

HHS Public Access

Author manuscript

Surg Innov. Author manuscript; available in PMC 2021 February 03.

Published in final edited form as:

Surg Innov. 2020 December; 27(6): 602–607. doi:10.1177/1553350620934931.

Surgery Task Load Index in Cardiac Surgery: Measuring Cognitive Load Among Teams

Lauren R. Kennedy-Metz, PhD^{1,2}, Hill L. Wolfe, MPA³, Roger D. Dias, MD, MBA, PhD^{4,5}, Steven J. Yule, PhD^{2,6}, Marco A. Zenati, MD, MSc, FETCS^{1,2}

¹Division of Cardiac Surgery, Veterans Affairs Boston Healthcare System, West Roxbury, MA, USA ²Department of Surgery, Harvard Medical School, Boston, MA, USA ³Boston University School of Public Health, Boston, MA, USA ⁴STRATUS Center for Medical Simulation, Brigham and Women's Hospital, Boston, MA, USA ⁵Department of Emergency Medicine, Harvard Medical School, Boston, MA, USA ⁶The University of Edinburgh, Scotland

Abstract

Background.—The most commonly used subjective assessment of perceived cognitive load, the NASA Task Load Index (TLX), has proven valuable in measuring individual load among general populations. The surgery task load index (SURG-TLX) was developed and validated to measure cognitive load specifically among individuals within a surgical team. Notably, the TLX lacks temporal sensitivity in its typical retrospective administration.

Objective.—This study sought to expand the utility of SURG-TLX by investigating individual measures of cognitive load over time during cardiac surgery, and the relationship between individual and team measures of cognitive load and proxies for surgical complexity.

Materials & Methods.—SURG-TLX was administered retrospectively in the operating room immediately following each case to approximate cognitive load before, during, and after cardiopulmonary bypass for cardiac surgery team members (surgeon, anesthesiologist, and perfusionist). Correlations were calculated to determine the relationship of individual and team measures of cognitive load over the entire procedure with bypass length and surgery length.

Results.—Results suggest that perceived cognitive load varies throughout the procedure such that cognitive load during bypass significantly differs compared to before or after bypass, across all 3 roles. While on bypass, results show that anesthesiologists experience significantly lower levels of perceived cognitive load than both surgeons and perfusionists. Correlational analyses reveal that perceived cognitive load of both the surgeon and the team had significant positive associations with bypass length and surgery length.

Corresponding Author: Lauren R. Kennedy-Metz, PhD, Division of Cardiac Surgery, VA Boston Healthcare System West Roxbury Campus, 1400 VFW Parkway, West Roxbury 02132-4927, MA, USA. Lauren.Kennedy-Metz@va.gov. Author Contributions

Study concept and design: Roger D. Dias, Steven J. Yule, and Marco A. Zenati

Acquisition of data: Hill L. Wolfe, Roger D. Dias, Steven J. Yule, and Marco A. Zenati

Analysis and interpretation: Lauren R. Kennedy-Metz, Hill L. Wolfe, Roger D. Dias, Steven J. Yule, and Marco A. Zenati Study supervision: Marco A. Zenati

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Conclusion.—Our findings support the utility of SURG-TLX in real cardiac cases as a measure of cognitive load over time, and on an individual and team-wide basis.

Keywords

cognitive load; surgery task load index; cardiac surgery

Background

Cardiac surgery teams face heightened intraoperative pressure due to multiple demands such as task complexity, distractions, and time pressure. In addition to technical skills and knowledge, human factors such as these can have tremendous influence on outcomes. When surgical team members perceive high levels of demand, cognitive load may increase, which could ultimately impair performance. Cognitive load, often referred to as mental load or mental effort, refers to the balance between one's cognitive resources and the demands imposed by a task. When an imbalance is present such that cognitive load is excessive, surgical team members' ability to adapt to changing work demands is diminished, and their likelihood of committing cognitive errors is enhanced. Recent work has shown that across 188 adverse events recorded over a 6 month period, 51.6% of all underlying human performance deficiencies contributing to those adverse events were due to cognitive errors.

Understanding the impact of surgical team members' cognitive load on their operative performance and patient safety outcomes requires tools with validity evidence that are designed to capture the complexity of the cognitive load construct. The most widely implemented self-report tool for measuring multiple components of cognitive load in surgery is the NASA Task Load Index (NASA-TLX).² Using 6 visual analog scales ranging from 0 to 100, the NASA-TLX enables respondents to subjectively assess their perceived cognitive load across distinct domains.⁹ Domains included in the original version of NASA-TLX are mental demand, physical demand, temporal demand, performance, effort, and frustration level.⁹

In an effort to generate a surgery-specific cognitive load measure, the surgery task load index (SURG-TLX) was developed and validated. The SURG-TLX scale presents a similar format as the NASA-TLX but substitutes 3 of the cognitive load dimensions with domains more relevant to surgical tasks. The adapted surgery-specific scale measures mental demand, physical demand, temporal demand, task complexity, situational stress, and distractions on visual analog scales ranging from 0 to $100.^1$ Importantly, having appropriate cognitive load assessment tools specific to surgery, such as the SURG-TLX, may enable valuable surgical educational opportunities, allowing targeted training corresponding to specific cognitive load domains.

Another version of the SURG-TLX further modified the included scales to measure 5 dimensions of cognitive load, including mental demand, physical demand, task complexity, distractions, and degree of difficulty. The first 4 domains from the SURG-TLX were included in this adaptation, along with the final domain, degree of difficulty, taken from the global operative assessment of laparoscopic skill questionnaire. ¹⁰ The modified scale

resulting from this effort was developed specifically to increase the scale's relevance in measuring intraoperative workload.⁷

Despite being widely used, the SURG-TLX instrument, along with the NASA-TLX which it was originally based on, has a number of limitations. The literature using the SURG-TLX as a measure of cognitive load has been tested primarily in surgical simulations, minimizing ecological validity and generalizability to actual workflow.² Due to the limited number of assessments occurring during live surgeries, the relationship between proxies for surgical complexity (ie, surgery length and bypass length) and cognitive load has been underexplored. In order to maintain natural behavior and minimize interruptions, self-report scales are often administered retroactively and globally, limiting the temporal sensitivity of their measures. Historically, measures of cognitive load through self-report scales have limited subsequent analyses to the assessment of individual load, allowing little interpretation into multi-person teams. To overcome these limitations, our approach relied on implementing the SURG-TLX to measure cognitive load derived naturalistically³ in the operating room immediately following actual cardiovascular surgery procedures, exploring the relationship between cognitive load and proxies for surgical complexity, evaluating cognitive load during 3 separate surgical phases, and calculating both individual and averaged team measures of cognitive load. Due to the specific interest in intraoperative load measurement, we chose to evaluate this construct via the modified SURG-TLX questionnaire developed for this purpose.⁷

Our aim was thus to explore how a modified version of the SURG-TLX tool captures perceived intraoperative cognitive load in this real clinical setting, both at the individual and team level, and examine how they are related to proxies for surgical complexity. We hypothesized (a) that there would be differences in SURG-TLX scores between individual team members and surgical phases, and (b) higher SURG-TLX scores positively associated with time on bypass and overall surgery length.

Method

This study was conducted in the operating room of a tertiary teaching hospital (VA Boston Healthcare System; VABHS), and the research protocol was approved by the institutional review boards of VABHS and Harvard Medical School (IRB#3047). All participants (patients and OR staff) provided written informed consent prior to the start of surgery. Immediately following each cardiac procedure, the attending surgeon, attending anesthesiologist, and perfusionist completed 3 SURG-TLX assessments (modified version)⁷ to assess their perceived cognitive load during each critical phase of the procedure: (1) precardiopulmonary bypass (CPB), (2) during CPB, and (3) post-CPB. Each assessment contained 5 subscales ranging from 0 to 100 (with 5-point steps): mental demand, physical demand, task complexity, distractions, and degree of difficulty. The modified version of the SURG-TLX was used to greater assess perceived difficulty intraoperatively and has been validated in the analogous literature. Assessments were administered using a pen-and-paper version for convenience and ease of use and were completed within a few minutes. No participants vocalized any objections nor difficulty completing the questionnaires.

Statistics

The data analyzed (N = 23 cases) represent a subset of data from a larger study. Statistical analyses were performed by computing average cognitive load for each SURG-TLX subscale (ie, cognitive load dimension) and total SURG-TLX score by surgical team role (surgeon, anesthesiologist, and perfusionist) and critical phase (pre-CPB, during CPB, and post-CPB). Length of time (surgery length and bypass length) was acquired from patient records. Multivariate ANOVA tests incorporating surgical team role as a between-subjects factor and critical phase as a within-subjects factor were used to examine differences across means, post hoc pairwise comparisons with the Bonferroni correction were calculated to identify significant relationships based on ANOVA results, and Spearman's rank-order correlation tests were used to measure associations among cognitive load and surgery length and bypass length. Surgery times were recorded by the circulating nurse, and bypass times were recorded on the CPB pump by the perfusionist. All surgery data were notated in the patient's electronic medical record. The statistical software SPSS (version 26.0) was used for analysis. Significance is reported as P < .05.

Results

A total of 207 assessments were collected, including one for each provider (3 providers per case) and each of the critical phases (3 phases per case) from 23 cardiac surgery procedures (12 isolated aortic valve replacement and 11 isolated coronary artery bypass grafting). Three attending cardiac surgeons, 5 attending anesthesiologists, and 3 perfusionists participated in this study over the course of the 23 total cases recorded.

ANOVA results revealed a significant interaction effect among anesthesiologists, perfusionists, and surgeons' average SURG-TLX scores pre-, during, and post-bypass (F[4132] = 40.447, P < .001). Post hoc analyses revealed that overall SURG-TLX scores were significantly different between anesthesiologists' (19.61, CI: 11.60-27.61) reports compared to those of perfusionists' (46.17, CI: 38.17-54.18, P < .001) and surgeons' (47.65, CI: 39.65-55.66, P < .001) during bypass. Notably, this separation is not observed among pre- or post-bypass phases of the surgeries analyzed, where the reported cognitive load converges across the 3 roles (Figure 1).

Within roles, multivariate tests revealed that perceived cognitive load was significantly different across phases for anesthesiologists (F[265] = 63.324, P<.001), perfusionists (F[265] = 17.040, P<.001), and surgeons (F[265] = 7.933, P=.001). Post hoc pairwise comparisons with the Bonferroni correction were calculated. Anesthesiologists' perceived cognitive load, according to the overall SURG-TLX score, significantly decreased by 22.739 between pre-bypass and bypass phases (P<.001), and significantly increased by 23.609 between bypass and post-bypass phases (P<.001). Among perfusionists, SURG-TLX scores increased by 11.826 between pre-bypass and bypass phases (P<.001). Surgeons reported a nonsignificant increase in the reported SURG-TLX score of 5.348 between pre-bypass and bypass phases (P=.087) and a significant decrease of 9.696 between bypass and post-bypass phases (P=.001).

Analysis of SURG-TLX scores, according to individual roles, revealed a strong positive correlation (>99% CI) between surgeons' total SURG-TLX scores and surgery length (r = .759, P< .001). Correlations between surgery length and total SURG-TLX values for anesthesiologists and perfusionists were not significant (Figure 2). A strong positive correlation (>99% CI) was also found between surgeons' total SURG-TLX scores and CPB length (r= .724, P< .001). Correlations between CPB length and total SURG-TLX values for anesthesiologists and perfusionists were not significant. All individual cognitive load dimensions reported by the surgeon were also positively correlated (>99% CI) with both surgery and CPB length (P< .05 for all relationships).

We also found that analysis of SURG-TLX dimensions collapsed across all team members and critical phases to arrive at a team SURG-TLX value for the entire procedure. Results reveal strong positive correlations (>99% CI) between surgery length and every team cognitive load dimension, individually and in total, combined across surgical team roles and surgery phases (P<.05). Similarly, bypass length was significantly correlated (>99% or >95% CI) with every team cognitive load dimension and the total team SURG-TLX score (P<<.05) (Table 1).

Discussion

The present human factor study supports that SURG-TLX assessments are capturing a variety of aspects of cognitive load in association with the individual workload between surgical phases and across team members, and elucidating relationships between cognitive load of the surgeon individually, the surgical team in combination, length of surgery, and duration of CPB. Our findings support feasibility of using the SURG-TLX in real cardiac cases as a measure of cognitive load on an individual and team-wide basis, as it significantly associates with known factors that both increase cognitive load and often represent proxies of overall surgical complexity, including surgery length and bypass length.

Our results show differences in perceived cognitive load according to both surgical phase and provider role, based on analysis of self-reported SURG-TLX scores. When considering team members individually, results reveal temporal differences in perceived cognitive load over the course of the procedure. Anesthesiologists reported lower cognitive load while on bypass compared to the periods of time preceding and following bypass, while perfusionists and surgeons reported the opposite trend: heightened cognitive load during bypass compared to levels before and after bypass. Notably, upon averaging SURG-TLX scores within each role across the entire procedure, temporal differences among individual providers become nonsignificant, resulting in mean values of 35.058, 38.159, and 42.638 for anesthesiologists, perfusionists, and surgeons, respectively This finding highlights the benefit of collecting perceptions reflective of multiple time periods to detect patterns of change, rather than relying on one global measure of cognitive load to capture an entire procedure.

Our analyses also show differences between individual roles, according to surgical phase. Specifically, when considering the bypass phase, surgeons and perfusionists perceive a greater cognitive load than anesthesiologists. Cognitive load reported before and after bypass reveals no differences between roles, supporting further investigation into factors

contributing to differences in perceived cognitive load during the bypass phase of cardiac surgery in particular. Reports of significantly lower demand of anesthetists compared to surgeons during bypass are congruent with other literature on cardiac surgery teams.⁶ Further inquiry should investigate potential common drivers of cognitive load specific to these roles and possible remedies to these drivers.¹¹ For example, one solution that may lessen cognitive load for surgeons and perfusionists may be implementing a ban on nonessential activities where only necessary discussions and tasks take place.

When SURG-TLX scores were averaged within each phase across all 3 team members, these previously interpersonal differences disappeared. Average SURG-TLX values collapsed across team members were 39.667 before bypass, 37.812 during bypass, and 38.377 after bypass, reflecting indistinguishable levels of cognitive load over time and emphasizing the importance of considering each role separately. It is critical to recognize and detect that primary tasks of each team member differ drastically at any given time in a sociotechnical system as complex as the cardiac surgery operating room. These results support that the multidimensionality of the SURG-TLX tool is capable of elucidating differences in cognitive load.

Although a specific load threshold predicting a negative impact on human performance is under debate, current available literature reports cognitive load scores "over 50-55" could lead to more performance errors. ^{7,12,13} Otherwise, establishing a cognitive load "redline" to identify when cognitive load is too high across applications and tasks is an ongoing process. ¹⁴ Given the available estimate within the domain of health care, however, SURG-TLX scores that meet or exceed a threshold of 5055 should be identified to investigate the corresponding frequency of errors. ³ Due to the scope and the scale of this study, we did not account for performance errors or near misses, or their relationship with individual or team cognitive load scores.

Results of our correlational analyses also indicate that longer surgeries are associated with greater amounts of perceived cognitive load for the surgical team, including higher perceived mental and physical demand, task complexity, degree of difficulty, distractions, and total cognitive load. Similarly, the longer the patients are on bypass, the higher the combined team rating of each cognitive load dimension collected. While these results may seem intuitive, this is the first empirical demonstration we know of in the literature confirming the relationship between perceived cognitive load and measures of surgical complexity derived from naturalistic settings. Given the nature of the data collection and the correlational analytic approach, we are unable to infer causality between cognitive load and duration from these data. Further research is needed to investigate additional contributing factors and to establish a causal link.

Despite some of the documented shortcoming associated with subjective self-assessment tools such as the SURG-TLX, this instrument, in particular, provides the opportunity to identify distinct contributors to overall cognitive load. Knowledge of primary contributors to overall load may enable targeted training and/or interventions to address coping with distinct types of demands. For example, if particular individuals or roles experience excessive physical demands, introduction of occasional micro-breaks for postural relief may be

valuable. ¹⁵ Identification of primary contributors to cognitive load could be enhanced with meaningful qualitative approaches, which is beyond the scope of this research, but should be considered in future applications.

This study has a number of limitations. The small sample size of surgical team members and total sample of assessments in this study limit generalizability of the predictive validity of the SURG-TLX. Furthermore, since the same surgical team members appear unsystematically in multiple cases, the 23 cases from this study are not independent. However, these are data gathered during real cardiac cases rather than simulations, and are ecologically valid. Accounting for other predictive patient outcomes, such as mortality, and morbidity risk scores, should also be included in future analyses. Further insight into specific factors contributing to perceptions of different cognitive load dimensions should be investigated.

While findings support temporal sensitivity compared to existing self-report approaches, conclusions could be greatly enhanced with further granularity. Even greater temporal sensitivity reflective of more dynamic and instantaneous changes in cognitive load includes continuous and objective measures, such as heart rate variability (HRV). HRV is the most commonly used psychophysiological objective measure of cognitive load in surgery² and could overcome many of the remaining limitations associated with self-report instruments.

Conclusion

In summary, our findings support the utility of a multidimensional self-report tool, the modified SURG-TLX, to assess cognitive load during cardiac surgery. Results demonstrate its temporal sensitivity such that significantly different perceived cognitive load was reported across distinct surgical phases. Findings also support the association between perceived cognitive load and approximations of surgical complexity but cannot determine the direction of causality within this relationship. Future research should continue to validate the SURG-TLX, specifically in intraoperative settings and among surgical teams.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health [grant number R01HL126896, PI: Zenati]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Wilson MR, Poolton JM, Malhotra N, Ngo K, Bright E, Masters RSW. Development and validation of a surgical workload measure: The surgery task load index (SURG-TLX). World J Surg. 2011;35(9):1961–1969. doi:10.1007/s00268-011-1141-4. [PubMed: 21597890]
- Dias RD, Ngo-Howard MC, Boskovski MT, Zenati MA, Yule SJ. Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. Br J Surg. 2018;105:491–501. doi:10.1002/bjs.10795. [PubMed: 29465749]
- 3. Wickens CD. Multiple resources and performance prediction. Theor Issues Ergon Sci. 2002;3(2):159–177.

 Carswell CM, Clarke D, Seales WB. Assessing mental workload during laparoscopic surgery. Surg Innov. 2005; 12(1):80–90. [PubMed: 15846451]

- 5. Weigl M, Antoniadis S, Chiapponi C, Bruns C, Sevdalis N. The impact of intraoperative interruptions on surgeons' perceived workload: An observational study in elective general and orthopedic surgery. Surg Endosc. 2015;29: 145–153. doi:10.1007/s00464-014-3668-6. [PubMed: 24986016]
- Wadhera RK, Parker SH, Burkhart HM, et al. Is the "sterile cockpit" concept applicable to cardiovascular surgery critical intervals or critical events? The impact of protocol-driven communication during cardiopulmonary bypass. J Thorac Cardiovasc Surg. 2010;139(2):312–319. doi:10.1016/j.jtcvs.2009.10.048. [PubMed: 20106395]
- Yu D, Lowndes B, Thiels C, et al. Quantifying intraoperative workloads across the surgical team roles: Room for better balance? World J Surg. 2016;40(7):1565–1574. doi:10.1007/ s00268-016-3449-6. [PubMed: 26952115]
- Suliburk JW, Buck QM, Pirko CJ, et al. Analysis of human performance deficiencies associated with surgical adverse events. JAMA Netw Open. 2019;2(7):e198067. doi:10.1001/ jamanetworkopen.2019.8067. [PubMed: 31365107]
- 9. Hart SG, Staveland LE. Development of NASA-TLX (task load index): Results of empirical and theoretical research. Adv Psychol. 1988;52:139–183.
- Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg. 2005;190(1):107–113. doi:10.1016/ j.amjsurg.2005.04.004. [PubMed: 15972181]
- 11. Wetzel CM, Kneebone RL, Woloshynowych M, et al. The effects of stress on surgical performance. Am J Surg. 2006; 191(1):5–10. doi:10.1016/j.amjsurg.2005.08.034. [PubMed: 16399098]
- 12. Mazur LM, Mosaly PR, Hoyle LM, Jones EL, Chera BS, Marks LB. Relating physician's workload with errors during radiation therapy planning. Pract Radiat Oncol. 2014;4(2): 71–75. doi:10.1016/j.prro.2013.05.010. [PubMed: 24890346]
- 13. Mazur LM, Mosaly PR, Hoyle LM, Jones EL, Marks LB. Subjective and objective quantification of physician's workload and performance during radiation therapy planning tasks. Pract Radiat Oncol. 2013;3(4):e171–e177. doi: 10.1016/j.prro.2013.01.001. [PubMed: 24674415]
- Hart SG. NASA-task load index (NASA-TLX); 20 years later. Proc Hum Factors Ergon Soc Annu Meet. 2006;50: 904–908.
- 15. Park AE, Zahiri HR, Hallbeck MS, et al. Intraoperative "Micro Breaks" with targeted stretching enhance surgeon physical function and mental focus: a multicenter cohort study. Ann Surg. 2017;265(2):1–7. doi:10.1097/SLA.000000000001665. [PubMed: 28009727]

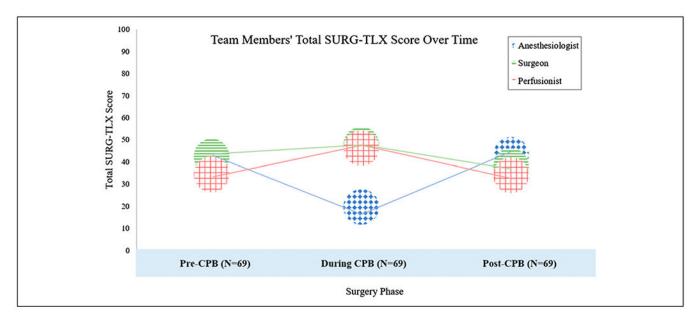


Figure 1. (N = 23 cases) Analysis of SURG-TLX scores, according to individual roles, revealed role-specific temporal patterns of cognitive load. Each phase included 69 SURG-TLX responses, consisting of 23 from each team member. Anesthesiologists (diamonds) reported significantly lower cognitive load than perfusionists (grid) and surgeons (horizontal lines) during CPB. Analyses also support that reported cognitive load is significantly lower within each role during bypass compared to both before or after bypass.

Note. SURG-TLX = surgery task load index, CPB = cardiopulmonary bypass.

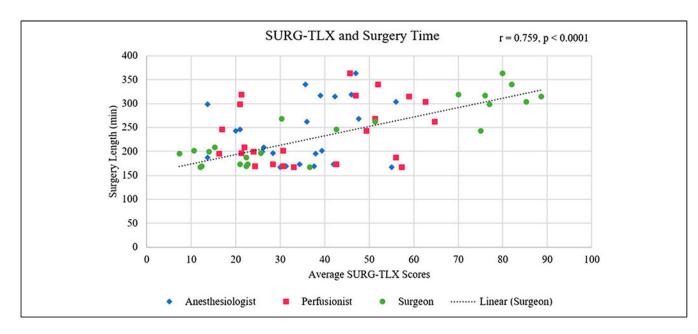


Figure 2. (N = 23 cases) Analysis of SURG-TLX scores, according to individual roles, revealed a strong positive correlation between surgeons' (circles) total SURG-TLX scores and surgery length (r= .759, P< .0001). Correlations for anesthesiologists (diamonds) and perfusionists (squares) were not significant. Each data point represents the average SURG-TLX across all 3 phases for the given team member within one case.

Note. SURG-TLX = surgery task load index.

Table 1.

(N=23~Cases) Correlation Coefficients (ρ) and 2-Tailed P-Values (sig.) of Surgery Length and Bypass Length Between Team SURG-TLX Subscale Scores and Overall Team SURG-TLX Scores. SURG-TLX Subscale Values Were Calculated by Averaging Across All 3 Team Members and All 3 Phases for the Given Cognitive Load Dimension Across Each of 23 Cases. The Total SURG-TLX Score Was Derived by Subsequently Averaging Each of the Subscale Scores.

	Surgery Length		Bypass Length	
	(2-Tailed)		(2-Tailed)	
SURG-TLX Subscale	Sig.	ρ	Sig.	ρ
Mental demand	.001	.626**	.002	.622 **
Physical demand	.001	.641 **	.002	.601 **
Task complexity	.001	.668**	.002	.641 **
Distractions	.006	.550**	.016	.498*
Degree of difficulty	.001	.651 **	.002	.607**
Total SURG-TLX score	.001	.638**	.003	.592**

Note:

Abbreviation: SURG-TLX = surgery task load index.

^{**}Correlation is significant at the .01 level,

^{*} correlation is significant at the .05 level.