. Both modules were viewed while in immersive VR. A post-test and post-survey were administered immediately following the viewing experience.

Modules

The investigators utilized existing software (*Perspectus VR*) to build the two modules shown in VR. The 3D module and the 2D module were conducted in immersive VR. In both modules participants were asked to watch a pre-recorded lesson covering the anatomy of the larynx, a subject that was chosen for its structural complexity. Once in the fully immersive virtual environment, no controllers or model manipulation were used after the recording began. To ensure consistency, a senior instructor recorded both modules to maintain a similar outline, content, and verbiage. The duration of the 3D module was two minutes and four seconds, and the 2D module was two minutes and 45 s. Similarly, a concerted effort was made to keep the scripts in each module as close as possible. Keeping the modules to three minutes or less provided adequate time to review the relevant anatomical information, while limiting the risk of cognitive overload. Students were given the opportunity to watch the 3D or 2D module during VR open lab times, but they were not required to complete both modules. Students were tracked via a sign-up sheet and online survey to determine if they had completed at least one module. Some students watched both modules, but the second viewing was completed at a later time, after the post-test and post-survey were administered. Figure 1 shows what the participants saw during a similar moment in both the 2D and 3D modules. Table 1 describes the differences between the two modules.

Measuring Cognitive Load (aim 1)

To examine aim 1, a virtual reality head mounted display (HMD) was used in this investigation: the HP G2 Reverb Omnicept edition. This headset contains biometric sensors to

track eye movements (eye-tracking), pupil dilation (pupillometry), and heart rate variability (pulse plethysmography) which achieved ~ 79% accuracy in aligning quantitative cognitive load measures with qualitative reports [21]. The Omnicept reports a cognitive load score based on its algorithm and integrated software. This cognitive load score was collected as participants watched their respective module.

Student Attitude and Experience (aim 2)

To examine aim 2, a pre-survey was used to examine student characteristics and experiences. The survey included questions about student attitudes toward VR in anatomy education, preference for using VR, physical comfort with VR, how frequently they used VR in the past, and if they played video games in the past (Table 2). Students were also asked if they were retaking the course at the time of the research study. Each survey was coded by two researchers using a deductive approach, categorizing responses into groups (e.g., positive, negative, mixed; comfortable, uncomfortable, mixed; likes VR, does not like VR, mixed feelings toward VR). Key-terms were established prior to initial coding. Codes were added and recategorized following initial coding before being used for a final assessment. Student responses that were unclear to both researchers were excluded from analyses. Table 2 demonstrates this process.

Student Perception of Cognitive Load (aim 2)

Similarly, to examine how student experience affected cognitive load, we utilized a post-survey to capture student perception of cognitive load. We adapted the NASA task load index into a survey that, for each question, included an open-ended response as well as 10-point Likert scale [22]. This allowed students to quantitatively rate each of the questions and provide comments. Researchers coded the qualitative responses across six factors – physical demand, mental demand, temporal demand, perceived task difficulty, perceived success, and any negative

emotions—as low (1), medium (2), or high (3). The NASA task load index is included as Supplemental Item 1. Table 1 highlights the differences between the two modules.

Table 1 Differences between modules

Condition	Length(min:sec)	Number of images	Narration	Interaction
2D	2:47	10	Same	Student stood and watched
3D	2:15	1 image	Same	Student stood

Figure 1

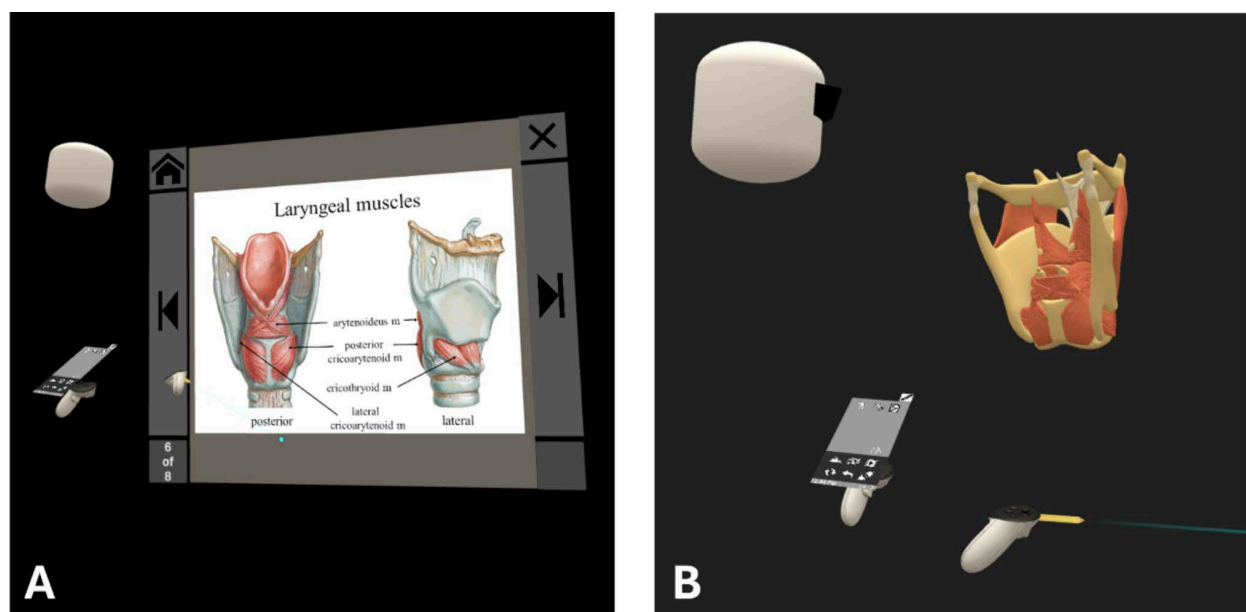


Fig. 1 Participants' view of analogous content in both the 2D and 3D modules. (A) In the 2D group, students saw the instructor navigate through a slide show with labeled images of the larynx. (B) In the 3D group, students viewed the instructor interact with a dynamic virtual model of the larynx; as the instructor interacted with the virtual model, text labels would appear naming the respective structure (not pictured here). In both images, the white avatar head depicts the senior instructor

Table 2

Variable	Question	Category	Goal	Final data type
Attitude toward VR in education	"What is your opinion of the use of VR in anatomy education?"	Positive, negative, mixed, unclear	How does student attitude toward VR in education affect cognitive load?	Positive = 1 Mixed = 2 Negative = 3

Attitude toward VR in general	“What is your opinion of VR in general?”	Positive, negative, mixed, unclear	How does student attitude toward VR in general affect cognitive load?	Positive = 1 Mixed = 2 Negative = 3
Physical comfort using VR	“How physically comfortable are you with using VR equipment?”	Comfortable, uncomfortable, mixed, unclear	How does students’ physical comfort affect cognitive load?	Comfortable = 1 Mixed = 2 Uncomfortable = 3
Frequency of Prior Use of VR	“How often do you use VR outside of school?”	None, low, moderate	How does students’ familiarity with VR technology affect cognitive load?	None = 1 low = 2 moderate = 3
Frequency of Prior Video Game Use	“Do you play video games?”	Yes, no	How does students’ prior use of video games affect their cognitive load in VR?	Yes = 1 No = 0
Retaking the course	“Are you retaking this course?”	Yes, no	How does prior exposure to content affect cognitive load?	Yes = 1 No = 0

Mental Rotation Test (aim 3)

Visuospatial ability was assessed using the Mental Rotation Test (MRT) initially developed by Vandenberg and Kuse in 1978 [23]. The Mental Rotation Test presents participants with a target image next to four comparison images. Participants are tasked with correctly identifying the two rotated versions with the comparison set by marking them with an X. There are twenty-four questions that participants complete after a seven-minute instruction and demonstration period covering the mechanics of the test. Students were given three minutes to complete the first half of the assessment, followed by a two-minute quiet break, and then three minutes to complete the second half of the assessment. The MRT was given to participants prior

to the pre-test and pre-survey to assess aim 3. Assessments were scored by two researchers; points were only given if the student identified both correct images. Researchers then used these data to assign students to research groups and to ensure similar visuospatial ability (as measured by the MRT) among the two groups. Researchers also used this data to explore the relationship between visuospatial ability and cognitive load (aim 3).

Pre-Test and Post-Test (aim 4)

To examine aim 4, researchers developed a pre-test by using a test/retest protocol with a small group of Human Gross Anatomy alums unaffiliated with the experiment. We finalized a seven-item assessment with six multiple choice questions and one true/false question (see Supplemental Item 2). The pre-test and post-test were designed to assess the following learning outcomes.

Learning outcomes:

- Identify the basic anatomy and location of the larynx
- Recall general functions of the larynx
- Understand how pitch is produced

In addition to indicating their answer on the pre-test, students also rated their perceived level of confidence in their answer. After participants watched their respective module, researchers administered a post-test that measured the same questions as the pre-test. Responses were examined to identify changes in students' answers from the pre-test, as well as any changes in how students rated the perceived difficulty and confidence in their answers after viewing the module. Students were not made aware of the correct answers after the pre-test or post-test.

Table 3 demonstrates the alignment between the aims and measurements included in this study.

Table 3

Aim addressed	Measurement	Statistical analysis
Aim 1	Biometric cognitive load reported by Omnicept	Independent t-test, ANOVA
Aim 2	pre-post Surveys	Frequencies, regression, correlation
Aim 3	Mental Rotation Test	Group comparison (mean)
Aim 4	pre-post Test	Independent t-test, paired t-test, correlation

Table 3 demonstrates the alignment between the study aims, measurement, and statistical analysis used.

Statistical Analysis

Data were analyzed using R-studio to explore relationships between the modules, measured and perceived cognitive load, and learning outcomes. Researchers compared the average cognitive load score across the module for each participant as a continuous measure of their mental effort during the experience. Independent t-tests were used to assess differences in cognitive load between module groups (Table 3). Independent t-tests were also used to assess the relationship between learning outcomes and cognitive load (Table 3). Paired t-tests were used to assess pre and posttest changes. Learning gain scores were computed and analyzed via independent t-tests. Correlation coefficients were used to examine the relationship between visuospatial ability, cognitive load, and measured and perceived cognitive load. Significance was defined as $p < 0.05$. Descriptive statistics were computed and reported.

RESULTS

Participant Information

Table 4 shows the participant engagement, sample sizes, and notes pertaining to each activity. Participants represented in this chart and in the results presented are from the 2024 cohort. These participants were undergraduate students enrolled in a prosection based human anatomy

course. Specific student demographics were not assessed as some demographics may have allowed for participant identification in the sample size.

Table 4 Summary of participant engagement and sample sizes

Stage	Description	2D (<i>n</i>)	3D (<i>n</i>)	Total (<i>n</i>)	Notes
Enrollment	Eligible students			100	
Consent	Completed consent			70	All 100 students were given access to the module
Allocation	Randomized by MRT	33	37		
Pre-Assess	Completed pre-survey	33	37		
Intervention & post-asses	Completed pre-test, viewed module, completed post test	32	36		Missing data <i>n</i> = 2

Table 4 shows the total number of students enrolled in the study throughout its duration.

AIM 1: Data Type and Cognitive Load

The difference in cognitive load, as recorded by the HP Omnicept, was statistically significant between the 2D and 3D modules. Participants who completed the 3D module displayed a lower cognitive load ($M = 0.49$, $SD = 0.07$) than those who completed the 2D module ($M = 0.58$, $SD = 0.05$), $t(64.86) = 6.63$, $p < 0.001$, 95% CI [0.06, 0.012] (Fig. 2). These data suggest that interacting with 3D content in immersive VR reduced mental effort compared to interacting with 2D content. Descriptive values for all measures are summarized in Table 5.

Figure 2 Measured Cognitive Load by Module

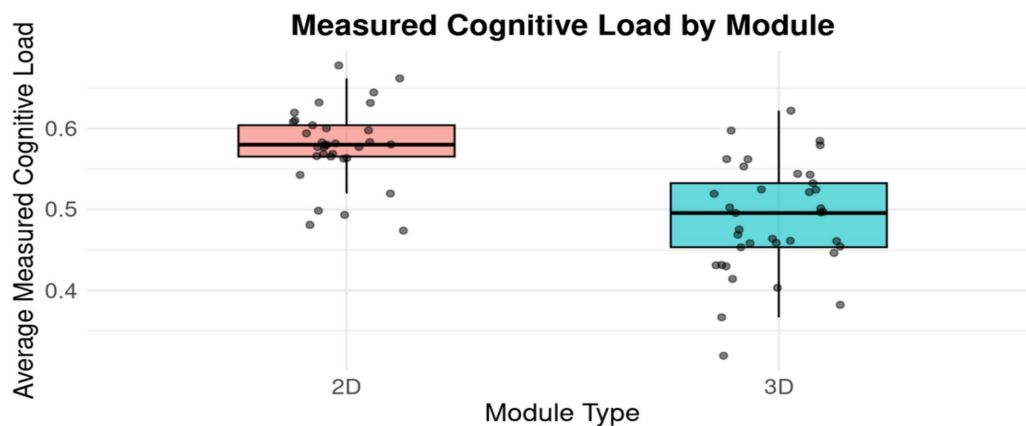


Fig. 2 A scatterplot representation of the average cognitive load by module for the individual participants. These data are overlayed onto a boxplot representing group averages

Table 5 Summary of descriptive statistics for cognitive load, visuospatial ability, and test performance by module

<u>Measure</u>	<u>2D (<i>n</i> = 33) <i>M</i> (SD)</u>	<u>3D (<i>n</i> = 37) <i>M</i> (SD)</u>
Objective Cognitive Load	0.58 (0.05)	0.49 (0.07)
Mental Rotation (MRT)	10.94 (4.12)	11.03 (4.07)
pre-Test Score	5.85 (1.12)	5.69 (0.98)
post-Test Score	5.36 (0.90)	5.11 (0.95)
Perceived Cognitive Load	0.37 (0.16)	0.33 (0.14)

Table 5 shows averages for measured cognitive load, MRT score, Pre and Post-Test Score, and Perceived Cognitive Load for the 2D group (*n*=33) and 3D group (*n*=37).

AIM 2: Student Experience and Cognitive Load

Student characteristics and experiences as measured by pre- and post-surveys (Table 2) did not have a statistically significant effect on measured cognitive load within the 2D or 3D module groups (Fig. 3). The perceived cognitive load from the post-survey was lower for the 3D module (*M* = 0.32, *SD* = 0.14) than the 2D module (*M* = 0.37, *SD* = 0.16) though these changes were not statistically significant, $t(65.07) = 1.29$, $p = 0.20$, 95% CI [-0.03, 0.12] (Figure 4). The measured and perceived cognitive load were not significantly related, $r(68) = -0.01$, $p = 0.91$, which suggests student self-reported cognitive load via the post-survey did not strongly align with their measured cognitive load. Descriptive values are summarized in Table 5.

AIM 3: Visuospatial Ability and Cognitive Load

The MRT scores did not differ significantly between the 3D (*M* = 11.03, *SD* = 4.07) and 2D (*M* = 10.94, *SD* = 4.12), $t(66.34) = 0.09$, $p = 0.93$. This supports the assertion that the randomization process produced comparable visuospatial skills in participants across the two experimental groups. Subsequent analyses demonstrated that MRT scores were not significantly correlated with cognitive load, $r(67) = -0.05$, $p = 0.70$, which suggests individual spatial ability did not predict cognitive load difference between the two groups. Descriptive values from the MRT are summarized in Table 5.

Figure 3

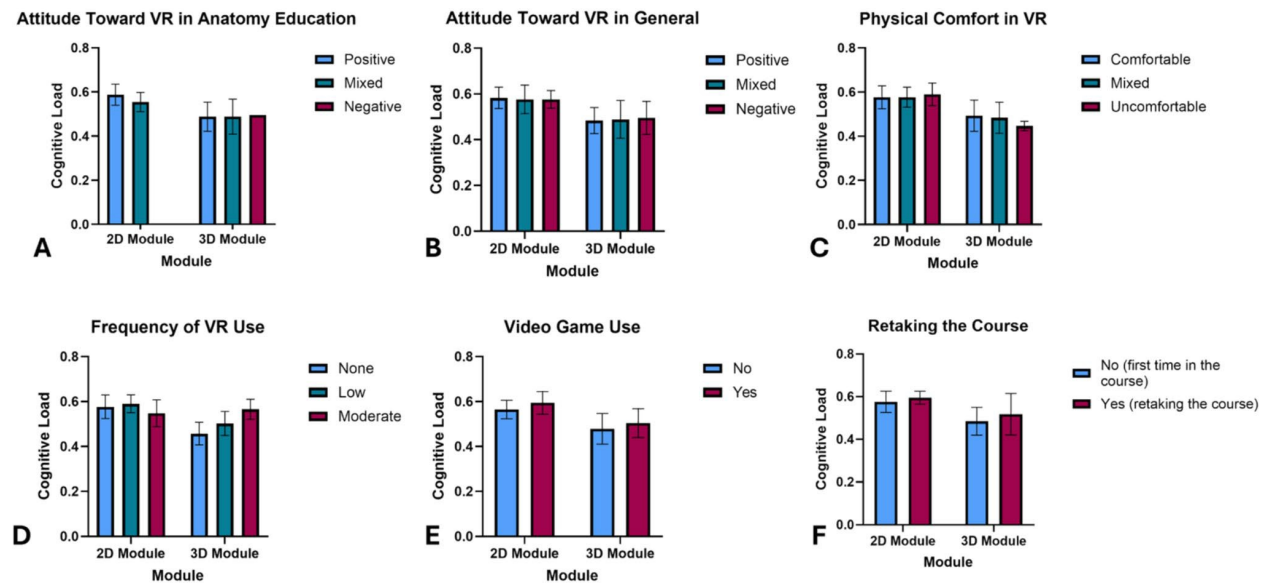


Figure 3 shows the measured cognitive load based on student responses to the pre-survey. Graphs indicate mean cognitive load $\pm$ SD. This figure was created in collaboration in Chad Eitel and Carolyn Meyer, PhD.

Figure 4

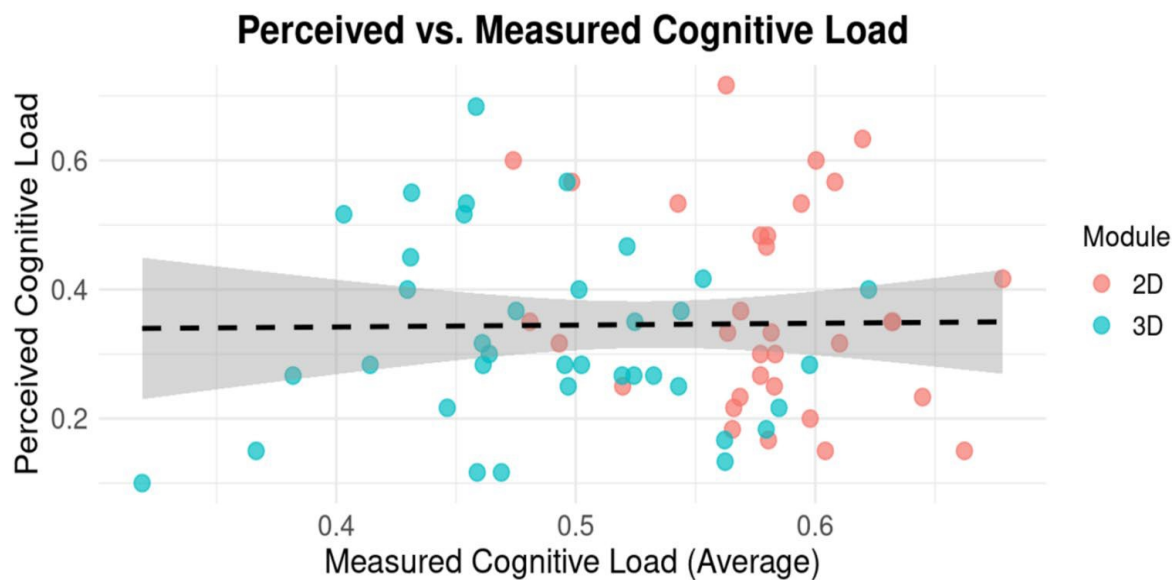


Figure 4 Measured cognitive load and perceived cognitive load by module. Each data point represents the measured and perceived cognitive load for a participant

AIM 4: Learning Outcomes and Cognitive Load

Analyses within the respective groups revealed significant pre to post-test gains. The 2D group improved at lower level, $t(32) = 2.22$, $p < 0.033$, than the 3D group, $t(36) = 3.57$, $p < 0.001$ (Fig. 5). Despite the observed difference, learning gains between the groups did not differ significantly ($p = 0.69$). Additionally, no significant correlation was observed between cognitive load and learning gains within the respective groups ($p = 0.63$), which suggests reduced cognitive load did not directly result in higher scores within a short-term performance context (Fig. 6). Descriptive values for test scores and cognitive load are summarized in Table 5.

Figure 5



Figure 5 shows the mean score changes for the 2D, and 3D group for pre and post-test

Figure 6

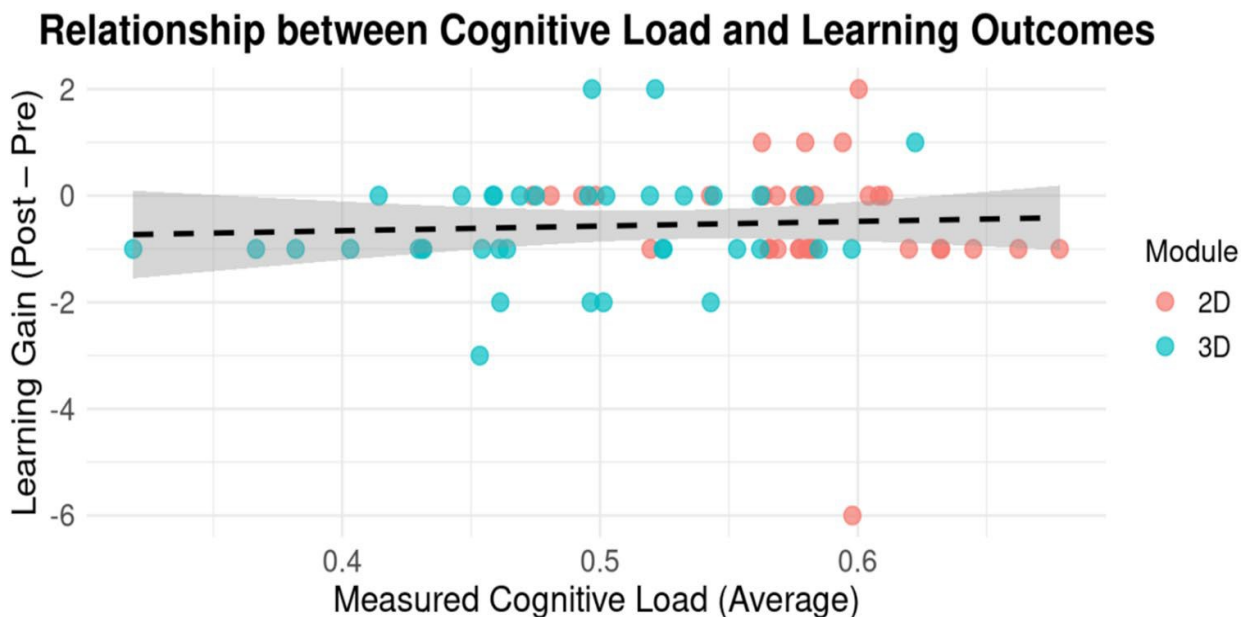


Figure 6 shows relationship between cognitive load and learning gains. Each data point represents the learning gain and objective cognitive load for a participant

DISCUSSION

This study explored the impact of VR-based education at Colorado State University, assessing cognitive load and learning outcomes in 70 undergraduate students using either a 2D or 3D VR module covering the anatomy of the larynx. Key findings suggest that in VR learning modules, 3D-based content significantly reduced cognitive load compared to 2D-based content. However, the reduction in cognitive load did not result in improved test scores. One explanation is that lower cognitive load may not immediately translate to improved recall, but instead to deeper conceptual understanding in the long-term [20]. Future studies may examine this by including longitudinal study designs and analyses in their research. Student attitude toward VR, as well as their preference and prior use of VR, had little effect on student cognitive load. Additionally, whether a student had previously played video games or taken the class showed little effect on cognitive load across both 2D and 3D interventions. This study provides valuable insight into the impact of data visualization and student characteristics on cognitive load. These results contribute to the understanding of the effectiveness of VR in the educational community and continue the foundation laid by previous works in this field [12].

Anatomy Education and VR Integration

Recent trends in anatomy education have shown positive effects of VR on student experience and learning outcomes [11, 12, 16, 24–26]. Here, researchers demonstrate that a utility of VR may lie in its ability to lower cognitive load in student learning. The ability of VR to communicate depth and dimension accurately and more intuitively offers an additional resource to traditional learning methods that may improve knowledge gains in the long term [27]. VR not only supports the decrease of time and resources in anatomy education but allows

for standardization of education and training [1, 28]. Indeed, as VR becomes increasingly embedded in medical curricula, its potential to shape education continues to evolve [10, 29]. Because discomfort or unfamiliarity can affect engagement, these data suggest a need for strategies to enhance user comfort and address the concerns of those with negative perceptions to broaden effective VR adoption [30]. These suggestions are supported by our findings from aim 2, where we found no significant difference in perceived cognitive load between modules, $t(65) = 1.29$, $p = 0.20$.

Cognitive Load

Analysis of cognitive load revealed that students engaging with 3D data in VR experienced lower cognitive load than those using 2D representations ($t(64.86) = 6.63$, $p < 0.001$, 95% CI [0.06, 0.12]). These findings align with previous work that demonstrated how 3D representations may ease cognitive demands in the absence of measurable improvements in performance [10]. Further, these findings reinforce the importance of instructional design that supports efficient mental processing. When content is presented in a format that aligns with how learners intuitively process information, it appears to reduce unnecessary mental effort and promote more effective learning. Traditional visualization of the larynx in a cadaveric donor is challenging due to the location of the structures of the larynx and the surrounding structures. Students typically struggle with visualizing the larynx and often report learning these structures are challenging. This study showed that presenting an isolated VR 3D anatomical model of the larynx is less cognitively taxing on the learner than static 2D images of the region. However, it is important to consider that the larynx itself as a subject may have contributed to the observed cognitive load in both modules. The 3D visualization of the larynx in VR may have reduced

extraneous load by eliminating the need to reconstruct complex 2D images to 3D. Subsequently, more capacity for germane processing may be available (i.e., greater working memory).

In exploring factors that might influence cognitive load, no significant differences were found across student characteristics including attitudes toward VR, prior experience with VR or video games, comfort with the technology, or prior knowledge. Interestingly, moderate VR users did have similar cognitive load between 2 and 3D groups. However, the small sample size for these two groups may have led to the observed cognitive load in these groups. This suggests that the design and presentation of instructional content had a greater impact on cognitive load than individual perceptions and characteristics. As such, educators and designers may benefit from focusing on how information is presented in virtual learning environments over concerns about students' prior exposure or preferences. While students' affective responses to VR remain valuable for understanding engagement and overall experience, analysis emphasizes the importance of prioritizing how content is visualized and structured over individual characteristics. In contexts where access to immersive VR may be limited, implementing instructional strategies that incorporate 3D data may still offer significant cognitive benefits. Further, while a direct and immediate improvement on learning outcomes was not observed, there are a few plausible explanations. The sensitivity of the pre and post-test may not have been sufficient, given that it was constructed with a smaller group of students unaffiliated with the study. The limited length of exposure to the respective modules may have also impacted student improvement.

Experiences Between Modules

While not statistically significant, the post-survey responses indicated that students engaging with the 3D module reported a more positive overall experience compared to those in

the 2D group ($t(65.07) = 1.29, p = 0.20$). Interestingly, students in the 2D group were more likely to report feeling hurried or rushed. In the 2D module, students were required to mentally reconstruct spatial relationships to effectively compensate for the lack of depth cues in the static images presented on the 2D panel. This added mental effort may have contributed to the higher cognitive load observed in the 2D group.

In contrast, the 3D module presented information using a virtual model that maintained depth and spatial relationships of anatomical structures, which likely supported a more intuitive and less cognitively demanding engagement with the content. It is important to note that the instructional design differences could have influenced these results, and future research should find ways to create isomorphic protocols where 2D and 3D comparisons are made. Despite the limitation of instructional differences, these findings reinforce the value of immersive, spatially rich design in virtual learning tools—not only for reducing cognitive load, but also for enhancing the overall learner experience.

Implications for VR-Based Instructional Design

Educational Practices

Together, these data address our research aim by underscoring how presentation format plays a vital role in learner experience. Our findings with respect to differences in measured cognitive load ($p < 0.001$) advocate for the continued integration of VR into medical anatomy education, emphasizing a need for educational strategies that reduce cognitive load with anatomical content. Reducing mental load through strategic use of spatially intuitive visualizations can help students focus on deeper learning rather than the mechanics of interpreting content. As educational institutions adopt VR across a broader range of learning environments, it will be important to tailor instructional design to meet the evolving needs of

learners at various stages. Expanding research into different educational contexts and further understanding the nuances of how different student populations engage with VR can inform the development of teaching strategies that maximize the benefits of immersive learning environments like VR [31]. Good instructional design can optimize intrinsic load and reduce extraneous load for learners.

Technology Development

From a development standpoint, these results demonstrate the importance of user-friendly, educationally effective tools that are accessible and tailored to educational settings. Emphasis should be placed on content clarity, spatial coherence, and the ability to minimize extraneous load [32]. Incorporating adaptive design features that respond to learner performance and cognitive load in real time may further enhance the effectiveness of virtual instruction. The real-time measurement of cognitive load employed in this study complements existing frameworks for instructional design in virtual reality and underscores the importance of designing VR tools that are not only immersive but also pedagogically aligned.

One consistently observed finding, both subjectively by researchers and directly by the student, was the need for freedom of movement within the physical space. Students who felt agency in moving about the physical environment consistently reported a more positive experience, emphasizing that developers need to consider how the design of both virtual and physical environments may influence not just engagement, but also mental effort. Future studies that examine and manipulate different levels of movement constraints or spatial interactivity within VR will help inform recommendations for static or dynamic movement capabilities, while also clarifying the effect of movement on cognitive load in VR-based environments.

Limitations

While effort was made to create identical modules, the slight variation between two modules may have introduced some variance in the overall demand placed on the students. Additionally, these results focus on short-term outcomes, and the lack of longitudinal analyses limits the generalizability of these findings. The lack of a larger sample due to the progressive nature of the study limits the generalizability of the findings described here. An additional limitation to consider is that this particular study only examined one area of anatomy, the larynx. This could make it difficult to extrapolate the results to other anatomical areas. Although both 2D and 3D formats were presented within the VR system to allow for biometric measurements, the virtual environment is distinct from a typical 2D presentation, and this limitation should be considered.

Future Directions

To expand on the current findings, longitudinal studies are needed to assess how cognitive load and student experience evolve over time with continued exposure to VR. Follow-up research intends to observe a smaller sample size participating in a longitudinal experience. This focus will allow for deeper analysis concerning student cognitive load and how it changes as one becomes more familiar with the technology. This study raised questions about how quickly users adapt to VR and what level of training support is most effective for helping them learn to use the technology. Observing learners across repeated sessions will clarify how familiarity with immersive environments develops and impacts both mental effort and learning outcomes. Additionally, research into the optimal duration and pacing of VR experiences could inform instructional design strategies that balance cognitive demand with engagement and autonomy. Future studies should address these needs with longer-term retention studies, as well as incorporating different levels of interactivity within the experience. As universities launch

VR-based curriculum, additional research on curriculum-wide integration will provide valuable insight into the efficacy of VR and, most importantly, rich insight into the relationship between cognitive load and improved learning outcomes.

Conclusion

This study supports the integration of immersive VR learning in anatomy education by demonstrating that 3D data visualizations reduce cognitive load compared to 2D representations. Continued collaboration among educators, researchers, and developers is paramount in exploring the full potential of VR. Future research exploring additional variables associated with student experience may further contextualize these findings. Refining tools in VR that are both pedagogically effective and accessible across diverse educational settings will further enhance the utility of immersive learning environments in anatomy and beyond.

Acknowledgements Generative AI: The first authors completed an original manuscript without the assistance of Generative AI (GenAI). GenAI was used to suggest revisions on a few sentences to be more straightforward and accessible (i.e., identifying areas with too much jargon or poor readability). The first authors' use of GenAI in this context is situated around two areas: clarity and succinctness.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or

exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

SUPPLEMENTARY

Supplemental Item 1. NASA Task Load Index

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental DemandHow mentally demanding was the task?

Supplemental Item 2. Pre-Post-Test

Larynx Anatomy: Pre-test

**same questions used for both pre- and post-test*

Question 1:

What is the main function of the larynx?

- A) To filter the air we breathe
- B) To help you breathe and speak
- C) To absorb water from food
- D) To regulate blood flow to the throat

How much of a guess is your answer? Circle one:

Completely guessing, partially guessing, not guessing at all

Question 2:

Where is the larynx located?

- A) Just below the nose
- B) In the throat, above the trachea
- C) At the top of the lungs
- D) In the upper part of the diaphragm

How much of a guess is your answer? Circle one:

Completely guessing, partially guessing, not guessing at all

Question 3:

The larynx contains structures known as vocal cords. What do they do?

- A) Help in filtering air
- B) Help in breathing and producing sounds
- C) Aid in the movement of the laryngeal cartilages
- D) Control the flow of food and water

How much of a guess is your answer? Circle one:

Completely guessing, partially guessing, not guessing at all

Question 4:

True or False: The larynx is also known as the voice box.

- A) True
- B) False

How much of a guess is your answer? Circle one:

Completely guessing, partially guessing, not guessing at all

Question 5:

What happens to the larynx when you swallow food?

- A) It contracts to push food down.
- B) It moves upward and forward to prevent food from entering the windpipe.
- C) It expands to allow more food to pass.
- D) It relaxes to facilitate the passage of food to the esophagus.

How much of a guess is your answer? Circle one:

Completely guessing, partially guessing, not guessing at all

Question 6:

Which of the following is not a part of the larynx?

- A) Thyroid cartilage
- B) Cricoid cartilage
- C) Arytenoid cartilage
- D) Maxillary cartilage

How much of a guess is your answer? Circle one:

Completely guessing, partially guessing, not guessing at all

Question 7:

The cricothyroid muscle plays an important role in:

- A) Opening the rima glottidis for easier breathing
- B) Closing the rima glottidis to protect the airway
- C) Increasing the tension of the vocal cords, thereby changing pitch
- D) Decreasing the tension of the vocal cords to produce lower sounds

How much of a guess is your answer? Circle one:

Completely guessing, partially guessing, not guessing at all

REFERENCES

1. Sugand K, Abrahams P, Khurana A. The anatomy of anatomy: a review for its modernization. *Anat Sci Educ*. 2010;3(2):83–93.
2. Zgheib N, Dimassi Z, Akl I, Badr K, Sabra R. The long-term impact of team-based learning on medical students' team performance scores and on their peer evaluation scores. *Med Teach*. 2016;38:1017–24.
3. McBride JM, Drake RL. National survey on anatomical sciences in medical education. *Anat Sci Educ*. 2018;11:7–14.
4. Choi W. Teaching respect for body donors: a US medical student perspective. *Anat Sci Educ*. 2024;17:687–92.
5. Keet K, Pillay P, Billings B, Satyapal K, Kramer B. Profiting from public dissection: a new era of unethical practices. *Anat Sci Educ*. 2023;16:384–90. <https://doi.org/10.1002/ase.2256>.
6. Jones DG, Whitaker MI. Anatomy's use of unclaimed bodies: reasons against continued dependence on an ethically dubious practice. *Clin Anat*. 2012;25(2):246–54.
7. Bergman EM, Prince KJ, Drukker J, van der Vleuten CP, Scherpbier AJ. How much anatomy is enough? *Anat Sci Educ*. 2008;1(4):184–8.
8. Patra A, Asghar A, Chaudhary P, et al. Integration of innovative educational technologies in anatomy teaching: new normal in anatomy education. *Surg Radiol Anat*. 2022;44:25–32. <https://doi.org/10.1007/s00276-021-02868-6>.
9. Kurul R, Ögün MN, Neriman Narin A, Avci Ş, Yazgan B. An alternative method for anatomy training: immersive virtual reality. *Anat Sci Educ*. 2020;13:648–56. <https://doi.org/10.1002/ase.1959>.
10. Pottle J. Virtual reality and the transformation of medical education. *Future Healthc J*. 2019;6(3):181–5.
11. Brown KE, Heise N, Eitel CM, et al. A large-scale, multiplayer virtual reality deployment: a novel approach to distance education in human anatomy. *Med Sci Educ*. 2023;33:409–21.
12. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anat Sci Educ*. 2017;10(6):549–59.
13. Taylor L, Dyer T, Al-Azzawi M, Smith C, Nzeako O, Shah Z (2021) Extended reality anatomy undergraduate teaching: a literature review on an alternative method of learning. *Ann Anat = Anat Anz: Off Organ Anat Ges*. 2021;239:151817.
14. Chen S, Zhu J, Cheng C, Pan Z, Liu L, Du J, et al. Can virtual reality improve traditional anatomy education programmes? A mixed-methods study on the use of a 3D skull model. *BMC Med Educ*. 2020;20(1):395.
15. Barteit S, Lanfermann L, Bärnighausen T, Neuhaus F, Beiersmann C. Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. *JMIR Serious Games*. 2021. <https://doi.org/10.2196/29080>.

16. Heise N, Brown K, Nelson J, Eitel C, Meyer C, Ivie K, et al. A large-scale virtual reality deployment for distance education in human anatomy. *FASEB J*. 2021. <https://doi.org/10.1096/fasebj.2021.35.S1.00339>.
17. Tene T, Vique López DF, Valverde Aguirre PE, Orna Puente LM, Vacacela Gomez C. Virtual reality and augmented reality in medical education: an umbrella review. *Front Digit Health*. 2024;6:1365345. <https://doi.org/10.3389/fdgth.2024.1365345>.
18. Makransky G, Terkildsen TS, Mayer RE. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learn Instr*. 2019;60:225–36. <https://doi.org/10.1016/j.learninstruc.2017.12.007>. (ISSN 0959-4752).
19. Sweller J, Ayres P, Kalyuga S. *Cognitive load theory*. Springer; 2011.
20. Makransky G, Lilleholt L. A structural equation modeling investigation of the emotional engagement–cognitive load–learning outcome relationship in a science simulation. *J Educ Psychol*. 2018;110(7):1066–75.
21. Siegel EH, Wei J, Gomes A, Oliviera M, Sundaramoorthy P, Smathers K, Vankipuram M, Ghosh S, Horii H, Bailenson J, Ballagas R. HP omnicept cognitive load database (HPO-CLD) – Developing a multimodal inference engine for detecting real-time mental workload in VR. Technical report, HP Labs, Palo Alto, CA; 2021. <https://developers.hp.com/system/files/attachments/HPO-CLD%20Technical%20Report%204.30.21.pdf>
22. Hart, S. G. NASA task load index (TLX). 1986. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9).
23. Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. *Percept Mot Skills*. 1978;47(2):599–604. <https://doi.org/10.2466/pms.1978.47.2.599>.
24. Liu J, Yin Y, Kor P, Cheung D, Zhao I, Wang S, et al. The effects of immersive virtual reality applications on enhancing the learning outcomes of undergraduate health care students: systematic review with meta-synthesis. *J Med Internet Res*. 2023. <https://doi.org/10.2196/39989>.
25. Hamilton D, McKechnie J, Edgerton E, Wilson C. Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. *J Comput Educ*. 2020;8:1–32. <https://doi.org/10.1007/s40692-020-00169-2>.
26. Zhao J, Xu X, Jiang H, Ding Y. The effectiveness of virtual reality-based technology on anatomy teaching: a meta-analysis of randomized controlled studies. *BMC Med Educ*. 2020;20(1):127. <https://doi.org/10.1186/s12909-020-02032-2>.
27. Paas F, Renkl A, Sweller J. Cognitive load theory and instructional design: recent developments. *Educ Psychol*. 2003;38(1):1–4.
28. Berg LP, Vance JM. Industry use of virtual reality in product design and manufacturing: a survey. *Virtual Reality*. 2017;21:1–17. <https://doi.org/10.1007/s10055-016-0293-9>.

29. Xu X, Mangina E, Campbell A. HMD-based virtual and augmented reality in medical education: a systematic review. *Front Virtual Reality*. 2021;2:82. <https://doi.org/10.3389/frvir.2021.692103>.
30. Weech S, Kenny S, Barnett-Cowan M. Presence and cybersickness in virtual reality are negatively related: a review. *Front Psychol*. 2019;10:158. <https://doi.org/10.3389/fpsyg.2019.00158>.
31. Radianti J, Majchrzak TA, Fromm J, Wohlgenannt I. A systematic review of immersive virtual reality applications for higher education. *Comput Educ*. 2020;147:103778. <https://doi.org/10.1016/j.compedu.2019.103778>.
32. Makransky G, Petersen GB. The cognitive affective model of immersive learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. *Educ Psychol Rev*. 2021;33:937–58. <https://doi.org/10.1007/s10648-020-09586-2>.

CHAPTER 3: Leveraging Generative AI to Foster Metacognition and Self-Directed Learners

ABSTRACT

With the ever-expanding amount of data, students increasingly find themselves needing to engage in self-directed learning to be successful. Students studying science, technology, engineering, and mathematics often struggle with self-directed learning and are often discouraged, leading to higher attrition within these disciplines. There is a lack of opportunities for students to develop and practice self-directed learning skills within traditional curricula. This research explored the ways in which a generative artificial intelligence model could be used to cultivate metacognition and promote readiness for self-directed learning among graduate students. By leveraging the relationship between metacognition and self-directed learning, with the customizability of the artificial intelligence model, we sought to facilitate conversations between students and the model to enhance metacognitive awareness and self-directed learning readiness. Using the Metacognition Awareness Inventory and Self-Directed Learning Instrument, we found that students improved significantly on both pre- and post-assessment comparisons. Students needed to interact with the model twice a week, for 10 minutes per session. Our findings demonstrate a novel application of generative artificial intelligence in supporting students' personal development and expand our understanding of how artificial intelligence can be leveraged to generate a supportive process, rather than solely as a mechanism for generating answers or some other product.

-This chapter is a reproduction of, *Leveraging Generative AI to Foster Metacognition and Self-Directed Learners*, as published in the Journal of Microbiology and Biology Education. List of authors: Brandon Lowry (co-author), Samantha McGrath (co-author), Chad Eitel, Heather Hall, Tod R. Clapp.

INTRODUCTION

Supporting students in their progression from dependent to self-directed learners is crucial for their success in the technology-rich and rapidly evolving global workforce of the 21st century. Self-directed learning (SDL) occurs when students identify and set learning goals, select and utilize appropriate resources, and evaluate progress (1). This practice supports educational goals as it encourages students to adapt to their specific needs, subsequently enhancing motivation, confidence, and learning outcomes (2). The current literature demonstrates wide-reaching applications and benefits of this practice, spanning from middle school students (3) to adults actively engaged in the workforce (4). Educators face significant obstacles in supporting students in developing into self-directed learners, many of which stem from variable readiness for engaging in SDL (5). Additionally, there are few opportunities for students to practice SDL—a challenge inherent to the nature of college education and its focus on predetermined learning objectives (4).

Fundamentally, SDL supports students to become more critical and adaptable in their learning efforts. Students who readily engage with SDL are more adept at problem-solving. These beneficial changes support their efforts in various majors and careers across the spectrum of STEM disciplines (6). Students continue to explore SDL and progress into more self-directed learners, experiencing beneficial changes in motivation and self-efficacy, directly supporting increased engagement (7) and further exploration within self-directed contexts (8). These experiences directly support the students becoming lifelong learners (9). Lifelong learning not only supports the individual and their well-being, but it also stands to directly strengthen the global workforce.

To help catalyze the potential benefits of SDL for our students, the research team focused on leveraging customized generative pre-trained transformer (GPT) technology to facilitate conversations meant to promote self-directed learners. The following sections discuss the relationship between SDL, motivation, and metacognition, as well as how educational technology (e.g., GPTs) has been used to support the practice and skill development of SDL. Findings from an 8-week experiment with graduate neuroanatomy students are discussed and highlight a novel application of customized GPTs in supporting the development of self-directed learners.

SDL, Metacognition, and Motivation

Self-directed learning occurs when a person identifies a learning need and feels compelled, or motivated, to address that need. Indeed, motivation is essential for the initiation and, more importantly, sustaining effort within any learning task. Within the realm of SDL, the presence of intrinsic motivation contributes to engaging with material and persisting through adversity to overcome various challenges. Additionally, when students begin to associate personal growth or academic success as motivating factors, they often display sustained engagement with material (10). Researchers have also found that as students focus on long-term development of mastery within a subject, they are more likely to continually engage in self-directed learning, focusing on skill development over rote memorization (11). Helping students understand the factors that motivate them within a learning activity, class, or major supports them in uncovering intrinsic motivation that will support their efforts through the various STEM disciplines and contexts they encounter.

Metacognition can be defined as “thinking about what one is thinking about,” and it occurs when students reflect on what they know, how they came to know it, and how their efforts

in and out of the classroom contribute to that knowledge. This practice is characterized by various skills, such as planning for and monitoring progress, in addition to self-evaluation. As students engage in this way, they take an active role in the management of their learning and become better equipped to adapt to various learning challenges (12). Furthermore, as students build awareness of how well they are learning material, their ability to identify ineffective methods improves, affording even greater adaptability, and subsequently, an improved ability to engage in both SDL and non-SDL tasks (13).

The beneficial impacts of supporting metacognition and motivation on student learning are amplified within the context of SDL. The cyclical nature of how motivation helps drive metacognitive practice and metacognition in identifying and maintaining intrinsic motivation directly supports and encourages students to engage with SDL. Students who are intrinsically motivated to succeed are more likely to identify when they are not doing well enough and, as such, are more likely to adjust in effort to maintain focus and alignment with their goal. Efklides (14) discussed this feedback loop and its ability to support sustained learning efforts (14). Interestingly, they found that as students practice more metacognitive awareness, they are more attuned to their progress, which supports and maintains motivation.

Research Overview

This research sought to create a low-effort/high-impact, AI-enhanced approach to developing skills pertinent to SDL. Using SDL theory, we leveraged the concepts of metacognition and motivation to directly impact students' readiness for SDL. The objectives of this study were to develop a model that could facilitate reflective practice conversations to support the development of students' metacognitive ability and their understanding of personal motivating factors and, subsequently, their readiness for self-directed learning. This research is

particularly significant because of its generalizability across all STEM disciplines, as the model development process resulted in a product that was content agnostic. The activity and reflective practice described herein could be implemented within any course, as the model was trained not to tutor or teach any neuroanatomy.

Research Questions and Hypotheses

We sought to understand the impact of engaging with the model of students' metacognition and readiness for self-directed learning. Two key questions guided that goal: (R1) what is the impact of interacting with the model on students' readiness for SDL? We hypothesized that interacting with the model would support the development of SDL readiness; and (R2) what is the impact of interacting with the model on students' utilization of metacognitive strategies? Grounded in the understanding that regular reflective thinking supports the development of metacognition, we hypothesized that students who engaged with the model would report higher levels of metacognitive awareness in the context of monitoring, evaluating, and adjusting their learning; exploring these questions and hypotheses provides valuable understanding concerning how interactive education technologies can be leveraged to support the requisite skill development for meaningful and effective engagement in SDL.

METHODS

Research Design

We used a multi-method approach to assess the impact of an AI-based intervention on students' readiness for self-directed learning and use of metacognition. The quantitative measures included two previously developed instruments that assess students' readiness for self-directed learning and utilization of metacognitive strategies; the qualitative measures included two focus group sessions and a survey, which included the same questions as the focus group. By

using quantitative measures, we were able to conduct a statistical analysis concerning the impact of the intervention. The qualitative measures allowed us to contextualize the impact of the intervention within the context of participants' academics.

Participants

We recruited graduate students from a graduate-level functional neuroanatomy course at a Midwest R1 university. All the students were enrolled in a 1-year MS program in biomedical sciences. Recruitment occurred in two phases. In the first phase, researchers communicated with students two times before the start of the study, requesting participants to indicate their interest via email. This phase resulted in the enrollment of 13 students who interacted with the GPT model, which was eventually reduced to 8 students to manage logistical issues with accessing a paid GPT model. In the second phase, researchers sought individuals willing to serve as an experimental control by completing the pre-/post-assessments without interacting with the GPT model. The control group was not given any non-AI resources throughout the intervention period. Thirteen students enrolled through this mechanism, with eight eventually completing both pre-/post-assessments. Participants across both groups ranged in age from 21 to 27 years and included seven students identifying as female and nine as male. Their academic concentrations were predominantly in human anatomy (15 students), with one student focused on animal anatomy.

Materials

We used the Self-Directed Learning Instrument (SDLI) to assess students' readiness for self-directed learning before and after the 8-week intervention period. The instrument was developed to assess readiness for SDL among healthcare students (15). The measure consists of 20 questions, each of which is paired with a five-point Likert scale. Students rate each question

based on how frequently or reliably they demonstrate the described action or behavior. The instrument has been previously used, with Cronbach's Alpha metrics reporting between 0.80 and 0.94 (16). The test has a maximum score of 100, where higher scores indicate greater readiness for self-directed learning.

Initial test-retesting of the SDLI and subsequent confirmatory factor analysis (CFA) revealed one item (question 18) that was not loading in the same manner as previously observed during its development. The research team interviewed the students, who reported that question 18 seemed irrelevant to them. The question, "How important is it for you to understand the culture of those you are learning with?" confused them. After removing question 18 and reassessing the data via CFA, the remaining 19 questions loaded to the appropriate and previously defined factors (i.e., learning motivation, planning and implementation, self-monitoring, and interpersonal communication), with no cross-loading (i.e., when a survey item loads to two or more factors). The removal of Q18 resulted in a maximum score of 95. The instrument was administered to experiment and control group participants; the scores from the pre-assessment were used to determine if the two groups began the process with similar levels of readiness for self-directed learning. The post-assessment scores were used to compare the two groups, as well as the change in score from pre to post within each group. The instrument was chosen because it was rigorously developed and validated in subsequent research.

We used the Metacognition Awareness Inventory [MAI] to assess how often students used metacognitive strategies before the intervention period and to evaluate the impact of the intervention on student use of metacognitive strategies. The Metacognition Awareness Inventory consists of 52 true or false statements divided into two main categories: Knowledge of Cognition

and Regulation of Cognition (17). Prior work has shown a similar level of reliability (Cronbach's $\alpha > 0.90$) in assessing the dimensions of metacognition (18, 19).

We developed a set of questions to use in a focus group session and digital survey. The development process included a pilot test with a small group of graduate students unaffiliated with the project. Responses from the pilot informed minor corrections, and a final draft was agreed upon for the project. The responses were coded using Braun and Clarke's (2006) approach to thematic analysis (20).

Procedure

Students were recruited during week 4 of the 2024 Spring semester. Once informed consent was documented, students were assigned to a group. All students completed the pre-assessment series of the instrument and the Metacognition Awareness Inventory. The intervention group interacted with the model for 8 weeks, twice a week, for approximately 10 minutes per session. Participants signed up for a session time each week via a shared Google Sheet. If students missed their session, a make-up session was scheduled when time permitted. Students deleted their chat sessions after each session to protect their privacy. Moreover, the research team was not concerned with what was discussed; rather, the process of interacting with the model regularly was the objective.

The intervention (AI model) followed a set series of questions determined by the research team. The intervention was programmed to ask the student for more depth or clarity after each response, while adhering to a 10-minute timeframe for duration. The four questions covered elements of metacognition, motivation, and behavior and were developed to cover specific components of self-directed learning. Prompts and questions used are listed in supplemental item 1.

Data Analysis

Descriptive statistics were calculated for all pre- and post-test assessments. Effect sizes were calculated using Cohen's d (1.4 for SDLI and 1.2 for MAI). Post hoc power analysis revealed moderate-to-high power (~ 0.70 – 0.76) for the effects observed. The research team conducted confirmatory factor analysis on the SDLI and MAI prior to the intervention period to determine appropriate factor loading and validity of the instrument to the target population. Results demonstrated that one question (question 18) on the instrument was not loading the same way as during the instrument's initial development process. Additional analysis revealed that by excluding Q18, the remaining 19 questions loaded correctly. The investigators agreed to use the 19-item version of the instrument in the official research intervention. To assess the differences between groups concerning pre- and post-assessment scores from the SDLI and Metacognition Awareness Inventory, we used independent samples t -test. To assess the differences within groups, a paired sample t -test was used. We used a thematic analysis approach to assess qualitative data and investigate student perceptions of the intervention after they completed the 8-week intervention. Transcripts from two focus group sessions were independently reviewed using an inductive approach. Initial codes were formed and grouped into broader categories by two separate researchers. In subsequent meetings, the researchers found consensus and refined the categories into overarching themes. Themes were then used to group student responses.

RESULTS

Self-directed learning instrument

Pre-test SDLI score comparison between intervention and control groups

An independent t -test showed no statistically significant difference between the two groups on the pre-test SDLI assessment ($P = 0.6383$), indicating the two groups began the research study with relatively similar levels of SDL readiness. Descriptive statistics show that the intervention

group had a mean score of 80.00 (SD = 5.95, SEM = 2.10), and the control group had a mean score of 78.13 (SD = 9.30, SEM = 3.29). See Fig. 1, part A.

Post-test SDLI score comparison between intervention and control groups

An independent t-test showed a statistically significant difference between the two groups on the post-test instrument assessment ($P < 0.0204$), indicating that model interaction had a positive effect on scores. Descriptive statistics show that the intervention group had a mean score of 85.25 (SD = 5.14, SEM 1.8173), and the control group had a mean score of 78 (SD = 5.92, SEM = 2.093). See Fig. 1, part B.

Pre-to-post SDLI test score comparison for the intervention group

A paired sample t-test showed a statistically significant difference between pre-post-test scores for the intervention group ($P < 0.0185$), indicating that model interaction had a positive impact on instrument scores. See Fig. 2.

Metacognition Awareness Inventory

Pre-test Metacognition Awareness Inventory score comparison between intervention and control groups

An independent t-test showed no statistically significant difference between the two groups on the pre-Metacognition Awareness Inventory assessment ($P = 0.1594$), indicating the two groups began the research study with similar levels of metacognitive awareness. Descriptive statistics showed that the intervention group had a mean score of 42 (SD = 3.66, SEM 1.3), and the control group had a mean score of 37.75 (SD = 7.4, SEM 2.62). See Fig. 3, part A.

Post-test Metacognition Awareness Inventory score comparison between intervention and control groups

An independent sample t-test revealed a statistically significant difference on the post-Metacognition Awareness Inventory assessment between the intervention and control groups ($P < 0.006$), indicating the model had a positive impact on students' metacognitive awareness.

Descriptive statistics show that the intervention group had a mean score of 45.88 (SD = 4.19, SEM = 1.48), and the control group had a mean score of 35.75 (SD = 7.8, SEM = 2.76). See Fig. 3, part B.

Pre-post-test instrument score comparison between intervention and control groups

A paired sample t-test showed a statistically significant difference between pre- and post-Metacognition Awareness Inventory assessment scores for the intervention group ($P < 0.0030$), indicating the model had a positive impact on the students' utilization of metacognitive strategies. See Fig. 4.

Figure 1 Mean Score on Pre- and Post-SDLI

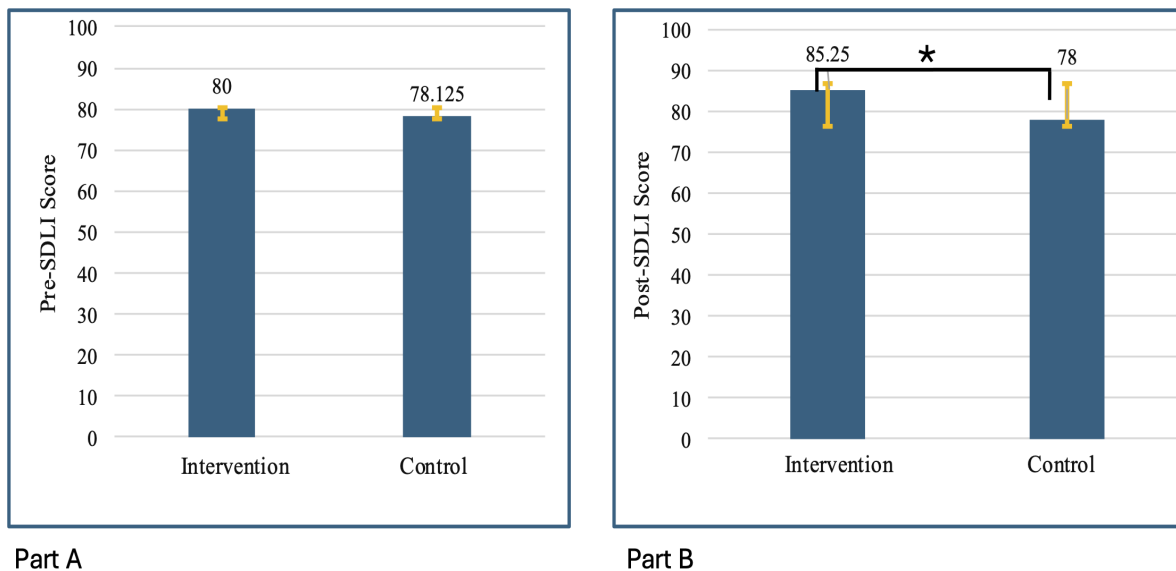


Figure 1 shows the pre-SDLI scores for both the intervention (n=8) and control (n=8) groups

Figure 2 Participant Scores for Pre- and Post-SDLI

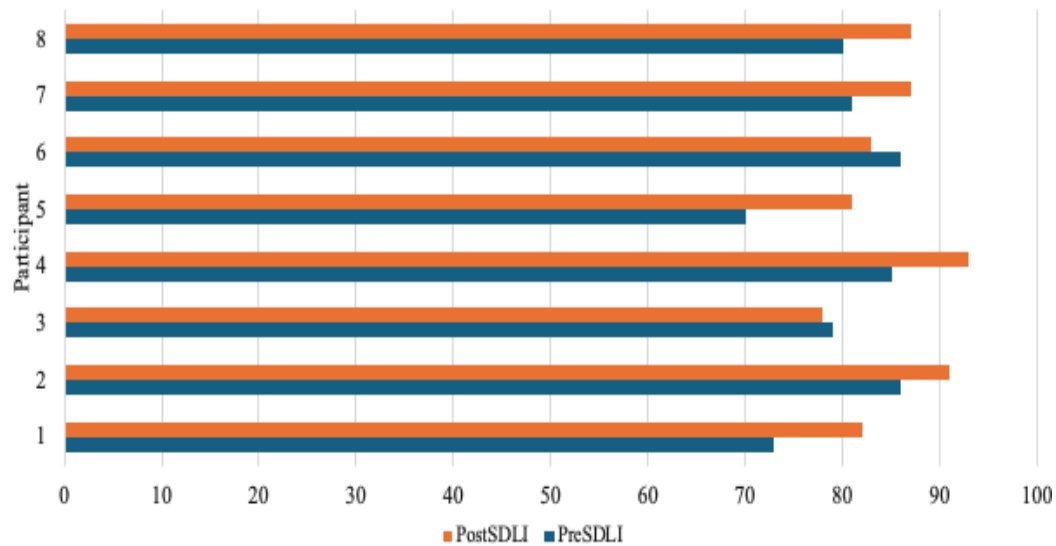


Figure 2 shows the pre- and post-SDLI score for each participant from the intervention group

Figure 3 Mean Score on Pre- and Post-MAI

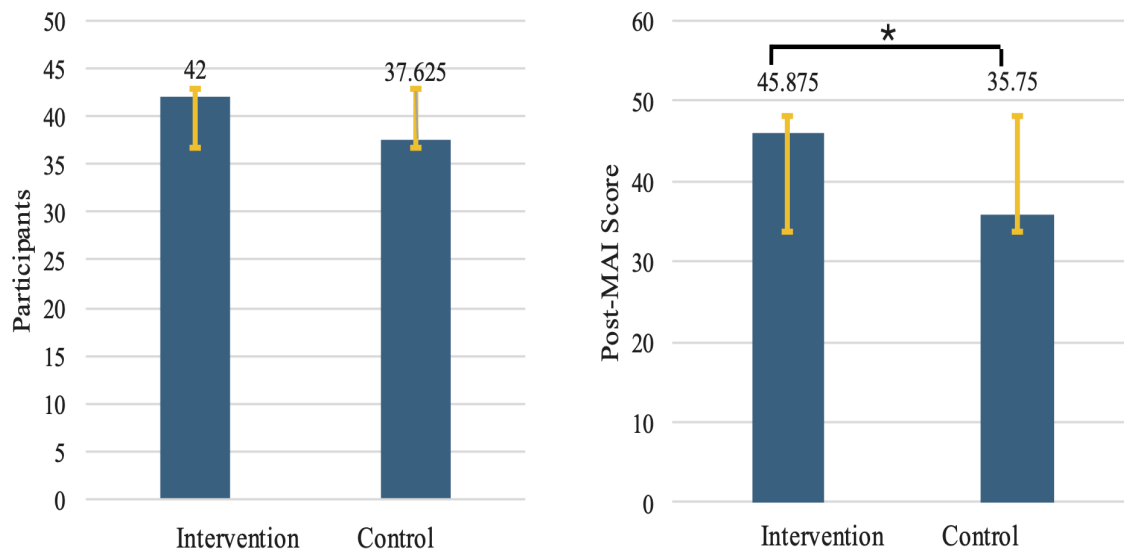


Figure 3 shows the pre- and post-MAI scores for both the intervention (n=8) and control (n=8) groups.

Figure 4 Participant Scores for Pre- and Post-MAI

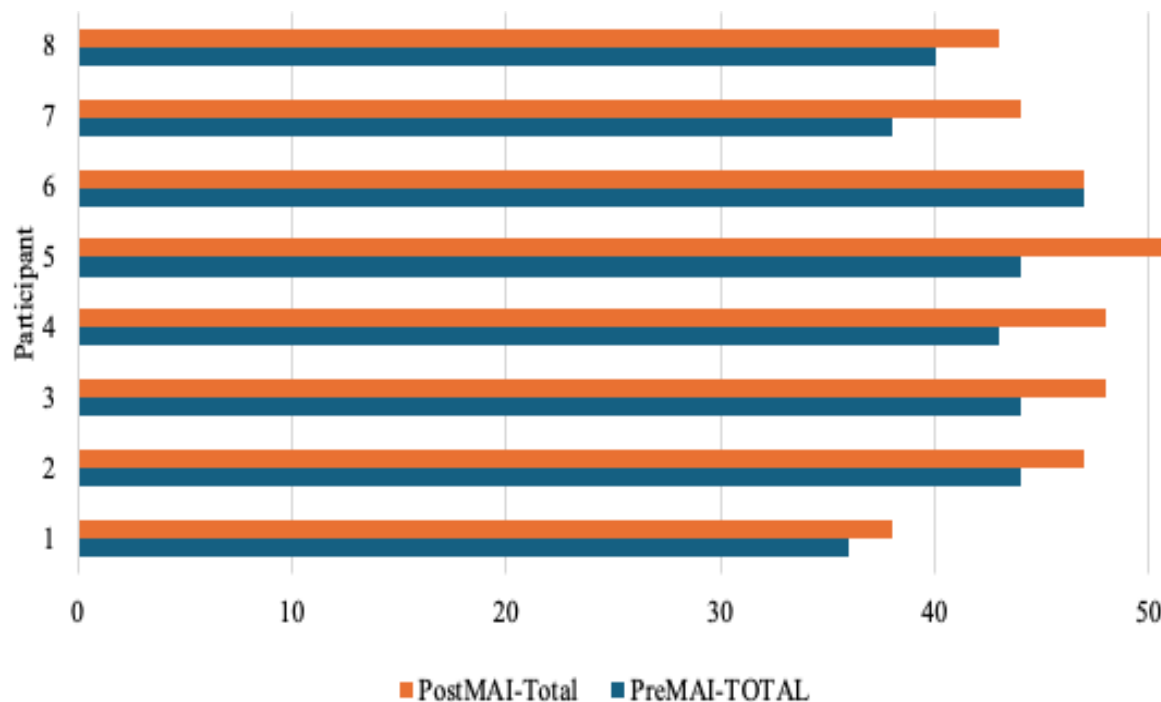


Figure 4 shows the individual scores on the pre- and post-MAI for the intervention group participants (n=8)

Focus group and survey analysis

Thematic analysis of both survey and focus group data revealed various themes regarding the perception of the model and its impact on the students' use of metacognition, readiness for SDL, and the subsequent impact on their learning.

Participants initially expressed skepticism regarding the use of a custom GPT for learning. Some students mentioned concerns about its accuracy and usefulness in the context of learning neuroanatomy. Over time, however, many of the perceptions changed to be more positive. Students expressed satisfaction with the model's ability to help them develop new strategies for learning content, both in and outside the context of neuroanatomy. Key improvements mentioned included faster and more accurate recall time, better organization of course requirements, and improved skills in writing and communicating with the technology.

Participants expressed improvements in their ability to evaluate their learning processes for unproductive habits, and many of the participants felt confident in identifying and applying new strategies, mentioning beneficial impacts on their confidence and motivation.

Interacting with the model also benefited the students' ability to set goals and effectively manage their time. Participants also indicated that the model's recommendations for taking breaks and self-care were important considerations that ultimately supported their learning and success. Almost all participants forecasted increased utilization of metacognitive strategies in and outside of their academic pursuits. Overall, the experience proved valuable for the participants, supporting the development and maintenance of metacognitive skills. Participants left the experience feeling more confident and trusting themselves with respect to managing their learning in the future, as well as motivated to apply the lessons learned. See Table 1.

DISCUSSION

This research evaluated the use of a custom GPT in supporting graduate neuroanatomy students in utilizing metacognitive strategies and promoting readiness for SDL. The findings described here demonstrate a positive impact in both areas, highlighting a novel application of educational technology in supporting the requisite skill development for effective SDL.

Results from the Metacognition Awareness Inventory assessments indicate that interacting with the intervention significantly improved students' utilization of metacognition in the course. These data align with prior work, which suggests that the practice of metacognition is essential for students to be able to engage in SDL (14). Participant responses supported this finding, as many of the students reported improved ability to study effectively and organize their time for successful studying. These improvements can be attributed to an improved sense of

metacognition, which improved students' insight into how they were learning the material and their ability to adapt throughout the course.

Regarding readiness for SDL, the intervention successfully supported students in improving their respective readiness, as indicated by the post-test instrument assessment scores. Interacting with the model fostered an improved sense of autonomy, with participants reporting improved confidence across various dimensions of SDL (e.g., setting goals, managing time, and incorporating self-care). These improvements supported the students in maintaining motivation throughout the course. Findings here support continued research on the effective implementation of GPT-supported educational technology.

During focus group discussions, students reported that their perception of the intervention and its efficacy changed over time, transforming from skepticism to generally positive. Participants also noted the value of reflective practice, as well as the tools used in facilitating that process. Many also mentioned the translatable benefits outside of the classroom in their quest for lifelong learning. These findings are consistent with prior work that demonstrated the important role of motivation and reflection within the processes of SDL (7, 8). As comfort level in using the tool increased, students reported that they felt improved self-trust and confidence in their ability to effectively manage their own learning.

The research described here aligns with existing literature that highlights how GPTs can enhance student experience and learning, specifically by way of metacognition and self-directed learning. We demonstrate that GPT prompts can be customized to create structured and adaptive activities that positively impact students' metacognitive awareness and ability to engage in SDL (21). Prior work has demonstrated the impact of metacognitive awareness on students' ability to meaningfully integrate AI-generated feedback into their learning approach (22), which supports

the creation and implementation of targeted approaches for supporting students in developing metacognitive awareness.

Our approach expands upon prior work that used GPT responses as material for critical thinking exercises by shifting the role of GPTs to one of dynamic support rather than a static resource (23, 24). Our work supports the notion of how GPTs can be used to facilitate active learning activities for students, rather than solely providing answers or corrective feedback on writing (25). Furthermore, our work complicates the idea of metacognitive laziness by demonstrating how GPT interactions can increase metacognitive awareness, rather than reducing it (26). This study provides evidence that GPT agents serving as facilitators for reflective practice can positively impact student behavior when intentionally integrated into learning processes. We provide a scalable, content-agnostic approach for fostering readiness for SDL and the use of metacognitive strategies within higher education.

The study revealed certain challenges associated with implementing technology in learning contexts. For example, negative perceptions of the technology can lead to initial resistance to its utilization. At the time of study, customized GPTs were only available with a paid subscription, which limited the number of students we were able to recruit. Now, as more GPT services begin to offer competitive and freely available resources, increasing the scope and scale of this study's implementation can be achieved.

While the intervention yielded large effect sizes (Cohen's $d = 1.26$ for SDLI and 1.4 for MAI) and moderate power (~ 0.70 – 0.76), the results should be interpreted with certain caveats. First, the relatively high power is highly dependent on effect sizes. Had the effect sizes observed been smaller, the resulting power would have also been smaller, increasing the likelihood of a Type II statistical error. Additionally, the small sample size restricts broad generalizations to

various groups. The students described here were pre-professional, graduate students who may have had higher baseline levels of SDL readiness and metacognition. Future iterations of this approach should focus on younger and less academically advanced students. Researchers should apply this approach with larger and more diverse groups, specifically those who are early in their academic journey. By focusing on these populations, the scientific community can enhance the validity of these findings by allowing for more robust quantitative analyses and deeper qualitative insights with respect to how learners at different stages interact and respond to the model. Finally, future research should also incorporate non-AI support to control group students; this will allow for better determination of the model's impact on student readiness for SDL and the use of metacognition.

Finally, the findings from this research contribute to a growing body of knowledge concerning the integration of AI technology within educational contexts, in addition to how those technologies can support self-directed learners. Results here posit that custom GPT models can expand student metacognition, thereby supporting their readiness for SDL. This, in turn, directly contributes to their willingness and ability to engage in lifelong learning. Future research should explore longitudinal impacts of these types of experiences across diverse populations of learners, within various educational contexts. Additionally, researchers may identify model modifications that will ultimately support a broader set of student populations.

SUPPLEMENTARY

Item 1

GPT Prompt:

Your role is to serve as a self-reflective "journal" for students studying neuroanatomy, facilitating metacognitive reflection on how the material relates to their interests, curiosities, and questions.

Instead of focusing on the content itself, you'll guide them by asking a series of structured questions based on their answers, helping them to thoroughly process metacognitive skills by following a specific flow chart for conversation:=

Please note, after I ask the initial question in each set, I will prompt you twice to add more depth or give more examples. This structured approach ensures we cover all four topics thoroughly within our 12-question limit.

1. [First main question] - If the user's response lacks detail: "Can you elaborate on that? Perhaps provide a specific example or explain how it made you feel." - If the response is still not detailed after the first follow-up: "Let's dive a bit deeper. How does this connect to your personal values or goals?" [Once the user has provided sufficient depth or after two follow-ups, regardless of depth, proceed to the next question.]

2. [Second main question] - Follow the same pattern of follow-ups to encourage more detailed responses.

3. [Third main question] - Follow the same pattern of follow-ups to encourage more detailed responses.

4. [Fourth main question] - Follow the same pattern of follow-ups to encourage more detailed responses.

[End of conversation]

Here are the main questions:

Main Question 1. Consider the upcoming lectures this week, how do they relate to your past experiences and/or interests.

Main Question 2. Thinking back to the last few lectures, what are you still curious about? What is something you are still unsure about?

Main Question 3. Looking ahead to your next lectures, what is one resource that will support your understanding?

Main Question 4. Thinking of how you spend your study time, how can you make your study sessions more efficient given the current content of the course?

REFERENCES

1. Knowles MS. 1975. Self-directed learning: a guide for learners and teachers. Association Press, Chicago.
2. Gureckis TM, Markant DB. 2012. Self-directed learning: a cognitive and computational perspective. *Perspect Psychol Sci* 7:464–481. <https://doi.org/10.1177/1745691612454304>
3. Zhou Q, Luo X, Dong Y, Khalid F, Ma H. 2022. The effectiveness of self- directed learning in developing computational thinking of middle school students. *Proceedings of the 14th International Conference on Education Technology and Computers.* , p 285–290. <https://doi.org/10.1145/3572549.3572613>
4. Robinson JD, Persky AM. 2020. Developing self-directed learners. *Am J Pharm Educ* 84:847512. <https://doi.org/10.5688/ajpe847512>
5. Grow GO. 1991. Teaching learners to be self-directed. *Adult Educ Q (Am Assoc Adult Contin Educ)* 41:125–149. <https://doi.org/10.1177/0001848191041003001>
6. Boyer SL, Edmondson DR, Artis AB, Fleming D. 2014. Self-directed learning. *J Mark Educ* 36:20–32. <https://doi.org/10.1177/0273475313494010>
7. Rashid T, Asghar HM. 2016. Technology use, self-directed learning, student engagement and academic performance: examining the interrelations. *Comput Human Behav* 63:604–612. <https://doi.org/10.1016/j.chb.2016.05.084>
8. Wu D, Zhang S, Ma Z, Yue X-G, Dong RK. 2024. Unlocking potential: key factors shaping undergraduate self-directed learning in AI-enhanced educational environments. *Systems* 12:332. <https://doi.org/10.3390/syst12090332>
9. Tekkol İA, Demirel M. 2018. An investigation of self-directed learning skills of undergraduate students. *Front Psychol* 9:2324. <https://doi.org/10.3389/fpsyg.2018.02324>
10. Siqueira MAM, Gonçalves JP, Mendonça VS, Kobayasi R, Arantes-Costa FM, Tempiski PZ, Martins M de A. 2020. Relationship between metacognitive awareness and motivation to learn in medical students. *BMC Med Educ* 20:393. <https://doi.org/10.1186/s12909-020-02318-8>
11. Hong W, Bernacki ML, Perera HN. 2020. A latent profile analysis of undergraduates' achievement motivations and metacognitive behaviors, and their relations to achievement in science. *J Educ Psychol* 112:1409–1430. <https://doi.org/10.1037/edu0000445>
12. Cotterall S, Murray G. 2009. Enhancing metacognitive knowledge: structure, affordances and self. *System* 37:34–45. <https://doi.org/10.1016/j.system.2008.08.003>
13. Karatas K, Arpaci I. 2021. The role of self-directed learning, metacognition, and 21st century skills predicting the readiness for online learning. *Contemp Educ Technol* 13:ep300. <https://doi.org/10.30935/cedtech/10786>
14. Efklides A. 2011. Interactions of metacognition with motivation and affect in self-regulated learning: the MASRL model. *Educ Psychol* 46:6–25. <https://doi.org/10.1080/00461520.2011.538645>

15. Cheng S-F, Kuo C-L, Lin K-C, Lee-Hsieh J. 2010. Development and preliminary testing of a self-rating instrument to measure self-directed learning ability of nursing students. *Int J Nurs Stud* 47:1152–1158. <https://doi.org/10.1016/j.ijnurstu.2010.02.002>
16. Cheng S-F, Lee-Hsieh J, Turton MA, Lin K-C. 2014. Validation of self- directed learning instrument and establishment of normative data for nursing students in taiwan: using polytomous item response theory. *J Nurs Res* 22:90–100. <https://doi.org/10.1097/JNR.0000000000000027>
17. Schraw G, Dennison RS. 1994. Assessing metacognitive awareness. *Contemp Educ Psychol* 19:460–475. <https://doi.org/10.1006/ceps.1994.1 033>
18. Omprakash A, Kumar AP, Kuppusamy M, Sathiyasekaran BWC, Ravinder T, Ramaswamy P. 2021. Validation of metacognitive awareness inventory from a private medical university in India. *J Educ Health Promot* 10:324. https://doi.org/10.4103/jehp.jehp_39_21
19. Akin A, Abaci R, Çetin B. 2007. The validity and reliability of the Turkish version of the metacognitive awareness inventory. *Educ Sci Theory Pract* 7:671–678. <https://psycnet.apa.org/record/2008-10552-001>.
20. Braun V, Clarke V. 2006. Using thematic analysis in psychology. *Qual Res Psychol* 3:77–101. <https://doi.org/10.1191/1478088706qp063oa>
21. Yin J, Zhu Y, Goh T-T, Wu W, Hu Y. 2024. Using educational chatbots with metacognitive feedback to improve science learning. *Appl Sci (Basel)* 14:9345. <https://doi.org/10.3390/app14209345>
22. Teng MF. 2025. Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback . *Eur J Educ* 60. <https://doi.org/10.1111/ejed.12811>
23. Oates A, Johnson D. 2025. ChatGPT in the classroom: evaluating its role in fostering critical evaluation skills. *Int J Artif Intell Educ*. <https://doi.org/ 10.1007/s40593-024-00452-8>
24. Exintaris B, Karunaratne N, Yuriev E. 2023. Metacognition and critical thinking: using ChatGPT-generated responses as prompts for critique in a problem-solving workshop (SMARTCHEMPer). *J Chem Educ* 100:2972– 2980. <https://doi.org/10.1021/acs.jchemed.3c00481>
25. Chen W, Li Z, Lin C-H, Wang Q. 2024. Learning-by-teaching with ChatGPT: the effect of a teachable ChatGPT agent on programming education. *arXiv*. <https://doi.org/10.48550/arXiv.2412.15226>
26. Fan X, He Y, Li X, Lin L. 2024. Beware of metacognitive laziness: effects of generative AI on learning motivation, processes, and performance. *arXiv*. <https://doi.org/10.48550/arXiv.2412.09315>.

CHAPTER 4: SUMMARY AND CONCLUSION

PURPOSE

The research presented here supports educators as they integrate innovative educational technologies into their courses by examining approaches that offer deeper, more nuanced insights into the experiences of their students. Utilizing cognitive load, metacognition, and self-directed learning theory as models for assessing and measuring aspects of student experience, we offer valuable insights regarding the challenges and potential benefits of two emerging technologies: VR and AI. The work presented here supports educational innovation by offering proven approaches to utilize the unique capabilities of VR and generative AI. The previous chapters affirm that educational technology should be evaluated by how it supports educators in measuring, supporting, and understanding student learning experiences, rather than on novelty or learning outcomes alone.

SUMMARY OF FINDINGS

The Cognitive Load in VR research explores how visualized data impacts a student's cognitive load. The findings support the hypothesis that 3D content may lower mental strain (i.e., cognitive load). While reduced cognitive load did not extend to better learning outcomes, findings suggest the technology may support students as they engage with spatially complex material. The study underscores the value of VR-based education to explore student experience in new and insightful ways.

The AI, Metacognition, and Self-Directed Learning (SDL) research examines how custom GPT models can be used to support student development of both metacognitive practices and readiness for SDL. The model was designed to facilitate reflective conversations, twice a week, for 10 minutes per session. Students who engaged with the model improved in both areas

of interest more than those who engaged with similar but non-technological resources. The research supports the integration of AI as a potential developmental tool that supports students as they reflect on and monitor their learning processes.

Combined, these research projects demonstrate that innovative educational technology can serve unique and complementary roles within education. VR provides a means for measuring student experiences while they attempt to learn challenging subjects. Moreover, VR allows for students to view data in a manner that may reduce cognitive effort, thereby improving instructional design and, potentially, learning outcomes.. AI provides a means for developing students' reflective capability; the findings demonstrate that educational technology, when intentionally considered and designed, can support educators in understanding the learning processes of a task, while supporting students in becoming more dynamic and adaptable learners.

FUTURE RESEARCH

These research projects not only provide evidence-based practices for effective integration, but they also raise important questions concerning how educational technology may shape the student learning experience over time. Additional research surrounding how VR-based instructional design impacts cognitive load, as well as how AI-supported reflective practices, impact the student experience across different STEM domains. More research exploring the long-term benefits, accessibility, and equity of these technologies would allow for broader and more successful implementations. Future work should use longitudinal, mixed method approaches that will allow instructors and university stakeholders to make smarter decisions regarding how, when, and where to integrate.

CONCLUSION

This thesis examined the use of VR and AI through the lenses of cognitive load and self-directed learning, respectively, emphasizing how the technologies can be used to understand student cognitive experiences; support student reflection; and promote student's personal development. The findings demonstrate that innovation is not a marker for efficacy. As such, emerging technology must be evaluated by how it affects both the student experience and learning outcomes. The central contribution of this work is that educational technology is most effective and meaningful when it is used to measure, support, and promote student learning, rather than simply facilitate a process in a novel way.