Young John Q. (Orcid ID: 0000-0003-2219-5657)

Features of the Learner, Task and Instructional Environment that Predict Cognitive Load Types during Patient Handoffs: Implications for Instruction

Predictors of Cognitive Load Types during Handoffs

John Q. Young, MD, MPP, PhD, Krima Thakker, Majnu John, MS, PhD, Karen Friedman, MD, Rebekah Sugarman, Justin L. Sewell, MD, MPH, PhD, Patricia S. O'Sullivan, EdD

Dr. Young is Professor and Vice Chair of Education, Department of Psychiatry, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell and Zucker Hillside Hospital at Northwell

Ms. Thakker is research coordinator, Department of Psychiatry, Zucker Hillside Hospital at Northwell Health

Dr. John is statistician, Division of Research, Zucker Hillside Hospital at Northwell Health

Dr. Friedman is Professor and Vice Chair of Medicine, Department of Medicine, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell

Ms. Sugarman is a medical student, University of Michigan School of Medicine

Dr. Sewell is Associate Professor, Department of Medicine, University of California, San Francisco School of Medicine

Dr. O'Sullivan is Professor, Department of Medicine, and Director of Research and Development in Medical Education, University of California, San Francisco School of Medicine

Conflicts of Interest: The authors have no conflicts of interest to report.
Data availability: Data available on request from the authors
Acknowledgments: None

Features of the Learner, Task and Instructional Environment that Predict Cognitive Load Types during Patient Handoffs: Implications for Instruction

Predictors of Cognitive Load Types during Handoffs

John Q. Young, MD, MPP, PhD, Krima Thakker, Majnu John, MS, PhD, Karen Friedman, MD, Rebekah Sugarman, Justin L. Sewell, MD, MPH, PhD, Patricia S. O'Sullivan, EdD

Dr. Young is Professor and Vice Chair of Education, Department of Psychiatry, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell and Zucker Hillside Hospital at Northwell

Ms. Thakker is research coordinator, Department of Psychiatry, Zucker Hillside Hospital at Northwell Health

Dr. John is statistician, Division of Research, Zucker Hillside Hospital at Northwell Health

Dr. Friedman is Professor and Vice Chair of Medicine, Department of Medicine, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell

Ms. Sugarman is a medical student, University of Michigan School of Medicine

Dr. Sewell is Associate Professor, Department of Medicine, University of California, San Francisco School of Medicine

Dr. O'Sullivan is Professor, Department of Medicine, and Director of Research and Development in Medical Education, University of California, San Francisco School of Medicine

Conflicts of Interest: The authors have no conflicts of interest to report.
Data availability: Data available on request from the authors
Acknowledgments: None

ABSTRACT

We used the Cognitive Load Inventory for Handoffs (CLIH) to identify predictors of cognitive load types during patient handoffs in order to identify opportunities to improve instruction. In 2019, out of a total of 1,807 residents and fellows within a 24-hospital health system, 693 (38.4%) completed the CLIH after a patient handoff. Multivariable regression yielded predictors for each cognitive load type. Intrinsic load associated with features of the learner (fatigue positively associated) and task (higher complexity clinical setting, number of patients, and handoff length positively associated). Extraneous load associated with learner (fatigue positively associated, and number of times trained in the verbal protocol negatively associated) and task design (number of sources of written information positively associated). Germane load associated with learner (level of training negatively associated, and fatigue positively associated) and instructional environment (interruptions negatively associated and formal feedback positively associated). Implications for instructional design are explored.

KEYWORDS: cognitive load theory; cognitive load types; patient handoffs; medical education; instructional design; patient safety

INTRODUCTION

Patient handoffs occur when the responsibility for the care of a patient or panel of patients is transferred from one clinician or team to another. Patient handoffs occur frequently and are vulnerable to communication errors that can lead to patient harm.[1] As a result, performing patient handoffs has been identified across the medical education continuum as an essential competency.[2-5] Despite the increased emphasis in both undergraduate and graduate medical education, patient handoffs remain a significant patient safety challenge, even in those studies reporting improvements.[6] Cognitive Load Theory has helped unpack the complexity of handoffs.[7] Originally developed by John Sweller[8], Cognitive Load Theory (CLT) focuses on the implications of limited working memory (WM) for learning.[9] While sensory and long-term memory have relatively infinite capacity, WM is highly constrained. In fact, WM can only actively process (i.e. organize, compare and contrast) two to four elements at any given moment.[10,11] When the cognitive load of a learning tasks such as a handoff exceeds the working memory capacity of the trainee, learning and performance suffers.

CLT envisions at least two, and possibly three, types of cognitive load (CL).[12] *Intrinsic load* (IL) arises from the information processing demands associated with performing the task itself. Both task complexity and learner expertise determine the intrinsic load imposed by a handoff. More expert learners will experience less IL for any given task; similarly, higher task complexity will impose higher IL. *Extraneous load* (EL) occurs when learners use working memory resources to process information not essential to the task, most commonly originating from the design of the task (e.g., multiple sources of information forcing one to toggle between screens), external environment (e.g., noise or interruptions), or internal environment (e.g., excessive preoccupation with an unrelated matter).[13-15] *Germane load* (GL) is imposed when learners use cognitive strategies (e.g., assess one's understanding, self-explanation) to refine existing schemata and enhance storage in long-term memory.[16] Recent work by Sweller and others has suggested that germane load may best be understood as a component of intrinsic load rather than a separate type of load.[17-19] Yet, some empirical work, especially within medical education, has found evidence for germane load as a separate type.[20-22] The relationship of germane and intrinsic loads and the arguments for two- versus three-factor models are active areas of theoretical and empirical research.

Prior theoretical research explored how characeristics of the learner, task, and instructional environment might modulate each cognitive load type during a patient handoff.[7] In this study, we examine these relationships empirically. We identify features of the learner, task and instructional environment that predict IL, EL, and GL to identify strategies to improve future handoff instructional design and protocol development.

METHOD

Study design

We performed a cross-sectional survey study to identify features of the learning and clinical environment that predict each CL type. We collected data for the current study contemporaneously with data for another study.[XXX] Both studies were planned a priori as separate efforts with different aims. The prior study focused on the instrument development process and developing evidence for validity for the Cognitive Load Inventory for Handoffs (CLIH) with respect to content, response process, and internal structure. This study focused on associations of learner, task, and environment features with each cognitive load type in order to identify implications for handoff instruction and protocol development. We modelled this approach after a series of studies examining cognitive load during procedural learning.[20,23]

The Institutional Review Board for XXX reviewed and deemed the study protocol exempt status as an educational research study with minimal risk.

Procedures

As described elsewhere, we prospectively enrolled a sample of 1,823 residents and fellows from a large, 24-hospital health system in XXX that sponsors 122 distinct nationally accredited residency and fellowship programs.[22] Email addresses for all residents and fellows were obtained from the health system's Office of Academic Affairs. Between January and March, 2019, each trainee received an email invitation from three study authors (JQY, RS, KF) with a link to the electronic survey hosted by RED-Cap, an academic software program that supports research surveys.[24] We asked participants to complete the survey after a handoff. Non-respondents received weekly emails over seven weeks in order to increase response rate.[25] Invitees could participate only once and could enter a drawing for one of four $250 gift cards.[26] Data collected included a measure of cognitive load types, demographic information, and features of the clinical and learning encounter, each of which is described below. Of the 1,823 trainees invited to participate, 16 had undeliverable email addresses, resulting in a pool of 1,807 potential participants. As previously reported, we received 693 responses (38.4%), representing all training programs in the health system.

Measures

Cognitive load measures. We used the CLIH to estimate the IL, EL, and GL that each trainee experienced during the patient handoff.[22] A prior study collected evidence for validity for the scores generated by the CLIH. Factor analyses supported a three factor rather than two factor model. The three factors had high internal consistently and associated with the items for IL, EL, and GL, respectively. Model fit parameters were strong and the scores for IL, EL, and GL associated, as predicted, with level of training and clinical setting.[22]

The CLIH includes 16 items (11 point scale, strongly agree to strongly disagree): 5 for IL, 7 for EL, and 4 for GL. To calculate the IL, EL, and GL experienced by each individual, items within each cognitive load type were averaged together to generate a score. Table 1 depicts the items used to measure each type of CL.

Demographic characteristics. Demographic data included trainee gender; specialty; specialty, hospital and clinical setting in which the handoff occurred; reason for the handoff; and role in the handoff. We had no theoretical reason to expect these data to influence cognitive load and we did not include them in our analyses.

Predictor variables. While prior research has examined the impact of the learning and clinical environment on handoff accuracy[27], previous explorations of the predictors of cognitive load types during a handoff have been theoretical.[7] Based on the findings from theoretical work and input from 9 international experts in cognitive load theory (5) and handoffs (4), we developed questions to measure our predictor variables hypothesized to influence each type of CL. For IL, questions focused on factors related to the two main drivers as conceived by CLT: learner knowledge (e.g., level of training and fatigue – which effects learner cognition) and the task's complexity or number of information elements (e.g., clinical setting, length of handoff, number of patients, and number of clinicians).

For EL, questions related to either the learner (e.g., level of training, fatigue, number of times trained in the protocol), task design (whether the verbal and written communication were standardized, and the number of sources of written information), and the environment (e.g., interruptions and noise and their impact on concentration). Standardization of verbal and written handoff procedures is thought to reduce EL by allowing both the sender and receiver to know in what order and format each type of information is to be communicated.[7]

For GL, questions addressed the learner (e.g., level of training and fatigue) and the instructional environment (were receiver's able to take notes on the written component of the sign-out, did the learner receive formal feedback from a senior, fellow or faculty member, and to what extent did interruptions impact concentration). Per CLT, feedback should facilitate schema refinement and interruptions should impair concentration, a key mediator of GL.

The study authors iteratively reviewed and refined these questions and then incorporated them into the survey. Table 2 describes, for each CL type, the learner, task, and instructional environment variables that were identified as relevant and then, for each variable, the measure (i.e., survey question), data type, and the response options.

Analysis

Analysis was conducted in RStudio (version 1.2.1335, build 1379). We calculated descriptive statistics using appropriate measures of central tendency and dispersion. Based on our knowledge of CLT, we used a consensus process amongst the authors to predict a priori whether each variable should have a positive or negative association with the relevant CL type.

We then developed three multivariable linear regression models – one each with IL, EL, and GL as the outcome variables, respectively. The predictor variables identified as relevant by CLT were entered into the models. Following univariate analyses, only variables with a p-value of less than 0.1 were included in the multivariable regression analysis. Our data included 3 ordinal predictor variables, which were dichotomized. Level of training, which included options ranging from PGY-1 to PGY-8 or higher, was dichotomized into PGY-1 and all others given the knowledge gap between PGY-1s and others. Clinical setting was dichotomized into ICU versus all other settings due to patient complexity between the ICU and other settings. Fatigue was dichotomized into rested and fatigued. Because the CLIH instrument itself includes items on noise and interruptions, these items related to environment were excluded from our analysis.

RESULTS

The sample characteristics were described previously.[xxx] Most learners were either PGY-1, PGY-2, or PGY-3 level. Approximately 60% reported being either very or somewhat fatigued at the time of the handoff. Table 3 shows the descriptive statistics for the predictor variables and IL, EL and GL. Table 4 indicates the hypothesized relationship and the outcomes of the univariate and multivariable analyses.

Intrinsic Load

The handoffs occurred in a variety of clinical settings, including the ICU (about 13%) and non-ICU inpatient (67%). The average length of a handoff was nearly 18 minutes (SD=15.3) and included 10 patients (SD=10.6) and nearly 2.5 other clinicians (SD=4.8). (Table 3) The multivariable linear regression model for IL indicated that ICU as a clinical setting was associated with significantly higher IL compared to other settings. In addition, as fatigue, number of patients, and length of the handoff increased so did IL. These findings were consistent with a priori predictions. However, contrary to our hypotheses, level of training did not influence IL. (Table 4)

Extraneous Load

Respondents indicated that the handoff followed a standardized verbal communication protocol only 27% of the time and, in those situations, trainees reported on average about 4 prior trainings in that specific protocol. Sixty two percent of the handoffs had a written component with an average of 2.3 (SD=1.03) different documents. Of those with a written component, 79% had a standardized template. Participants were interrupted on average 1.6 times (SD=2.25) and about 56% of the learners reported negative impact of those interruptions on their concentration. (Table 3) In the multivariable linear regression model, the number of prior trainings in the handoff protocol predicted lower EL while the number of written documents and fatigue predicted higher EL. These findings were as predicted. Several variables did not influence EL as predicted; namely, level of training and whether the verbal and written components of the handoff were standardized. (Table 4)

Germane Load

About 14 percent of the time, the respondents reported receiving formal feedback from a senior resident, fellow or a faculty member. When a respondent was a receiver, they had the ability to take written notes almost all of the time. (Table 3) 56% of the respondents indicated that the interruptions made it somewhat, very, or extremely difficulty to concentrate. In the multivariable regression models, higher level of training predicted lower GL while interruptions, fatigue and formal feedback predicted higher GL. (Table 4) These findings were all consistent with a-priori predictions. Surprisingly, the ability to take notes during the handoff did not influence GL.

DISCUSSION

In this study, we identified features of the learner, task, and instructional environments that predicted cognitive load types during patient handoffs. Higher IL was associated with fatigue, working in a higher complexity clinical setting, more patients, and longer handoffs. Higher EL was associated with fatigue and number of sources of written information and lower EL with training in the verbal protocol being used. Higher GL was associated with fatigue and formal feedback and lower GL with more advanced stage of training and interruptions.

While the above findings were consistent with a-priori predictions based on CLT, several findings were contrary to our predictions. CLT predicts that IL should decrease as the level of training increases. In our study, level of training was not associated with IL. There are several possible explanations. Interns may perform so many handoffs from the outset that they have acquired relative mastery by the mid-point of the first year, which is when the survey was administered. Alternatively, the finding may relate to how roles change with increased responsibilities associated with higher levers of training in medical education. This increase in information elements may account for the fact that IL was similar across years of training. If true, then in the setting of this study, the IL of the task seemed well matched to trainee experience. Also, while the evidence for the validity of the IL items is strong, it is possible that the construct of IL is not fully captured by the existing items on the CLIH and, as a result, the IL scores do not relate to level of training as they should and would with additional items added. Finally, it is important to consider that level of training may not the best measure of expertise, i.e., knowledge about the clinical medicine and handoff process may vary more within a given level of training than between. For example, a study of simulated handoffs found that learner knowledge as measured by illness script maturity was a better predictor of IL than level of training.[27] While harder to measure, future studies might consider using illness script maturity or some other more specific measure of knowledge.

Two additional findings were contrary to our expectations. First, the capacity to take notes did not increase GL for receivers. Note taking induces schema refinement and should increase GL. While all receivers had the capacity to take notes, it could be that most did not which would then explain the finding. In the authors' experience, whether receivers take notes is highly variable and depends on the modelled behavior (e.g., do seniors take notes?) and the architecture of the space (e.g., a learner may have the capacity to take notes but chooses not to if they are standing and not in a comfortable position). Second, while fatigue's association with higher IL and EL was expected and similar to what was found in a study of colonoscopy[23], we did not expect fatigue to predict higher GL. In CLT, fatigue is understood to reduce working memory capacity and slow information retrieval and encoding.[19] In essence, the learner becomes 'less expert' and the same task then imposes higher IL. If the IL (and EL) of the task exceeds the working memory capacity of the fatigued learner, then you would expect GL to decrease as working memory resources are re-allocated to manage the IL and EL. The positive association of fatigue with GL in our study may reflect that working memory capacity was not exceeded. This possibility is supported by the fact that mean IL was moderate and EL low. If so, then the same GL inducing strategies (e.g., compare and contrast, monitoring understanding) would require more effort in the 'slowed down' brain of the fatigued learner. Alternatively, this observed relationship between fatigue and GL may reflect the challenges of self-report. Finally, while the germane load items formed a factor separate from IL in prior research, it could be that this factor represents something other than germane load such as 'effort', as suggested by Leppink et al when their measure of cognitive load types during classroom learning yielded three factors.[17,28] If so, then increased fatigue would lead to increased effort on the same task and a higher score for this factor. Most importantly, it is clear that fatigue, even within an environment with duty hour restrictions, still needs to be actively managed to optimize learning.

Overall, these findings provide additional evidence supporting the validity of the CLIH. Prior research collected evidence supporting validity with respect to the CLIH's content, response process, and internal structure.[29] The results of this study provide evidence for validity with respect to how the scores associate with other variables. For the most part, the IL, EL, and GL scores varied as predicted with the learner, task, and instructional features. These findings lend added confidence that a measure such as the CLIH can measure cognitive load types during handoffs and can be used to deepen our understanding of the cognitive mechanisms of handoff errors and improve our instruction and protocols. Future validity research should explore how the CLIH scores associate with actual learning and clinical outcomes (e.g., information loss or distortion during the handoff).

These findings have important implications for handoff instruction. Faculty can titrate IL to the learner's expertise by modulating the number of cases in the handoff, the length of the handoff, and the clinical setting. Moreover, when this is not possible and the IL will exceed the learner's development stage, partial task completion and worked examples, both well-studied methods for managing IL, can be used.[12] In part-task approaches, the learner could perform only certain components of the handoff. In worked examples, the supervisor can take over the handoff (rather than permitting the learner to struggle unproductively) and then narrate how they are managing the complexity. To support these instructional elements, protocols could be developed that trigger the provision of more time for the handoff or more senior resident or attending support when the number of patients or complexity of patients exceeds a predetermined threshold.[30]

Prior research has argued that verbal protocols reduce EL by conserving working memory resources that otherwise would be consumed with anticipating when and in which order different types of information will be communicated.[7] It is noteworthy that standardization of the protocols did not influence EL when also including number of trainings in that protocol; familiarity with the protocol was more important than the mere existence of the protocol. Standardization is critical but additional gains can be achieved through training in the protocol. Similarly, EL increased when the trainee had to use more than one written document to obtain the essential clinical information. When working memory resources are expended on information search (e.g., looking back and forth between two or more documents), EL increases and less working memory is available for the core task. Decreasing the number of written sources is a key goal of standardizing the written process and can also be used to vary the EL for purposes of training.

Finally, the predictors of GL are instructive. Germane load relates to the use of working memory resources dedicated to cognitive activities that enhance schema refinement and automation during the performance of the task. Consistent with CLT, more experienced learners utilized less GL; as expertise increases, the need for schema refinement decreases. While the number of interruptions had no influence, interruptions that impacted concentration did reduce GL. This distinction is important; some interruptions are relatively benign while others are not. Meanwhile, the provision of feedback by a senior, fellow, or attending induced GL. While seniors and even attendings may participate in some handoffs, it is not common for trainees to be directly observed and provided structured feedback. In fact, only 14% of the learners in this study received feedback. Most curricula are likely heavy on practice and light on feedback. Given the large proportion of medical errors associated with handoffs, both UME and GME training programs should prioritize creating opportunities for structured observation and feedback.

Limitations

The study has several limitations. The study was performed in a single health system. However, this single health system is diverse and participants in the study came from multiple specialties and hospitals. Participants included residents and fellows, so we do not know if students or faculty would respond in the same way. In addition, the CLIH only measures learners' perceptions of IL, EL, and GL as recalled after completion of the handoff. These perceptions are undoubtedly biased by recall. However, asking learners to complete the CLIH during actual handoffs was not thought to be feasible. Future research could compare the post-hoc CLIH scores with in-the-moment physiologic methods, though, the physiologic methods to date cannot differentiate between cognitive load types. Finally, while the CLIH has three factors, it is important to acknowledge that given the current debate within the CLT research community, the third factor that we understand to be GL may in fact be something other than GL. Future research on the CLIH should collect additional evidence to better understand this third factor.

Conclusion

In summary, this study identifies features of the learner, task, and instructional environment that modulate cognitive load types. These findings and the approach taken in this study in general can help inform future improvements to handoff curricula and protocols.

REFERENCES

1. Horwitz LI, Moin T, Krumholz HM, Wang L, Bradley EH. Consequences of inadequate sign-out for patient care. *Arch Intern Med.* 2008;168(16):1755-1760.
2. Lomis K, Amiel JM, Ryan MS, et al. Implementing an Entrustable Professional Activities Framework in Undergraduate Medical Education: Early Lessons From the AAMC Core Entrustable Professional Activities for Entering Residency Pilot. *Acad Med.* 2017;92(6):765-770.
3. Lane-Fall MB, Davis JJ, Clapp JT, Myers JS, Riesenberg LA. What Every Graduating Resident Needs to Know About Quality Improvement and Patient Safety: A Content Analysis of 26 Sets of ACGME Milestones. *Acad Med.* 2018;93(6):904-910.
4. Caverzagie KJ, Cooney TG, Hemmer PA, Berkowitz L. The development of entrustable professional activities for internal medicine residency training: a report from the Education Redesign Committee of the Alliance for Academic Internal Medicine. *Acad Med.* 2015;90(4):479-484.
5. Young JQ, Hasser C, Hung EK, et al. Developing End-of-Training Entrustable Professional Activities for Psychiatry: Results and Methodological Lessons. *Acad Med.* 2018;93(7):1048-1054.
6. Starmer AJ, Spector ND, Srivastava R, et al. Changes in medical errors after implementation of a handoff program. *N Engl J Med.* 2014;371(19):1803-1812.
7. Young JQ, Ten Cate O, O'Sullivan PS, Irby DM. Unpacking the Complexity of Patient Handoffs Through the Lens of Cognitive Load Theory. *Teach Learn Med.* 2016;28(1):88-96.
8. Sweller J. Cognitive load during problem solving: Effects on learning. *Cogn Sci.* 1988;12(2):257-285.
9. Sweller J, van Merrienboer JJG. Cognitive Load Theory and Instructional Design for Medical Education. In: Walsh K, ed. *The Oxford Textbook of Medical Education.* Oxford, UK: Oxford University Press; 2013:74-85.
10. Baddeley A. Working memory: theories, models, and controversies. *Annu Rev Psychol.* 2012;63:1-29.
11. Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci.* 2001;24(1):87-114.
12. 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):371-384.
13. Young JQ, Sewell JL. Applying cognitive load theory to medical education: construct and measurement challenges. *Perspectives on medical education.* 2015;4(3):107-109.
14. Feldon DF. Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educ Psych.* 2007;42(3):123-137.
15. Choi H-H, van Merriënboer JJG, Paas F. Effects of the physical environment on cognitive load and learning: Towards a new model of cognitive load. *Educ Psychol Rev.* 2014;26(2):225-244.
16. Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive architecture and instructional design. *Educ Psychol Rev.* 1998;10(3):251-296.
17. Leppink J, Paas F, van Gog T, van der Vleuten CPM, van Merrienboer JJG. Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction.* 2014;30:32-42.
18. Sweller J, Ayres PL, Kalyuga S. *Cognitive load theory.* New York: Springer; 2011.
19. Sweller J, van Merriënboer JJG, Paas F. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ Psychol Rev.* 2019;31(2):261-292.
20. Sewell JL, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. Measuring cognitive load during procedural skills training with colonoscopy as an exemplar. *Med Educ.* 2016;50(6):682-692.

21. Young JQ, Irby DM, Barilla-LaBarca ML, Ten Cate O, O'Sullivan PS. Measuring cognitive load: mixed results from a handover simulation for medical students. *Perspectives on medical education.* 2016;5(1):24-32.
22. Young JQ, John M, Thakker K, et al. Evidence for Validity for the Cognitive Load Inventory for Handoffs. *Med Educ.* 2020.
23. Sewell JL, Boscardin CK, Young JQ, Ten Cate O, O'Sullivan PS. Learner, Patient, and Supervisor Features Are Associated With Different Types of Cognitive Load During Procedural Skills Training: Implications for Teaching and Instructional Design. *Acad Med.* 2017;92(11):1622-1631.
24. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics.* 2009;42(2):377-381.
25. Dillman DA, Smyth JD, Christian LM. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method.* Wiley Publishing; 2014.
26. Stovel RG, Ginsburg S, Stroud L, Cavalcanti RB, Devine LA. Incentives for recruiting trainee participants in medical education research. *Med Teach.* 2018;40(2):181-187.
27. Young JQ, van Dijk SM, O'Sullivan PS, Custers EJ, Irby DM, Ten Cate O. Influence of learner knowledge and case complexity on handover accuracy and cognitive load: results from a simulation study. *Med Educ.* 2016;50(9):969-978.
28. Leppink J, Paas F, Van der Vleuten CP, Van Gog T, Van Merrienboer JJ. Development of an instrument for measuring different types of cognitive load. *Behav Res Methods.* 2013;45(4):1058-1072.
29. Young JQ, Thakker K, John M, Friedman K, Sugarman R, Sewell JL. Evidence for Validity for the Cognitive Load Inventory for Handoffs. *Med Educ.* In Press.
30. Young JQ, Wachter RM, Ten Cate O, O'Sullivan PS, Irby DM. Advancing the next generation of handover research and practice with cognitive load theory. *BMJ quality & safety.* 2016;25(2):66-70.

Table 1: Items for the Cognitive Load Inventory for Handoffs[a,b]
INTRINSIC LOAD: Please rate your agreement with the following statements regarding the handoff you have completed:
1. The patient problems were complex
2. The handoff included significant clinical decision(s) that needed to be made
3. The handoff included significant diagnostic and/or treatment uncertainty
4. I had to consider multiple or complex interactions between diseases
5. I had to consider multiple or complex interactions between treatments
EXTRANEOUS LOAD: Please rate your agreement with the following statements regarding the handoff. These statements are about the environment and your mindset during the handoff:
6. The other clinician used jargon out of context
7. I was distracted by the other clinician's attitude
8. I was self-conscious due to who was present
9. I was frequently interrupted (e.g., pages, phone calls, people, etc...)
10. Noise made it difficult to concentrate
11. During the handoff, important information was not easily available when I needed it
12. I was thinking about things unrelated to the sign-out
GERMANE LOAD: Please rate your agreement with the following statements regarding your mental effort during the handoff you have completed:
13. I had to work hard to connect my own medical knowledge to the patient problems
14. I had to work hard to organize the patient information into a coherent clinical picture
15. During the sign-out, I had to work hard to concentrate on how well I understood the information
16. I had to take steps to clarify points of confusion
a. Participants were asked to rate their level of agreement with each item using a 0 ('strongly disagree' to 10 ('strongly agree') scale. *b. Young JQ, John M, Thakker K, et al. Evidence for Validity for the Cognitive Load Inventory for Handoffs. Med Educ. 2020.*

Table 2: Measurement of Handoff Features Predicted to be Associated with Cognitive Load Types				
Cognitive Load Construct	**Variable**	**Data Type**	**Survey Question**	**Response Options**
Intrinsic Load				
Learner Related	1. Level of training[a]	Dichotomous (R1 v all other levels of training)	Indicate your year of training	PGY-1, PGY-2, PGY-3, PGY-4, PGY-5, PGY-6, PGY-7, PGY-8 or higher
	2. Fatigue[a]	Dichotomous (fatigued v rested)	Rate your level of fatigue/rest	Very fatigued, Somewhat fatigued, Somewhat rested, Very rested
Task Related	1. Clinical Setting[a]	Dichotomous (ICU v all other settings)	Indicate the clinical setting in which the handoff occurred	Inpatient ICU, Inpatient Non-ICU, Emergency Department, Ambulatory, Peri-operative Setting, Other
	2. Length of handover	Continuous	Roughly how long did the handoff take (in minutes)	Less than 5, 5 to 10, 11 to 20, 21 to 30, 31 to 40, 41 to 50, 51 to 60, 61 to 70, 71 to 80, More than 80
	3. Number of patients	Continuous	Estimate the number of patients that you signed-out or received	1, 2, 3, …,48, 49, 50, more than 50
	4. Number of clinicians	Continuous	How many clinicians (trainees, attendings, nurses etc..) did you receive information from or send information to during the handoff	Text box to enter number
Extraneous Load				
Learner Related	1. Level of training	Dichotomous (R1 v all other levels of training)	Indicate your year of training	PGY-1, PGY-2, PGY-3, PGY-4, PGY-5, PGY-6, PGY-7, PGY-8 or higher
	2. Fatigue	Dichotomous (fatigued v rested)	Rate your level of fatigue/rest:	Very fatigued, Somewhat fatigued, Somewhat rested, Very rested
	3. Standardization of verbal communication	Dichotomous	Did the handoff follow a protocol (e.g., IPASS) for the presentation of verbal information?	Yes/No
Task Design	1. Standardization of verbal communication	Dichotomous	Did the handoff follow a protocol (e.g., IPASS) for the presentation of verbal information?	Yes/No

Table 2: Measurement of Handoff Features Predicted to be Associated with Cognitive Load Types				
Cognitive Load Construct	**Variable**	**Data Type**	**Survey Question**	**Response Options**
	2. Protocol includes written component	Dichotomous	Did the handoff include a written handoff document?	Yes/No
	3. Standardization of written communication	Dichotomous	Did the written document follow a standard template?	Yes/No
	4. Number of sources for written information	Continuous	How many different sources of written information did you have to use?	0, 1, …, 9, 10
Environment[b]	1. Number of times interrupted	Continuous	Roughly how many times were you interrupted during the handoff (e.g., page, person, etc…)	0, 1, …, 9, 10, more than 10
	2. Impact of interruptions on concentration	Continuous	To what extent did the interruptions make to concentrate?	Extremely difficult, Very difficult, Somewhat difficult, Not so difficult, Not at all difficult
	3. Impact of noise on concentration	Continuous	To what extent did noise make it difficult to concentrate:	Extremely difficult, Very difficult, Somewhat difficult, Not so difficult, Not at all difficult
Germane Load				
Learner Related	1. Level of training	Dichotomous (R1 v all other levels of training)	Indicate your year of training	PGY-1, PGY-2, PGY-3, PGY-4, PGY-5, PGY-6, PGY-7, PGY-8 or higher
	2. Fatigue	Dichotomous (fatigued v rested)	Rate your level of fatigue/rest:	Very fatigued, Somewhat fatigued, Somewhat rested, Very rested
Environment	1. If receiver, were you able to take notes	Dichotomous	If you were a receiver, were you able to take notes?	Yes/No
	2. Impact of interruptions on concentration	Continuous	To what extend did the interruptions make it difficult to concentrate	Extremely difficult, Very difficult, Somewhat difficult, Not so difficult, Not at all difficult
	3. Formal feedback	Dichotomous	Did you receive formal feedback from a senior resident, fellow, or faculty[c]	Yes/No

Table 2: Measurement of Handoff Features Predicted to be Associated with Cognitive Load Types				
Cognitive Load Construct	**Variable**	**Data Type**	**Survey Question**	**Response Options**

a. Our data included 3 ordinal predictor variables, which were dichotomized. Level of training, which included options ranging from PGY-1 to PGY-8 or higher, was dichotomized into PGY-1 and all others because the authors believed that the difference in knowledge between PGY-1s and others was much greater than the difference between any subsequent level of training. Similarly, clinical setting was dichotomized into ICU versus all other settings because the difference in patient complexity between the ICU and other settings was thought to be far greater than any difference between the other settings. Fatigue was dichotomized because the option set included only four options making treatment as a continuous variable problematic.

b. For extraneous load, the Cognitive Load Inventory for Handoffs asks about frequency of interruptions and impact of noise on concentration. Therefore, all of the items pertaining to the environment were not included in this analysis since the CLIH itself already incorporates.

c. Our survey asked separate questions about whether the respondents received formal feedback from faculty and then from a senior resident or fellow. For our analysis, we combined these two items to make a single item for receiving formal feedback from any of the three.

Table 3. Learner, Task, and Environmental Features Among Trainees Performing Handoffs during the 2018-2019 Academic Year	
Characteristic	**Measure**
All Models	
Number of trainees, no.	693
Year in training, no. (%)	
PGY-1	215 (31.02%)
PGY-2	180 (25.97%)
PGY-3	144 (20.78%)
PGY-4 Residents	59 (8.51%)
PGY-4 Fellows	20 (2.89%)
PGY-5 and higher	74 (10.68%)
Missing	1 (0.14%)
Fatigue, No (%)	
Very fatigued	71 (10.25%)
Somewhat fatigued	344 (49.64%)
Somewhat rested	191 (27.56%)
Very rested	77 (11.11%)
Missing	10 (1.44%)
Intrinsic Load Model	
Intrinsic Load, mean (SD)	4.72 (2.06)
Length of the handoff (minutes), mean (SD)	17.9 minutes (15.3)
Number of patients, mean (SD)	10.25 (10.55)
Number of clinicians, mean (SD)	2.46 (4.83)
Extraneous Load Model	
Extraneous Load, mean (SD)	2.55 (1.79)
Verbal protocol standardized, no. (%)	187 (26.98%)
Number times trained in verbal protocol, mean (SD)	3.94 (2.37)
Handoff includes written component, no. (%)	429 (61.90%)
Written component standardized, no. (%)	340 (79.25%)
Number of different written documents, mean (SD)	2.31 (1.03)
Number of times interrupted, mean (SD)	1.65 (2.25)
Extent to which interruptions made it difficult to concentrate[b], no. (%)	See GL Model below
Impact of noise on concentration, no. (%)	
Not at all difficult	187 (26.98%)
Not so difficult	251 (36.22%)
Somewhat difficult	175 (25.25%)
Very difficult	49 (7.07%)
Extremely difficult	22 (3.17%)
Missing	9 (1.30%)
Germane Load Model	
Germane Load, mean (sd)	3.40 (2.25)
If receiver, ability to take notes, no. (%)	111 (88.80%)
Extent to which interruptions made it difficult to concentrate[b], no. (%)	
Not at all difficult	21 (4.90%)
Not so difficult	154 (35.90%)

Table 3. Learner, Task, and Environmental Features Among Trainees Performing Handoffs during the 2018-2019 Academic Year	
Characteristic	**Measure**
Somewhat difficult	172 (40.09%)
Very difficult	50 (11.66%)
Extremely difficult	21 (4.90%)
Missing	11 (2.56%)
Received formal feedback from a senior resident or fellow, no. (%)	72 (10.39%)
Received formal feedback from a faculty member, no. (%)	22 (3.17%)
Received formal feedback from a senior resident, fellow, or faculty member[a], no. (%)	79 (11.40%)

d. Our survey asked separate questions about whether the respondents received formal feedback from faculty and then from a senior resident or fellow. For our analysis, we combined these two items to make a single item for receiving formal feedback from any of the three.
e. This question was asked only of those who indicated they experienced interruptions (429).

Table 4: Measurement of Handoff Features Predicted to be Associated with Cognitive Load Types

Cognitive Load Construct	Feature (directionality)	Predicted pattern of association	Univariate P value	Coefficient (95% CI)	ß Coefficient	P value
Intrinsic Load Model						
Learner	Level of training (increasing; 0=R1, 1=all others)	↓	0.496	-	-	-
	Fatigue (increasing; 0=rested, 1=fatigued)	↑	0.0002	0.46 (0.15, 0.77)	0.11	0.004
Task	Clinical Setting (increasing complexity; 0=all other, 1=ICU)	↑	<0.0001	1.03 (0.57, 1.49)	0.17	<0.0001
	Length of handover (increasing)	↑	<0.0001	0.18 (0.08, 0.28)	0.15	0.0002
	Number of patients (increasing)	↑	<0.0001	0.02 (0.0, 0.04)	0.11	0.006
	Number of clinicians (increasing)	↑	0.03	0.02 (-0.01, 0.05)	0.04	0.35
Extraneous Load						
Learner	Level of training (increasing; 0=R1, 1=all others)	↓	0.01	0.43 (-0.29, 1.15)	0.10	0.25
	Number of times trained in the protocol (increasing)	↓	0.002	-0.21 (-0.36, -0.06)	-0.25	0.006
	Fatigue (increasing; 0=rested; 1=fatigued)	↑	<0.0001	0.94 (0.28, 1.6)	0.24	0.006
Task Design[a]	Standardization of verbal communication (0=No, 1=Yes)	↓	0.29	-	-	-
	Protocol includes written component (0=No, 1=Yes)	↓	0.58	-	-	-
	Standardization of written communication (0=No, 1=Yes)	↓	0.36	-	-	-
	Number of sources for written information (increasing)	↑	0.02	0.97 (0.18, 1.76)	0.21	0.02
Germane Load						
Learner	Level of training (increasing; 0=R1, 1=all others)	↓	<0.0001	-0.88 (-1.36, -0.40)	-0.18	0.0003
	Fatigue (increasing; 0=rested, 1=fatigued)	↓	<0.0001	0.55 (0.07, 1.04)	0.11	0.03
Instruction	If receiver, were you able to take notes (0=No, 1=Yes)	↑	0.98	-	-	-
	Number of interruptions	↑	0.0007	0.003 (-0.11, 0.12)	0.002	0.96
	Impact of interruptions on concentration (worsening)	↓	<0.0001	-0.45 (-0.71, -0.19)	-0.18	0.0007
	Formal feedback (0=No, 1=Yes)	↑	0.03	0.75 (0.05, 1.44)	0.11	0.04

a. For extraneous load, the Cognitive Load Inventory for Handoffs asks about frequency of interruptions and impact of noise on concentration. Because the extraneous load score incorporates these items, they were not included in this study of what learner, task, or environment features are associated with the EL score.