COMPARING NOVICE NURSES' COGNITIVE LOAD IN ROUTINE AND NON-ROUTINE SIMULATIONS WITH MIXED METHODS

by

Jeremy W. Hutson, MSN, RN

A Dissertation

Presented to

Texas Christian University

Harris College of Nursing & Health Sciences

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

April 3, 2025

APPROVAL

COMPARING NOVICE NURSES' COGNITIVE LOAD IN ROUTINE AND NON-ROUTINE SIMULATIONS USING MIXED METHODS

by

Jeremy William Hutson

Dissertation approved:

Ashley Franklin
box SIGN 15773Y71-195YZ2Q8
Ashley E. Franklin, PhD, Major Professor

Beth A. Rogers
box SIGN 19QJYJL1-195YZ2Q8
Beth A. Rogers, PhD, Committee Member

Yan Zhang
box SIGN 4R73XQK1-195YZ2Q8
Yan Zhang, PhD, Committee Member

Adam C. King
box SIGN 1VLL26Z1-195YZ2Q8
Adam C. King, PhD.,
Health Sciences PhD Program Director

PREFACE

I am forever grateful to my mom, Char; my sister, Vickki; and my brother, Jason, for their love and support as I embarked on my journey through graduate education. Their visits, encouragement, and willingness to just listen fed my soul along the way.

Thank you to my teaching mentor, colleague, and dear friend, Carol Thorn. Your wisdom and passion for nursing education ignited a fire within me that has carried my farther than I ever imagined. I have always been overwhelmed by your patience and kindness in helping me discover my path as an educator.

I would also like to thank the HPLC staff at TCU. You welcomed me with open arms and were so integral in helping me achieve my goals. Your hard work and dedication to your students is an inspiration.

Saying thank you to Dr. Ashley Franklin cannot fully capture my thoughts and feelings of gratitude. Your kindness and patience at every turn has been constant. Having had the privilege to be a part of your dissertation research team, it is so special for me to complete my graduate education journey with you. Thank you for your encouragement, wisdom, and mentorship. I would also like to thank others on my dissertation committee who have provided support and guidance along the way. To Dr. Yan Zhang, thank you for helping me embrace statistics. Your kindness and wisdom have overwhelmed me from the start and guided my learning to a deeper level. To Dr. Beth Rogers, thank you for being open to my ideas on cognitive load and for our many deep conversations on nursing simulation.

ABSTRACT

COMPARING NOVICE NURSES' COGNITIVE LOAD IN ROUTINE AND NON-ROUTINE SIMULATIONS WITH MIXED METHODS

Jeremy W. Hutson

Ashley E. Franklin, PhD, RN, CNE, CHSE-A, ANEF
Dissertation Chair

Background: A readiness-for-practice gap exists within new graduate nurses entering the workforce. Educators use simulation to boost practice-readiness. Researchers are exploring how cognitive load impacts novice nurses' learning and performance during simulation. A variety of measures have been used to capture cognitive load during skill-based and routine, holistic simulations; however, the impact of increasing client complexity within holistic simulations has yet to be explored. As such, the purpose of this body of research was to examine how cognitive load manifests in novice nurses during holistic simulations.

Methods: Three investigations were performed. First, a conceptual paper summarized methodological choices pertaining to simulation cognitive load research. Second, psychometric analysis of the National Aeronautics & Space Administration – Task Load Index, a subjective instrument used across disciplines to quantify perceived cognitive load – was performed. Third, a mixed methods study compared physiologic and behavioral responses to increased cognitive load during routine and non-routine, holistic simulations. Participants were junior, undergraduate novice nurses enrolled in a dedicated simulation course at a nursing school in the Southern region of the United States.

Results: This body of work contributes new knowledge to help researchers appreciate that triangulation of objective and subjective measures improves rigor and understanding. The NASA-TLX survey is a valid and reliable instruction for measuring novice nurses' cognitive load in simulation. There was a difference in cognitive load between routine and non-routine simulation types ($F(1, 10) = 23.99$, $p < 0.001$, η^2=.706). In non-routine simulation, participants' pupil sizes were larger (B = 0.238, SE = 0.096, t = 2.47, p = 0.015). Novice nurses' emotional experiences of increased cognitive load occurred on a spectrum.

Conclusion: This body of research provides evidence that increases in client complexity impacts novice nurses' cognitive load and that cognitive load decreases with exposure. Moreover, this research has uncovered additional questions regarding how individual characteristics may contribute to mitigating cognitive load, providing direction for future research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

Chapter I – INTRODUCTION

Introduction and Background

Following completion of the nursing curricula, 77 percent of new graduate nurses entering the workforce lack basic competence to provide safe patient care (Kavanagh & Szweda, 2017). Nurse educators are trying to improve new graduates' practice readiness; simulation is an effective and efficient healthcare pedagogy, and it has become an integral part of most nursing curricula. Because the theory-to-practice gap is growing, simulationists are motivated to explore conditions which underpin errors in learning and performance (Eppich & Reedy, 2022). Human factors psychology helps nurse educators and simulationists understand that high cognitive load (CL) impairs learning and performance (Haji et al., 2015; Paas & van Merriënboer, 2020; Sweller, 2010, 2020; Young et al., 2016). As such, a body of literature is growing from nursing academic and practice settings where researchers describe both novice and practicing nurses' cognitive load in simulation. In this body of work, *novice nurse* refers most often to a prelicensure nursing student.

CL describes the mental effort required to process information in working memory (Sweller, 2010). Researchers explore the capacity of working memory relative to task complexity, prior knowledge, and extraneous distractors (Chen et al., 2023; Sweller, 2010; Sweller et al., 2011). Working memory's ability to process information is limited, and learning and performance are impaired when working memory is overwhelmed (Sweller et al., 2011).

The overarching purpose of this body of work is to examine how CL manifests in novice nurses during holistic simulations. The specific aims of this program of research were to:

1. Identify how researchers measure CL in healthcare simulation research.

2. Determine the validity and reliability of NASA-TLX as a subjective measure of CL with novice nurses in simulation.
3. Compare novice nurses' CL during routine and non-routine, holistic simulations.

To achieve this aim, we will consider the research question: "What physiologic responses do novice nurses experience during routine and non-routine simulations?" We hypothesize pupil diameter will trend positively and heart rate variability will trend negatively with novice nurses' CL because the sympathetic nervous system is sensitive to stressors such as increased CL (Knight et al., 2020). When activated, the fight-or-flight response is reflected in major organ system changes, providing researchers direct and real-time measures of CL.

4. Describe novice nurses' lived experience of high cognitive load and the factors in simulation which contribute to its increase. The research questions guiding this qualitative portion of the study are:
    1. How do novice nurses describe their lived experience of high cognitive load during simulation?
    2. How do simulation components contribute to novice nurses' experience of high cognitive load?
    3. What behaviors do novice nurses perceive they demonstrate when experiencing high cognitive load?

Theoretical Framework

Cognitive Load Theory (CLT) underpins this investigation. Sweller (2011) operationalizes three CL types and describes cognitive architecture during information processing. CLT has been applied across disciplines to inform instructional design and is of particular interest to nursing simulation education as the amount and complexity of information novices must process can overwhelm working memory. Simulation provides a safe environment to learn and practice safety-critical skills (Al-Elq, 2010), and the ability to control and standardize the simulation environment is well-suited for empirical inquiry; therefore, it is logical to investigate novice nurses' CL in simulation.

Working memory and long-term memory are the cognitive structures responsible for information processing and retrieval (de Jong, 2010; Sweller et al., 2011). Working memory describes the short-term cognitive resources which are limited and can only process three to seven information units at a time (Adam, 2017; Cowan, 2010; Miller, 1956). In cognitive architecture, working memory processes new information first and then ideally information transfers and is stored in long-term memory (Guo & Wang, 2022; Webb & Dennis, 2020). Long-term memory has unlimited capacity which allows for additions and revisions to cognitive architecture (Forsberg et al., 2021). Individuals can retrieve information from long-term memory, reducing CL to optimizing efficiency (Cowan, 2008).

Within working memory, three types of CL impact learning and performance outcomes. First, intrinsic load — the cognitive effort inherent in a given task — is determined by task complexity and prior knowledge. Since task complexity is largely influenced by prior knowledge stored in schema and individual characteristics, novice nurses are especially susceptible to higher intrinsic load (Gonzalez & Kardong-Edgren, 2017). Their lack of cognitive architecture adds to task complexity, which increases cognitive demands (Sweller et al., 2011).

Second, extraneous load is the effort involved in processing irrelevant or distracting information. Most often, increases in extraneous load are due to the manner in which information is presented or other environmental distractors (Sweller et al., 2011). Much like intrinsic load, extraneous load increases with element interactivity (Paas et al., 2004; Sweller et al., 2011; van Merriënboer & Sweller, 2005). For example, nurses must discern the location of a clinical alarm, its significance, and decide whether or not to respond. If the alarm and patient are familiar to the nurse, information processing involves less element interactivity. However, the novice nurse who is unfamiliar with this unit's equipment and patient population will experience higher element interactivity and, therefore, higher extraneous load from the same situation because they lack cognitive architecture. Since extraneous load is easily manipulated, researchers introduce distractors such as time pressure and interruptions as conditions under which to examine CL and ultimately optimize learning.

Finally, germane cognitive load (GCL) represents mental resources dedicated to processing and storing information (Cierniak et al., 2009; Sweller, 2010; Sweller et al., 2011). Its designation as a "load" suggests a negative impact on total CL; however, Greenberg and Zheng (2023) noted GCL did not have an additive effect on ICL. Instead, they found effective management of high CL was relative to the amount of germane resource available at the time of information processing. Thus, learning is optimized when germane resource availability matches or exceeds cognitive demands relative to the learning task. Given working memory's limited capacity, educators use their understanding of load types to inform efforts towards maintaining to optimize novice nurses' CL levels and improve performance.

Cognitive Load Measurement Approaches

Understanding cognitive architecture functionality and the interplay between each of the three load types informs research design and measurement approach. Subjective and objective measures quantify CL and enable researchers to examine relationships between load source, its impact on CL, and the resulting behavioral and performance effects (Bartels & Marshall, 2012; Leppink et al., 2013; Sweller, 2018; Zagermann et al., 2016). Subjective measures have been popular owing to their ease of administration and cost-effectiveness (Ayres, 2017). Hart and Staveland's (1988) National Aeronautics and Space Administration's Task Load Index (NASA-TLX) has been most widely used in nursing CL research (Campoe & Giuliano, 2017; Kataoka et al., 2011; Matsushima & Kadohama, 2021); however, Paas' (1992) Cognitive Load Rating Scale (Schlairet et al., 2015), and Josephsen's Self-Report Cognitive Load Measurement Tool 2.0 (Roh et al., 2022) have also been used. Two meaningful limitations of subjective measures are the potential for recall and selection biases (Naismith et al., 2015; Skulmowski, 2023).

Objective, reliable, and real-time CL measures improve scientific rigor in CL research (Ayres et al., 2021; Eppich & Reedy, 2022). For example, physiologic measures track changes in the eyes, skin, brain, heart, and lungs when the sympathetic nervous system responds to stressors such as CL increases (Korbach et al., 2017; Park et al., 2022). Eye-tracking technology documents changes in pupil diameter and patterns of gaze (Das Chakladar & Roy, 2023; Phitayakorn et al., 2015; Zakeri et al., 2020). Researchers have used eye-tracking measures in both simulation (Cabrera-Mino et al., 2019; Kataoka et al., 2011) and professional practice (Ahmadi et al., 2024) environments with novice and experienced nurses. Electrodermal sensors measure electrical conductance of the skin related to eccrine sweat glands as they fill and drain (Phitayakorn et al., 2015); electrodermal monitoring has only been used in professional practice with experienced nurses (Ahmadi et al., 2024). Kataoka et al., (2011)
included heart rate (HR) and heart rate variability (HRV) in their study comparing CL differences between novice nurses, inexperienced nurses, and experienced nurses.

All measurement approaches have limitations. For example, neither subjective nor objective measures can differentiate between CL types with acceptable reliability (Korbach et al., 2018). Some subjective and objective measures have acceptable reliability and validity, but the scientific community has not reached consensus as to which measure(s) are the gold standard.

The Impact of Complexity Across Routine and Non-Routine Simulations on Cognitive Load

Simulationists scaffold simulation complexity to optimize CL and improve learning. For example, simulationists theoretically decrease intrinsic load by designing simulation objectives with novice nurses' previous experience in mind. Accordingly, simulationists facilitate more routine experiences for early learners and non-routine experiences for advanced learners. Scaffolding routine and non-routine experiences provides researchers an opportunity to compare novice nurses' CL between simulations and over time.

The terms routine and non-routine are borrowed from organizational psychology to describe the type of work or situation to be encountered (Flin et al., 2008). For our purposes, routine simulations include tasks which are performed in the normal course of non-emergent care. In the context of nursing simulation, novice nurses carry out their a priori plan of care in a routine experience without having to alter the plan in a meaningful way. For example, novices might prioritize a focused assessment and pain management during pre-briefing for a simulated post-operative hip-replacement patient and subsequently carry out their plan during a routine simulation scenario. Researchers have described novice nurses' CL during routine simulations related to medication administration (Kataoka et al., 2011; Matsushima & Kadohama, 2021) and physical assessments (Schlairet et al., 2015).

Non-routine simulations most often require nurses to pivot their thinking in a response to an unanticipated cue from the simulation patient or environment. In our post-operative hip replacement scenario example, having the patient experience acute chest pain as a post-operative complication, which may or may not have been discussed during pre-briefing, characterizes the simulation as non-routine. Cabrera-Mino et al., (2019) used a non-routine simulation and compared novice nurses' and experienced nurses' pupil diameter as they provided care for a simulated patient experiencing respiratory distress.

Review of Literature Uncovers Methodological Gaps In Nursing Cognitive Load Research

There are seven cross-sectional, quantitative CL studies, involving novice nurses (Matsushima & Kadohama, 2021; Roh et al., 2022; Schlairet et al., 2015), experienced nurses (Ahmadi et al., 2024; Campoe & Giuliano, 2017), or comparing both samples (Cabrera-Mino et al., 2019; KATAOKA et al., 2011). Researchers often examine nurses' CL in the context of skill performance. For example, two research teams have published investigations of nurses' CL during routine simulations involving medication administration with an intravenous pump (KATAOKA et al., 2011a; Matsushima & Kadohama, 2021); another research team introduced interruptions during medication administration and documented trends in nurses' CL in the non-routine simulation (Campoe & Giuliano, 2017). Interestingly, novice nurses are typically juniors or senior students during data collection, and most simulations are routine performance which require lower level thinking and are inherently less complex. There is a gap in the CL literature related to novice nurses' experiences with non-routine, holistic simulations and group simulations.

There is a smaller body of literature measuring nurses' CL during holistic scenarios involving multiple skills such as assessments, procedures, medication administration, patient education, and interprofessional communication (Ahmadi et al., 2024a; Cabrera-Mino et al., 2019; Roh et al., 2022). Two research teams have published CL outcomes after routine, holistic simulations with individual nurses providing care for a simulated patient (Cabrera-Mino et al., 2019; Schlairet et al., 2015). Roh and colleagues (2022) published a study reporting novice nurses' CL after completing an academic course with routine, holistic simulations involving groups of four novice nurses (Roh et al., 2022). Holistic simulations require higher-level thinking and are more complex owing to task complexity. Further, group dynamics contribute to high CL (Seddigh et al., 2015; Skuballa et al., 2019), because people, background noise, and conversations are additional distractors. When groups of novice nurses work collectively to provide patient care (Au et al., 2023), novices may be able to apply knowledge and solve clinical problems more effectively (Burgess et al., 2020; Whitley et al., 2015). More research with novice and experienced nurses is needed in holistic simulation to describe CL and to lay a foundation for intervention research to optimize learning and performance.

Two qualitative studies add to our understanding of nurses' CL related to skill performance (Aldridge & Hummel, 2019) and interprofessional collaborative reasoning (Blondon et al., 2017). While simulation design varied, qualitative descriptions add to the body of literature pertaining to sources of CL and consequences of high CL on performance. Conceptually, simulationists can infer high CL negatively influences =novice nurses' readiness-for-practice.

Next Steps

This body of work will fill important knowledge gaps in extant literature pertaining to novice nurses' CL experiences. The overarching aim is to compare novice nurses' CL during routine and non-routine simulations using mixed methods. There are six research questions that guide this dissertation (see **Table 1.1**).

Table 1.1. Research Questions and Corresponding Titles of Papers in Manuscript Dissertation

Research Questions	Title
1) How do researchers measure CL in healthcare simulation research?	**Chapter II:** Methodological Considerations for Healthcare Simulation Cognitive Load Research
2) Is NASA-TLX a valid and reliable cognitive load measure for use with novice nurses in simulation?	**Chapter III:** Psychometric Testing of NASA-TLX to Measure Learners' Cognitive Load in Individual and Group Nursing Simulations
3) What physiologic responses do novice nurses experience in routine and non-routine simulations? 4) How do novice nurses describe their experience of high cognitive load during simulation? 5) How do simulation components contribute to novice nurses' experience of high cognitive load? 6) What behaviors do novice nurses perceive they demonstrate when experiencing high cognitive load?	**Chapter IV:** Comparing Novice Nurses' Cognitive Load in Routine and Non-Routine Simulations Using Mixed Methods

Researchers use subjective and objective measures to quantify CL. Chapter II explores the strengths and limitations of common measurement approaches. Subjective methods such as self-report surveys and think-aloud strategies are discussed, as are physiologic measures including eye-tracking, heart rate variability, and electrodermal activity. The methodological considerations explored in Chapter II led to selection of pupil size and heart rate variability as the physiologic measures and NASA-TLX as a subjective measure for the dissertation study described in chapter IV.

NASA-TLX is a subjective measure widely used in nursing research. Its popularity is likely due to ease of administration and cost-effectiveness; however, extant literature demonstrates a paucity of reliability and validity evidence of NASA-TLX in simulation research with novice nurses. In short, we cannot assume an instrument originally tested in samples of professional pilots will perform similarly in the context of nursing simulation education. As such, Chapter III explores psychometric evidence supporting NASA-TLX with novice nurses in group and individual simulations.

Summary

In brief, this collective body of work will move the science of nursing simulation forward and lay the foundation for intervention research to optimize learning and performance. This body of work contributes new knowledge to help researchers appreciate that triangulation of objective and subjective measures improves rigor and understanding of how cognitive load increases relative to instructional design. More specifically, the NASA-TLX survey is a valid and reliable instrument for measuring novice nurses' CL during simulation. A strength of the mixed methods approach is consideration of multiple data sources to compare novice nurses' CL between simulations and over time. Including first-person reports directly from novice nurses (Wasti et al., 2022) adds richness to the comparison of CL experiences during routine and non-routine simulations. Since the phenomenon of high CL is not well-defined in nursing, the mixed methods approach will provide a breadth and depth of understanding. Once study results have been finalized, this body of work will conclude with overall themes. Recommendations for future research and application to nursing education will be provided at that time.

References

Ahmadi, N., Sasangohar, F., Yang, J., Yu, D., Danesh, V., Klahn, S., & Masud, F. (2024). Quantifying workload and stress in intensive care unit nurses: Preliminary evaluation using continuous eye-tracking. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *66*(3), 714–728. https://doi.org/10.1177/00187208221085335

Aldridge, M. D., & Hummel, F. (2019). Nursing students' perceptions of skills learning. *Nurse Educator*, *44*(3), 170–174. https://doi.org/10.1097/NNE.0000000000000569

Al-Elq, A. (2010). Simulation-based medical teaching and learning. *Journal of Family and Community Medicine*, *17*(1), 35. https://doi.org/10.4103/1319-1683.68787

Au, M. L., Tong, L. K., Li, Y. Y., Ng, W. I., & Wang, S. C. (2023). Impact of scenario validity and group size on learning outcomes in high-fidelity simulation: A systematic review and meta-analysis. *Nurse Education Today*, *121*, 105705. https://doi.org/10.1016/j.nedt.2022.105705

Ayres, P. (2017). Subjective Measures of Cognitive Load. In *Cognitive Load Measurement and Application* (pp. 9–28). Routledge. https://doi.org/10.4324/9781315296258-2

Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. J. G. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.702538

Bartels, M., & Marshall, S. P. (2012). Measuring cognitive workload across different eye tracking hardware platforms. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 161–164. https://doi.org/10.1145/2168556.2168582

Blondon, K. S., Maître, F., Muller-Juge, V., Bochatay, N., Cullati, S., Hudelson, P., Vu, N. V., Savoldelli, G. L., & Nendaz, M. R. (2017). Interprofessional collaborative reasoning by residents and nurses in internal medicine: Evidence from a simulation study. *Medical*

Teacher, *39*(4), 360–367. https://doi.org/10.1080/0142159X.2017.1286309

Burgess, A., van Diggele, C., Roberts, C., & Mellis, C. (2020). Team-based learning: Design, facilitation and participation. *BMC Medical Education*, *20*(S2), 461. https://doi.org/10.1186/s12909-020-02287-y

Cabrera-Mino, C., Shinnick, M. A., & Moye, S. (2019). Task-evoked pupillary responses in nursing simulation as an indicator of stress and cognitive load. *Clinical Simulation in Nursing*, *31*, 21–27. https://doi.org/10.1016/j.ecns.2019.03.009

Campoe, K. R., & Giuliano, K. K. (2017). Impact of frequent interruption on nurses' patient-controlled analgesia programming performance. *Human Factors*, *59*(8), 1204–1213. https://doi.org/10.1177/0018720817732605

Chen, O., Paas, F., & Sweller, J. (2023). A Cognitive Load Theory Approach to Defining and Measuring Task Complexity Through Element Interactivity. *Educational Psychology Review*, *35*(2), 63. https://doi.org/10.1007/s10648-023-09782-w

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, *25*(2), 315–324. https://doi.org/10.1016/j.chb.2008.12.020

Cowan, N. (2008). *Chapter 20 What are the differences between long-term, short-term, and working memory?* (pp. 323–338). https://doi.org/10.1016/S0079-6123(07)00020-9

Das Chakladar, D., & Roy, P. P. (2023). Cognitive workload estimation using physiological measures: a review. *Cognitive Neurodynamics*. https://doi.org/10.1007/s11571-023-10051-3 de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, *38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0

Debue, N., & van de Leemput, C. (2014). What does germane load mean? An empirical contribution to the cognitive load theory. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01099

Eppich, W., & Reedy, G. (2022). Advancing healthcare simulation research: innovations in theory, methodology, and method. *Advances in Simulation*, *7*(1), 23. https://doi.org/10.1186/s41077-022-00219-y

Flin, R., O'Connor, P., & Crichton, M. (2008). *Safety At The Sharp End: A Guide to Non-Technical Skills*. CRC Press.

Forsberg, A., Adams, E. J., & Cowan, N. (2021). *The role of working memory in long-term learning: Implications for childhood development* (pp. 1–45). https://doi.org/10.1016/bs.plm.2021.02.001

Fraser, K. L., Ayres, P., & Sweller, J. (2015). Cognitive Load Theory for the Design of Medical Simulations. *Simulation in Healthcare : Journal of the Society for Medical Simulation*, *10*(5), 295–307. https://doi.org/10.1097/SIH.0000000000000097

Gonzalez, L., & Kardong-Edgren, S. (2017). Deliberate practice for mastery learning in nursing. *Clinical Simulation in Nursing*, *13*(1), 10–14. https://doi.org/10.1016/j.ecns.2016.10.005

Greenberg, K., & Zheng, R. (2023). Revisiting the debate on germane cognitive load versus germane resources. *Journal of Cognitive Psychology*, *35*(3), 295–314. https://doi.org/10.1080/20445911.2022.2159416

Guo, L., & Wang, C. (2022). Enabling automatic retrieval of schemas from long-term memory in English grammar practice. *Asia Pacific Education Review*, *23*(2), 361–373. https://doi.org/10.1007/s12564-022-09762-9

Haji, F. A., Rojas, D., Childs, R., de Ribaupierre, S., & Dubrowski, A. (2015). Measuring cognitive load: performance, mental effort and simulation task complexity. *Medical Education*, *49*(8), 815–827. https://doi.org/10.1111/medu.12773

Kataoka, J., Sasaki, M., & Kanda, K. (2011). Effects of mental workload on nurses' visual behaviors during infusion pump operation. *Japan Journal of Nursing Science*, *8*(1), 47–56. https://doi.org/10.1111/j.1742-7924.2010.00158.x

Kavanagh, J. M., & Szweda, C. (2017). A Crisis in Competency: The Strategic and Ethical Imperative to Assessing New Graduate Nurses' Clinical Reasoning. *Nursing Education Perspectives*, *38*(2), 57–62. https://doi.org/10.1097/01.NEP.0000000000000112

Knight, E. L., Giuliano, R. J., Shank, S. W., Clarke, M. M., & Almeida, D. M. (2020). Parasympathetic and sympathetic nervous systems interactively predict change in cognitive functioning in midlife adults. *Psychophysiology*, *57*(10). https://doi.org/10.1111/psyp.13622

Korbach, A., Brünken, R., & Park, B. (2017). Measurement of cognitive load in multimedia learning: a comparison of different objective measures. *Instructional Science*, *45*(4), 515– 536. https://doi.org/10.1007/s11251-017-9413-5

Korbach, A., Brünken, R., & Park, B. (2018). Differentiating different types of cognitive load: A comparison of different measures. *Educational Psychology Review*, *30*(2), 503–529. https://doi.org/10.1007/s10648-017-9404-8

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load.

Behavior Research Methods, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Matsushima, M., & Kadohama, H. (2021). Factors behind why nursing students do not recognize patient needs occurring while confirming intravenous infusion even when they observe them: A comparative descriptive design study. *Nurse Education Today*, *101*, 104886. https://doi.org/10.1016/j.nedt.2021.104886

Naismith, L. M., Cheung, J. J. H., Ringsted, C., & Cavalcanti, R. B. (2015). Limitations of subjective cognitive load measures in simulation-based procedural training. *Medical Education*, *49*(8), 805–814. https://doi.org/10.1111/medu.12732

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science*, *32*(1/2), 1–8. https://doi.org/10.1023/B:TRUC.0000021806.17516.d0

Paas, F., & van Merriënboer, J. J. G. (2020). Cognitive-Load Theory: Methods to Manage Working Memory Load in the Learning of Complex Tasks. *Current Directions in Psychological Science*, *29*(4), 394–398. https://doi.org/10.1177/0963721420922183

Park, S. H., Goldberg, S. A., Al-Ballaa, A., Tayeb, B., Basurrah, M., Abahuje, E., Miccile, C., Pozner, C. N., Yule, S., & Dias, R. D. (2022). Objective measurement of learners' cognitive load during simulation-based trauma team training: A pilot study. *Journal of Surgical Research*, *279*, 361–367. https://doi.org/10.1016/j.jss.2022.06.023

Phitayakorn, R., Minehart, R. D., Pian-Smith, M. C. M., Hemingway, M. W., & Petrusa, E. R. (2015). Practicality of using galvanic skin response to measure intraoperative physiologic autonomic activation in operating room team members. *Surgery*, *158*(5), 1415–1420. https://doi.org/10.1016/j.surg.2015.04.009

Rogers, B. A., & Franklin, A. E. (2021). Cognitive load experienced by nurses in simulation-based learning experiences: An integrative review. *Nurse Education Today*, *99*, 104815. https://doi.org/10.1016/j.nedt.2021.104815

Roh, Y. S., Jang, K. I., & Issenberg, S. B. (2022). Gender differences in psychological safety, academic safety, cognitive load, and debriefing satisfaction in simulation-based learning. *Nurse Educator*, *47*(5), E109–E113. https://doi.org/10.1097/NNE.0000000000001179

Schlairet, M. C., Schlairet, T. J., Sauls, D. H., & Bellflowers, L. (2015). Cognitive load, emotion, and performance in high-fidelity simulation among beginning nursing s: A pilot study. *Journal of Nursing Education*, *54*(3). https://doi.org/10.3928/01484834-20150218-10

Seddigh, A., Stenfors, C., Berntsson, E., Bååth, R., Sikström, S., & Westerlund, H. (2015). The association between office design and performance on demanding cognitive tasks. *Journal of Environmental Psychology*, *42*, 172–181. https://doi.org/10.1016/j.jenvp.2015.05.001

Skuballa, I. T., Xu, K. M., & Jarodzka, H. (2019). The impact of Co-actors on cognitive load: When the mere presence of others makes learning more difficult. *Computers in Human Behavior*, *101*, 30–41. https://doi.org/10.1016/j.chb.2019.06.016

Skulmowski, A. (2023). Guidelines for Choosing Cognitive Load Measures in Perceptually Rich Environments. *Mind, Brain, and Education*, *17*(1), 20–28. https://doi.org/10.1111/mbe.12342

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J. (2018). Measuring cognitive load. *Perspectives on Medical Education*, *7*(1), 1–2. https://doi.org/10.1007/S40037-017-0395-4

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory* (M. J. Spencer & S. P. Lajoie, Eds.). Springer Science + Business Media.

Tran, T. Q., Boring, R. L., Dudenhoeffer, D. D., Hallbert, B. P., Keller, M. D., & Anderson, T. M. (2007). Advantades and Disadvantages of Physiological Assessment for Next Generation Control Room Design. *2007 IEEE 8th Human Factors and Power Plans and HPRCT 13th Annual Meeting*, 259–263.

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0

Webb, C. E., & Dennis, N. A. (2020). Memory for the usual: the influence of schemas on memory for non-schematic information in younger and older adults. *Cognitive Neuropsychology*, *37*(1–2), 58–74. https://doi.org/10.1080/02643294.2019.1674798

Whitley, H. P., Bell, E., Eng, M., Fuentes, D. G., Helms, K. L., Maki, E. D., & Vyas, D. (2015). Practical Team-Based Learning from Planning to Implementation. *American Journal of Pharmaceutical Education*, *79*(10), 149. https://doi.org/10.5688/ajpe7910149

Young, J. Q., Irby, D. M., Barilla-LaBarca, M.-L., Ten Cate, O., & O'Sullivan, P. S. (2016). Measuring cognitive load: mixed results from a handover simulation for medical students. *Perspectives on Medical Education*, *5*(1), 24–32. https://doi.org/10.1007/s40037-015-0240-6

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring Cognitive Load using Eye Tracking Technology in Visual Computing. *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 78–85. https://doi.org/10.1145/2993901.2993908

Zakeri, Z., Mansfield, N., Sunderland, C., & Omurtag, A. (2020). Physiological correlates of cognitive load in laparoscopic surgery. *Scientific Reports*, *10*(1), 12927. https://doi.org/10.1038/s41598-020-69553-3

Chapter II – METHODOLOGICAL CONSIDERATIONS FOR HEALTHCARE SIMULATION COGNITIVE LOAD RESEARCH

Authors: Jeremy W. Hutson, Ashley E. Franklin, Beth A. Rogers

Jeremy Hutson was the primary author on this paper; Dr. Ashley Franklin was the senior author. This paper is ready for submission to *Clinical Simulation in Nursing* (Impact Factor 3.4)

Institutional Review Board approval was not required for this paper.

Abstract

Background: In this conceptual article, we describe strengths and opportunities for using cognitive load (CL) research to explore conditions that underpin simulation learning and performance. We aimed to explore subjective and objective measures, provide suggestions for implementing CL research, and identify implications of CL research for simulation design.

Methods: A journal club approach was implemented in fall 2023 to identify and synthesize CL literature. Each author independently evaluated articles, and eight collaborative sessions were held to discuss the advantages and limitations of measurement approaches and relative applications to healthcare simulation research.

Results: Cognitive Load Theory (CLT) parses out intrinsic, extraneous, and germane components of working memory. Higher CL is typically associated with impaired learning; therefore, CL data adds depth to simulation assessments by considering mental effort as a condition that underpins learning and performance. Subjective measures, including self-report surveys and think-aloud protocols, provide first-person accounts of perceived CL. Objective measures, including eye- tracking and heart rate variability, monitor physiologic responses of the sympathetic nervous system as indirect CL representations.

Conclusion: There is not yet consensus about which CL measure is most accurate. A mixed methods approach may improve rigor and answer the variety of research questions posed in healthcare simulation research. Robust subjective and objective CL data illuminates individual learner, cohort-level, and system-level simulation outcomes.

Introduction

Educators use cognitive load (CL) research to inform instructional design across disciplines, including healthcare simulation, and improve learning outcomes. There is much attention for managing learners' emotions and unnecessary stress during simulation, as both reduce CL (Reedy, 2015). Prioritizing cues provided during simulation, while still providing opportunities for learners to notice salient information, further reduces CL by avoiding split attention (Fraser et al., 2015). The International Nursing Association for Clinical Simulation and Learning's Healthcare Simulation Standards of Best Practice™ (Watts et al., 2021) and Chickering and Gamson's seminal work provides a framework for balanced simulation design (Chickering & Gamson, 1987). For example, by leveling objectives to learners' previous experience and minimizing insignificant distractors, educators optimize CL (Lapierre et al., 2023; Rogers & Franklin, 2021). Given its experiential nature, healthcare simulation is primed for CL research to describe how simulation design influences CL and in turn how CL impacts performance.

CL represents the amount of mental effort required from working memory during learning and performance (Sweller, 2011). Generally, high CL impairs learning and task performance (Korbach et al., 2018). CL is the sum of intrinsic load, extraneous load, and germane load (Van Merrienboer & Sweller, 2010). Intrinsic load represents effort required to perform tasks. Extraneous load is effort associated with information unrelated to learning tasks. Germane load is effort required to retrieve knowledge and create new schemas in long-term memory. Considering the interplay of three CL sources is paramount in healthcare simulation given how inherent critical thinking and task complexity are to healthcare practice.

Understanding conditions under which simulation promotes learning is key for advancing healthcare simulation science (Eppich & Reedy, 2022). Researchers use several CL measures to understand conditions underpinning learning (Vieira, 2017) and accomplish study aims feasibly. The purpose of this conceptual paper is to provide tentative suggestions for measuring CL.

Background

Through CL research, simulationists explore how environmental cues and task complexity impact learning. Tremblay et al., (2017) found immersive simulated environments' technical and equipment aspects were associated with significantly higher CL in pharmacy students compared to peers working with fewer extraneous stimuli. Similarly, novice nurses experienced increased CL while performing routine physical assessments and interventions such as titrating oxygen compared to expert nurses performing the same task (Cabrera-Mino et al., 2019a). Among pre-hospital providers, Bahr and colleagues (2023) recently used physiologic measures to identify clinical events with high CL in a pediatric cardiac arrest scenario. Collectively, these examples highlight how CL research informs simulation design.

Extant clinical judgment research explores situation awareness by evaluating learners' eye movements. Henneman et al. (2017) noted nurses who correctly recognized circulatory overload during blood transfusions spent more time examining salient data than extraneous cues. Amster et al., (2015) found nursing students who correctly identified an allergy error noticed salient cues more than peers with inaccurate clinical judgments (Amster et al., 2015). These studies provide additional insight into the intersection of simulation design, CL, and clinical judgment outcomes.

Educators use simulation assessments to measure competency. Assessments typically consider completion time and performance accuracy; however, task performance may overlook mental effort. Considering mental effort as part of assessment adds depth to simple evaluations. The Conscious Competence Matrix provides a framework for positioning learners in four quadrants moving toward self-awareness and competence (Cannon et al., 2010). Wearable eye-tracking technology, with a first-person video and audio recording, provides educators with more information about mental effort and how CL underpins competent performance. Triangulation of eye-tracking data with valid and reliable performance measures creates a more holistic evaluation than typical time and performance evaluation alone.

CL research provides simulationists a tangible approach for system improvement through a human factors lens. Using human factors science in healthcare supports safe patient care, especially by considering human-system interactions (Russ et al., 2013). Among the top ten human factors essential to patient safety identified by Flin et al., (2009) communication, stress, fatigue, situation awareness, decision-making, and environment are especially relevant to CL and simulation education. CL research represents a mechanism to identify human factors.

Theoretical Foundation

Cognitive Load Theory (CLT) posits knowledge acquisition increases with efficient use of cognitive resources (Korbach et al., 2017). Three working memory components are responsible for knowledge acquisition and transfer (Paas et al., 2004; Sweller, 1988, 2011). CL is determined by element interactivity, defined as number of informational units (i.e., concepts, ideas, procedures, etc.) needing interaction or connection for learning to occur (Korbach et al., 2018; Sweller, 2010). Element interactivity influences the three components of working memory—intrinsic, extraneous, and germane loads—relative to a given task and drives CL fluctuations.

Intrinsic CL describes inherent task complexity of new knowledge (Sweller, 2011) Intrinsic load is not modifiable, unless a learner recalls previous information from long-term memory and applies it to learning new material (Korbach et al., 2018; F. Paas et al., 2004; Van Merrienboer & Sweller, 2010). *Extraneous CL* specifically relates to challenges imposed on working memory by information formatting and delivery. When there are too many extraneous factors, learners have less working memory available to attend to new knowledge and learning is impaired. *Germane CL* describes the mental effort involved in processing new information; a learner's background and previous experiences influence germane load (Sweller, 2011).

Situated Cognition fits with simulation and CL research, owing to the experiential nature of simulation and how learning is embedded within a scenario context and participants' interactions. Durning et al., (2011) applied Situated Cognition Theory to demonstrate how contextual factors influenced clinical outcomes. For example, clinical reasoning relies on interplays between the care provider/team, patient, and situation. Contextual factors such as care provider's experience level, patient's health literacy, and presenting signs/symptoms influence a providers' intrinsic, extraneous, and germane CL. Therefore, CL fluctuations relative to the situational context may promote or impair learning and performance (Zagermann et al., 2016).

Methods

Setting and Participants

In the fall semester of 2023, three collaborators in a research lab at Texas Christian University (Fort Worth, TX, USA) met twice monthly using a journal club approach to critique application of measurement approaches to healthcare simulation.

Journal Club Format and Products

There were eight journal club meetings over the course of a semester, and each meeting lasted two hours. Each member of the group independently evaluated between six and eight published articles before each journal club meeting. During journal club, one member summarized each articles' introduction, methods, results, conclusion and highlighted advantages and limitations of measurement approaches. Before moving to a subsequent presentation, rich discussion unfolded with input, guidance, and questions from peers. Over time, group members organized papers by common measurement approaches in a cloud-based content management platform and developed an outline of advantages and limitations. By the end of the semester, the content management platform contained 57 full-text papers using a variety of CL measurement approaches from within healthcare simulation settings and external disciplines, such as education, public safety, and marketing.

Results

Measures for CL

Choosing a CL measure is a key methodological consideration. Objective measures capture psychophysiologic responses to the sympathetic nervous system, whereas subjective measures, such as self-report surveys or think-aloud strategies, rely on participants' appraisal of how they experience CL. Subjective and objective measures have unique benefits in terms of validity, reliability, and feasibility. Here, we present common CL measures and highlight their strengths and limitations.

Objective Physiologic Measures

In response to calls for objective CL measurement using direct, reliable, and real-time instruments (Brunken et al., 2003), researchers document physiologic responses to stimuli on the heart, lungs, eyes, skin, and brain (Ayres et al., 2021; Chu et al., 2023) as CL proxies. *Eye-tracking* (ETG) technology software reliably detects eye changes (Popa et al., 2015). ETG technology is based on eye-mind and immediacy assumptions, which describe how the mind processes information received by the eyes without delay (Just & Carpenter, 1980). ETG measurements include pupil dilation, fixations (number of fixations and fixation duration), and saccades (Pauszek, 2023). Pupil dilation is an involuntary physiologic response where dilation correlates with additional cognitive effort (Bartels & Marshall, 2012). *Fixation* eye movements can be both voluntary (knowledge driven attention shifts) and involuntary (environmental cues such as blinking lights or noise) and are motivated by the allocation of attention and cognitive activity spent on processing information (Gog et al., 2009; Zagermann et al., 2016).

Researchers

Table 2.1. Common Eye-tracking Measures

Pupil dilation	Increased pupil sizes indicate increased task difficulty and mental effort (Chen et al., 2011). Researchers use pupil size changes to explore how expertise impacts CL during task performance (Cabrera-Mino et al., 2019b; Naik et al., 2022)
Fixation	A state where the eye remains still for between 200 milliseconds up to several seconds (Korbach et al., 2017). Researchers define salient areas of interest (AOIs) a priori, and analysis focuses on number of fixations and total fixation duration. Typically, increased number of fixations (Canham & Hegarty, 2010; de Koning et al., 2010) and total fixation time correspond to higher levels of CL (van Gog & Scheiter, 2010). Due to their voluntary aspect, researchers use fixation to explore situation awareness and attention related to clinical
	reasoning (Amster et al., 2015; Henneman et al., 2017; Shinnick, 2022)

Saccade	Voluntary eye movements the represent a shift in focus from one fixation to another. This time between fixations typically lasts between 30 and 80 milliseconds and is most visualized using scanpath maps of gaze direction (Zagermann et al., 2016). Visual input is reduced during saccades; however, saccades help individuals maintain perception (Ibbotson & Krekelberg, 2011) and can lead to improved reaction times (Ibbotson & Krekelberg, 2011; Johns et al., 2009).
Saccade velocity	An indicator of task difficulty, where decreased velocity indicates tiredness and increased velocity indicates task difficulty (Zagermann et al., 2016). Di Nocera et al., (2006) used visual scanpaths of airline pilots to determine CL was increased during times of highest workload (take-off and landing), when multiple AOI required attention and processing (Di Nocera et al., 2006).

prefer eye tracking measures for continuous real-time, objective CL measurement. ETG glasses are portable and unobtrusive to learning environments. **Table 2.1** presents common pupil change and eye movement measures.

Heart rate and heart rate variability represent CL's physiologic effect on the sympathetic nervous system. Heart rate (HR) measured over time increases with additive CL. For example, Mauriz et al., (2021) noted significant increases in pre/post-test HR among nursing students participating in a cardiopulmonary resuscitation simulation. Similarly, heart rate variability (HRV) describes small changes in time between heart beats, measured by examining an electrocardiogram (ECG); HRV is inversely related with CL (Solhjoo et al., 2019). Technological advances allow wireless ECG monitoring, increasing usability in simulation environments. Researchers use HRV to explore how CL impacts psychomotor skills and memory (Cao et al., 2019; Durantin et al., 2014), clinical reasoning (Park et al., 2022; Solhjoo et al., 2019), and peer learning (Nakayama et al., 2021). *Respiratory rate* (RR) may be captured by an inductive effort belt worn during simulation (Ayres et al., 2021; Hidalgo-Muñoz et al., 2019; Rendon-Velez et al., 2016). Increases in task complexity causes RR increases (Grassmann et al., 2016) . Tidal volume, sigh rate, partial pressure minute ventilation, and partial pressure of end tidal carbon dioxide are additional respiratory proxies for CL (Grassmann et al., 2016); however, equipment required may not be feasible owing to movement restrictions.

Electrodermal Activity (EDA) measures skin electrical conductance related to eccrine gland sweat production, which are sensitive to sympathetic stimuli (Ayres et al., 2021; Posada Quintero et al., 2018). Abrupt changes in skin conductance, called galvanic skin response, more specifically reflect CL-related changes which influence eccrine sweat glands filling and draining (Phitayakorn et al., 2015). Researchers use EDA technology incorporated into smart wristbands for measuring participants' galvanic skin responses to evaluate a video-based learning intervention (Hoogerheide et al., 2019) and to identify CL in operating room team members (Phitayakorn et al., 2015).

Scalp electrodes measure CL through either an electroencephalography (EEG) or near-infrared spectrometry using non-invasive headbands. EEG measures CL with theta, alpha, and beta frequency brain waves as proxy (Chikhi et al., 2022). Wireless EEG technology allows researchers to explore CL and create predictive models based on EEG results (Yoo et al., 2023) while participants experience multimedia (Sarailoo et al., 2022) and virtual reality (Peterson et al., 2018) environments and create predictive models based on EEG results (Yoo et al., 2023). Bahr and colleagues (2023) used near-infrared spectrometry and a wireless headband and converted changes in absorbed light to changes in deoxygenated and oxygenated hemoglobin concentrations in the prefrontal cortex and measured task-evoked hemodynamic changes concurrently (Bahr et al., 2023).

Subjective Measures

Researchers use *self-report surveys* extensively as valid and reliable CL measures in research across disciplines. Surveys most often include self-ratings of perceived task difficulty, engagement, or effort (Korbach et al., 2017), based on the assumption participants introspect effectively and then accurately report cognitive effort in terms of numeric scores (Paas, 1992). Researchers typically use subjective CL measures after simulation but before debriefing. Some researchers; however, pause simulation to collect participant perceptions and then resume simulation (Di Nocera et al., 2006). It is important to note, however, there is uncertainty regarding how pausing during task performance impacts CL.

The National Aeronautics and Space Administration Task Load Index (NASA-TLX) is the most widely used CL measure (Hart & Staveland, 1988) by military, healthcare, aviation, and education disciplines (Grier, 2015). This scale divides task load into six single item subscales: mental demand, physical demand, temporal demand, perceptions of performance, effort, and frustration. TLX uses a summary score to represent workload from 0 (low) to 100 (high). Similarly, Paas's original workload scale (Paas et al., 1994; van Gog et al., 2012) used a 9-point Likert scale to rate mental effort from 1 (very, very, low effort) to 9 (very, very, high effort) on a single item scale. Researchers use the TLX and Paas scale independently or in conjunction with objective methods to represent CL. The high reliability and validity of these methods combined with ease of administration has made them popular choices among CL researchers.

Think-aloud (TA), or verbal protocol, is a qualitative strategy useful for exploring cognitive processes related to behavior and decision-making. Concurrent think-aloud (CTA) strategies require participants to verbalize thoughts during task performance (McDonald et al., 2013), whereas retrospective think-aloud (RTA) protocols involve participants recalling their thoughts afterwards (Salmerón et al., 2017). Researchers use CTA strategies to explore clinical reasoning and critical thinking processes with nurses. For example, Laukvik et al., 2023 used CTA strategy in the setting of electronic health record documentation. Similarly, Burbach et al., (2015) employed both CTA and RTA strategies while exploring nursing students' clinical reasoning during simulation experiences. Both TA strategies offer rich qualitative data.

Discussion

Limitations

Objective measures capture physiologic responses to evoked stimuli. It is important to recognize how factors unrelated to simulation trigger the sympathetic nervous system into producing the same physiologic CL proxies. For example, medications, physical conditions, environment and unrelated emotional stress impact HR, pupil size, and sweat production (Ayres et al., 2021). There is a gap in procedural standards for CL research because consensus is not established for which objective measure(s) are most valid and reliable. Subjective reports, however, rely on learners' ability to introspect, recall events retrospectively, and assign an appropriate numeric value representing perceived mental effort. The subjectivity inherent to self-report measures may result in missed or less than accurate CL accounts (Ayres, 2017). Further, CTA may actually increase CL because it requires participants' dual processing to perform a task and verbalize thoughts simultaneously.

This paper provides theoretical and methodological insight into how CL research advances healthcare simulation science and simulation design. CLT provides an understanding of the functional cognitive architecture involved in learning and task performance. This paper provided an overview of common objective and subjective CL measures and their strengths, limitations, and feasibility. Although subjective measures such as NASA-TLX and Paas' Workload Scale are valid and reliable, multiple objective measures may add depth to understanding because they capture real-time physiological responses to CL fluctuations. As validity and reliability for objective methods continues to be explored, questions emerge regarding method selection.

There is a lack of consensus surrounding which CL measure is most valid and reliable. For example, researchers use HR, RR, HRV, pupil dilation, and subjective measures independently to indicate CL fluctuations. However, each method has limitations which could impact validity. Triangulating CL measures (i.e., objective and subjective) reaches a more comprehensive understanding of CL (Heale & Forbes, 2013). For example, Lee et al., (2019used pupil diameter and gaze metrics combined with a self-report scale to investigate the effects of pausing a medical simulation game on CL (Lee et al., 2019). Similarly, Cai et al., (2022) combined pupil fixation and saccade metrics with NASA-TLX to explore CL differences between novice and expert orthopedic surgeons performing procedures in simulation. Supporting physiologic data with subjective responses clarifies whether results are true reflections of CL fluctuations as opposed to influences by environmental, physical, or emotional factors. Triangulation may be the solution for improving rigor in CL research until consensus is reached about which objective measures are most valid and reliable.

Conclusion

It is clear CL research adds significant knowledge to simulation literature by exploring human factors involved in learning. The intersection of individual participant, physical environment, learning tasks, and human interactions domains, as described by Situated Cognition, impacts learning and behavior. As such, researchers must ask questions which explore cognitive functioning within these domains to identify salient components supporting learning and performance. For example, “What is the influence of peer learning on CL in healthcare simulation? and Can it be optimized?” “Do cues delivered in simulation overwhelm learners’ CL such that they cannot recognize salient data?” CL research is well-positioned to shape simulation design, curriculum sequencing, and ongoing professional development following licensure.

While subjective measures, such as NASA-TLX, were the gold standard for quantifying CL for decades, objective physiologic CL measures such as ETG, HRV, and RR add reliability and validity. Measuring CL objectively during various conditions—such as task complexity, environmental conditions, simulation design, or human interactions—provides insight into information processing, knowledge retention, and simulation performance.

References

Amster, B., Marquard, J., Henneman, E., & Fisher, D. (2015). Using an Eye Tracker During Medication Administration to Identify Gaps in Nursing Students' Contextual Knowledge: An Observational Study. *Nurse Educator*, *40*(2), 83–86. https://doi.org/10.1097/NNE.0000000000000097

Ayres, P. (2017). Subjective Measures of Cognitive Load. In *Cognitive Load Measurement and Application* (pp. 9–28). Routledge. https://doi.org/10.4324/9781315296258-2

Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. J. G. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology*, *12*, 702538. https://doi.org/10.3389/fpsyg.2021.702538

Bahr, N., Ivankovic, J., Meckler, G., Hansen, M., Eriksson, C., & Guise, J.-M. (2023). Measuring cognitively demanding activities in pediatric out-of-hospital cardiac arrest. *Advances in Simulation*, *8*(1), 15. https://doi.org/10.1186/s41077-023-00253-4

Bartels, M., & Marshall, S. P. (2012). Measuring cognitive workload across different eye-tracking hardware platforms. *Proceedings of the Symposium on Eye Tracking Researchand Applications*, 161–164. https://doi.org/10.1145/2168556.2168582

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 53–61. https://doi.org/10.1207/S15326985EP3801_7

Burbach, B., Barnason, S., & Thompson, S. A. (2015). Using "Think Aloud" to Capture Clinical Reasoning during Patient Simulation. *International Journal of Nursing Education Scholarship*, *12*(1), 1–7. https://doi.org/10.1515/ijnes-2014-0044

Cabrera-Mino, C., Shinnick, M. A., & Moye, S. (2019a). Task-Evoked Pupillary Responses in

Nursing Simulation as an Indicator of Stress and Cognitive Load. *Clinical Simulation in Nursing*, *31*, 21–27. https://doi.org/10.1016/j.ecns.2019.03.009

Cabrera-Mino, C., Shinnick, M. A., & Moye, S. (2019b). Task-Evoked Pupillary Responses in Nursing Simulation as an Indicator of Stress and Cognitive Load. *Clinical Simulation in Nursing*, *31*, 21–27. https://doi.org/10.1016/j.ecns.2019.03.009

Cai, B., Xu, N., Duan, S., Yi, J., Bay, B. H., Shen, F., Hu, N., Zhang, P., Chen, J., & Chen, C. (2022). Eye tracking metrics of orthopedic surgeons with different competency levels who practice simulation-based hip arthroscopic procedures. *Heliyon*, *8*(12), e12335–e12335. https://doi.org/10.1016/j.heliyon.2022.e12335

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*(2), 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Cannon, H., Feinstein, A., & Friesen, D. (2010). Managing Complexity: Applying the Conscious Competence Model to Experiential Learning. *Development in Business Simulation and Experiential Learning*, *37*, 172–182.

Cao, X., MacNaughton, P., Cadet, L., Cedeno-Laurent, J., Flanigan, S., Vallarino, J., Donnelly-McLay, D., Christiani, D., Spengler, J., & Allen, J. (2019). Heart Rate Variability and Performance of Commercial Airline Pilots during Flight Simulations. *International Journal of Environmental Research and Public Health*, *16*(2), 237. https://doi.org/10.3390/ijerph16020237

Chen, S., Epps, J., Ruiz, N., & Chen, F. (2011). Eye activity as a measure of human mental effort in HCI. *Proceedings of the 16th International Conference on Intelligent User Interfaces*, 315–318. https://doi.org/10.1145/1943403.1943454

Chickering, A. W., & Gamson, Z. F. (1987). *Seven Principles for Good Practice in Undergraduate Education.* https://eric.ed.gov/?id=ED282491

Chikhi, S., Matton, N., & Blanchet, S. (2022). EEG power spectral measures of cognitive workload: A meta-analysis. *Psychophysiology*, *59*(6), e14009-n/a. https://doi.org/10.1111/psyp.14009

Chu B, Marwaha K, Sanvictores T, Awosika AO, Ayers D. Physiology, Stress Reaction. 2024 May 7. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan–. PMID: 31082164.

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, *20*(2), 111–122. https://doi.org/10.1016/j.learninstruc.2009.02.010

Di Nocera, F., Camilli, M., & Terenzi, M. (2006). Using the Distribution of Eye Fixations to Assess Pilots' Mental Workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(1), 63–65. https://doi.org/10.1177/154193120605000114

Durantin, G., Gagnon, J.-F., Tremblay, S., & Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural Brain Research*, *259*, 16–23. https://doi.org/10.1016/j.bbr.2013.10.042

Durning, S., Artino, A. R., Pangaro, L., van der Vleuten, C. P., & Schuwirth, L. (2011). Context and clinical reasoning: understanding the perspective of the expert's voice. *Medical Education*, *45*(9), 927–938. https://doi.org/10.1111/j.1365-2923.2011.04053.x

Eppich, W., & Reedy, G. (2022). Advancing healthcare simulation research: innovations in theory, methodology, and method. *Advances in Simulation*, *7*(1), 23. https://doi.org/10.1186/s41077-022-00219-y

Flin, R., Winter, J., Sarac, C., & Raduma Tomas, M. A. (2009). *Human Factors in Patient Safety: Review of Topics and Tools*. World Health Organization. http://www.who.int/patientsafety/research/methods_measures/human_factors/human_factors_review.pdf

Fraser, K. L., Ayres, P., & Sweller, J. (2015). Cognitive Load Theory for the Design of Medical Simulations. *Simulation in Healthcare : Journal of the Society for Medical Simulation*, *10*(5), 295–307. https://doi.org/10.1097/SIH.0000000000000097

Gog, T. van, Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior*, *25*(2), 325–331. https://doi.org/10.1016/j.chb.2008.12.021

Grassmann, M., Vlemincx, E., von Leupoldt, A., Mittelstädt, J. M., & Van den Bergh, O. (2016). Respiratory Changes in Response to Cognitive Load: A Systematic Review. *Neural Plasticity*, *2016*, 1–16. https://doi.org/10.1155/2016/8146809

Grier, R. A. (2015). How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1727–1731. https://doi.org/10.1177/1541931215591373

Hart, S. G., & Staveland, L. E. (1988). *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research* (pp. 139–183). https://doi.org/10.1016/S0166-4115(08)62386-9

Heale, R., & Forbes, D. (2013). Understanding triangulation in research. *Evidence Based Nursing*, *16*(4), 98–98. https://doi.org/10.1136/eb-2013-101494

Henneman, E. A., Gawlinski, A., Nicholas, C., McAfee, K., Marquard, J. L., & Andrzejewski, C. (2017). Identification of Transfusion-Associated Circulatory Overload: An Eye-Tracking Study. *Clinical Simulation in Nursing*, *13*(12), 675–679. https://doi.org/10.1016/j.ecns.2017.05.004

Hidalgo-Muñoz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., Tattegrain, H., & Gabaude, C. (2019). Respiration and Heart Rate Modulation Due to Competing Cognitive Tasks While Driving. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00525

Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & van Gog, T. (2019). Enhancing example-based learning: Teaching on video increases arousal and improves problem-solving performance. *Journal of Educational Psychology*, *111*(1), 45–56. https://doi.org/10.1037/edu0000272

Ibbotson, M., & Krekelberg, B. (2011). Visual perception and saccadic eye movements. *Current Opinion in Neurobiology*, *21*(4), 553–558. https://doi.org/10.1016/j.conb.2011.05.012

Johns, M., Crowley, K., Chapman, R., Tucker, A., & Hocking, C. (2009). The effect of blinks and saccadic eye movements on visual reaction times. *Attention, Perception, & Psychophysics*, *71*(4), 783–788. https://doi.org/10.3758/APP.71.4.783

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Korbach, A., Brünken, R., & Park, B. (2017). Measurement of cognitive load in multimedia learning: a comparison of different objective measures. *Instructional Science*, *45*(4), 515–536. https://doi.org/10.1007/s11251-017-9413-5

Korbach, A., Brünken, R., & Park, B. (2018). Differentiating Different Types of Cognitive Load: A Comparison of Different Measures. *Educational Psychology Review*, *30*(2), 503–529. https://doi.org/10.1007/s10648-017-9404-8

Lapierre, A., Lavoie, P., Castonguay, V., Ann-Marie, L., & Arbour, C. (2023). The influence of the simulation environment on teamwork and cognitive load in novice trauma professionals at the emergency department: Piloting a randomized controlled trial. *INTERNATIONAL EMERGENCY NURSING*, *67*. https://doi.org/10.1016/j.ienj.2022.101261

Laukvik, L. B., Rotegård, A. K., Lyngstad, M., Slettebø, Å., & Fossum, M. (2023). Registered nurses' reasoning process during care planning and documentation in the electronic health records: A concurrent think-aloud study. *Journal of Clinical Nursing*, *32*(1–2), 221–233. https://doi.org/10.1111/jocn.16210

Lee, J. Y., Donkers, J., Jarodzka, H., & van Merriënboer, J. J. G. (2019). How prior knowledge affects problem-solving performance in a medical simulation game: Using game-logs and eye-tracking. *Computers in Human Behavior*, *99*, 268–277. https://doi.org/10.1016/j.chb.2019.05.035

Mauriz, E., Caloca-Amber, S., Cordoba-Murga, L., & Vazquez-Casare, A. M. (2021). Effect of Psychophysiological Stress and Socio-Emotional Competencies on the Clinical Performance of Nursing Students during a Simulation Practice. *International Journal of Environmental Research and Public Health*, *18*(10), 5448.

https://doi.org/10.3390/ijerph18105448

McDonald, S., Zhao, T., & Edwards, H. M. (2013). Dual Verbal Elicitation: The Complementary Use of Concurrent and Retrospective Reporting Within a Usability Test. *International Journal of Human-Computer Interaction*, *29*(10), 647–660. https://doi.org/10.1080/10447318.2012.758529

Naik, R., Kogkas, A., Ashrafian, H., Mylonas, G., & Darzi, A. (2022). The Measurement of Cognitive Workload in Surgery Using Pupil Metrics: A Systematic Review and Narrative Analysis. *Journal of Surgical Research*, *280*, 258–272. https://doi.org/10.1016/j.jss.2022.07.010

Nakayama, N., Ejiri, H., Arakawa, N., & Makino, T. (2021). Stress and anxiety in nursing students between individual and peer simulations. *Nursing Open*, *8*(2), 776–783. https://doi.org/10.1002/nop2.680

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Paas, F. G. W. C., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of Cognitive Load in Instructional Research. *Perceptual and Motor Skills*, *79*(1), 419–430. https://doi.org/10.2466/pms.1994.79.1.419

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science*, *32*(1/2), 1–8. https://doi.org/10.1023/B:TRUC.0000021806.17516.d0

Park, H. J., Choi, D., Park, H. A., & Lee, C. A. (2022). Nurse evaluation of stress levels during CPR training with heart rate variability using smartwatches according to their personality: A prospective, observational study. *PloS One*, *17*(6), e0268928. https://doi.org/10.1371/journal.pone.0268928

Pauszek, J. R. (2023). An introduction to eye tracking in human factors healthcare research and medical device testing. *Human Factors in Healthcare*, *3*, 100031. https://doi.org/10.1016/j.hfh.2022.100031

Peterson, S. M., Furuichi, E., & Ferris, D. P. (2018). Effects of virtual reality high heights exposure during beam-walking on physiological stress and cognitive loading. *PLOS ONE*, *13*(7), e0200306. https://doi.org/10.1371/journal.pone.0200306

Phitayakorn, R., Minehart, R. D., Pian-Smith, M. C. M., Hemingway, M. W., & Petrusa, E. R. (2015). Practicality of using galvanic skin response to measure intraoperative physiologic autonomic activation in operating room team members. *Surgery*, *158*(5), 1415–1420. https://doi.org/10.1016/j.surg.2015.04.009

Popa, L., Selejan, O., Scott, A., Mureşanu, D. F., Balea, M., & Rafila, A. (2015). Reading beyond the glance: eye tracking in neurosciences. *Neurological Sciences*, *36*(5), 683–688. https://doi.org/10.1007/s10072-015-2076-6

Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., & Chon, K. H. (2018). Electrodermal Activity Is Sensitive to Cognitive Stress under Water. *Frontiers in Physiology*, *8*. https://doi.org/10.3389/fphys.2017.01128

Reedy, G. B. (2015). Using Cognitive Load Theory to Inform Simulation Design and Practice. *Clinical Simulation in Nursing*, *11*(8), 355–360 https://doi.org/10.1016/j.ecns.2015.05.004

Rendon-Velez, E., van Leeuwen, P. M. ., Happee, R., Horváth, I., van der Vegte, W. F., & de Winter, J. C. F. (2016). The effects of time pressure on driver performance and physiological activity: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *41*, 150–169. https://doi.org/10.1016/j.trf.2016.06.013

Rogers, B. A., & Franklin, A. E. (2021). Cognitive load experienced by nurses in simulation-based learning experiences: An integrative review. *Nurse Education Today*, *99*, 104815. https://doi.org/10.1016/j.nedt.2021.104815

Russ, A. L., Fairbanks, R. J., Karsh, B.-T., Militello, L. G., Saleem, J. J., & Wears, R. L. (2013). The science of human factors: separating fact from fiction. *BMJ Quality & Safety*, *22*(10), 802–808. https://doi.org/10.1136/bmjqs-2012-001450

Salmerón, L., Naumann, J., García, V., & Fajardo, I. (2017). Scanning and deep processing of information in hypertext: an eye tracking and cued retrospective think-aloud study. *Journal of Computer Assisted Learning*, *33*(3), 222–23 https://doi.org/10.1111/jcal.12152

Sarailoo, R., Latifzadeh, K., Amiri, S. H., Bosaghzadeh, A., & Ebrahimpour, R. (2022). Assessment of instantaneous cognitive load imposed by educational multimedia using electroencephalography signals. *Frontiers in Neuroscience*, *16*. https://doi.org/10.3389/fnins.2022.744737

Shinnick, M. A. (2022). Situational Awareness Differences Between Novice and Expert Nurses: Is There a Correlation With Clinical Judgment? *Clinical Simulation in Nursing*, *62*, 57–65. https://doi.org/10.1016/j.ecns.2021.10.002

Solhjoo, S., Haigney, M. C., McBee, E., van Merrienboer, J. J. G., Schuwirth, L., Artino, A. R., Battista, A., Ratcliffe, T. A., Lee, H. D., & Durning, S. J. (2019). Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and Self-Reported Measures of Cognitive Load. *Scientific Reports*, *9*(1), 14668. https://doi.org/10.1038/s41598-019-50280-3

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J. (2011). *Cognitive Load Theory* (pp. 37–76). https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tremblay, M.-L., Lafleur, A., Leppink, J., & Dolmans, D. H. J. M. (2017). The simulated clinical environment: Cognitive and emotional impact among undergraduates. *Medical Teacher*, *39*(2), 181–187. https://doi.org/10.1080/0142159X.2016.1246710

van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and Frequency of Mental Effort Measurement: Evidence in Favour of Repeated Measures. *Applied Cognitive Psychology*, *26*(6), 833–839. https://doi.org/10.1002/acp.2883

van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction*, *20*(2), 95–99. https://doi.org/10.1016/j.learninstruc.2009.02.009

Van Merrienboer, J. J. G., & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical Education*, *44*(1), 85–93.

https://doi.org/10.1111/j.1365-2923.2009.03498.x

Vieira, L. N. (2017). Cognitive effort and different task foci in post-editing of machine translation: A think-aloud study. *Across Languages and Cultures*, *18*(1), 79–105. https://doi.org/10.1556/084.2017.18.1.4

Watts, P. I., McDermott, D. S., Alinier, G., Charnetski, M., Ludlow, J., Horsley, E., Meakim, C., & Nawathe, P. A. (2021). Healthcare Simulation Standards of Best PracticeTM Simulation Design. *Clinical Simulation in Nursing*, *58*, 14–21. https://doi.org/10.1016/j.ecns.2021.08.009

Yoo, G., Kim, H., & Hong, S. (2023). Prediction of Cognitive Load from Electroencephalography Signals Using Long Short-Term Memory Network. *Bioengineering*, *10*(3), 361. https://doi.org/10.3390/bioengineering10030361

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring Cognitive Load using Eye Tracking Technology in Visual Computing. *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 78–85. https://doi.org/10.1145/2993901.2993908

Chapter III – PSYCHOMETRIC TESTING OF NASA-TLX TO MEASURE LEARNERS' COGNITIVE LOAD IN INDIVIDUAL AND GROUP NURSING SIMULATIONS

Authors: Jeremy W. Hutson, Ashley E. Franklin, Beth A. Rogers, Danielle Walker

Corresponding Author: Jeremy W. Hutson

Jeremy W. Hutson was the primary author on this paper and conducted analysis under the direction of Dr. Danielle Walker. Dr. Ashley Franklin was senior author. This paper was accepted by *Clinical Simulation in Nursing* (impact factor 3.4), an indexed and peer-reviewed journal with a larger readership of those interested in nursing education, as a major article.

The Institutional Review Board determined this study was exempt

Abstract

Background: The National Aeronautics & Space Administration-Task Load Index (NASA-TLX) is a subjective instrument for measuring cognitive load. Originally designed for aviation, it has proven a valid and reliable instrument across disciplines. Despite extensive use, its ability to measure cognitive load in group and individual nursing simulation has yet to be explored.

Methods: This secondary analysis investigates NASA-TLX psychometric properties among 488 surveys completed by learners from a Bachelor's nursing program during individual and group simulation. Psychometric tests included item analysis, Cronbach's alpha, inter-item correlations, and exploratory factor analysis.

Results: NASA-TLX demonstrated acceptable reliability and validity in individual simulation; however, there was less reliability in group simulations.

Conclusion: Administration variances across individual and group simulation impact NASA-TLX validity and reliability.

Background

Researchers seek to describe cognitive load (CL) to optimize learning in healthcare simulation. The interest in measuring simulation CL stems from recognizing CL underpins behavior and knowledge application errors (Cowan, 2014). Measuring CL is relevant to simulation more so than other pedagogies because the experiential nature and realism create extraneous distractors (Reedy, 2015). Educational psychology literature is replete with data pinpointing extraneous distractors as CL contributors (de Jong, 2010; PAAS et al., 2004; van Merriënboer & Sweller, 2005). In simulation, external distractions including intravenous pump alarms, moulage, time pressure, and increased cue volumes increase learners' extraneous CL, resulting in behavior and knowledge application errors (Reedy, 2015). CL is sensitive to emotions as internal distractors (LeBlanc & Posner, 2022). CL trends alongside emotion, so learners use cognitive resources more rapidly when emotions are high, leaving more room for error. Since individual learner factors, scenario content, and peer groups induce emotional responses, it is worthwhile for simulationists to consider how high CL decreases information processing and memory recall (LeBlanc & Posner, 2022).

Cognitive Load Theory (CLT) describes working memory as the functional aspect of cognitive architecture responsible for information processing and memory recall (Sweller, 2011). Short-term cognitive resources called working memory are limited to process 3-7 information units at one time (Cowan, 2010; Miller, 1956), highlighting the importance for intentional instructional design from developing learning objectives, identifying cues, and building schemas that scaffold as learners progress through a curriculum. Information processing demands are divided between intrinsic, extraneous, and germane loads, with their sum effort comprising global CL (Sweller, 2010). Intrinsic load is the cognitive effort inherent in the task; extraneous load represents effort allocated to less pertinent, or distracting information; and germane load describes cognitive resources allocated to process intrinsic load (van Merriënboer & Sweller, 2005). Effective learning and optimal performance occur when learners have more cognitive resources available to process intrinsic load (Sweller, 1988). Once working memory has processed information, it is transferred to long-term memory where it is stored and available for future recall (Sweller, 2010).

Element interactivity and prior knowledge determine task complexity (Chen et al., 2023). The number of items requiring processing to learn or perform a task represents element interactivity (Sweller, 2010). For example, memorizing oxygen's chemical symbol as two elements ("O" and "2"), and therefore, is not complex. However, analyzing vital signs to inform clinical decisions requires learners process multiple elements and interpret their collective value, representing more complex tasks. Learners must know the significance of tachypnea in long-term memory, for example, to process cues quickly. Possessing knowledge in long-term memory explains differences in CL between novices and experts during complex tasks. Less knowledge available in long-term memory means novices use more cognitive resources and require additional time compared to experts (Cabrera-Mino et al., 2019).

Within the context of element interactivity, task complexity impacts intrinsic and extraneous loads for learners in simulation (Chen et al., 2023). For example, processing cues from electronic medical records, patients, and peers, combined with acuity, increases intrinsic CL. Having increased intrinsic load decreases cognitive resources available for extraneous distractors (Cierniak et al., 2009). Attention has been given towards limiting extraneous distractors in simulation design (Reedy, 2015). For example, simulationists turn off alarms, limit distractors in patient charts, and present worked-out models to decrease scenario complexity in an effort to match objectives with learners' previous experience (Fraser et al., 2015). Efforts to minimize extraneous distractors in turn decrease split attention and free up cognitive resources. To improve management of cognitive burden imposed by intrinsic and extraneous loads during simulation, researchers seek to quantify CL and inform simulation design.

Researchers developed the National Aeronautics and Space Administration Task Load Index (NASA-TLX) in 1988 to measure astronaut trainees' perceived workload (Hart, 2006; Hart & Staveland, 1988), but other fields (i.e., aviation, medicine, nursing, and military) subsequently adopted NASA-TLX. NASA-TLX captures ratings of mental demand, physical demand, temporal demand, performance, effort, and frustration. Many efforts build on Hart and Staveland's (1988) seminal work. Paas (1992) first simplified mental workload ratings to a one item survey but having one item limits reliability and does not capture CL contributors. Leppink et al., (2013) expanded the CL rating scale for measuring intrinsic, extraneous and germane cognitive load in samples of psychology students. Wilson et al., (2011) and Harris et al., (2020) modified the NASA-TLX questions to reflect medical students' CL during in-person and virtual surgical simulation CL. Additionally, Josephsen, (2018) revised Leppink's questions for nursing simulation, but found poor reliability for measuring extraneous CL. Most recently, Greer et al., (2023) refined a tool for measuring intrinsic, extraneous, and germane CL with pediatric medical residents but found simulation designs could impact extrinsic and germane load items so the tool is not ready for widespread use. Most nursing researchers measure CL using NASA-TLX (Rogers & Franklin, 2021). Researchers in medicine and psychology use NASA-TLX combined with objective measures such as eye-tracking, heart rate variability, and cortisol levels (Ayres et al., 2021), where triangulating data from multiple sources provides a holistic understanding of CL and improves rigor (Jones & Bugge, 2006).

Reliability and Validity in Different Contexts

Validity and reliability characteristics reflect how a scale performs relative to administration processes and sample (Barry et al., 2014; DeVellis & Thorpe, 2022; Streiner & Kottner, 2014). For example, NASA-TLX was designed for aeronautics, and researchers provided context-specific instructions, definitions, and performed a pilot test before data collection (Hart & Staveland, 1988). Assuming this instrument will perform exactly the same under different conditions introduces instrumentation bias. NASA-TLX may provide accurate CL data with health sciences learners; however, validity and reliability characteristics must be reported for each use.

Pre-licensure nursing programs use group and individual simulations to accommodate large cohort sizes and match educational activities to learners' previous experience (Rogers et al., 2020). Task complexity and extraneous distractors change in individual versus group simulations (Rogers & Franklin, 2021). We cannot assume NASA-TLX will perform similarly across individual and group simulation because group dynamics impact CL and because individuals and groups may interpret instructions and definitions uniquely (Lapierre et al., 2023).

The NASA-TLX is most frequently used to capture nursing learners' CL in training and research (Rogers & Franklin, 2021). Healthcare researchers have adapted NASA-TLX, however, a paucity of reliability and validity evidence may undermine results. Therefore, the purposes of this study are to: 1) examine the reliability of NASA-TLX in both group and individual simulation settings; and 2) explore how NASA-TLX items represent CL components among prelicensure nursing students during simulation.

Methods

A secondary analysis of 488 NASA-TLX surveys completed by learners in a prelicensure baccalaureate nursing program at a liberal arts university in the southern United States was completed. Participants were enrolled in either a traditional or accelerated nursing program. Inclusion criteria involved those who 1) participated in simulation as a part of regular nursing coursework 2) were at least 18 years old and 3) consented to using data in future research. 150 participants each completed two surveys, one following each of two formative individual simulations separated by four weeks in 2018-2019, as part of a previously published study (Franklin et al., 2020). 188 additional participants each completed one survey in a separate study following one formative group simulation in 2020-2022. Participants completed surveys on a Qualtrics XM (Provo, UT, USA) platform immediately after simulation and before debriefing; participants self-reported demographic data and provided consent for researchers to gather GPA and Assessment Technologies Institute (ATI) standardized test scores from academic records. Our Institutional Review Board determined this study was not human subjects research. The individual simulation protocol involved one learner completing a simulation activity while the group protocol involved three learners working together. Group participants performed tasks specific to their roles (RN1=safety check, RN2=physical assessment, and RN3=medication administration).

Measure

National Aeronautics & Space Administration Task Load Index (NASA-TLX)

A total of six-items assess participants' CL perceptions with a visual analog scale (Hart & Staveland, 1988). Participants self-rate each item on a 0-21 scale. Some researchers determine which item contributes most using pairwise comparisons. However, most nursing researchers use global NASA-TLX scores (Rogers & Franklin, 2021). Five of the six items are anchored with

"Low" on the left, representing low numerical ratings, and "High" on the right. The original performance item, however, displays its right anchor (low score) as "Good" and left anchor (high score) as "Poor"; reverse-scoring may be required.

Analysis

Individual Simulation

All analyses were performed by a statistician using IBM SPSS Statistics (Version 29). We used descriptive statistics to summarize global CL. Next, a paired-samples t-test was used to compare item means and global CL between Time 1 and Time 2. Then, average scale correlations and inter-item correlations were calculated. Correlations between 0.2 and 0.5 indicate measurement of the same construct (Röschel et al., 2021). Finally, we used Cronbach's α to determine internal consistency. DeVellis and Thorpe (2022) noted an $\alpha \geq .70$ is acceptable.

To explore how items represent CL from a validity perspective, exploratory principal components analysis (PCA) was conducted. A cross-sectional approach was used to establish initial validity evidence among nursing students after an individual simulation (Time 1). PCA identifies the minimum number of components that adequately explain the total observed variance (Grimm & Yarnold, 1995). In social science research, the goal is to explain more than half of the observed variance. First, data were examined to check if items were well-suited for PCA based on the Kaiser-Meyer-Olkin (KMO) criteria, Bartlett's test for sphericity, and the individual correlations among the six items. A KMO measurement above 0.5 indicates that an item is suitable, although a value above 0.8 is more ideal (Kaiser, 1974). Since reverse-scoring potentially impacts KMO, both reverse and non-reverse scored items were examined. Bartlett's test for sphericity was also used to determine the suitability of PCA through examining correlations among the six items. With Bartlett's test, strong correlations between items indicate suitability for PCA (Plitchta & Kelvin, 2013). Finally, individual correlations among the six items were examined. Generally, each item needs to have at least one correlation with another item that is above 0.3 (Kaiser, 1974).

Group Simulation

We used descriptive statistics to summarize global CL. Taking into consideration how group dynamics impact CL, a one-way ANOVA was used to compare global CL means between RN1, RN2, and RN3. Cronbach's α determined internal consistency for the entire group sample. Since CL contributors may be role specific, we also calculated Cronbach's α by role to determine NASA-TLX's sensitivity in group simulation. Choosing to use role as the unit of analysis (i.e., RN1, RN2, RN3) functionally limited sample size and validity testing; however, it aligns most closely to Cognitive Load Theory.

Results

Individual Simulation

The final sample of 150 prelicensure nursing students completed NASA-TLX immediately following two individual simulations, totaling 300 surveys. See **Table 3.1** for demographic characteristics.

Table 3.1. Demographic Characteristics

Sociodemographic Characteristics	Individual Sim	Group Sim
	N(%)	

Gender		
Female	142(92.8)	170(91.4)
Male	10(6.5	15(8.1)
Age Range		
18-21	45(29.4)	84(45.2)
22-26	93(60.8)	88(47.3)
27-32	11(7.2)	8(4.3)
33-38	3(2.0)	4(2.2)
39+	1(0.7)	1(0.5)
Race/Ethnicity		
Caucasian	138(90.8)	153(80.3)
Asian	5(3.3)	12(6.5)
African	5(3.3)	10(5.4)
American	1(0.7)	2(1.1)
Native	3(2.0)	8(4.3)
American		
Other		
Previous Degree		
Yes	34(22.2)	26(14.0)
No	119(77.8)	159(85.5)
Previous		
Healthcare		
Work Experience	37(24.2)	114(61.3)
Yes	116(75.8)	69(37.1)
No		

Characteristics of NASA-TLX in Individual Simulation

Descriptive statistics for Time 1 and Time 2 are presented in **Table 3.2**. Two items demonstrated significant change across time. There was a statistically significant increase in physical demand between Time 1 (M=5.64, ±4.17) and Time 2 (M=6.44, ±3.93), t_{149}= -2.77, p= .006; representing a small effect Cohen's d = .047. There was also a statistically significant decrease in frustration scores between Time 1 (M=11.31, SD ± 5.89) and Time 2 (M=9.6, SD ± 5.71), t_{149}=3.08, p=.002 with small effect Cohen's d=.251. Remaining individual items were similar between Time 1 and Time 2.

Table 3.2. NASA-TLX Descriptive Statistics

	Time 1		Time 2	
Item	Mean (SD)	Range	Mean (SD)	Range
Mental	12.16 (4.19)	2-21	11.96 (3.92)	1-21
Physical	5.64 (4.17)	0-16	6.44 (3.93)	0-18
Temporal	11.57 (4.69)	2-21	12.00 (4.86)	1-21
Effort	13.38 (4.11)	2-21	13.36 (4.10)	0-21
Frustration	11.31 (5.89)	0-21	9.60 (5.71)	0-21
Performance	11.23 (4.32)	1-21	10.44 (3.90)	0-19
Total	65.28 (18.02)	12-121	63.80 (16.96)	14-104

NASA-TLX items have acceptable relationships, except performance, which is poorly related to all items except frustration (see **Table 3.3**). The average scale correlation (M=.309, range=.639) suggests reasonable homogeneity among scale items overall (Piedmont, 2014) despite poor correlations with performance item. Alpha for NASA-TLX was acceptable at Time 1 (α=.729) and Time 2 (α=.702). NASA-TLX was well-suited for PCA (KMO=0.755). The Measure for Sampling Adequacy (MSA) for individual items was acceptable; however, the performance item (MSA=.481) fell just below the minimum threshold of 0.5. KMO and MSA were run with performance item reverse-scored and non-reverse-scored, and results were identical. Bartlett's test indicated PCA was suitable with the six workload items ($p<.001$) because there was at least one correlation above 0.3 for each item.

The first PCA model included all six NASA-TLX items and resulted in two emerging components (i.e., cognitive and emotional) with eigenvalues above 1. The cumulative proportion explained by both components was 65.89%, with the first component contributing 45.08% and the second contributing 20.81%. NASA-TLX items represented similar CL components in both Varimax and Promax rotations. Table 4 summarizes results for Varimax component loadings.

Table 3.3. NASA-TLX Inter-item Correlations

	Mental	Physical	Temporal	Effort	Frustration	Performance
Mental	1.00	.409	.541	.537	.368	.067
Physical		1.00	.423	.342	.396	.019
Temporal			1.00	.478	.456	.116
Effort				1.00	.209	-.098
Frustration					1.00	.373
Performance						1.00

A second PCA model with only five items was run, leaving out performance due to its low MSA. Results converged into only one component having an eigenvalue above 1, explaining a cumulative proportion of 53.59% of the total variation (see **Table 3.4**).

Table 3.4. Factor Loading 2 Models

	Model A Component		Model B Component
	1	2	1
Mental	.801	.101	.796
Physical	.672	.156	.693
Temporal	.770	.241	.807
Effort	.793	-.195	.706
Frustration	.476	.700	.645
Performance	-.099	.888	-

Note. Model A, Varimax orthogonal rotation; Model B, Performance item removed, no rotation

Group Simulation

The final sample of 188 prelicensure learners completed NASA-TLX immediately following a group simulation. Each participant self-selected the role they performed during the group simulation (RN1=safety check nurse, RN2=assessment nurse, or RN3=medication nurse).

Characteristics of NASA-TLX in Group Simulation

There was no statistically significant difference in global CL between roles (RN1, RN2, and RN3), $F_{(2, 185)}=.759$, $p=.470$. The alpha for NASA-TLX global sample was acceptable (α=.720). Alpha was also acceptable at RN3 role (α=.802); however, alpha was questionable at RN2 role (α=.678) and RN1 role (α=.684).

Discussion

Researchers use NASA-TLX to describe CL and optimize learning in healthcare simulation. Although NASA-TLX is widely used across multiple disciplines, this is the first study to examine reliability and validity of NASA-TLX in pre-licensure nursing learners during individual and group simulation. Findings from this secondary analysis of 488 surveys indicate NASA-TLX is a reliable measure of CL in individual simulation settings; however, more exploration into its sensitivity in group simulation is needed.

We expected NASA-TLX would differentiate between learners at different time points, owing to maturation. Comparison of means of mental demand, temporal demand, effort, frustration, performance, and global CL trended as expected, though results were not significant. These findings are congruent with leveled progression standards scaffolded across a curriculum.

In other words, an increased dose of simulation should reduce CL. While the "practice effect," in which improvements are related to prior exposure to the same clinical experience, could be a threat to internal validity (Steadman et al., 2006), the individual simulation protocol protected against practice effect by changing patient gender and diagnosis while retaining conceptual simulation design elements. The group simulation protocol included one simulation experience, mitigating any risk of practice effect.

The significant increase in perceived physical demand ratings in individual simulation was unexpected. Individual differences may impact interpretations of "physical load". For example, NASA-TLX (Hart & Staveland, 1988) describes physical load related to "pulling, pushing, turning…" and whether these were "strenuous" or "demanding." Therefore, higher ratings in the absence of physical exertion seem counterintuitive. Braarud (2021) reported inconsistent physical demand ratings among participants completing computer-based nuclear training simulations and suggested the human-machine interaction may have been perceived as physical load. Given the low physical exertion demand in individual simulation, we hypothesize participants conflated the definition of physical demand with electronic health record navigation, performing physical assessments, and walking to the medication room. As such, ratings of physical demand may not truly reflect physical load as intended by Hart & Staveland (1988).

Factor analysis revealed mental demand, physical demand, temporal demand, and effort items strongly related to the same CL construct, while frustration and performance items did not. The frustration item shows cross-loadings on two different components in both PCA models, suggesting frustration is linked to cognitive and emotional CL components. It is difficult to clearly distinguish what cognitive and emotional components measure. These results are similar to extant literature. Tubbs-Cooley et al. (2018) found intensive care nurses' performance correlated with frustration but was unrelated to remaining items. Helton et al., (2022) concluded frustration and performance scales were emotion-based responses unrelated to CL and should be removed from NASA-TLX as they comprise a separate, yet related, construct. Relative to its practical applications, educators must consider emotions are inextricably related to learning (Anine et al., 2022; Madsgaard et al., 2022); therefore, we must acknowledge differences in how frustration impacts CL between learning and professional environments. While the relationship of frustration and performance to CL seems tenuous, Hart and Staveland (1988) contend self-appraisal affects CL in a variety of ways. Hart and Staveland also noted frustration can provide insight into specific sources of task-related CL reflected in inter-item correlations.

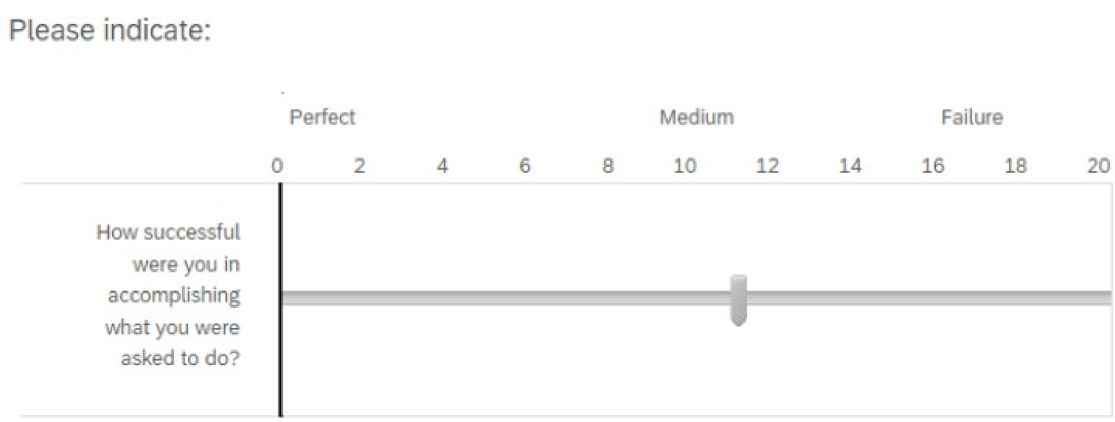
Selection bias is an important consideration. Five NASA-TLX items use “Low” as the left anchor descriptor and “High” as the right anchor descriptor, while the performance item anchors “Good” on the left and “Poor” on the right (**Figure 3.1**). The NASA-TLX instruction manual explains lower numerical scores represent lower CL and assumes a higher self-appraisal for the performance item indicates less CL. The nuances of anchor formats, however, require continued discussion regarding best practice considerations for NASA-TLX administration; nuances either require researchers reverse code the performance item or provide participants with explicit instructions, definitions, and use a pilot test prior to data collection. Despite the potential limitation of selection bias on validity, the KMO of 0.755 indicates individual simulation data is reasonably suited for PCA. Further, individual item correlations suggest good factorability.

Questions regarding the effectiveness of performance and frustration items as contributors to global CL have motivated creation of alternate CL instruments. Creators often

Figure 3.1. Reversed Anchors Examples-Individual and Group Simulations

Individual Simulation

Group Simulation



retain core items from NASA-TLX and add additional items which aim to be sensitive to discipline-specific contributors. For example, the Simulation Task Load Index (SIM-TLX) measures CL in virtual simulation contexts (Harris et al., 2020) and Surgery Task Load Index (SURG-TLX) measures CL in surgical contexts (Wilson et al., 2011). Both instruments retained mental demand, physical demand, and temporal demand items from NASA-TLX and added some items, such as task complexity, situational stress, and distractions. SIM-TLX also retained the frustration item from NASA-TLX despite its inconsistent performance across disciplines (Harris et al., 2020). While SIM-TLX and SURG-TLX demonstrate sensitivity with additional dimensions, arguably, task complexity, situational stress, and distractions exist within the original NASA-TLX. For example, the mental demand item description asks, "How much mental and perceptual activity was required?" and "Was the task easy or demanding? Simple or complex?" (Hart & Staveland, 1988). Similarly, the frustration item description asks "How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?" (Hart & Staveland, 1988). Frustration could align with situational stress. Nuances in language and interpretation require researchers provide context-specific instructions and definitions and perform a pilot test prior to data collection.

While NASA-TLX has shown acceptable reliability in individual simulation, it has less reliability in group simulations. Our supposition is that administration differences across individual and group simulation negatively impact validity and reliability, especially the nuances of anchors on the performance item and the fact that our sample was imbalanced with more individual simulation surveys. It is quite possible simulation design and element interactivity in group simulation impact intrinsic and extraneous CL components so much that they also take away from validity and reliability. Finally, group simulation design elements (e.g. role clarity, report to start the scenario, changing patient condition, and inherent time pressure from a 20-minute scenario) may explain the difference in validity and reliability comparing individual and group simulation in our sample. NASA-TLX data collected in group simulation had reliability approaching an acceptable range. We did not perform EFA owing to the smaller sample of surveys from RN1, RN2, and RN3 roles.

We have four recommendations for improving future research using NASA-TLX. First, researchers should examine the influence of role on CL in group simulation. Second, we recommend exploration of the frustration item to better distinguish it's function as an emotional or cognitive component. Third, researchers should triangulate NASA-TLX with objective measures to improve rigor. Finally, we recommend consistent reporting of reliability statistics, procedures for handling administration and reverse-scoring, simulation context and number of scenario participants.

The convenience sample drawn from a single university for both protocols, and self-selection of roles in group simulation, present potential limitations. While limitations could impact generalizability, this secondary analysis allowed for psychometric testing.

Conclusion

This secondary analysis provides empirical evidence of NASA-TLX as a valid and reliable instrument in individual simulation and a reliable instrument in group simulation. While performance and frustration items create variability, researchers may improve validity by including context-specific instructions and definitions and by using a pilot test. Researchers must also consider visual analog scale anchor formats in the design and analysis plans to reduce selection bias. Triangulating NASA-TLX results with objective CL data may improve rigor. Future research is needed to explore nuances of NASA-TLX characteristics in group simulations and the sensitivity of NASA-TLX to discriminate CL contributors.

References

Anine, M., Kari, R., Monica, Ø., & Hilde, S.-S. (2022). Health professional students' self-reported emotions during simulation-based education: An interpretive descriptive study. *Nurse Education in Practice*, *63*, 103353. https://doi.org/10.1016/j.nepr.2022.103353

Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. J. G. (2021). The Validity of Physiological Measures to Identify Differences in Intrinsic Cognitive Load. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.702538

Barry, A. E., Chaney, B., Piazza-Gardner, A. K., & Chavarria, E. A. (2014). Validity and Reliability Reporting Practices in the Field of Health Education and Behavior. *Health Education & Behavior*, *41*(1), 12–18. https://doi.org/10.1177/1090198113483139

Braarud, P. Ø. (2021). Investigating the validity of subjective workload rating (NASA TLX) and subjective situation awareness rating (SART) for cognitively complex human–machine work. *International Journal of Industrial Ergonomics*, *86*, 103233. https://doi.org/10.1016/j.ergon.2021.103233

Cabrera-Mino, C., Shinnick, M. A., & Moye, S. (2019). Task-Evoked Pupillary Responses in Nursing Simulation as an Indicator of Stress and Cognitive Load. *Clinical Simulation in Nursing*, *31*, 21–27. https://doi.org/10.1016/j.ecns.2019.03.009

Chen, O., Paas, F., & Sweller, J. (2023). A Cognitive Load Theory Approach to Defining and Measuring Task Complexity Through Element Interactivity. *Educational Psychology Review*, *35*(2), 63. https://doi.org/10.1007/s10648-023-09782-w

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, *25*(2), 315–324.

https://doi.org/10.1016/j.chb.2008.12.020

Cowan, N. (2010). The Magical Mystery Four. *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review, 26*(2), 197-223. https://doi.org/10.1007/s10648-013-9246-y.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, *38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0

DeVellis, R. F., & Thorpe, C. T. (2022). *Scale Development: Theory and Applications.* (Fifth Edition). SAGE Publications, Inc.

Franklin, A. E., Thielke, L., Gilbert, G. E., & Waller, M. (2020). TIDES: examining the influence of temporal individual differences on multitasking in educational simulation. *Advances in Simulation*, *5*(1), 31. https://doi.org/10.1186/s41077-020-00144-y

Fraser, K. L., Ayres, P., & Sweller, J. (2015). Cognitive Load Theory for the Design of Medical Simulations. *Simulation in Healthcare : Journal of the Society for Medical Simulation*, *10*(5), 295–307. https://doi.org/10.1097/SIH.0000000000000097

Greer, S. K., Jeffe, D. B., Manga, A., Murray, D. J., & Emke, A. R. (2023). Cognitive Load Assessment Scales in Simulation Validity Evidence for a Novel Measure of Cognitive Load Types. *Simulation in Healthcare: Journal of the Society for Medical Simulation*, *18*(3), 172–180. https://doi.org/10.1097/SIH.0000000000000665

Grimm, L. G., & Yarnold, P. R. (1995). *Reading and Understanding Multivariate Statistics* (L. G. Grimm & P. R. Yarnold, Eds.). American Psychological Association.

Harris, D., Wilson, M., & Vine, S. (2020). Development and validation of a simulation workload measure: the simulation task load index (SIM-TLX). *Virtual Reality : The Journal of the Virtual Reality Society*, *24*(4), 557–566. https://doi.org/10.1007/s10055-019-00422-9

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research* (pp. 139–183). https://doi.org/10.1016/S0166-4115(08)62386-9

Helton, W. S., Jackson, K. M., Näswall, K., & Humphrey, B. (2022). The National Aviation and Space Agency Task Load Index (NASA-TLX): Does it Need Updating? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 1245–1249. https://doi.org/10.1177/1071181322661370

Jones, A., & Bugge, C. (2006). Improving understanding and rigour through triangulation: an exemplar based on patient participation in interaction. *Journal of Advanced Nursing*, *55*(5), 612–621. https://doi.org/10.1111/j.1365-2648.2006.03953.x

Josephsen, J. (2018). Cognitive Load Measurement, Worked-Out Modeling, and Simulation. *Clinical Simulation in Nursing*, *23*, 10–15. https://doi.org/10.1016/j.ecns.2018.07.004

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02291575

Lapierre, A., Lavoie, P., Castonguay, V., Lonergan, A. M., & Arbour, C. (2023). The influence of the simulation environment on teamwork and cognitive load in novice trauma professionals; A pilot-controlled study. *Science Talks, 5,* 100133. https://doi.org/10.1016/j.sctalk.2023.100133

LeBlanc, V. R., & Posner, G. D. (2022). Emotions in simulation-based education: friends or foes of learning? *Advances in Simulation*, *7*(1), 3. https://doi.org/10.1186/s41077-021-00198-6

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Madsgaard, A., Smith-Strøm, H., Hunskår, I., & Røykenes, K. (2022). A rollercoaster of emotions: An integrative review of emotions and its impact on health professional students' learning in simulation-based education. *Nursing Open*, *9*(1), 108–121. https://doi.org/10.1002/nop2.1100

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Mouzé-Amady, M., Raufaste, E., Prade, H., & Meyer, J.-P. (2013). Fuzzy-TLX: using fuzzy integrals for evaluating human mental workload with NASA-Task Load indeX in laboratory and field studies. *Ergonomics*, *56*(5), 752–763. https://doi.org/10.1080/00140139.2013.776702

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

PAAS, F., RENKL, A., & SWELLER, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science*, *32*(1/2), 1–8. https://doi.org/10.1023/B:TRUC.0000021806.17516.d0

Plitchta, S. B. & Kelvin, E. (2013). *Munro's Statistical Methods for Health Care Research* (H. Surrena, Ed.; 6th Edition). Walters Kluwer Health, Lippincott Williams &Wilkins.

Reedy, G. B. (2015). Using Cognitive Load Theory to Inform Simulation Design and Practice. *Clinical Simulation in Nursing*, *11*(8), 355–360. https://doi.org/10.1016/j.ecns.2015.05.004

Rogers, B. A., & Franklin, A. E. (2021). Cognitive load experienced by nurses in simulationbased learning experiences: An integrative review. *Nurse Education Today*, *99*, 104815. https://doi.org/10.1016/j.nedt.2021.104815

Rogers, B., Baker, K. A., & Franklin, A. E. (2020). Learning Outcomes of the Observer Role in Nursing Simulation: A Scoping Review. *Clinical Simulation in Nursing*, *49*, 81–89. https://doi.org/10.1016/j.ecns.2020.06.003

Röschel, A., Wagner, C., & Dür, M. (2021). Examination of validity, reliability, and interpretability of a self-reported questionnaire on Occupational Balance in Informal Caregivers (OBI-Care) – A Rasch analysis. *PLOS ONE*, *16*(12), e0261815. https://doi.org/10.1371/journal.pone.0261815

Steadman, R. H., Coates, W. C., Huang, Y. M., Matevosian, R., Larmon, B. R., McCullough, L., & Ariel, D. (2006). Simulation-based training is superior to problem-based learning for the acquisition of critical assessment and management skills*. *Critical Care Medicine*, *34*(1), 151–157. https://doi.org/10.1097/01.CCM.0000190619.42013.94

Streiner, D. L., & Kottner, J. (2014). Recommendations for reporting the results of studies of instrument and scale development and testing. *Journal of Advanced Nursing*, *70*(9), 1970–1979. https://doi.org/10.1111/jan.12402

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J. (2011). *Cognitive Load Theory* (pp. 37–76). https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tubbs-Cooley, H. L., Mara, C. A., Carle, A. C., & Gurses, A. P. (2018). The NASA Task Load Index as a measure of overall workload among neonatal, paediatric and adult intensive care nurses. *Intensive and Critical Care Nursing*, *46*, 64–69. https://doi.org/10.1016/j.iccn.2018.01.004.

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, *17*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0

Wilson, M. R., Poolton, J. M., Malhotra, N., Ngo, K., Bright, E., & Masters, R. S. W. (2011a). Development and Validation of a Surgical Workload Measure: The Surgery Task Load

Index (SURG-TLX). *World Journal of Surgery*, *35*(9), 1961–1969. https://doi.org/10.1007/s00268-011-1141-4

Chapter IV – COMPARING NOVICE NURSES' COGNITIVE LOAD IN ROUTINE AND NON-ROUTINE SIMULATIONS USING MIXED METHODS

Authors: Jeremy W. Hutson, Ashley E. Franklin, Beth A. Rogers, Yan Zhang

Corresponding Author: Jeremy W. Hutson

Jeremy W. Hutson was the primary author of this paper and conducted analysis under Dr. Franklin during his tenure at TCU; Dr. Ashley Franklin was senior author on this paper. This paper is ready for submission to *Clinical Simulation in Nursing* (impact factor 3.4).

Institutional Review Board determined this study was exempt.

Abstract

Background: Cognitive load impacts learning and performance. Researchers use both objective and subjective measures to quantify cognitive load. Simulationists should consider the impact of cognitive load on learning, performance, and readiness-for-practice.

Methods: This mixed methods design compared novice nurses' cognitive load in two routine and two non-routine, holistic simulations. We collected objective and subjective quantitative data and qualitative survey data from eleven participants. Two-way repeated measures ANOVA and linear mixed effects models evaluated changes in cognitive load between routine and non-routine simulations and over time.

Results: There was a difference in cognitive load between routine and non-routine simulation types ($F(1, 10) = 23.99$, $p < 0.001$, $\eta^2 = .706$). In routine simulation, participants' pupil sizes were larger ($B = 0.238$, $SE = 0.096$, $t = 2.47$, $p = 0.015$). Novice nurses' emotional experiences of increased cognitive load occurred on a spectrum.

Conclusions: Novice nurses rate cognitive load higher in non-routine simulations; however, cognitive load decreases with exposure. Despite high cognitive load levels, novice nurses describe individual differences impacting their lived experience.

Background

The Nursing Executive Center's 2007 survey of nurse leaders found new graduate nurses met performance expectations in only two of 36 competencies evaluated (Berkow et al., 2009). Ten years later, Kavanagh and Szweda (2017) reported only 23 percent of new graduate nurses possessed entry-level competency needed for safe patient care. Novice nurses' readiness-for-practice gap contributes to approximately 100,000 deaths and $20 billion annual healthcare costs (Rodziewicz et al., 2025). To address this gap, educators use simulation to boost practice readiness (Watts et al., 2021). Simulation researchers use both objective and subjective measures to quantify cognitive load (CL). Simulationists should consider the impact of cognitive load on learning, performance, and readiness-for-practice. (Cabrera-Mino et al., 2019; Kataoka et al., 2011).

CL represents the mental effort required for information processing. Three load-types comprise global CL; excessive demand in any type can overwhelm CL and impair learning (Cierniak et al., 2009; Paas, 1992; Paas & van Merriënboer, 2020). Intrinsic load, the mental effort associated with the task, is fixed and related to an individual's knowledge or experience level. Germane load, resource availability for processing intrinsic load, has limited capacity. However, extraneous load, is variable and imposed by distractions (Sweller et al., 2011). Instructional design, physical environment, and individual characteristics such as previous experience, thoughts, or emotions contribute to extraneous and germane loads (PAAS et al., 2004). Since 2011, researchers have measured nurses' CL in simulation and practice environments in seven quantitative studies (Ahmadi et al., 2024; Cabrera-Mino et al., 2019; Campoe & Giuliano, 2017; Jung & Roh, 2022; Kataoka et al., 2011; Matsushima & Kadohama, 2021; Schlairet et al., 2015). The body of qualitative CL literature is much smaller, with one study describing nurses' CL experience of interprofessional reasoning (Blondon et al., 2017) and another about learning technical skills (Aldridge & Hummel, 2019).

Four studies have explored the effects of high CL on novice nurses' behavioral performance (KATAOKA et al., 2011; Matsushima & Kadohama, 2021; Roh et al., 2022); others utilized samples of experienced nurses (Campoe & Giuliano, 2017) and compared novice with experienced nurses' CL (Cabrera-Mino et al., 2019). One study explored the comparative effect of day shift versus night shift on CL (Ahmadi et al., 2024). In novice nurses, high CL impairs behavioral performance (Kataoka et al., 2011; Matsushima & Kadohama, 2021; Roh et al., 2022 Schlairet et al., 2015).

The aim of this body of work was to compare novice nurses' CL during routine and non-routine, holistic simulations using mixed methods. We hypothesized there would be a difference in CL during routine and non-routine simulations and that physiologic measures would trend with subjective CL.

Methods

Design

We used a two-group, repeated measures, convergent mixed methods design to provide a breadth and depth of understanding.

Measures

The National Aeronautics and Space Administration – Task Load Index (NASA-TLX) quantified perceived CL (Hart & Staveland, 1988). Six dimensions are rated on a bipolar numeric scale from 0-100, with higher scores indicating increased CL (Hart & Staveland, 1988). The NASA-TLX has acceptable internal consistency (Cronbach's α= 0.7 – 0.8) with novice nurses during simulation (Hutson et al., 2024).

Tobii Pro Glasses 3 (Tobii, 2023) measured pupil size, which typically increases when the autonomic nervous system (ANS) is triggered (Ahmadi et al., 2024; Tobii, 2023; Zagermann et al., 2016). Biometric vests captured heart rate (HR), standard deviation of normal-to-normal beats (SDNN), and root mean square of successive differences (RMSSD), as heart rate variability (HRV) metrics. Similar to pupil size, HR typically increases during times of stress, however, SDNN and RMSSD usually decrease (Hexoskin, 2024; Kim et al., 2024; Shaffer & Ginsberg, 2017; Solhjoo et al., 2019).

Qualitative data was collected through Qualtrics. We purposively surveyed novice nurses who reported high CL on the NASA-TLX survey — defined as a global score of 45 or higher — and asked participants to answer six semi-structured questions that guide reflection.

Sample

A purposive, voluntary, convenience sample was recruited from third year, pre-licensure novice nurses enrolled in a dedicated simulation course at a university in the southern United States. A convenience sample of 11 novice nurses was recruited. G*Power suggested 12 participants were needed to detect a difference in CL between routine and non-routine simulations, with a 0.5 effect size, 90% power, and four data collection points using the ANOVA tests.

To be included, participants were enrolled in the dedicated simulation course during fall semester of 2024, able to participate in simulation while wearing eye tracking glasses, and comfortable wearing a snug-fitting biometric vest. Exclusion criteria included need for prescriptive eye glasses, inability to calibrate eye tracking glasses, or inability to capture HRV.

Data Collection

This study received IRB approval. After informed consent, we captured baseline pupil size at stations with unique light sources.

Study Groups

Novice nurses completed two routine and two non-routine simulations, following the same study procedures. Novice nurses first participated in a routine, holistic simulation involving a group of three; only one novice nurse out of the group of three participated in the study. The second and third data collection points involved non-routine, holistic simulations and a group of three. The final data collection point was a routine, holistic simulation where novice nurses provided care independently.

Study Procedures

We scheduled four data collection sessions on four different days, and approximately two weeks separated data collection points. On the day of data collection, novice nurses donned a biometric vest and Tobii Pro Glasses 3. Novice nurses wore eye tracking glasses and the biometric vest for the 20-minute simulation.

HRV data were collected continuously; pupil size was captured for specific behaviors indicative of increased CL. (see **Figure 4.1**). Immediately after simulation, novice nurses

Figure 4.1. Observed Behaviors Guide

1) Learner frozen during simulation: no movement or speaking for more than 10 seconds
2) Cognitive aide use > 3 times
3) Pacing: at least 2 walking passes between 2 points without taking action
4) Nervous laughter: forced laugh response
5) Short communication: blunt, gruff, or rudely short communication to peers or simulated patient
6) Performance error (ex. medication administration error)
7) Longer task completion time: > 10 mins to complete any specific tasks (ex. safety check, medication administration, IV pump programming, focused assessment)
8) Missed response to salient patient cues

completed a NASA-TLX survey on Qualtrics before debriefing. Debriefing was not a part of study procedures.

Participants who reported CL score of 45 of higher on the NASA-TLX survey received an additional Qualtrics survey previously described (see **Figure 4.2**).

Analysis

We used IBM SPSS Statistics (Version 29) for descriptive and comparative analysis. A two-way repeated measures ANOVA explored the relationship between study group and global NASA-TLX scores over four simulation days. Kubios HRV Scientific software (Kubios, 2025) was used to process HRV. Separate linear mixed models examined the effect of routine and non-routine simulation types on average pupil size in millimeters and HRV data over time with fixed

Figure 4.2. Qualitative Survey Prompts

1. Tell me what parts of your simulation were more challenging for you.
2. Explain why that part of simulation felt challenging (didn't know what to do, I didn't know how to do xyz, I was nervous to try xyz, there was so much happening etc, I knew time was sensitive).
3. Describe how you felt when you experienced this challenging part of simulation? (ex. lost, nervous, overwhelmed, etc).
4. How did these feelings affect your thinking during simulation?
5. Describe the physical behaviors were you aware of you were doing during this time of feeling challenged, overwhelmed, etc? (ex. frozen, stared at nothing, paced, bit my lip, etc).
6. Explain what, if anything, during the simulation made it difficult for you to focus or concentrate and why?

and random effects. Models included fixed effects for simulation types and time, with participant as a random intercept. Spearman's rank-order correlation coefficients examined associations between objective measures.

We used Braun and Clark's (2006) thematic analysis to analyze qualitative data (Nowell et al., 2017). Two researchers (JWH and AEF) kept reflexive journals and debriefed weekly. Prolonged engagement by one researcher (JWH) who attended the majority of data collection provided a deeper contextual understanding. Two researchers (JWH and AEF) inductively coded qualitative data independently and compared their findings. Results of thematic analysis were triangulated with NASA-TLX scores, eye-tracking, and HRV data to improve validity.

Results

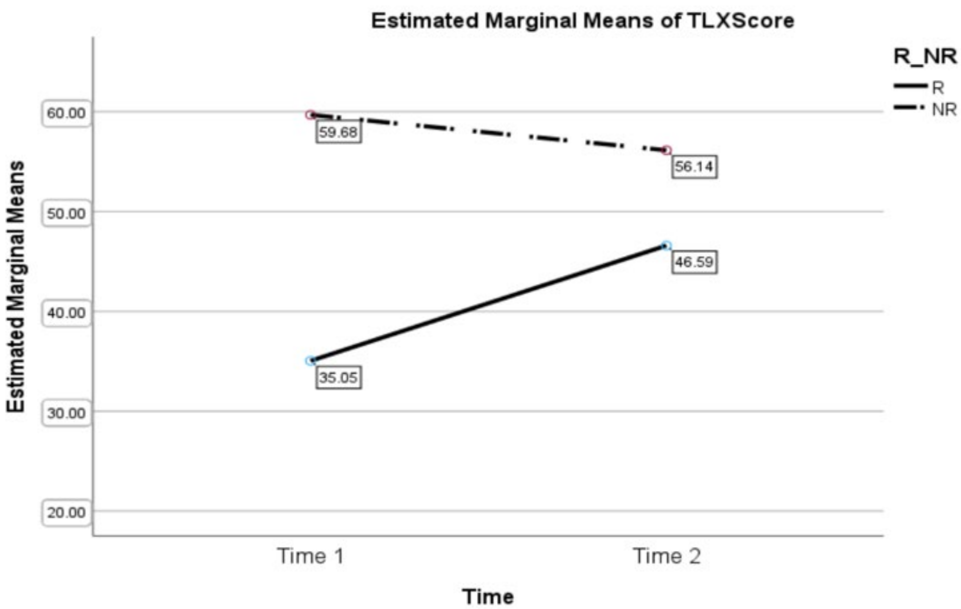
Quantitative Results

Eleven prelicensure nursing students completed the study protocol. See **Table 4.1** for Characteristics of Pre-licensure Nursing Students.

NASA-TLX

A two-way repeated measures ANOVA revealed differences in CL between routine and non-routine simulation types ($F[1, 10] = 23.99$, $p < 0.001$, η^2=.706) and an interaction between simulation types and simulation times ($F(1, 10) = 13.48$, $p = 0.004$, η^2=.574). Posts hoc analysis with Bonferroni adjustment revealed increased CL between routine time 1 and non-routine time 1, (24.64 [95% CI, 14.34 to 34.93] points, $p < 0.001$) and also between routine-time 2 and non-routine time 2, (9.55 [95% CI, 2.01 to 17.09] points, p = 0.018). These results support our hypothesis that novice nurses would report increased cognitive load during non-routine simulations compared to routine simulations. See **Figure 4.3** for Estimated Marginal Means Plot.

Figure 4.3 NASA-TLX Estimated Marginal Means Plot



Pupil Size

Pupil sizes were larger (B = 0.238, SE = 0.096, t = 2.47, p = 0.015) in routine simulations. There was a positive fixed effect of time on pupil size (B = 0.207, SE = 0.084, t = 2.48, p = .015). There was an interaction between routine simulation type and time. However, pupil size decreased over time during non-routine simulations (B = -.285, SE = 0.143, t = -1.99, p = 0.049). The random intercept variance was 0.224, p = 0.033 and Interclass Correlation Coefficient = 0.638. Restricted -2 Log Likelihood (-2LL) indicated the random effects model (2LL=135.55) was the best fit compared to fixed effects model (-2LL= 223.71).

Table 4.1. Characteristics of Pre-Licensure Nursing Students

Sociodemographic Characteristics	N(%)
Age	
18-21	8(72.7)
22-26	2(18.2)
27-32	1(9.1)
Gender	
Female	9(81.8)
Male	2(18.2)
Race	
Asian	2(18.2)
White	9(81.8)
Ethnicity	
Hispanic or Latino	8(72.7)
Non-Hispanic or Latino	3(27.3)
Previous Degree	
Yes	2(18.2)
No	9(81.8)

Heart Rate Variability

A fixed effect model revealed HR differences in routine simulation (B = 12.39, SE = 4.73, t = 2.62, p = 0.017). The random intercept variance was 257.92 and Interclass Correlation Coefficient = 0.743. Restricted -2 Log Likelihood (-2LL) indicated the random effects model (-2LL = 233.99) was a better fit compared to fixed effects model (-2LL = 252.93). Similar models for SDNN and RMSSD yielded comparable patterns (see **Table 4.2**). HRV results did not support our hypothesis.

Spearman's rank-order correlations revealed a negative relationship between both HR and SDNN (r_s= -.915, $p < 0.001$), and between HR and RMSSD (r_s= -.893, $p < 0.001$). There was a positive relationship between SDNN and RMSSD (r_s= .957, $p < 0.001$).

Table 4.2. Heart Rate Variability Measures Mixed Effects Model

	HR					SDNN					RMSSD				
Predictor	***B***	***SE***	***t***	***p***	***95% CI***	***B***	***SE***	***t***	***p***	***95% CI***	***B***	***SE***	***t***	***p***	***95% CI***
(Intercept)	90.15	6.00	15.02	<.001	77.27, 103.04	52.18	7.57	6.89	<.001	35.61, 68.75	46.45	9.01	5.16	<.001	26.63, 66.26
Sim Type	12.39	4.73	2.62	.017	2.50 - 22.78	-9.67	4.56	-2.13	.047	-19.22, -.156	-14.14	4.95	-2.86	.01	-24.49, -3.78
Time	257.92	134.78	.747	.056	92.62– 718.27	-3.78	4.8	-.788	.440	-13.81, 6.25	-9.77	5.21	-1.87	.076	-20.67, 1.13
Random Effects															
Residual	89.209					82.40					97.08				
Participant Variance	257.92					478.99					700.39				
ICC	.743					.853					.878				
N	32					32					32				
Model Fit															
-2LL Fixed	252.93					268.28					277.77				
-2LL Fixed+Random	233.99					237.55					243.99				

**Model Fit: -2LL = -2 Restricted Log Likelihood*

Themes

Simulation Design as Inputs

Simulation design elements, including client-specific elements and knowledge application opportunities, helped novice nurses meet a priori objectives (**Table 4.3**). Similarly, technology (e.g., manikins, electronic health records) as structural elements of simulation imposed increased CL. Novice nurses described time pressure as an extraneous source of CL. Novice nurses experienced pressure related to limited simulation duration. For example, "I knew I had 20 mins but…in the clinical setting it takes me more than 20 mins to do a full assessment." Further, time pressure related to client-specific elements was reflected in comments like, "I knew I was taking too long regarding my response to the patient's subjective and objective data" and "Balancing safety and expediency of epinephrine administration during our patient's anaphylactic episode." Time pressure related to change in client condition may impose more CL when novice nurses work to prevent patients' deterioration.

Novice nurses described challenges communicating when experiencing increased CL. One learner stated, "It's challenging when you try to come up with topics to talk with clients after finishing the initial assessment." Similarly, "I noticed (stress) led to my inability to communicate with my team members and our client."

Learner Experiences as Outputs

Novice nurses identified cognitive, physical, and emotional outputs when experiencing high CL. In situations with perceived high CL, learners described feeling "frozen," "stuck," and even "unable to think clearly." Novice nurses felt non-productive signs, such as "heart racing and hands were shaking a bit" and "my breathing was heavy, and I felt a bit nauseous." Other novice nurses, however, revealed searching behaviors connected to increased CL such as

"I…was just standing there looking at the patient for a while because I was trying to think about what we should be doing but couldn't think of anything." Some novice nurses described searching for cues, like "I kept looking around for things that could help my team."

Novice nurses' emotional experiences of increased CL occurred on a spectrum. Anxiety, stress, and a sense of feeling overwhelmed were common and somewhat anticipated. However, comments such as "I was a bit nervous and anxious merely because I was afraid of messing up" and "I felt nervous that I had forgot even more or that I was going to continue to forget steps" demonstrate how self-monitoring further increases CL. Conversely, some novice nurses described positive emotions, including "I felt confident in the fact that I did everything I had to do," "I felt more confident in my ability," "I felt like our team was more prepared," and "I felt very excited to have confidence in an emergency situation for the first time." Comments related to positive emotions hint towards maturation. Interestingly, these novice nurses described positive emotions despite rating perceived CL as high.

Table 4.3 Qualitative Themes and Representative Quotations

Simulation Design as Inputs: Simulation design includes predetermined elements regarding both client case and activity structure which impact learners' cognitive load. Case-related components highlight client acuity, knowledge application, communication, and medication safety; structural elements time limits, number of learners, and technology used in simulation	
Sub-theme	**Learner Quote**
Change in client status	"*the challenging part was when the patient started to rapidly decline.*" *"One thing that was challenging for me was the sudden change in health status."* *"…the only part that made my heart rate go up was when our patient started to have a hard time breathing and it progressively got worse…"*
Time pressure	"*Acquiring and administering medications in a timely manner was challenging.*" *"…pressured to perform quickly."* *"I knew that I was taking a little too long regarding my response to the patient's subjective and objective data."* *"I knew I had 20 mins but…in the clinical setting it takes me more than 20 mins to do a full assessment…."* *"I knew time was sensitive and felt like we were running out of time."* *"Immediate dosage calculation."* *"Balancing safety and expediency of epinephrine administration during our patient's anaphylactic episode."* *"The race against time made doing calculations feel as if it took forever."* *"I knew time was sensitive and I place great importance on this aspect given my prior deficiencies in this department…"*
Communication	*"It's challenging when you try to come up with topics to talk with them after finishing the initial assessment."* *"Trying to link the communication to the current medical diagnosis and how they should manage the symptoms was challenging."* *"I noticed that (stress) lead to my inability to communicate with my team members and patient."*

Knowledge application	*"Conducting patient education on the patient's conditions regarding their pathophysiology and relationship with one another was challenging."* *"…when it came to treating this, we were all at a loss on what to do."* *"I felt that applying knowledge and the time constraint was the most challenging for me."*
Technology	"…pretending the manikin was real was slightly challenging." *"I get lost when looking at the MAR…I was confused on what time it was and which ones were due…"* *"…not seeing orders populate in MAR."* *"…not seeing new orders in the MAR felt like we were going blindsided."* *"I think the manikin threw me off a little bit."*
Learner Experiences as Outputs: Learners' interaction with simulation design inputs evokes a spectrum of cognitive, physical, and emotional states as well as thinking processes. The way learners describe their experience changes over time despite having high cognitive load ratings.	
Sub-theme	**Learner Quote**
Cognitive states	*"These feelings caused me to freeze…"* *"I think I froze for a little bit because I was unsure of what I was supposed to be doing…."* *"I just froze…"* *"I was stunned…"* *"I tried hard to think of interventions but it was like it was blanking for me."* *"I think it for sure clouded my thinking."* *"I blanked out and couldn't come up with what to say…."* *"I felt like I had blinders on and the perspective narrowed to just the patient."*
Emotional states	*"I felt unsure and confused."* *"I didn't feel overwhelmed but a bit challenged."* *"I felt overwhelmed and anxious."* *"I felt nervous and a little anxious."*

	"Nervous, overwhelmed, anxious." *"Nervous and frustrated that I did not know what else to do…."* *"Pressured and anxious."* *"I felt confident in the fact I did everything I had to do."* *"I kept my cool…."* *"I felt more confident this time."* *"Despite the nervous and pressured feelings, I felt I was most engaged and excited during this scenario."*
Thinking process	*"I couldn't remember."* *"I couldn't think entirely clearly…."* *"I lost my train of thought…"* *"…I was trying to think about what we should be doing but couldn't think of anything."* *"It surprisingly made me think better."* *"…being stressed made my thinking a little bit stronger."* *"It made me more focused."* *"It made me speed up my thinking process regarding the nursing process."*
Physical states	*"I stared at nothing from time to time."* *"I believe I stared at the patient until I figured out what to do."* *"I kept looking around."* *"I started sweating a lot, my heart rate was probably through the roof, & my eyes started to swell with tears."* *"I felt fidgety and disorganized."* *"My breathing was heavy and I felt a bit nauseous."* *"…I moved/paced from one side of the bed to the other too frequently."*

	"I could feel my heart racing and my hands were shaking a little bit." *"I noticed myself touching my hair a lot…."* *"I noticed a decrease in my peripheral vision."* *"I noticed my hand started shaking…."*
Individual Characteristics as Throughputs: Learners' mindsets, past experience, skills, and beliefs shape their lived experience of cognitive load within simulation.	
Sub-theme	**Learner Quote**
Mitigating high CL	*"...practicing before sim helped calm my nerves & reaffirm everything I learned in lab."* *"At one point I just had to take a step back and call timeout to readjust."* *"I reminded myself that I have other meds to give & that I should focus on giving the meds I was certain about rather than getting hung up on a medication I could figure out…in the patient's room."*
Performance monitoring	"I felt nervous that I had forgot even more of that I was going to continue to forget steps."

Individual Characteristics as Throughputs

Individual characteristics help to explain the variance among novice nurses who all rated their CL as high. For example, while one learner explained "…I froze and stared at the meds," another mentioned "…I couldn't think clearly but I was determined to figure out the best things I could do in the moment." Some novice nurses were able to persevere while another became stuck. Some novice nurses described coping mechanisms including "…deep breaths bring me back" and "I was able to calm myself the further we got into the simulation."

Discussion

Healthcare simulation activities inherently involve multiple sources of CL. This is the first study to compare novice nurses' CL between routine and non-routine, holistic simulations using mixed methods. Findings from this study indicate novice nurses experience CL on a spectrum, with higher load during non-routine simulations compared to routine simulations.

Novice nurses scoring at or above our threshold of 45 points on NASA-TLX described unique experiences across routine and non-routine simulations.

Our findings correspond with previous research where novice nurses reported increased CL while providing holistic care (Cabrera-Mino et al., 2019) and during skills-based (SB) simulations under time pressure (Matsushima & Kadohama, 2021). Further, our results support findings by Salo et al., (2025) who described a spectrum of emotional experiences owing to individual characteristics.

A degree of pressure is inherent in simulation design related to *a priori* time limits and learning objectives. Our findings align with previous research investigating the impact of time pressure during SB scenarios (Kataoka et al., 2011; Matsushima & Kadohama, 2021) and holistic scenarios where time imposed extraneous load. In our study, novice nurses self-monitoring during simulation recognized time pressure.

Time pressure may contribute toward intrinsic CL when novice nurses respond to emergent changes in client condition. Within the context of holistic scenarios, time pressure relates to both client condition and subsequent need for interventions. Our results add to the literature by identifying time pressure as intrinsic load.

NASA-TLX findings support our hypothesis that novice nurses experience increased CL during non-routine simulation compared to routine simulation. Results align with Cognitive Load Theory (Sweller, 2011), which posits CL increases relative to intrinsic and extraneous influences. Given opportunities to make clinical judgments in non-routine simulations, increases in CL were expected. Pupil size and HRV results, however, indicated higher CL was experienced during routine simulations. Since pupil dilation and heart rate changes are involuntary physiologic responses, these results were unexpected. Physiologic responses did not trend with NASA-TLX. It is possible analysis did not capture significant non-routine simulation influences on CL because Kubios software analyzes the first five minutes of data meeting quality thresholds. CL triggers in non-routine simulations likely occurred after the first five minutes. Also, the sequence and timing of group versus individual simulations may have influenced results. Participants were naïve to the routine simulation formats, which consisted of one group and one individual scenario. However, participants had time to become familiar with the group format as both non-routine scenarios were the last group simulations. Additionally, the emergent nature of the non-routine simulations resulted in collaboration among peers, potentially distributing CL amongst the group rather than contained within one participant.

We used objective and subjective measures over time to compare CL between routine and non-routine simulations. Additional open-ended prompts provided rich data on the lived experience of increased CL. Our results build upon previous work using non-routine holistic scenarios by exploring CL with qualitative data (Cabrera-Mino et al., 2019). Our findings add to the literature around increased CL during simulation, specifically by highlighting the spectrum of CL experience to frame future intervention research testing CL mitigation strategies (Lee et al., 2020; Wheeler et al., 2021).

Our findings from reflexive thematic analysis correspond with previous research (Salo et al., 2025; Schlairet et al., 2015) by highlighting a spectrum of CL experiences based on individual characteristics. Despite all novice nurses indicating increased CL on NASA-TLX, relative emotional experiences were unique. Emotions contribute to CL and influence behaviors (Anine et al., 2022; Madsgaard et al., 2022). For example, some novice nurses described themselves as feeling “overwhelmed” and becoming “stuck” when responding to emergent clinical situations while others were “excited” and actively worked to improve client condition. Polarity in emotional experiences corresponds with the literature exploring novice nurses’ emotions during simulation (Salo et al., 2025) and debrief (Schlairet et al., 2015). Our study of CL in routine versus non-routine simulation adds to the literature by highlighting differences in CL experiences during emergent situations. Future research exploring emotion regulation to mitigate increased CL could benefit simulation design.

Quantitative results support qualitative findings that individual characteristics are significant contributors to the experience of CL during simulation. For example, increases in perceived CL correspond with feelings of being “anxious,” “overwhelmed,” and “frozen,” though some novice nurses described feeling “excited” and “thinking more clearly.” Novice nurses experience emotions along a spectrum even when survey data indicates increased CL.

Our qualitative findings beg the question: Why does such variance exist among learner experiences when all were exposed to the same stimuli? It is possible individual characteristics such as emotional maturity, past experience, and beliefs influence the lens through which novices experience and respond to clinical situations (Fraser et al., 2012; LeBlanc & Posner, 2022). In our study, novices experienced more CL when communicating with clients compared to performing tasks like physical assessments or administering medications. Qualitative findings help us understand the spectrum of difficulty novice nurses perceive when leading dialogue unrelated to their specific tasks. Our simulation design has novice nurses provide care in one of three specific nursing roles (e.g., safety-check, assessment, and medication nurse). Role assignments may limit their ability to see the patient holistically, therefore, creating conversation on relevant topics outside the scope of an individual nursing role may be challenging for novices. Additionally, an inability to view the patient holistically may cause impaired communication among team members who are focused solely on role-specific tasks.

Our study adds to previous literature about CL and skill performance (KATAOKA et al., 2011; Matsushima & Kadohama, 2021) by situating medication administration within holistic simulations. In holistic simulations, multiple sources of CL are inherent. Similarly, our comparison of routine and non-routine simulations adds to previous CL research using holistic scenarios (Roh et al., 2022; Schlairet et al., 2015) to understand differences in CL experiences. Finally, our study was the first to collect data longitudinally. Our study sets the foundation for future research to investigate possible effects of maturation on CL.

Strengths and Limitations

One of the main strengths of this study is its mixed methods approach. Combining objective data with rich, first-person descriptions provided a comprehensive understanding of CL during routine and non-routine, holistic scenarios. Triangulation provided quantitative support for qualitative findings and improved overall rigor.

The first limitation of this study is convenience sampling. Findings could differ if participants included novice nurses from multiple universities. Thus, our findings may not generalize to all novice nurse populations. Conceptually, a learner who is unconsciously incompetent may not experience CL because they are unable to accurately assess the situation (Williams & Nel, 2023). Findings could be different if novice nurses were conscious of their (in)competence. Regarding research design, a strength of this study is that novice nurses assumed roles consistently (e.g., safety-check nurse in routine scenario 1, assessment nurse in non-routine scenario 1, medication nurse in non-routine scenario 3); it could be that role assignment added to difficulty of a scenario in a way our research team did not anticipate. As such, findings could differ if novice nurses were assigned a different role in the same scenario. Despite this limitation, our findings generated a large effect size estimate of the difference between CL in routine and non-routine scenarios. From a measurement standpoint, we arbitrarily set a cutoff score for increased CL at 45 on NASA-TLX. Results could differ with another threshold to define increased CL. Further, using asynchronous qualitative prompts did not allow for follow up or clarification. Therefore, results could differ if we used synchronous interviews or focus groups.

There was some missing data related to pupil size and HRV related to technology, though it was a random pattern of missingness. Our mixed-effects regression accommodated for missing data. While statistical significance was found within pupil and HRV measures, the small sample size may affect generalizability of both fixed and random effects model estimates. To improve robustness of inferences and confirm trends, future research should aim to replicate this study with larger sample sizes. Kubios software is a limitation for this design as it calculates HRV data from the first five minutes of data meeting threshold standards. Simulations in this study were all twenty minutes long, therefore, HRV analyses may not reflect physiologic changes occurring outside the five-minute processing time frame. However, HRV analysis is well-suited for SB simulations with shorter durations. Given that results demonstrated individual characteristics were present in HRV data, future research using HRV in nursing simulation is warranted.

A final limitation is that we did not measure simulation performance or learner satisfaction. As such, we are unable to make inferences about how increased CL related to simulation performance and satisfaction. Despite this limitation, this study adds to our understanding of how novice nurses experience CL in routine and non-routine simulations.

Conclusion

The aim of this study was to compare novice nurses' CL during routine and non-routine simulations. Our mixed methods design produced quantitative evidence that non-routine simulations impose more CL than routine simulations. In addition, qualitative prompts elicited rich descriptions of novice nurses' lived experience and informed conclusions about emotions occurring on a spectrum.

References

Ahmadi, N., Sasangohar, F., Yang, J., Yu, D., Danesh, V., Klahn, S., & Masud, F. (2024). Quantifying Workload and Stress in Intensive Care Unit Nurses: Preliminary Evaluation Using Continuous Eye-Tracking. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *66*(3), 714–728. https://doi.org/10.1177/00187208221085335

Aldridge, M. D., & Hummel, F. (2019). Nursing Students' Perceptions of Skills Learning. *Nurse Educator*, *44*(3), 170–174. https://doi.org/10.1097/NNE.0000000000000569

Anine, M., Kari, R., Monica, Ø., & Hilde, S.-S. (2022). Health professional students' self-reported emotions during simulation-based education: An interpretive descriptive study. *Nurse Education in Practice*, *63*, 103353. https://doi.org/10.1016/j.nepr.2022.103353

Berkow, S., Virkstis, K., Stewart, J., & Conway, L. (2009). Assessing New Graduate Nurse Performance. *Nurse Educator*, *34*(1), 17–22. https://doi.org/10.1097/01.NNE.0000343405.90362.15

Blondon, K. S., Maître, F., Muller-Juge, V., Bochatay, N., Cullati, S., Hudelson, P., Vu, N. V., Savoldelli, G. L., & Nendaz, M. R. (2017). Interprofessional collaborative reasoning by residents and nurses in internal medicine: Evidence from a simulation study. *Medical Teacher*, *39*(4), 360–367. https://doi.org/10.1080/0142159X.2017.1286309

Cabrera-Mino, C., Shinnick, M. A., & Moye, S. (2019). Task-Evoked Pupillary Responses in Nursing Simulation as an Indicator of Stress and Cognitive Load. *Clinical Simulation in Nursing*, *31*, 21–27. https://doi.org/10.1016/j.ecns.2019.03.009

Campoe, K. R., & Giuliano, K. K. (2017). Impact of Frequent Interruption on Nurses' Patient-Controlled Analgesia Programming Performance. *Human Factors*, *59*(8), 1204–1213.

https://doi.org/10.1177/0018720817732605

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, *25*(2), 315–324. https://doi.org/10.1016/j.chb.2008.12.020

Fraser, K., Ma, I., Teteris, E., Baxter, H., Wright, B., & McLaughlin, K. (2012). Emotion, cognitive load and learning outcomes during simulation training. *Medical Education*, *46*(11), 1055– 1062. https://doi.org/10.1111/j.1365-2923.2012.04355.x

Hart, S. G., & Staveland, L. E. (1988). *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research* (pp. 139–183). https://doi.org/10.1016/S0166-4115(08)62386-9

Hutson, J. W., Franklin, A. E., Rogers, B. A., & Walker, D. (2024). Psychometric Testing of NASA-TLX to Measure Learners' Cognitive Load in Individual and Group Nursing Simulations. In *[Manuscript submitted for publication]*. Harris College of Nursing & Health Sciences, Texas Christian University.

Jung, M. J., & Roh, Y. S. (2022). Mediating Effects of Cognitive Load on the Relationship Between Learning Flow and Clinical Reasoning Skills in Virtual Simulation Learning. *Clinical Simulation in Nursing*, *64*, 16–23. https://doi.org/10.1016/j.ecns.2021.12.004

Kataoka, J., Sasaki, M., & Kanda, K. (2011). Effects of mental workload on nurses' visual behaviors during infusion pump operation. *Japan Journal of Nursing Science*, *8*(1), 47–56. https://doi.org/10.1111/j.1742-7924.2010.00158.x

Kavanagh, J. M., & Szweda, C. (2017). A Crisis in Competency: The Strategic and Ethical Imperative to Assessing New Graduate Nurses' Clinical Reasoning. *Nursing Education*

Perspectives, *38*(2), 57–62. https://doi.org/10.1097/01.NEP.0000000000000112

Kim, H. J., Park, Y., & Lee, J. (2024). The Validity of Heart Rate Variability (HRV) in Educational Research and a Synthesis of Recommendations. *Educational Psychology Review*, *36*(2), 42. https://doi.org/10.1007/s10648-024-09878-x

LeBlanc, V. R., & Posner, G. D. (2022). Emotions in simulation-based education: friends or foes of learning? *Advances in Simulation*, *7*(1), 3. https://doi.org/10.1186/s41077-021-00198-6

Lee, J. Y., Donkers, J., Jarodzka, H., Sellenraad, G., & van Merriënboer, J. J. G. (2020). Different effects of pausing on cognitive load in a medical simulation game. *Computers in Human Behavior*, *110*, 106385. https://doi.org/10.1016/j.chb.2020.106385

Madsgaard, A., Smith-Strøm, H., Hunskår, I., & Røykenes, K. (2022). A rollercoaster of emotions: An integrative review of emotions and its impact on health professional students' learning in simulation-based education. *Nursing Open*, *9*(1), 108–121. https://doi.org/10.1002/nop2.1100

Matsushima, M., & Kadohama, H. (2021). Factors behind why nursing students do not recognize patient needs occurring while confirming intravenous infusion even when they observe them: A comparative descriptive design study. *Nurse Education Today*, *101*, 104886. https://doi.org/10.1016/j.nedt.2021.104886

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional

Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science*, *32*(1/2), 1–8. https://doi.org/10.1023/B:TRUC.0000021806.17516.d0

Paas, F., & van Merriënboer, J. J. G. (2020). Cognitive-Load Theory: Methods to Manage Working Memory Load in the Learning of Complex Tasks. *Current Directions in Psychological Science*, *29*(4), 394–398. https://doi.org/10.1177/0963721420922183

Rodziewicz TL, Houseman B, Vaqar S, Hipskind JE. Medical Error Reduction and Prevention. 2024 Feb 12. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan–. PMID: 29763131.

Roh, Y. S., Jang, K. I., & Issenberg, S. B. (2022). Gender Differences in Psychological Safety, Academic Safety, Cognitive Load, and Debriefing Satisfaction in Simulation-Based Learning. *Nurse Educator*, *47*(5), E109–E113. https://doi.org/10.1097/NNE.0000000000001179

Salo, V., Rikala, P., Vähäsantanen, K., & Hämäläinen, R. (2025). Emotions in nursing students' simulations: A qualitative case study. *Clinical Simulation in Nursing*, *101*, 101701. https://doi.org/10.1016/j.ecns.2025.101701

Schlairet, M. C., Schlairet, T. J., Sauls, D. H., & Bellflowers, L. (2015). Cognitive Load, Emotion, and Performance in High-Fidelity Simulation Among Beginning Nursing Students: A Pilot Study. *Journal of Nursing Education*, *54*(3). https://doi.org/10.3928/01484834-2015021810

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*. https://doi.org/10.3389/fpubh.2017.00258

Solhjoo, S., Haigney, M. C., McBee, E., van Merrienboer, J. J. G., Schuwirth, L., Artino, A. R., Battista, A., Ratcliffe, T. A., Lee, H. D., & Durning, S. J. (2019). Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and Self-Reported Measures of Cognitive Load. *Scientific Reports*, *9*(1), 14668. https://doi.org/10.1038/s41598-019-502803

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory* (M. J. Spencer & S. P. Lajoie, Eds.). Springer Science + Business Media.

Tobii. (2023). *Technical Specifications: Eye tracking*. https://www.tobii.com/resource-center/learn-articles/why-choose-tobii-pro-glasses-3-for-eye-tracking-research

Watts, P. I., Rossler, K., Bowler, F., Miller, C., Charnetski, M., Decker, S., Molloy, M. A., Persico, L., McMahon, E., McDermott, D., & Hallmark, B. (2021). Onward and Upward: Introducing the Healthcare Simulation Standards of Best PracticeTM. *Clinical Simulation in Nursing*, *58*, 1–4. https://doi.org/10.1016/j.ecns.2021.08.006

Wheeler, J., Dudas, K., & Brooks, G. (2021). Anxiety and a Mindfulness Exercise in Healthcare Simulation Prebriefing. *Clinical Simulation in Nursing*, *59*, 61–66. https://doi.org/10.1016/j.ecns.2021.05.008

Williams, G. W., & Nel, B. P. (2023). Teachers' professional development shaped through self-reflection using video-stimulated recall. *African Journal of Teacher Education and Development*, *2*(1). https://doi.org/10.4102/ajoted.v2i1.21

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring Cognitive Load using Eye Tracking Technology in Visual Computing. *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 78–85. https://doi.org/10.1145/2993901.2993908

Chapter V — Discussion

This chapter is similar to the discussion chapter of a traditional dissertation, with the exception that it will address themes across manuscripts and expand on some areas that were not addressed in previous chapters due to publication constraints.

Introduction

New graduate nurses are entering professional practice lacking competence to provide safe and effective patient care. The Nursing Executive Center's 2007 survey of nurse leaders found new graduate nurses met performance expectations in only two of the 36 competencies evaluated (Berkow et al., 2009). Ten years later, (Kavanagh & Szweda, 2017) reported only 23 percent of new graduate nurses entering the workforce possessed entry-level competency needed to provide safe patient care. New nurses' readiness-for-practice gap contributes to approximately 100,000 deaths and $20 billion in additional healthcare costs annually (Rodziewicz et al., 2023). To address this gap, educators turn to simulation to standardize learning experiences, assess competency, and improve patient safety (Watts et al., 2021).

Researchers use simulation to explore cognitive load (CL) sources, the impact of high CL on learning, and both physiologic and behavioral responses to high CL. The overarching purpose of this body of work was to examine how CL manifests in novice nurses during holistic simulations. The specific aims of this program of research were to 1) identify how researchers measure CL in healthcare simulation research, 2) determine the validity and reliability of NASA-TLX as a subjective measure of CL with novice nurses in simulation, 3) compare novice nurses' CL during routine and non-routine, holistic simulations, and 4) describe novice nurses' lived experience of increased CL during routine and non-routine simulations. Each specific aim generated new knowledge which will contribute to extant literature. Inferences drawn from this body of research will support educators' consideration of CL in simulation teaching and research; findings will support future research in both academic and practice settings. This final chapter begins with a summary of principal findings for each aim. Key topics selected for further discussion will follow. This chapter concludes with recommendations for future research.

Summary and Principal Findings

Objective and Subjective Measures of Cognitive Load Are Used in Healthcare Simulation Research

To address the first specific aim of this body of work, we explored quantitative and qualitative methods for measuring CL in healthcare simulation education. Researchers use both objective and subjective quantitative measures, either alone or in combination, to identify conditions which increase CL and, therefore, impair learning and performance. There were three specific findings regarding CL measures (**Table 5.1**). First, objective measures reflect physiologic responses to simulation stimuli. While objective measures may improve scientific rigor, the autonomic nervous system (ANS) does not discern between different sources of CL, therefore, objective measures may not completely reflect CL increases relative to instructional design. Second, subjective measures are popular in nursing due to ease of use and cost-effectiveness, however they are subject to reporting bias. Third, in the absence of consensus among researchers regarding preferred measurement methods, triangulation of both objective and subjective measures likely improves rigor and understanding of how CL increases relative to instructional design.

Conclusions from this conceptual paper contribute to nursing simulation research. First, we identified that researchers use both objective and subjective measures to quantify the impact of simulation on CL. Second, we uncovered a lack of consensus regarding preferred measurement methods Therefore, researchers embarking in CL research should choose measures to match their specific aims. Owing to limitations inherent in both objective and subjective measures, we recommend triangulation of multiple CL measures as best-practice to improve scientific rigor.

Table 5.1. Principal Findings: Specific Aim 1

Specific Aim 1: Identify how researchers measure cognitive load in healthcare simulation research.	
Chapter Title	Principal Findings
Methodological Considerations for Healthcare Simulation Cognitive Load Research	1) Objective measures capture physiologic responses to autonomic nervous system stimuli and are used as CL proxies. 2) Subjective measures require recall and reporting of subjective cognitive experiences after simulation. 3) Triangulation of quantitative and qualitative data likely improves rigor.

***Note*: CL = Cognitive Load**

Validity and Reliability of NASA-TLX for Use in Nursing Simulation Cognitive Load Research

The second aim of this body of research was to determine the validity and reliability of NASA-TLX as a subjective measure of CL with novice nurses in simulation. To address reliability, a secondary analysis of 150 participants' mean NASA-TLX scores were compared between two individual simulations. We hypothesized novice nurses would have lower NASA-TLX scores over time, possibly owing to maturation. Further, item means were compared between two individual simulations, and average scale correlations were calculated. We also used Cronbach's α to determine internal consistency. Exploratory principal component analysis (PCA) examined how items represent CL from a validity perspective in the same sample.

To further examine the reliability of the NASA-TLX in group simulation, a secondary analysis from 188 participants explored how learners assigned different roles rated their CL. We also used Cronbach's α to determine internal consistency in group simulation (**Table 5.2**).

Table 5.2. Principal Findings: Specific Aim 2

Specific Aim 2: Determine the validity and reliability of NASA-TLX as a subjective measure of CL with novice nurses in simulation.

Study	Principal Findings
Psychometric Testing of NASA-TLX to Measure Learners' Cognitive Load in Individual and Group Nursing Simulation	1) Novice nurses perceived an increase in physical demand between time 1 (M = 5.64, ± 4.17) and time 2 (M = 6.44, ± 3.93), t_{149} = -2.77, p = 0.006, d = 0.047 during individual simulations. 2) Frustration level decreased over time: time 1 (M = 11.31, ± 5.89), time 2 (M = 9.6, ± 5.71), t_{149} = 3.08, p= 0.002, d = 0.251. 3) In individual simulation, the NASATLX demonstrates reasonable homogeneity among scale items (M = 0.309, range = 0.639), except for performance item. Internal consistency was acceptable for both time 1 (α = 0.729) and time 2 (α = 0.702). 4) The mental, physical, temporal, effort, and frustration items load onto cognitive and emotional constructs of CL. Performance item may have less impact on CL in nursing simulations. 5) Internal consistency for global group simulation surveys was acceptable (α =

		0.720); however, reliability varied across safety check, assessment, and medication nurse roles.

***Note*: NASA-TLX = National Aeronautics & Space Administration – Task Load Index**

NASA-TLX has been widely used across disciplines, including nursing. This is the first study to examine its validity and reliability with novice nurses during simulation. Psychometric analysis revealed all items and overall CL trended as expected. Results for perceived physical demand were unexpected and may be attributed to novice nurses' differences in understanding of the meaning of "physical demand" within this context. Regarding validity, four of the six items strongly related to CL while two items did not. Frustration was strongly linked to both cognitive and emotional CL components, while performance linked only to emotions. Removing the performance item did not improve the model, indicating performance may be linked to CL and another related construct.

Findings from this study indicate NASA-TLX is a valid and reliable instrument for measuring CL during individual nursing simulations. Results of this study make a meaningful contribution to nursing simulation in that researchers can have confidence when using NASA-TLX in individual simulations to measure novice nurses' perceived CL. Future research should investigate the reliability of NASA-TLX in group simulations.

Individual Characteristics May Explain Variance in Cognitive Load Perceptions

The third aim of this body of work was to compare novice nurses' CL during routine and non-routine, holistic simulations. Eleven participants completed the study protocol using a two-group, repeated measures design. Each participant completed two routine and two non-routine scenarios, with two weeks in between each data collection day. Global NASA-TLX scores were compared between groups using a two-way, repeated measures ANOVA. Separate linear mixed models examined the effect of routine and non-routine simulation types on average pupil size in millimeters and HRV measures over time with fixed and random effects. We hypothesized the NASA-TLX, pupil size, and heart rate variability would reflect increased CL during non-routine simulations. While novice nurses rated CL higher during non-routine simulations, pupil size and HRV metrics indicated increased CL during routine simulations, which was unexpected. The mixed effects model indicates variance in pupil size and HRV may be attributed to characteristics unique to each individual (**Table 5.3**). Our results align conceptually with previous research exploring CL differences between novice and expert nurses (Cabrera-Mino et al., 2019), where experience and knowledge directly influenced CL. For example, some novice nurses' perception of a clinical situation may be more accurate than others and influence their perception of CL. Similarly, some novices may cope with cognitive demands more effectively than their peers. While quantitative results indicate individual characteristics may contribute to how CL is experienced by novices, we cannot infer which individual characteristics may contribute to the experience of CL and to what extent from these data alone.

Findings from this study make meaningful contributions to nursing simulation research. In particular, NASA-TLX scores supported Cognitive Load Theory based on simulation design in routine versus non-routine simulations (Sweller, 2020). Researchers can have confidence in the theoretical connectedness of NASA-TLX. This research highlights the influence of individual characteristics on CL. Future research should explore how individual characteristics influence learning and performance.

Table 5.3. Principal Findings: Specific Aim 3

Specific Aim 3: Compare novice nurses' CL during routine and non-routine, holistic simulations	
Study	Principal Findings
Comparing Novice Nurses' Cognitive Load in Routine and Non-Routine Simulations Using Mixed Methods	1) Overall NASA-TLX scores were higher in non-routine versus routine simulations ($F[1, 10] = 23.99$, $p < 0.001$, $\eta^2 = 0.706$). Scores also increased between routine 1 and non-routine 1 simulations (24.64 [95% CI, 14.34 to 34.93] points, $p < 0.001$) as well as between routine 2 and nonroutine 2 simulations (9.55 [95% CI, 2.01 to 17.09] points, $p = 0.018$). 2) Pupil sizes were larger in routine simulations (B = 0.238, SE = 0.096, t = 2.47, p = 0 .015), which was unexpected. Pupil sizes decreased over time (B = -0.285, SE = 0.143, t = -1.99, p = 0.049). Participant as random effect explained a large portion of variance (ICC = 0.638). 3) HRV metrics trended unexpectedly; however, large proportions of variances attributed to participant as random effect. Means results should be interpreted with caution. 4) HR correlated with SDNN ($r_s = -0.915$, $p < 0.001$), and RMSSD ($r_s = -0.893$, $p < 0.001$). There was a significant positive

relationship between SDNN and RMSSD ($r_s = 0.957$, $p < 0.001$).

***Note:* NASA-TLX = National Aeronautics & Space Administration – Task Load Index; HRV = Heart Rate Variability; HR = Heart Rate; SDNN = Standard Deviation of Normal-to-Normal Intervals; RMSSD = Root Mean Square of Successive Differences**

Novice Nurses' Experience of Increased CL Occurs on a Spectrum

To address the fourth specific aim of this body of research, novice nurses scoring 45 points or higher on NASA-TLX were invited to answer six interview-style prompts to elaborate on their CL experience. Reflexive analysis yielded three overarching themes representing the lived experience (**Table 5.4**). Simulation design elements can contribute to novice nurses' intrinsic and extraneous CL. More importantly, novice nurses experience emotional responses to increased CL across a spectrum. Individual characteristics shape how novice nurses perceive and respond to clinical situations. Findings correspond with previous studies describing the role emotions play in learning and responding in healthcare environments (Anine et al., 2022; Madsgaard et al., 2022).

Overall, this body of research provides a significant contribution to nursing simulation. There is now evidence that novice nurses experience emotional responses to increased CL on a spectrum, and this spectrum is a result of unique individual characteristics.

Table 5.4. Principal Findings: Specific Aim 4

Specific Aim 4: Describe novice nurses' lived experience of increased CL during routine and nonroutine simulations.

Study	Principal Findings
Comparing Novice Nurses' Cognitive Load in Routine	1) Analysis revealed three overarching themes: Simulation Design as and Inputs, Learner Experiences as Outputs, and Individual Characteristics as Throughputs 2) Time as a design element can impose both intrinsic and extraneous load based on simulation objectives in holistic versus skills-based simulations. 3) Novice nurses experience a spectrum of emotions in response to increased CL. 4) Individual characteristics inform how experiences are perceived and interpreted. 5) Our findings highlighting the spectrum of CL experiences create opportunities for future intervention research testing CL mitigation strategies.

***Note*: CL = Cognitive Load**

Recommendations to reduce CL in nursing simulation have targeted simulation design, especially minimizing extraneous distractors and matching objectives to learners' knowledge level (Watts et al., 2021). Educators following simulation design best practices typically address CL proactively; however, this study adds new knowledge by demonstrating that novices experience a spectrum of emotions in response to increased

CL. Given emotions influence cognitive functioning, our results provide a foundation for future interventional research on CL mitigation strategies. Managing CL proactively and in real time will optimize learning and performance in simulation.

Continued Discussion on Selected Themes

Cognitive Load in Simulation

The American Association of Colleges of Nursing has given much attention to improving novice nurses' competence to increase readiness-for-practice (Lewis et al., 2022). Simulation CL research is well situated to add insight to contributors and gaps in competence, especially because researchers pinpoint CL and its impact on learning (Eppich & Reedy, 2022).

Previous research has shed light on CL experiences during focused skills performance (Campoe & Giuliano, 2017; KATAOKA et al., 2011; Matsushima & Kadohama, 2021) and within holistic routine care contexts (Cabrera-Mino et al., 2019; Roh et al., 2022; Schlairet et al., 2015). However, this body of research is the first to investigate CL experienced by novices during holistic simulations with varying complexity. Since the professional practice environment is complex and dynamic, simulation strategies should also investigate how changing client needs within a holistic context affect novice nurses. Changes in client needs likely impact learning and, ultimately, performance. More research is needed exploring novice nurses' CL and performance during realistic, holistic situations.

Group Simulation and Cognitive Load

Providing care in peer groups has varying effects on participants based on level of expertise. Walker et al. (2022) noted experts working in groups may have the capacity to manage CL increases effectively, whereas novices with less experience may not. In other words, novices working in groups may experience higher CL compared to experts. However, Collaborative Working Memory Effect explains how multiple individuals apply their limited working memory processing together to solve a problem (Kirschner et al., 2008; Sweller et al., 2019).

Results from the current body of research support positive effects of collaborative efforts in simulation with novices. We hypothesized group simulation would impose extraneous load. Instead, novice nurses experienced a sense of validation and improved confidence from peers. As such, it is important to consider the potential impact of peer groups. Although not experts yet, novice nurses reach a level of comfort and competence in simulation (maturation) that could yield a redundancy effect when working with peers, ultimately increasing their CL, signaling their readiness for working independently. This natural progression from collaborative learning to independent thinking follows tenets of a scaffolding curricula to benefit learning outcomes. Further research is needed to understand the comprehensive effects of group learning in simulation on CL and its implications to instructional design.

Emotion Regulation

Emotion regulation is a participant construct reflecting that individuals' experiences of emotions and their ability to regulate them during simulation occur across a spectrum. Individuality is salient to educators, owing to the significance emotions have on learning and performance. For example, emotions influence our focus, attention, memory, and motivation (LeBlanc & Posner, 2022). Previous research has explored the impact of anxiety and frustration on performance outcomes (Brazil et al., 2023; DeMaria et al., 2016; Fraser et al., 2014). The current body of research adds to extant literature by describing emotions experienced as a result of increased situational CL. Since emotions can change moment-to-moment, educators should consider both the emotional impact from previous experiences and how evolving emotions during simulation impact learning and performance.

The spectrum of experiences described in this body of research warrant investigation into the conditions which produce positive performance outcomes. Much attention is placed on structural design elements of simulation to mitigate CL. However, emotions as a personal construct piques questions regarding CL mitigation interventions to build coping skills.

Researchers have begun to explore the efficacy of CL pre-simulation mindfulness activities (Takhdat et al., 2024) and a reflective pause during simulation to mitigate CL (Lee et al., 2024). Further research is needed to investigate how novices can regulate their emotional responses to optimize learning and performance.

Recommendations for Future Research

Cognitive Load and Performance in Routine versus Non-Routine Simulation Types

A limitation of this study was that performance outcomes were not measured relative to simulation type, therefore, causal relationships could not be established. CLT posits too much intrinsic and/or extraneous load will result in decreased performance (Paas & van Merriënboer, 2020). While this body of research compared CL in routine and non-routine simulations, the next logical step is to determine causal relationships between CL increases relative to simulation type (routine versus non-routine and group versus individual) and performance outcomes.

Another limitation of this study was that we classified simulation as routine or nonroutine a priori according to perceived complexity. It is possible our expert understanding of complexity differs from novice nurses' perspectives. Findings could be different with a more scientific system to classify simulation as routine or non-routine and score complexity. There is a gap in the literature related to nursing simulation complexity and learning outcomes. Pharmacy researchers have manipulated complexity and measured CL (Tremblay et al., 2023); however, more research is needed to determine if manipulating complexity in holistic simulation impacts novice nurses' performance outcomes.

Emotion Regulation Strategies

This body of research identified novice nurses' emotional experiences occur on a spectrum, suggesting emotions can be regulated to decrease extraneous load. To optimize learning, it is important educators consider sources of CL to inform instructional design. This body of work lays the groundwork for intervention research incorporating emotional regulation strategies. Given positive results from intervention studies with nursing and medical students (Lee et al., 2024b; Takhdat et al., 2024), the next logical step is to test interventions against control groups and with longitudinal designs to identify feasibility and efficacy over time.

Design

To help determine causal relationships between performance and interventions, future research utilizing a pre/post design is recommended. Further, using two groups in a control comparison longitudinally can differentiate intervention effectiveness versus maturation.

Conclusion

The purpose of this body of work was to compare how novice nurses' CL during routine and non-routine simulations. Quantitative results from this body of work provide evidence that non-routine simulations impose more CL that routine simulations. In addition, qualitative results describe an emotional spectrum resulting from increased CL. Novice nurses describe feelings of being "anxious" and "overwhelmed" but also "excited" and "able to think more clearly" during times of increased CL. Both quantitative and qualitative results support the concept of individual characteristics playing a significant role in the experience of CL during simulation. Our mixed methods design was a key strength to the body of work. Subjective and objective measures quantified CL, while qualitative prompts elicited rich descriptions of novice nurses' lived experience.

References

Anine, M., Kari, R., Monica, Ø., & Hilde, S.-S. (2022). Health professional students' self-reported emotions during simulation-based education: An interpretive descriptive study. *Nurse Education in Practice*, *63*, 103353. https://doi.org/10.1016/j.nepr.2022.103353

Berkow, S., Virkstis, K., Stewart, J., & Conway, L. (2009). Assessing New Graduate Nurse Performance. *Nurse Educator*, *34*(1), 17–22. https://doi.org/10.1097/01.NNE.0000343405.90362.15

Brazil, V., Orr, R., Canetti, E. F. D., Isaacson, W., Stevenson, N., & Purdy, E. (2023). Exploring participant experience to optimize the design and delivery of stress exposure simulations in emergency medicine. *AEM Education and Training*, *7*(2). https://doi.org/10.1002/aet2.10852

Cabrera-Mino, C., Shinnick, M. A., & Moye, S. (2019). Task-Evoked Pupillary Responses in Nursing Simulation as an Indicator of Stress and Cognitive Load. *Clinical Simulation in Nursing*, *31*, 21–27. https://doi.org/10.1016/j.ecns.2019.03.009

Campoe, K. R., & Giuliano, K. K. (2017). Impact of Frequent Interruption on Nurses' Patient-Controlled Analgesia Programming Performance. *Human Factors*, *59*(8), 1204–1213. https://doi.org/10.1177/0018720817732605

DeMaria, S., Silverman, E. R., Lapidus, K. A. B., Williams, C. H., Spivack, J., Levine, A., & Goldberg, A. (2016). The impact of simulated patient death on medical students' stress response and learning of ACLS. *Medical Teacher*, *38*(7), 730–737.

http://lib.tcu.edu/PURL/EZproxy_link.asp?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,ip,uid&db=hch&AN=116662485&site=ehost-live

Eppich, W., & Reedy, G. (2022). Advancing healthcare simulation research: innovations in theory, methodology, and method. *Advances in Simulation*, *7*(1), 23. https://doi.org/10.1186/s41077-022-00219-y

Fraser, K., Huffman, J., Ma, I., Sobczak, M., McIlwrick, J., Wright, B., & McLaughlin, K. (2014). The emotional and cognitive impact of unexpected simulated patient death: a randomized controlled trial. *CHEST*, *145*(5), 958–963. https://doi.org/10.1378/chest.13-0987

Kataoka, J., Sasaki, M., & Kanda, K. (2011). Effects of mental workload on nurses' visual behaviors during infusion pump operation. *Japan Journal of Nursing Science*, *8*(1), 47–56. https://doi.org/10.1111/j.1742-7924.2010.00158.x

Kavanagh, J. M., & Szweda, C. (2017). A Crisis in Competency: The Strategic and Ethical Imperative to Assessing New Graduate Nurses' Clinical Reasoning. *Nursing Education Perspectives*, *38*(2), 57–62. https://doi.org/10.1097/01.NEP.0000000000000112

Kirschner, F., Paas, F., & Kirschner, P. A. (2008). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, *21*, 31–42.

LeBlanc, V. R., & Posner, G. D. (2022). Emotions in simulation-based education: Friends or foes of learning? *Advances in Simulation*, *7*(1), 3. https://doi.org/10.1186/s41077021-00198-6

Lee, J. Y., Donkers, J., Jarodzka, H., Sellenraad, G., Faber, T. J. E., & van Merriënboer, J. J. G. (2024). The Effects of Reflective Pauses on Performance in Simulation Training. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, *19*(2), 82–89. https://doi.org/10.1097/SIH.0000000000000729

Lee, J. Y., Donkers, J., Jarodzka, H., Sellenraad, G., Faber, T. J. E., & van Merriënboer, J. J. G. (2024). The Effects of Reflective Pauses on Performance in Simulation Training. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, *19*(2), 82–89. https://doi.org/10.1097/SIH.0000000000000729

Lewis, L. S., Rebeschi, L. M., & Hunt, E. (2022). Nursing Education Practice Update 2022: Competency-Based Education in Nursing. *SAGE Open Nursing*, *8*, 23779608221140776. https://doi.org/10.1177/23779608221140774

Madsgaard, A., Smith-Strøm, H., Hunskår, I., & Røykenes, K. (2022). A rollercoaster of emotions: An integrative review of emotions and its impact on health professional students' learning in simulation-based education. *Nursing Open*, *9*(1), 108–121. https://doi.org/10.1002/nop2.1100

Matsushima, M., & Kadohama, H. (2021). Factors behind why nursing students do not recognize patient needs occurring while confirming intravenous infusion even when they observe them: A comparative descriptive design study. *Nurse Education Today*, *101*, 104886. https://doi.org/10.1016/j.nedt.2021.104886

Paas, F., & van Merriënboer, J. J. G. (2020). Cognitive-Load Theory: Methods to Manage Working Memory Load in the Learning of Complex Tasks. *Current Directions in Psychological Science*, *29*(4), 394–398. https://doi.org/10.1177/0963721420922183

Rodziewicz TL, Houseman B, Vaqar S, Hipskind JE. Medical Error Reduction and Prevention. 2024 Feb 12. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan–. PMID: 29763131.

Roh, Y. S., Jang, K. I., & Issenberg, S. B. (2022). Gender Differences in Psychological Safety, Academic Safety, Cognitive Load, and Debriefing Satisfaction in Simulation-Based Learning. *Nurse Educator*, *47*(5), E109–E113. https://doi.org/10.1097/NNE.0000000000001179

Schlairet, M. C., Schlairet, T. J., Sauls, D. H., & Bellflowers, L. (2015). Cognitive Load, Emotion, and Performance in High-Fidelity Simulation Among Beginning Nursing Students: A Pilot Study. *Journal of Nursing Education*, *54*(3). https://doi.org/10.3928/01484834-20150218-10

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Sweller, J., van Merrienboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*, 261–292.

Takhdat, K., Rebahi, H., Rooney, D. M., Ait Babram, M., Benali, A., Touzani, S., Lamtali, S., & El Adib, A. R. (2024). The impact of brief mindfulness meditation on anxiety, cognitive load, and teamwork in emergency simulation training: A randomized controlled trial. *Nurse Education Today*, *132*, 106005. https://doi.org/10.1016/j.nedt.2023.106005

Tremblay, M., Rethans, J., & Dolmans, D. (2023). Task complexity and cognitive load in simulation-based education: A randomised trial. *Medical Education*, *57*(2), 161–169. https://doi.org/10.1111/medu.14941

Walker, K., Asoodar, M., Rudolph, J., Meguerdichian, M., Yusaf, T., Campbell-Taylor, K., & van Merrienboer, J. (2022). Learning practices of experience healthcare teams and dyads in acute care settings: A scoping review. *BMJ Open*, *12*(7).

Watts, P. I., McDermott, D. S., Alinier, G., Charnetski, M., Ludlow, J., Horsley, E., Meakim, C., & Nawathe, P. A. (2021). Healthcare Simulation Standards of Best Practice™ Simulation Design. *Clinical Simulation in Nursing*, *58*, 14–21. https://doi.org/10.1016/j.ecns.2021.08.009

APPENDICES

Appendix A

National Aeronautics & Space Administration – 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 Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Appendix B

Institutional Review Board Documentation

IRB #: IRB#2024-79 Title: Describing Cognitive Overload in Pre-Licensure Nursing Simulation

Creation Date: 2-15-2024 End Date:

Status: Approved

Principal Investigator: Ashley Franklin

Review Board: TCU IRB-1

Sponsor: Study History

Submission: Type Initial

Review: Type Expedited

Decision: Approved

Submission Type: Modification

Review Type :Expedited

Decision: Approved

Key Study Contacts:

Member Jeremy Hutson Role Co-Principal Investigator Contact jeremy.hutson@tcu.edu Member Ashley Franklin Role Principal Investigator Contact A.B.EDGE@tcu.edu Member Ashley Franklin Role Primary Contact Contact A.B.EDGE@tcu.edu Member Beth Rogers Role Investigator Contact b.a.rogers@tcu.edu

Appendix C

Consent Form for Human Subjects Research

Informed Consent for Research

Start of Block: Block 1

Q6 Informed Consent to Participate in Research Title of Research: Delineating the Impact of Simulation Role and Motivation on Cognitive Load Principal Investigator: Ashley Franklin [Co-investigators:] Jeremy Hutson and Beth Rogers Overview: You are invited to participate in a research study. In order to participate, you must be 18 years or older, enrolled in NURS 30671 in Fall 2024, and have successful calibration of pupil size and eye tracking using our research equipment (Tobii Pro Glasses 3). To participate, you must not: require wearing glasses for normal vision (soft contact lenses are okay), use a medical device with infrared sensitivity, or have history of photosensitive epilepsy. Study Details: This study is being conducted at the Texas Christian University Health Professions Learning Center in Fort Worth, Texas. We will also use Qualtrics for collecting consent, demographic and research surveys. The purpose of this study is to describe learners' cognitive load (CL) in simulation. We anticipate this study will require a 15-30 minute screening session, a 15- 30 minute consent and questionnaire at the beginning of the study. Then, there will be three 20-minute group simulations, and one 20 minute individual simulation as a part of routine class activities. We are asking you to share audio and video from routine class activities for research purposes only (eg. audio and video collected by our research equipment and audio/visual system). On each class day, we will also send you a link to answer some questions about your simulation experience. We think this will take 15 minutes per simulation. Participants: You are being asked to take part because you have no previous experience in simulation at TCU. We want to understand how to help novice nurses learn in simulation. If you decide to be in this study, you will be one of 20 participants in this research study at TCU. Voluntary Participation: Your participation is voluntary. You should only take part in this study if you want to volunteer. You should not feel that there is any pressure to take part in the study. You are free to participate in this research or withdraw at any time. You do not have to participate and may stop your participation at any time. Deciding to participate or not to participate will not affect your student status or course grade at TCU. You will continue to participate in the regularly scheduled classroom simulation requirements. However, all study activities must be completed to receive the Amazon gift card. Confidentiality: Even if we publish the findings from this study, we will keep your information private and confidential. Anyone with authority to look at your records must keep them confidential. What is the purpose of the research? The purpose of this research study is to describe learners' cognitive load (CL) in simulation. What is my involvement for participating in this study? You will be asked to meet with researchers to calibrate our research equipment to capture your pupil size and eye tracking. We will also ask you to try on our Hexoskin shirts to ensure you can wear them comfortably. We will complete a baseline screening that involves 15-30 minutes of your time. After screening, you will wear the research equipment during schedule simulation activities while you complete NURS 30671. We will ask you to complete surveys during simulation and after simulation day for approximately 15 minutes to answer questions about your experience in simulation. There are minimal risks involved with this research, and we do not anticipate any additional risks than participating in normal class activities. Your involvement in this research involves 4 additional simulation experiences that are part of normal class activities and take place over a semester. We are asking you to wear our eye tracking glasses and a Hexoskin shirt during your normal class activities and share your simulation videos and eye tracking data with the research team. During simulations, we will record your eye movements. These recordings will only be accessible to the research team. We will name the recordings using de-identified coding system to ensure privacy. We will also collect your vital signs using the Hexoskin shirts. The surveys, interviews, video recordings and vital signs will be kept in double password-protected and encrypted electronic files containing no personal identifiers. Videos will be stored for five years after the study is published. After 2031, the videos will be deleted and removed from the database unless you consent to allowing us to retain them for longer as part of a repository. We may learn information about you as part of the research. We will not share this information with you because our intent is to make inferences based on group data. Are there any alternatives and can I withdraw? You do not have to participate in this research study. You should only take part in this study if you want to volunteer. You should not feel that there is any pressure to take part in the study. You are free to participate in this research or withdraw at any time. Your decision to participate or not participate will not affect your student status, course grade, recommendations, or access to future courses or training opportunities. What are the risks for participating in this study and how will they be minimized? There are minimal risks involved with this research, and we do not anticipate any additional risks than participating in normal class activities. The eye tracking glasses used as research equipment may distract you. We will give you time to adjust to wearing the glasses each time you don them. Most people say they do not notice the glasses at all. There is a risk related to infrared light from the glasses that can interfere with other medical devices. To reduce the risk, we will exclude participants who use those medical devices based on self-report during screening. Through eye tracking glasses, do not increase to risk of seizure, the focus of gaze on a patient monitor with colored, blinking indicators of a patient's physical status could trigger seizure activity in individuals who have previous epilepsy. There is also a small risk of seizure when wearing eye tracking glasses. To reduce the risk, we will exclude participants with photosensitivity and previous epilepsy based on self-report during screening. We will coach participants to verbalize any sensation while wearing glasses, especially change in vision or smell that could represent a pre-seizure aura. If participants experience the sensation, we will have them remove the eye tracking glasses immediately. There is a small risk of shock from the electrical equipment in the glasses. To reduce the risk, the battery pack will be attached to your waistband and glasses secured with a strap around the circumference of participants' heads to prevent the glasses from dropping. If glasses are dropped, the research team will disconnect the battery to prevent electrical shock. There are no known risks for the Hexoskin device. Participating in simulation as a part of normal class activities may induce unpleasant feelings about performance in a simulation role, such as frustration, anxiety, or disappointment. As a part of normal class activities, a trained facilitator will lead a reflective discussion after simulation, though that discussion is not part of the study protocol. If participants are still upset, investigators will help them find a counselor. Although there are steps the investigators have taken to protect your identity, there is a small risk of loss of confidentiality. The following steps will be taken to protect your personal information: • Documents completed as part of the study by faculty are immediately de-identified and coded with a unique code that does not contain health or student information. • Data will be kept in a password-protected and encrypted electronic files. Paper copies will be kept in a locked cabinet in a locked investigator's office (Rogers). Electronic files will contain no personal identifiers. • Consent forms will be kept in a password-protected and encrypted electronic files. • Investigators will maintain one file that contains personal information and study codes (name, email address, and phone number) to follow up with participants during data collection. This file will be double password-protected. Only two investigators (Franklin and Hutson) will have access to the file. The file will be destroyed at the conclusion of the study. • All information will be stored for 5 years after the study is published in a locked file cabinet in a locked investigator's office (Franklin) in a secure building; only the investigators will have access to those files. After 5 years, paper files will be shredded, and electronic files will be deleted. • The only Information retained longer than 5 years after the study is published will be in a data repository for use in possible future research if you indicate agreement with a separate signature at the end of this consent document. What are the benefits of participating in this study? Although you will not directly benefit from being in this study, others might benefit because the knowledge generated may help us design better simulations and improve learning outcomes during simulation. Will I be compensated for participating in this study? You will receive a payment of $50 Amazon Gift card for your participation. This card will be dispersed after all study surveys and simulations are completed. Please tell the researchers if you have any injuries or other problems related to your participation in the study. You should contact your primary care physician for treatment. If your injury or sickness is an emergency, you should call 911 for an ambulance to take you to the emergency room. You or your insurance will be billed for whatever care you receive. Texas Christian University does not provide compensation or payment for any injury or physical harm that may occur as a result of being in this study. Also, Texas Christian University does not provide compensation for loss of employment, income, or emotional duress that may result from your injury or harm. By signing this form, you do not give up your right to seek payment if you are harmed as a result of being in this study. What are my costs to participate in the study? There will be no additional costs to you as a result of being in this study. How will my confidentiality be protected?

Every effort will be made to limit the use and disclosure of your personal information, including research study records, to people who have a need to review this information. We cannot promise complete secrecy. Your records may be reviewed by authorized University personnel, and other individuals who will be bound by the same provisions of confidentiality. Your identifiers might be removed from your private records or your videos. Your information or videos could be used and/or distributed to another investigator for future research studies without additional consent from you or your Legally Authorized Representative. We may publish what we learn from this study. If we do, we will not include your name. We will not publish anything that would let people know who you are. Access to your health information is required to be part of this study. If you choose to take part in this study, you are giving us the authorization (i.e., your permission) to use the protected health information and information collected during the research that can identify you. The health information that we may collect and use for this research may include medical history that may be considered sensitive.   What will happen to the information collected about me after the study is over? We will keep your research data to use for future research. Your name and other information that can directly identify you will be kept secure and stored separately from the research data collected as part of the project. We will not share your research data with other investigators.   Who should I contact if I have questions regarding the study or concerns regarding my rights as a study participant? You can contact Ashley Franklin at a.b.edge@tcu.edu or by phone at 817-929-7986 with any questions that you have about the study. Dr. Brie Diamond, Chair, TCU Institutional Review Board, (817) 257-6152, b.diamond@tcu.edu; or Dr. Floyd Wormley, Associate Provost of Research, research@tcu.edu HIPAA Authorization This research uses or discloses Protected Health Information as defined by the Health Insurance Portability and Accountability Act (HIPAA).   By signing this form, you are permitting Texas Christian University to use your health information for research purposes. The names of the TCU researchers who will gather this information from you are listed at top of this document, including the lead researcher. We will collect the following health information from you: • Name • Telephone number • Email address • Phone number • Protected Health information: o History from any previous healthcare service (i.e., diagnosis or treatment) for seizures, medical devices impacted by infrared lighting, and vision screening • Protected Research information: o Heart rate, respiratory rate, minute ventilation, step count, pupillometry (i.e., constriction and dilation), eye movements (i.e., fixations, saccades) and eye metrics (i.e., time to first fixation, scan path, fixation counts, fixation duration). By selecting "Agree to participate" below, you are agreeing to be in this study. Make sure you understand what the study is about before you agree. You will be given a copy of this document for your records upon request. If you have any questions about the study after you agree to participate, you can contact the study team using the information provided above. “Agree to Participate”

○ Agree to participate (1)

○ Decline (2)

End of Block: Block 1

Start of Block: Default Question Block

Page Break

Q2 Consent to be audio/video recorded

- I agree to be audio/video recorded (1)
- I decline to be audio/video recorded (2)

Page Break

Q3 Consent to Use Data for Future Research: I agree that my information may be shared with other researchers for future research studies that may be similar to this study or may be completely different. The information shared with other researchers will not include any information that can directly identify me. Researchers will not contact me for additional permission to use this information.

○ I agree to have data used for future research (1)

○ I decline to have data used for future research (2)

Page Break

Q4 Consent to be Contacted for Participation in Future Research: I give the researchers permission to keep my contact information and to contact me for future projects.

◯ I agree to be contacted for participation in future research (1)

◯ I decline to be contacted for participation in future research (2)

Page Break

Q5 Please type your first and last name

____

End of Block: Default Question Block