

Using Cognitive Load Theory to Improve Posthospitalization Follow-Up Visits

Elizabeth M. Harry^{1,2} Grace H. Shin¹ Bridget A. Neville¹ Stuart R. Lipsitz¹ Gennady M. Gorbovitsky³
David W. Bates^{1,2} Jeffrey L. Schnipper^{1,2}

¹Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, United States

²Harvard Medical School, Harvard University, Boston, Massachusetts, United States

³Partners Healthcare Information Systems, Boston, Massachusetts, United States

Address for correspondence Elizabeth M. Harry, MD, Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, United States (e-mail: EHarry@bwh.harvard.edu).

Appl Clin Inform 2019;10:610–614.

Background and Significance

Cognitive load theory states humans have limited capacity for information processing via working memory.^{1,2} As health care complexity increases, processing each patient medical record is now a considerable challenge. Working memory can be overwhelmed and information can be lost, impacting patient safety. Wasting working memory on inefficient information presentation carries high risk. Providers experiencing high cognitive load are hypothesized to provide poorer care for patients and be at higher risk of providing care influenced by stereotypes and bias.³ The period after hospital discharge is a particularly high-risk period for patients, and studies have shown deficiencies in documentation of hospitalizations, thus adding to the challenges of providers caring for these patients.^{4,5} Electronic health record (EHR) interfaces designed based on information needs of intensive care physicians demonstrated decreased task load, time to completion, and cognitive errors.⁶ Strengthening this end user-based design by applying the science of cognitive load theory may further improve providers' ability to process information, reducing potential errors and fatigue.^{6–9}

Methods

In this randomized controlled nonclinical trial, we developed a posthospitalization dashboard (PHD), randomized internal medicine residents and faculty to the PHD or usual care for a simulated standardized posthospital follow-up visit, and collected data on performance and cognitive load. To design the PHD, we conducted interviews with internists to identify key content for follow-up visits, consulted with human

factors and cognitive load specialists to design the user interface, and prototyped the PHD for feedback and refinement. Using cognitive load theory, information presentation was optimized to minimize extraneous cognitive load due to split attention¹⁰ (e.g., having to consult several information sources for a comprehensive picture of the patient's status). We pulled information via live data services from our EHR (Epic, Verona, Wisconsin, United States) to populate the PHD (→Fig. 1). Due to time and budget constraints, we simulated data extraction in a few cases where live data services were not yet available. This study was approved by the Brigham and Women's human subjects research committee.

The PHD (→Fig. 1) consists of three columns. The leftmost column consists of historical information including the team that cared for the patient during the hospitalization, the active hospital problems addressed, and a clear list of medications that were discontinued, changed, or continued during the hospitalization. The middle column displays information from the hospitalization and postdischarge period, including vital sign trends and laboratory abnormalities that were unresolved at the time of discharge. The final column includes documentation, including items critical for follow-up, incidental findings, critical goals-of-care information (including any code status change), and links to other critical pieces of documentation, such as visit notes from consultants and home care services. This column also includes future appointments to help with coordinating care over the hospital-to-home continuum. The PHD was designed to minimize redundant information and group-relevant information to paint a cohesive picture of the hospital stay across information types. This was augmented using the column format to display historical information, data, and care coordination/follow-up items.

received
April 4, 2019
accepted after revision
June 30, 2019

© 2019 Georg Thieme Verlag KG
Stuttgart · New York

DOI <https://doi.org/10.1055/s-0039-1694748>.
ISSN 1869-0327.

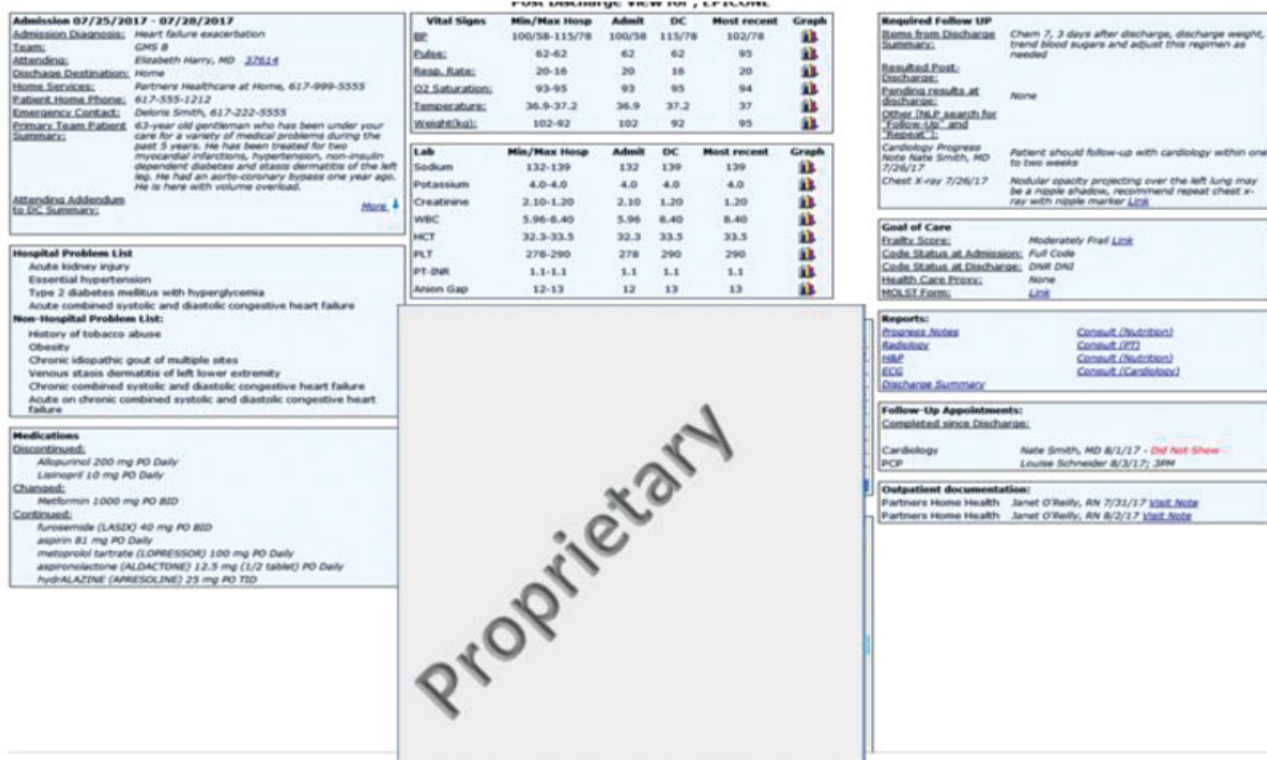


Fig. 1 Screenshot of posthospitalization dashboard. Data are compiled using live data services from the electronic health record (EHR) and displayed on a screen that can be accessed from within the EHR environment. Underlined text in blue represents links to additional documentation. Under the box labeled “Proprietary” is a report from the EHR demonstrating laboratory tests whose latest values remain abnormal, with results and date.

We developed a fictitious virtual patient with a congestive heart failure exacerbation. The content for the hospitalization and subsequent home health visits were included in the simulated medical record and PHD. Eight elements of patient safety risk were built into the record, each with a desired appropriate action to be taken by the provider during the simulated observed postdischarge follow-up visit. These risk elements were based on expert interviews with providers who frequently conduct posthospital follow-up visits and were noted to be kinds of information needed for patient care but were sometimes missing or difficult to find. The risk elements included identifying increasing patient volume, inadequate support at home, incidental pulmonary nodule, missed cardiology appointment, a medication error, health proxy documentation missing, a code status change, and acute kidney injury. Our primary outcome was number of appropriate actions taken (out of 8 predetermined correct actions), based on the documentation written by each participant and as adjudicated independently by two clinician investigators (E.M.H., J.L.S.), both blinded to the study arm; discrepancies were resolved by consensus and we did not calculate interrater reliability. Our secondary outcomes were identification (based on direct observation by a trained research assistant using a think-aloud protocol) and documentation of each safety risk (based on participant documentation and adjudicated by the principal investigator [E. M.H.]); time to complete the follow-up visit; and task load. Task load was measured using the National Aeronautics and

Space Administration Task Load Index (NASA TLX), reported as an overall workload score from 0 to 100 based on a weighted average of six subscales: mental demands, physical demands, temporal demands, performance, effort, and frustration which has been shown to have good reliability and validity.^{11,12} Per TLX protocol, the total score is weighted based on subjective importance to raters of each subtype of workload. For example, frustration is evaluated using the prompt: “How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?” and scored on a 0 to 100 scale.

With institutional review board approval, we recruited medical resident and primary care physician faculty at a single academic medical center via email, obtained informed verbal consent, and randomized them to usual care (EHR process alone, as is currently in use at our medical center) or intervention (PHD + EHR) via a random number generator iOS app from October to December of 2017. Participants who received the PHD were briefly oriented to its features. All clinical content in the PHD was available in the EHR. Participants were instructed to review the hospitalization and any postdischarge patient information, including history and exam information from the current visit from a previsit note given to the participant from our research assistant, and document their findings and planned actions as a visit note while thinking aloud. Participants in the usual care arm were told to use the EHR just as they normally would during a postdischarge clinic visit, while participants in the

intervention arm were told to use the EHR plus the PHD. Participants completed the cognitive load assessment on an iPad (Apple Inc., Cupertino, California, United States) using the NASA TLX application immediately following the simulated visit. All participants received the same case from the research assistant, unblinded to study arm. We hypothesized a priori that performance would improve, and perceived workload would decrease with the PHD compared with usual care.

Data Analysis

Bivariate analyses were conducted for participant characteristics in the two study arms. We used Fisher's exact test for dichotomous characteristics (sex and faculty/resident status), and Wilcoxon rank sum exact test for continuous characteristics (postgraduate year). Outcomes (both primary and secondary) were all continuous (and not normally distributed), and we compared them between arms using Wilcoxon rank sum exact test.¹³ We calculated unadjusted differences (PHD minus usual care) of medians and then inverted the Wilcoxon rank sum test to generate 95% confidence intervals (CIs). Similar analyses were performed for the secondary outcomes. Two-sided *p*-values of < 0.05 were considered statistically significant. All analyses were performed using SAS software, version 9.4 (SAS Institute Inc., Cary, North Carolina, United States).

Using preliminary data for the study, we estimated baseline results and the interquartile range (IQR) for median number of appropriate actions taken. With 20 participants per arm, using a two-sided Wilcoxon rank-sum test with a 5% type I error, this study had 80% power to detect a 1.75 difference in the number of appropriate actions taken between arms, that is, an increase from a median of 3.0 actions in the control arm to 4.75 actions in the intervention (IQR 2.5).

Results

Twenty participants were in the PHD arm and 21 in the usual care arm (out of 140 approached). There were no significant differences in sex, resident status, or postgraduate year by arm (►Appendix A). The results are summarized in ►Table 1. Participants using the PHD demonstrated a median of 5.0 out of 8 appropriate actions taken versus 3.0 out of 8 in the usual care arm (difference in medians 2.0, 95% CI, 0.9–3.1, *p* < 0.001). The median number of issues identified was 5.0 in the PHD versus 2.0 in the usual care arm (*p* < 0.001). Median weighted cognitive load was 50 (PHD) versus 63 (usual care) out of 100, *p* = 0.02. There were no statistically significant differences in subscales for mental demand, physical demand, temporal demand, performance, or effort by arm, or in time to complete the task. The median frustration subscale was 3.5 (PHD) versus 8.0 (usual care), *p* = 0.04.

Table 1 Outcomes

Outcome	Dashboard (<i>n</i> = 20) Median (IQR)	Usual care (<i>n</i> = 21) Median (IQR)	Unadjusted difference in medians (95% CI)	<i>p</i> -Value
Primary outcome				
Number of appropriate actions taken ^a	5.0 (4.0–5.0)	3.0 (2.5–4.0)	2.0 (0.9, 3.1)	< 0.001
Secondary outcomes				
Number of safety issues identified ^a	5.0 (4.0–6.0)	2.0 (2.0–3.0)	3.0 (1.8, 4.2)	< 0.001
Number of safety issues documented ^a	4.0 (3.0–6.0)	3.0 (2.0–4.0)	1.0 (0.2, 1.8)	0.01
Weighted cognitive load ^b	50.0 (41.3–60.2)	63.3 (46.3–68.7)	–13.3 (–25.0, –1.6)	0.02
Adjusted components of cognitive load ^c				
Mental demand	14.8 (9.7–22.3)	18.7 (12.0–21.3)	–3.8 (–12.8, 5.2)	0.41
Physical demand	0.0 (0.0–0.0)	0.0 (0.0–0.0)	0.0 (0.0, 0.0)	NA ^d
Temporal demand	9.3 (4.0–16.7)	10.0 (8.0–15.0)	–0.7 (–5.4, 4.1)	0.79
Performance	5.0 (3.0–7.7)	4.0 (3.3–6.7)	1.0 (–2.8, 4.8)	0.61
Effort	10.3 (4.3–13.2)	12.0 (9.3–18.7)	–1.7 (–3.8, 0.5)	0.12
Frustration	3.5 (1.0–8.3)	8.0 (3.3–17.3)	–4.5 (–8.9, –0.1)	0.04
Time to complete task (minutes)	12.4 (8.5–14.6)	12.8 (9.8–16.8)	–0.5 (–1.6, 0.7)	0.42

Abbreviations: CI, confidence interval; IQR, interquartile range; NASA TLX, National Aeronautics and Space Administration Task Load Index.

^aOut of 8 predetermined actions: identifying increasing volume, inadequate support at home, nodule on chest X-ray, allopurinol missing on discharge medication list, no health care proxy listed, a chance in code status, and development of acute kidney injury during hospitalization.

^bUsing NASA TLX, 0–100 scale, lower score is better.

^cAdjusted components: Scale is weighted based on subjective importance to raters of each subtype of workload, computed by multiplying each rating by the weight given to that factor by that subject. The sum of the weighted ratings for each task is divided by 15 to restore the 0–100 scale for each component, same as for the total cognitive load.

^dAll observations for physical demand in the dashboard arm (and all except 3 observations in the usual care arm) had a value of zero, making statistical testing fairly uninterpretable.

Discussion

We found that a PHD built using cognitive load theory was associated with an increase in the likelihood providers will take appropriate actions in a simulated posthospitalization follow-up visit. This improvement was seen with no increase in visit time, a significantly lower level of frustration and overall perception of task load, and superior performance in identifying and documenting patient safety risks present in the medical record. These data suggest that this tool, informed by human factors and cognitive load theories, was able to improve information presentation, decrease extraneous cognitive load, decrease provider frustration, and therefore decrease the likelihood that critical information was missed.

Limitations of this proof of concept study include some simulated data elements in the PHD not currently available from live data sources (but available soon), the simulated patient and visit (which decreased realism but had the advantage of standardizing the evaluation), the relatively small sample size, and the participants' lack of familiarity with the PHD (which if anything could bias the results in favor of usual care). The relatively low participation rate may limit the generalizability (but not the internal validity) of the findings. The participants could not be blinded to intervention status, which could affect subjective outcomes such as cognitive load, but the primary outcome was adjudicated by two blinded investigators. One might argue that the case was designed to emphasize the strengths of the PHD, but there is nothing unusual about the case or the PHD elements to suggest that this actually occurred. Finally, a predesigned evaluation such as this one might hide potential limitations of the intervention; future studies should include qualitative input from users to avoid such bias.

Health care today faces two critical challenges: information overload and provider frustration and burnout.^{14,15} Using cognitive science to navigate information overload to decrease frustration while increasing performance can impact patient outcomes and provider experience. Understanding cognitive science when designing information presentation systems is critical to future health care delivery as complexity increases. Further randomized controlled trials with all live data and real patient interactions are necessary to draw further conclusions.

Clinical Relevance Statement

Acknowledging cognitive limitations and designing information presentation with these limitations in mind allows providers to perform most optimally. Considering cognitive load theory and consulting human factors experts when developing information technology is important as the amount of information providers are processing increases.

Multiple Choice Questions

1. Cognitive load theory states humans have _____ capacity for information processing.

- a. Unlimited.
- b. Limited.
- c. Limited only when under stress.
- d. Modifiable.

Correct Answer: The correct answer is option b. Humans have a limited capacity for processing this information and this capacity is known as working memory. This capacity can be narrowed when stressed physiologically or emotionally. Practicing memory skills does not expand working memory, but rather enhances the capacity to retrieve information from long-term memory, a resource not known to have limitations, in contrast to short term working memory.

2. Working memory is used unnecessarily by
- a. Data standardization.
 - b. Inefficient data presentation.
 - c. Workflow standardization.
 - d. Consolidation of all critical data pieces to one easily accessed format.

Correct Answer: The correct answer is option b. Extraneous cognitive load unnecessarily uses working memory. The three largest contributors to extraneous cognitive load are lack of standardization, split attention, and redundancy. Efforts to standardize workflows or data presentation, consolidate data, and eliminate redundant data presentation will all decrease extraneous cognitive load and free up working memory.

Protection of Human and Animal Subjects

All human subject involvement of this study was approved by the local IRB.

Conflict of Interest

None declared.

References

- 1 Kahneman D. Attention and Effort. Englewood Cliffs, NJ: Prentice-Hall; 1973:88
- 2 Hamilton V, Warburton DM, Mandler G. Human stress and cognition: an information processing approach. In: Hamilton V, Warburton DM, eds. Thought Processes, Consciousness, and Stress. New York: John Wiley & Sons; 1979:179-201
- 3 Burgess D. Are providers more likely to contribute to healthcare disparities under high levels of cognitive load? How features of the healthcare setting may lead to biases in medical decision making. *Med Decis Making* 2010;30(02):246-257
- 4 Sockolow PS, Bowles KH, Adelsberger MC, Chittams JL, Liao C. Impact of homecare electronic health record on timeliness of clinical documentation, reimbursement, and patient outcomes. *Appl Clin Inform* 2014;5(02):445-462
- 5 Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. *JAMA* 2007;297(08):831-841
- 6 Ahmed A, Chandra S, Herasevich V, Gajic O, Pickering BW. The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance. *Crit Care Med* 2011;39(07):1626-1634

- 7 Phansalkar S, Edworthy J, Hellier E, et al. A review of human factors principles for the design and implementation of medication safety alerts in clinical information systems. *J Am Med Inform Assoc* 2010;17(05):493–501
- 8 Zachariah M, Phansalkar S, Seidling HM, et al. Development and preliminary evidence for the validity of an instrument assessing implementation of human-factors principles in medication-related decision-support systems–I-MeDeSA. *J Am Med Inform Assoc* 2011;18(Suppl 1):i62–i72
- 9 Harry E, Pierce R, Kneeland P, Huang G, Stein J, Sweller J. Cognitive load theory and its implications for health care complexity - NEJM catalyst. *New England Journal of Medicine: Catalyst*. Published 2018. Available at: <https://catalyst.nejm.org/cognitive-load-theory-implications-health-care/>. Accessed March 21, 2018
- 10 Ayres P, Editor A. From the archive: 'Managing split-attention and redundancy in multimedia instruction' by S. Kalyuga, P. Chandler, & J. Sweller (1999). *Applied Cognitive Psychology*, 13, 351–371 with commentary. *Appl Cogn Psychol* 2011;25(S1): S123–S144
- 11 Xiao YM, Wang ZM, Wang MZ, Lan YJ. The appraisal of reliability and validity of subjective workload assessment technique and NASA-task load index [in Chinese]. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi* 2005;23(03):178–181
- 12 Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv Psychol* 1988;52:139–183
- 13 Warmuth E, Bhattacharya GK, Johnson RA. Statistical concepts and methods. J. Wiley & Sons, New York-Santa Barbara-London-Sydney-Toronto 1977. xv, 639 s., £ 11.25; \$ 19. *Biometrical J* 1980;22(04):371–371
- 14 Wilson TD. Information overload: implications for healthcare services. *Health Informatics J* 2001;7(02):112–117
- 15 Dyrbye LN, West CP, Satele D, et al. Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. *Acad Med* 2014;89(03):443–451

Appendix A Respondent characteristics by study arm

	Dashboard		Usual care		Chi square, <i>p</i> -value	Fisher's exact, <i>p</i> -value	Wilcoxon's rank sum, <i>p</i> -value	Wilcoxon's exact, <i>p</i> -value
Characteristic	<i>n</i> = 21		<i>n</i> = 20					
Gender (<i>n</i> , %)					0.8665	1.0000		
Female	9	42.86%	10	50.00%				
Male	11	52.38%	11	55.00%				
Faculty versus Resident (<i>n</i> , %)					0.6537	0.7557		
Faculty	9	42.86%	8	40.00%				
Resident	11	52.38%	13	65.00%				
Years of experience (PGY) (median, IQR)	3	2, 9	3	2, 6.50			0.843	0.8368
Min, max	1	16	1	15				

Abbreviation: IQR, interquartile range.