Validity Evidence for an Instrument for Cognitive Load for Virtual Didactic Sessions

Grace Hickam MD[1], Jaime Jordan MD, MAEd[2], Mary R C Haas MD, MHPE[3], Jason Wagner MD[4], David Manthey MD[5], Stephen John Cico MD, MEd[6], Margaret Wolff, MD, MHPE[7], Sally A Santen MD, PHD[8]

[1]Medical Education Fellow, Clinical Instructor, Department of Emergency Medicine, Virginia Commonwealth University

[2]Associate Professor of Clinical Emergency Medicine, Associate Director of Residency Training Program, Vice-Chair, Acute Care College
Department of Emergency Medicine
David Geffen School of Medicine at UCLA, Los Angeles, CA
Ronald Reagan UCLA Medical Center, Los Angeles, CA

[3]Assistant Residency Director, Instructor
Department of Emergency Medicine
University of Michigan Medical School
Ann Arbor, MI

[4]Residency Program Director, Assistant Professor of Emergency Medicine
Washington University in St. Louis School of Medicine

[5]Professor of Emergency Medicine
Wake Forest School of Medicine

[6]Assistant Dean for Graduate Medical Education, Associate Professor of Clinical Emergency Medicine & Pediatrics, Indiana University School of Medicine, Indianapolis, IN.

[7]Associate Professor of Emergency Medicine and Pediatrics, University of Michigan

[8]Senior Associate Dean, Assessment, Evaluation, and Scholarship and professor, Emergency Medicine, Virginia Commonwealth University School of Medicine, and professor of Emergency Medicine and Medical Education, University of Cincinnati College of Medicine.

Funding for this research: None

Potential financial conflicts of interest: VCU receives funding from the American Medical Association and CTSA award No. UL1TR002649 from the National Center for Advancing Translational Sciences for some of Dr. Santen's effort.

Acknowledgements: John Cyrus for literature review, Meagan Rawls for statistical support, and Collyn Murray, MD for providing the virtual lecture.

DR. GRACE HICKAM (Orcid ID : 0000-0002-6247-4530)

DR. JAIME JORDAN (Orcid ID : 0000-0002-6573-7041)

DR. MARY R C HAAS (Orcid ID : 0000-0002-9506-5928)

DR. JASON WAGNER (Orcid ID : 0000-0001-8702-0706)

DR. MARGARET WOLFF (Orcid ID : 0000-0002-3637-2653)

DR. SALLY SANTEN (Orcid ID : 0000-0002-8327-8002)

Article type : Original Contribution

Corresponding author mail id:- Grace.Hickam@vcuhealth.org

Validity Evidence for an Instrument for Cognitive Load for Virtual Didactic Sessions

Grace Hickam MD[1], Jaime Jordan MD, MAEd[2], Mary R C Haas MD, MHPE[3], Jason Wagner MD[4], David Manthey MD[5], Stephen John Cico MD, MEd[6], Margaret Wolff, MD, MHPE[7], Sally A Santen MD, PHD[8]

Abstract

Background

COVID necessitated the shift to virtual resident instruction. The challenge of learning via virtual modalities has the potential to increase cognitive load. It is important for educators to reduce cognitive load to optimize learning, yet there are few available tools to measure cognitive load. The objective of this study is to identify and provide validity evidence following Messicks' framework for an instrument to evaluate cognitive load in virtual emergency medicine didactic sessions.

Methods

This study followed Messicks' framework for validity including content, response process, internal structure, and relationship to other variables.

Content validity evidence included: 1) engagement of reference librarian and literature review of existing instruments; 2) engagement of experts in cognitive load, and relevant stakeholders to review the literature and choose an instrument appropriate to measure cognitive load in EM didactic presentations.

Response process validity was gathered using the format and anchors of instruments with previous validity evidence and piloting amongst the author group.

A lecture was provided by one faculty to four residency programs via Zoom™. Afterwards, residents completed the cognitive load instrument. Descriptive statistics were collected; Cronbach's alpha assessed internal consistency of the instrument; and correlation for relationship to other variables (quality of lecture).

Results

The 10-item Leppink Cognitive Load instrument was selected with attention to content and response process validity evidence. Internal structure of the instrument was good (Cronbach's alpha= .80). Subscales performed well- intrinsic load (α=.96, excellent), extrinsic load (α=.89, good) and germane load (α=.97, excellent). Five of the items were correlated with overall quality of lecture ($p<.05$).

Conclusions

The 10-item Cognitive Load instrument demonstrated good validity evidence to measure cognitive load and the subdomains of Intrinsic, Extraneous, and Germane load. This instrument can be used to provide feedback to presenters to improve the cognitive load of their presentations.

Introduction

The SARS-CoV-2 global pandemic prompted an unprecedented pivot to online medical education. In a relatively short period of time, online learning has moved from the fringes to the cornerstone of medical education.[1] Educators globally have shared their experiences providing how-to guides and lessons learned.[2,3] This initial literature has largely focused on practical elements to help programs transition to online learning.[4,5] Given the differences in instructional approaches and environment between the classroom and virtual settings, it is important to consider learning theories within this virtual context to improve effectiveness of learning.[6–8]

One important premise for learning is Cognitive Load Theory which examines the relationships between working memory and long-term memory.[9] The amount of information working memory can attend to is finite (i.e. cognitive load) and affected by three different factors: intrinsic cognitive load, extrinsic cognitive load, and germane cognitive load.[9–11]

Intrinsic cognitive load refers to the inherent difficulty of understanding a given topic.[12] Although instructors cannot control the difficulty of content presented, they can modify the way they structure and sequence presentation of the material to facilitate understanding and reduce intrinsic load.[12] Suggested strategies to optimize intrinsic learning during lectures include: activate prior learner knowledge; limit the amount of material covered; align content with learner level and experience; and tailor content to flow from simple to complex.[11]

Extrinsic cognitive load refers to resources devoted to the processing of content delivered and represents the component of cognitive load most readily controlled by the instructor.[12] Strategies for reducing extrinsic load have included: minimize environmental distractions; ensure optimal room set-up and audio visual support; focus content only on the learning objectives; utilize visual aids that emphasize imagery rather than text; and rehearsing the session in advance.[11]

Germane cognitive load refers to the process of consolidating newly acquired information from working memory into long-term memory.[12] During this process, the brain organizes new data through the formation of *schema*. Strategies for promoting germane load have included utilizing schema to present information; grouping information in meaningful ways; incorporating concept mapping; and decreasing the level of support as learners advance.[11]

When one of these cognitive load components increases, there is less capacity in the working memory for the other components. In other words, given the limited capacity of working memory, learning and performance will be impaired if working memory is overloaded with activities that don't directly contribute to learning.[9,12] Therefore, instructional design should consider the role and limitations of working memory to maximize learning.

Understanding the influence of cognitive load on the process of learning is key to enhancing virtual instruction. One approach to optimize cognitive load is to provide feedback through the utilization of cognitive load measurement tools. This can help identify strategies that are augmenting and inhibiting learning and retention.[8] Existing measurements of cognitive load commonly fall under three categories: self-report measures, dual-task measures and measures of physiological parameters.[13] Several approaches to measuring cognitive load have previously been undertaken, including those that rely on subjective (self-reported), behavioral and/or physiologic data. Subjective measures such as the Paas scale are the most common, and often inquire about the mental effort required during a learning task.[17,18] The NASA Task Load Index (NASA-TLAX) represents another commonly used subjective cognitive load measure containing six question items related to mental demand, physical demand, temporal demand, performance, effort and frustration.[19] Other measures have included reduced performance on secondary tasks and other physiologic measures such as pupillometry.[20] While each approach to measuring cognitive load carries strengths and weaknesses, many of these commonly used tools do not account for all three of the different components of cognitive load. While measuring individual components of cognitive load may be beneficial, given the pivotal role cognitive load plays in learning, we sought a tool that provides a more complete picture of cognitive load in teaching settings.

Although several different cognitive load measurement instruments have been developed, there is not an instrument with validity evidence designed for measuring cognitive load in the virtual didactic setting for medical trainees. The objective of this study is to identify and provide validity evidence for an instrument to evaluate cognitive load in virtual emergency medicine didactic sessions.

Methods

Study Design: This was a prospective observational study to collect validity evidence on a cognitive load instrument.

<u>Instrument Selection</u>

We employed several processes to select an instrument including engagement of reference librarian, extensive literature review of existing instruments to measure cognitive load, engagement of cognitive load experts and relevant stakeholders to review the literature and choose an instrument appropriate to measure cognitive load in emergency medicine (EM) didactic presentations.

A search was conducted by a research librarian in APA PsycTests, APA PsycInfo, and PubMed. In PsycTests the term *cognitive load* was used to identify validated instruments mentioning the concept. In PsycInfo, a combination of keywords and controlled vocabulary was used to search for the concepts “cognitive load” and “lecture-based instruction” in order to identify instruments used in existing research on the topic. For example, variations on the following search were employed in PsycInfo: (MM "Human Channel Capacity" OR TI "cognitive load") AND (lecture OR didactic). In Pubmed, keywords and phrases were used to create a similar search as there is no specific controlled vocabulary for cognitive load.

The author team reviewed all available instruments and chose a 10-item instrument by Leppink et al. that has only been used for in-class college population in a non-virtual setting.[16] Leppink et al. previously developed the 10-item cognitive load tool with the intention of measuring all three components of cognitive load; although not previously applied to medical residents, the tool had validity evidence in the context of statistics lectures delivered to university students in the social and health sciences.[16] Thus it was important to collect validity evidence with a resident population while using the virtual platform.

<u>Collection of Validity Evidence</u>

We followed Messicks’ framework[14] for validity including content, response process, internal structure, and relationship to other variables. We chose Messick's framework because it is advocated by the American Educational Research Association, the American Psychological Association, the National Council on Measurement in Education, and the Joint Committee on Standards for Educational and Psychological Testing in the 2014 Standards for Educational and Psychological Testing.[15] This study was deemed exempt by the Institutional Review Board of Virginia Commonwealth University School of Medicine.

Content validity was based on the use of an existing instrument and the opinion of our expert author group. We made one word change to appropriately reflect the content of EM didactics to two items on the instrument to be more general and applicable to any topic/lecture as the Leppink instrument specifically addressed the topic of statistics. It contains 3 subscales- intrinsic load (items 1,2,3), extrinsic load (items 4,5,6), and germane load (items 7,8,9,10). The response options are scaled (0 meaning not at all the case and 10 meaning completely the case). We also included a question regarding the overall quality of the lecture with ratings of Poor, Fair, Good, Excellent, Outstanding.

Response process validity evidence was collected by using the original scale and items with previously published validity evidence. Further, the instrument was piloted and read aloud amongst the author group to ensure clarity and agreement of instrument items among the author group.

Piloting Instrument and Study Protocol:

Once the steps were completed to confirm the content and response process of the instrument, we initiated a pilot study to collect further validity evidence. The study setting and participants for the pilot were four Accreditation Council for Graduate Medical Education (ACGME) accredited emergency medicine residency programs. Study participants were emergency medicine residents, post-graduate years one through four.

An EM faculty member who is not part of the author group delivered a lecture virtually via an online platform to four residency programs on two separate dates. The lecture topic was chosen by the guest speaker and focused on local “home remedies” that are seen in the emergency department. Immediately following the lecture, we invited residents in attendance to complete an online survey consisting of the cognitive load instrument. Additional information regarding how to fill out the survey was not provided other than the link to the survey. The sample population was a convenience sample of residents participating in educational resident conference for ease of obtaining initial pilot data for the purpose of this study. Study data was collected and managed using REDCap electronic data capture tools hosted at Virginia Commonwealth University.[21,22] REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources.[21,22]

Data Analysis:

We calculated and reported descriptive statistics. *Internal Structure validity evidence* was analyzed with Cronbach's alpha and confirmatory factor analysis using the three-factor structure of Leppink.[16] Confirmatory factor analysis allows the testing of a priori models of latent constructs. The purpose of this analysis is to determine whether the subscales suggested by Leppink are reproducible among medical trainees. *Evidence of relationship to other variables validity* was determined through Pearson's correlation to compare cognitive load scores to overall lecture ratings by residents.

Results

A total of 124 residents participated in the virtual lecture conference, of these, a total of 54 residents participated in the study with completion of the instrument. Characteristics of participants are shown in Table 1. Mean scores for each item of the cognitive load instrument are displayed in Table 2. Evidence for *internal structure* included Cronbach's alpha (α) was 0.78 indicating good agreement. Subscales also performed well including intrinsic load (α = 0.96, excellent agreement), extrinsic load (α = 0.87, very good agreement), and germane load (α = 0.94, excellent agreement). In addition, a confirmatory factor analysis was performed to determine the fit of each of the subscales. Intrinsic load and germane load had good fit with root mean square error of approximation (RMSEA) below .05, comparative fit index (CFI), and Tucker–Lewis index (TLI) above .95, and standardized root mean squared Error (SRMR) below .08. However, extrinsic load showed a poor fit using all criteria.

*Evidence for relationship to other variable*s. Seven of the items were correlated with overall quality of lecture including: item 2 ($r = .293$, $p = 0.034$), item 5 ($r = -.392$, $p = 0.004$), item 6 ($r = -0.405$, $p = 0.003$), item 7 ($r = .418$, $p = 0.002$), item 8 ($r = .547$, $p < 0.001$), item 9 ($r = 0.619$, $p < 0.001$), item 10 ($r = 0.665$, $p < 0.001$) (Table 3).

Discussion

Instructors with a robust understanding of cognitive load theory can optimize various components during didactic sessions to enhance learning outcomes. This study provides initial validity evidence for an instrument that assesses cognitive load during virtual didactics. Such a tool may allow lecturers to evaluate the impact of different educational strategies on the cognitive load of their learners. The Cronbach's alpha overall indicated good agreement for internal structure and subscales performed well, although the fit demonstrated by confirmatory factor analysis varied by the type of cognitive load examined.

Intrinsic load, or the inherent difficulty in understanding a given topic, can be controlled in a presentation by building on prior knowledge of learners and sequencing material in natural order.[11,12] During the lecture being evaluated, concepts were presented in this fashion. The questions in the instrument intended to assess intrinsic load included #1-3, and specifically commented on the complexity of the topics, formulas, concepts, and definitions covered. It is logical then that responses to these questions using the assessment tool demonstrated high internal consistency, and confirmatory factor analyses demonstrated a good fit.

Extrinsic cognitive load, minimized by decreasing distractions and focusing on optimizing the learning environment, demonstrated the lowest internal consistency and had the weakest validity evidence in our virtual didactic presentation. Reviewing the specific wording of questions #4-6, which aimed to assess extrinsic load specifically, may illuminate this finding. Ambiguity over the meaning of the terms "instructions" or "explanations" may have negatively impacted internal consistency. Additionally, all three questions are negative statements, in contrast to the other statements, which read in a complimentary fashion. Due to social desirability bias, raters may be less likely to agree with negative statements. Additionally, external distractions, either within the environment or within the delivery of the lecture, can significantly impact extrinsic load and this data was not captured as part of the study.

Germane load can be minimized by organizing materials in meaningful groupings to aid in the formation of long-term memories. Deliberate organization of the material in the study presentation attempted to help learners organize concepts into meaningful and natural associations. Questions #7-10 in this instrument intended to measure germane load. These questions referenced the lecture's enhancement of the learner's understanding of the topic covered, the data related to the topic, and of concepts and definitions covered. Our results demonstrated high internal consistency regarding measurements of germane load.

Our study has several limitations. We applied our cognitive load instrument to a single lecture, which was rated to be an overall high-quality lecture, without a poorer quality lecture for comparison. Some of the residents evaluating the lecture also know the faculty speaker on a personal level, which may bias evaluation of the lecture. Not all residents present completed the instrument which may have created response bias. Although this was a multi-institutional study, multi-institutional, our results may have been limited by the small sample size and regional variation which may have impacted our data. Applying this tool to multiple lectures may help to draw additional conclusions relating to the overall use of this instrument as an assessment tool. Although there is low level evidence regarding the quality of lecture and its association with overall cognitive load, this is an opportunity for future work and additional research.

Next steps include determination of consequential validity by applying the tool during a variety of lectures of varying quality to determine if it can differentiate a high- versus low-quality lecture. In addition, we intend to apply a Delphi method of education experts within EM to optimize the tool for the emergency medicine virtual learning environment. Once adapted to this educational context, the tool has potential to become a key component of speaker evaluation forms. We also aim to investigate whether the tool can be utilized to evaluate cognitive load optimization strategies previously described,[11] and if use of this instrument to provide feedback to speakers improves the quality of future lectures.

Conclusion

A novel cognitive load assessment tool utilized during a virtual emergency medicine didactic demonstrated evidence of internal validity for intrinsic and germane loads, with poorer internal consistency for extrinsic load. Use of this instrument may provide important feedback to guide instructors of virtual didactic activities to maximize learning.

References:

1. Woolliscroft JO. Innovation in Response to the COVID-19 Pandemic Crisis. Acad Med [Internet] 2020 [cited 2021 Apr 27];Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7188042/
2. Hickam G, Santen SA, Cico SJ, et al. Rapid Adaptation to Remote Didactics and Learning in GME. AEM Educ Train [Internet] [cited 2021 Apr 27];n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/aet2.10528

3. Daniel M, Gordon M, Patricio M, et al. An update on developments in medical education in response to the COVID-19 pandemic: A BEME scoping review: BEME Guide No. 64. Med Teach 2021;43(3):253–71.
4. Gordon M, Patricio M, Horne L, et al. Developments in medical education in response to the COVID-19 pandemic: A rapid BEME systematic review: BEME Guide No. 63. Med Teach 2020;42(11):1202–15.
5. Sandars J, Correia R, Dankbaar M, et al. Twelve tips for rapidly migrating to online learning during the COVID-19 pandemic. MedEdPublish [Internet] 2020 [cited 2021 Apr 27];9. Available from: https://www.mededpublish.org/manuscripts/3068
6. Huynh R. The Role of E-Learning in Medical Education. Acad Med J Assoc Am Med Coll 2017;92(4):430.
7. Iwai Y. Online Learning during the COVID-19 Pandemic [Internet]. Sci. Am. Blog Netw. [cited 2021 Apr 27];Available from: https://blogs.scientificamerican.com/observations/online-learning-during-the-covid-19-pandemic/
8. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008;300(10):1181–96.
9. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. Med Teach 2014;36(5):371–84.
10. Venkat MV, O'Sullivan PS, Young JQ, Sewell JL. Using Cognitive Load Theory to Improve Teaching in the Clinical Workplace. MedEdPORTAL J Teach Learn Resour [Internet] [cited 2021 Apr 27];16. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7549387/
11. Jordan J, Wagner J, Manthey D, Wolff M, Santen S, Cico S. Optimizing Lectures From a Cognitive Load Perspective. AEM Educ Train 2019;4(3):306–12.
12. Chandler P, Sweller J. Cognitive Load Theory and the Format of Instruction. Cogn Instr 1991;8(4):293–332.
13. Klepsch M, Schmitz F, Seufert T. Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. Front Psychol 2017;8:1997.
14. Messick S. Validity. In: Linn R, editor. Education Measurement. New York: Macmillan; 1989. p. 13–103.
15. American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing. The Standards for Educational and Psychological

Testing [Internet]. 2014 [cited 2021 Apr 28];Available from: https://www.apa.org/science/programs/testing/standards

16. Leppink J, Paas F, Van der Vleuten CPM, Van Gog T, Van Merriënboer JJG. Development of an instrument for measuring different types of cognitive load. Behav Res Methods 2013;45(4):1058–72.
17. Paas FGWC. Training strategies for attaining transfer of problem solving skills in statistics: a cognitive-load approach. J Educ Psychol 1992;84(4):429–34.
18. Sweller J. Measuring cognitive load. Perspect Med Educ 2018;7(1):1–2.
19. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research [Internet]. In: Hancock PA, Meshkati N, editors. Advances in Psychology. North-Holland; 1988 [cited 2021 Apr 28]. p. 139–83.Available from: https://www.sciencedirect.com/science/article/pii/S0166411508623869
20. Szulewski A, Gegenfurtner A, Howes DW, Sivilotti MLA, van Merriënboer JJG. Measuring physician cognitive load: validity evidence for a physiologic and a psychometric tool. Adv Health Sci Educ 2017;22(4):951–68
21. PA Harris, R Taylor, R Thielke, J Payne, N Gonzalez, JG. Conde, Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support, J Biomed Inform. 2009 Apr;42(2):377-81.
22. PA Harris, R Taylor, BL Minor, V Elliott, M Fernandez, L O'Neal, L McLeod, G Delacqua, F Delacqua, J Kirby, SN Duda, REDCap Consortium, The REDCap consortium: Building an international community of software partners, J Biomed Inform. 2019 May 9 [doi: 10.1016/j.jbi.2019.103208]

Table 1. Characteristics of Participants

Demographics	
PGY-1	N= 16
PGY-2	N=14
PGY-3	N=13
PGY-4	N=11
Total Sample Size	N= 54
Participating Residency Programs	
WASHU	19
VCU	8
UMich	16
Wake Forest	11
Total	54

Table 2. Mean Item Scores for Leppink Instrument

	QS1	QS2	QS3	QS4	QS5	QS6	QS7	QS8	QS9	QS10
Mean	3.5	3.1	3.0	1.4	1.5	0.8	6.7	6.8	6.9	6.8
S.Dev	2.23	2.2	2.2	2.3	2.5	1.6	2.4	2.2	2.2	2.1

Table 3.Correlations with Each Question and Quality of Lecture

	QS1	QS2	QS3	QS4	QS5	QS6	QS7	QS8	QS9	QS10
Pearson Correlation	.237	.293*	.201	-.186	-.392	-.405	.418	.547	.619	.665
Sig. (2-tailed)	.087	.034	.149	.183	.004	.003	.002	.000	.000	.000