

Supplemental table B. Included studies.

Author and year	Location(s)	Study methodology	Participants ^a	Activity, Simulation or Workplace	Data sources	CL measurement technique(s) ^b	CLT-related study aim(s) ^b	Major finding(s) related to CLT ^b
Andersen 2015 ¹	Denmark	Quantitative, experimental	40 “novice” medical students	Mastoidectomy surgery, Simulation	Objective observations or tests	Visual secondary task	To compare CL during virtual reality simulated mastoidectomy training using distributed versus massed practice	Reaction time was lower among those learning mastoidectomy using a distributed versus massed practice approach
Andersen 2016 ²	Denmark	Quantitative, experimental	40 otorhinolaryngology residents	Mastoidectomy surgery, Simulation	Objective observations or tests	Visual secondary task	To compare CL during mastoidectomy training using in traditional dissection versus virtual reality simulation	Reaction time higher in cadaveric than in simulated mastoidectomy
Andersen 2016 ³	Denmark	Quantitative, experimental	36 medical students	Mastoidectomy surgery, Simulation	Objective observations or tests	Visual secondary task	To compare retention of mastoidectomy skills after virtual reality mastoidectomy training in distributed versus massed practice groups and to investigate CL during retention procedures	No difference in reaction time when performing mastoidectomy at transfer after learning from massed versus distributed practice
Andrade 2012 ⁴	United States	Quantitative, non-experimental	10 “senior” medical students, 10 internal medicine residents, 10 geriatric fellows	Geriatric home safety assessment, Simulation	Quantitative surveys, Interviews	NASA-TLX	To determine whether CL was associated with performance when learning from a 3D geriatric home safety simulation	CL not associated with performance when measured concurrently
Arico 2016 ⁵	France, Italy	Quantitative, experimental	12 air traffic control students	Air traffic management, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, EEG	To determine whether adjusting task complexity via adaptive automation would impact CL or performance during simulated air traffic management	Adaptive automation was associated with shorter reaction time and better performance, but subjective CL did not differ
Attrill 2016 ⁶	Australia	Qualitative	20 speech language pathology clinical supervisors	Speech pathology, Workplace	Focus groups	CL not measured	To qualitatively evaluate the experience of supervisors in teaching international students	Teachers felt that international students had overall higher CL than non-international students, particularly related to language issues
Avansino 2012 ⁷	United States	Quantitative, experimental	3 surgical residents, 3 surgical fellows, 1 surgical attending	Computerized provider order entry, Simulation	Quantitative surveys, Documents/ artifacts	NASA-TLX	To compare CL associated with using a systematically designed order set compared with existing ad hoc order set for simulated entering of post-operative patient orders	Systematically designed order set was easier to use, was associated with lower CL and better performance
Bertram 1990 ⁸	United States	Quantitative, non-experimental	39 internal medicine residents, 9 internal medicine attendings	Ambulatory patient care, Workplace	Quantitative surveys	Investigator-developed multi-item instrument	To develop an instrument to measure CL of physicians working in an internal medicine ambulatory setting, and to identify factors associated with CL	CL positively associated with number of patients seen and fatigue; CL negatively associated with self-rated quality of care and satisfaction with care provided
Bertram 1992 ⁹	United States	Quantitative, non-experimental	22 internal medicine residents	Ambulatory patient care, Workplace	Quantitative surveys	Investigator-developed multi-item instrument	To assess reliability and validity of adapted version of previously developed instrument (Bertram 1990) to measure CL in ambulatory internal	CL positively associated with fatigue; CL negatively associated with level of experience, self-rated quality of care

Supplemental table B. Included studies.

							medicine setting, and to identify factors associated with CL	provided, and faculty-rated quality of care provided
Bharathan 2013 ¹⁰	United Kingdom	Quantitative, non-experimental	25 gynecology residents, 9 gynecology attendings	Laparoscopic gynecologic surgery, Simulation	Objective observations or tests	NASA-TLX, Subjective Mental Effort Questionnaire	To evaluate construct validity for a laparoscopic surgical simulator, to evaluate effectiveness as a training tool, and to examine factors associate with CL	CL was higher for more complex/difficult procedures; Experience positively associated with performance
Blissett 2017 ¹¹	Canada	Quantitative, experimental	86 internal medicine residents	Cardiac auscultation, Simulation	Quantitative surveys, Objective observations or tests	Paas’ Subjective Rating of Mental Effort single item scale	To compare how terminally versus branching diagnostic schemas impacted performance and CL during simulated cardiac auscultation	Terminally branching diagnostic schemas (which should require less CL) were associated with lower CL, fewer errors, and greater diagnostic accuracy than hybrid diagnostic schemas
Boet 2017 ¹²	Canada	Quantitative, experimental	20 surgical residents	Post-operative crisis management, Simulation	Objective observations or tests	Vibrotactile secondary task	To examine whether debriefing after an initial post-operative crisis simulation impacts CL during subsequent simulated post-operative crisis situations	Debriefing after an initial crisis scenario was associated with lower reaction time in a second scenario, compared with a control setting without debriefing
Britt 2015 ¹³	United States	Quantitative, experimental	14 surgical assistant students	Laparoscopic intracorporeal suturing, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, Visual secondary task	To determine whether a new spatial secondary task (“ball-and-tunnel” signal detection task) would identify differences in CL when performing cadaveric versus simulated suturing	Secondary task performance lower and CL higher during cadaveric versus simulated laparoscopic suturing
Broyles 2011 ¹⁴	United States	Mixed methods, experimental	27 career and technical education pre-service teachers	Teaching, Workplace	Objective observations or tests, Focus groups	CL not measured	To determine whether degree of CL impacted pre-service teachers’ ability to engage in reflection	Higher CL state interfered with ability to engage in reflection
Byrne 1998 ¹⁵	United Kingdom	Quantitative, non-experimental	10 anesthesiology residents	Critical anesthesia incident management, Simulation	Objective observations or tests	CL not measured	To examine whether critical incidents during simulated anesthesia negatively impacted accuracy of charting	Accuracy of anesthesia charting was lower during periods of higher CL (simulated patient becomes critically ill), but inaccuracies varied depending on nature of critical incident (e.g., during pulmonary crisis, carbon dioxide level charted more accurately than heart rate)
Byrne 2013 ¹⁶	United Kingdom	Quantitative, non-experimental	20 anesthesiology residents and attendings	Administering anesthesia, Workplace	Objective observations or tests	Vibrotactile secondary task	To characterize inter- and intra-person variation in CL during actual administration of anesthesia	Greater variation in reaction time within subjects than between subjects
Cao 2007 ¹⁷	United States	Quantitative, experimental	24 surgical residents, 6 surgical fellows	Surgical ball transfer task, Simulation	Objective observations or tests	CL not measured	To determine whether adding haptic feedback during simulated surgical tasks mitigated performance decline associated with cognitively loading subjects	Cognitive loading reduced speed, but not accuracy, on surgical ball transfer task; Adding haptic feedback improved speed and accuracy when cognitively loaded; More experienced

Supplemental table B. Included studies.

								surgeons were faster and more accurate
Chen 2015 ¹⁸	Canada	Quantitative, experimental	60 “senior-level” nursing students	Cardiac and pulmonary auscultation, Simulation	Objective observations or tests	CL not measured	To compare impact of low- versus high-fidelity instructional techniques on performance during simulated cardiac and pulmonary auscultation	Lower fidelity simulation associated with better performance and recall than high-fidelity simulation; Both simulations associated with better performance than control group
Chen 2015 ¹⁹	Canada	Quantitative, experimental	77 “senior-level” nursing students	Cardiac and pulmonary auscultation, Simulation	Objective observations or tests	CL not measured	To determine whether adjustments intended to impact intrinsic, extraneous, and germane cognitive load had any impact during simulated cardiac and pulmonary auscultation	Less complex task (lower intrinsic load) associated with better performance; Mixed (versus blocked) teaching associated with better performance (implying greater germane load); Authors also concluded that multiple versus single representations of heart sounds did not impact extraneous load
Chowriappa 2015 ²⁰	United Kingdom, United States	Quantitative, experimental	22 surgical residents, 30 surgical fellows	Robotic surgery tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether an augmented-reality-based training (“Hands-on Surgical Training”) for simulated robotic assisted surgery impacted CL and performance compared with control condition	Intervention group performed better and had lower CL compared with control group
Chui 2014 ²¹	United States	Qualitative	8 practicing pharmacists	Pharmacy handoffs, Simulation	Interviews	CL not measured	To describe and characterize “information hazards” present during pharmacy handovers, including whether these hazards were related to types of CL	“Information hazards” present during pharmacy handovers included those related to CL: information overload (i.e., intrinsic load) and information scatter (i.e., extraneous load)
Crosby 1979 ²²	United States	Quantitative, experimental	Experiment 1: 12 student pilots, 12 pilot instructors Experiment 2: 11 student pilots	Flight, Simulation	Objective observations or tests	Visual secondary task	To determine whether differences in CL in single- versus dual-task conditions were impacted by student versus instructor status during simulated flight	Differences in response time in single-versus dual-task condition were greater for students than for instructors; Instructors performed better than students; Later-phase students performed similar to instructors
Cummings 2009 ²³	United States	Quantitative, experimental	31 student, recreational, and professional pilots	Flight, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether a heads-up informational display impacted performance and CL during simulated flight	No difference in CL or performance when using heads-up informational display during simulated flight
Dahlstrom 2009 ²⁴	Sweden	Quantitative, experimental	11 student pilots, 8 pilot instructors	Flight, Simulation and Workplace	Objective observations or tests	Single item subjective rating of CL, Heart rate	To determine whether CL differed in simulated versus actual flight, and in particular flight segments	Neither subjective CL nor overall heart rate variability differed in in simulated compared with actual

Supplemental table B. Included studies.

						variability, Eye movement		flight; Heart rate variability lower for some flight segments, suggesting different degree of CL was required
Dankbaar 2016 ²⁵	Netherlands	Quantitative, experimental	79 4 th -year medical students	Emergency medical care, Simulation	Quantitative surveys, Objective observations or tests	Investigator-developed multi-item instrument	To compare intrinsic and germane load experienced by medical students during simulated emergency medical care when instructional groups offered differing levels of fidelity and complexity	Comparing control group (e-module only) with two intervention groups (e-module with low fidelity text-based simulations, e-module with high-fidelity simulation game): intrinsic load of e-module was lowest of any element; game group had higher intrinsic load, germane load, and engagement; Performance on skills assessment was same among three groups
Davis 2009 ²⁶	UK	Quantitative, non-experimental	10 anesthesiology residents	Anesthesia crisis management, Simulation	Objective observations or tests	Vibrotactile secondary task	To pilot the use of a “new secondary task measure of CL” (vibrotactile device) in simulated anesthesia crisis management	During crisis portion of anesthesia simulation, response time increased and then went back to baseline
de Man 2014 ²⁷	Netherlands	Quantitative, experimental	17 anesthesiology residents, 10 anesthesiology attendings	Administering anesthesia, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether an “alarm-rich” simulated anesthesia environment would be associated with higher CL and reduced performance (time to detection of adverse events)	No difference in CL performance in alarm-free and alarm-rich simulations
DiStasi 2017 ²⁸	Spain	Quantitative, experimental	8 surgical residents, 8 surgical attendings	Laparoscopic tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, Gaze entropy and velocity	To compare cognitive demand of single- versus multi-site simulated laparoscopic surgery	CL (as measured by both gaze entropy and NASA-TLX) was greater with single- compared with multi-port approach; CL as measured by NASA-TLX was higher among residents compared with attendings
Doig 2011 ²⁹	United States	Quantitative, experimental	15 nursing students, 15 critical care nurses	Arterial blood gas interpretation, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To create and test a computer-aided tool to improve performance and reduce CL associated during simulated arterial blood gas interpretation	Accuracy and performance were better using tool compared with control group; CL was lower using the tool among nursing students, but not among practicing nurses
Dominessy 1991 ³⁰	United States	Quantitative, experimental	16 helicopter pilots	Helicopter flight, Simulation	Objective observations or tests	Reaction time to visual instructions, NASA-TLX	To compare CL during simulated helicopter flight when tactical information was presented in three formats: text, graphic, and numeric, and to determine whether performing	CL was highest for text and lowest for graphic; Secondary task reduced performance and increased CL

Supplemental table B. Included studies.

							an additional task (target acquisition search) impacted performance and CL	
Donato 2014 ³¹	United States	Quantitative, experimental	26 residents and 57 practicing physicians (all non-cardiologists)	Cardiac auscultation, Simulation	Quantitative surveys, Objective observations or tests	CL not measured	To compare CL and performance in simulated cardiac auscultation for groups trained with a “lower CL” self-study group or a “higher CL” multimedia lecture group	Participants in the “lower CL” self-study group improved more than the “higher CL” multimedia lecture group
Durning 2012 ³²	Netherlands, United States	Mixed methods, non-experimental	25 attending internists	Clinical reasoning, Simulation	Quantitative surveys, Surveys (open-ended questions), Interviews	Utterances and pauses were used to infer CL	To assess how contextual factors anticipated to increase CL (non-English-speaking patient, emotionally volatile patient, diagnostic suggestion, atypical presentation) impact clinical reasoning performance	Contextual factors anticipated to increased CL were associated with poorer diagnostic reasoning performance
Durning 2011 ³³	United States	Qualitative	25 attending internists	Patient interviewing, Simulation	Surveys (open-ended questions), Interviews	CL not measured	To qualitatively examine how contextual factors were felt to potentially impact clinical reasoning and CL	Three primary themes arose: (1) Components influencing impact of contextual factors, (2) Mechanisms for addressing contextual factors, and (3) Consequences of contextual factors for patient care; These factors were also felt likely to impact CL
Fenik 2013 ³⁴	Germany	Quantitative, experimental	14 final-year medical students, 16 medical residents	Central line insertion, Simulation	Objective observations or tests	CL not measured	To examine whether use of an “all-inclusive” central line kit during simulated central line insertion would reduce mistakes and improve performance compared with control condition in which subject has to retrieve all supplies separately	The intervention group made fewer mistakes and met more checklist items (better performance), and required less time to complete the insertion, than the control group
Flindall 2015 ³⁵	UK	Quantitative, experimental	20 medical students	Patient handoff, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare deterioration in recall of clinical information 2 hours after simulated patient handoffs for medical students who had completed a “mentally fatiguing high cognitive load task” compared to a control condition	High cognitive loading associated with reduced information freely recalled about cases
Fraser 2012 ³⁶	Canada	Quantitative, non-experimental	84 1 st -year medical students	Heart sound auscultation, Simulation	Quantitative surveys, Objective observations or tests	Paas’ Subjective Rating of Mental Effort single item scale	To determine whether emotion impacts CL during simulated cardiac auscultation, or post-training performance	Invigoration was positively correlated with CL, and tranquility negatively correlated; Higher CL associated with lower odds of correctly diagnosing a cardiac murmur previously learned, but no impact on diagnosing novel cardiac murmur

Supplemental table B. Included studies.

Fraser 2014 ³⁷	Canada	Quantitative, experimental	116 final-year medical students	Aspirin overdose management, Simulation	Quantitative surveys, Objective observations or tests	Paas’ Subjective Rating of Mental Effort single item scale	To determine the impact of unexpected patient death during simulated aspirin overdose management on CL during simulation, and performance on an observed structured clinical encounter 3 months later	Emotions were more negative and CL higher when the simulated patient unexpectedly died; performance 3 months later was lower in the group in which simulated patient died
Gaba 1990 ³⁸	United States	Quantitative, non-experimental	9 anesthesiology residents	Administering anesthesia, Workplace	Objective observations or tests	Investigator-developed multi-item instrument, Mathematics problem secondary task	To examine the use of a secondary task methodology (solving simple mathematics problems) to assess workload of anesthesiology residents during actual administration of anesthesia	Response times were higher during period requiring subjects to perform manual tasks; Response times did not correlate with level of training or case complexity
Gardner 2016 ³⁹	United States	Quantitative, experimental	41 1 st -year, 37 2 nd -year, 9 3 rd -year, and 2 4 th -year medical students	Laparoscopic suturing, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To examine how “seductive details” (interesting but tangential bits of information) affect workload and skill acquisition and transfer during simulated laparoscopic suturing	Control group had lower CL than “seductive details” group, and performed better at immediate testing and transfer
Guru 2015 ⁴⁰	United States	Quantitative, non-experimental	2 surgical residents, 3 surgical fellows, 5 surgical attendings	Robotic surgical tasks, Simulation	Objective observations or tests	EEG	To compare performance and CL when performing simulated robotic surgical tasks across groups with different levels of experience	Performance was positively associated and CL negatively associated with experience
Haji 2015 ⁴¹	Canada	Quantitative, non-experimental	8 1 st - and 2 nd -year medical students, 5 neurosurgical residents and attendings	Surgical knot tying, Simulation	Objective observations or tests	Paas’ Subjective Rating of Mental Effort single item scale, Visual secondary task	To compare CL as measured by secondary task performance and Paas’ single item during simulated surgical knot tying, among subjects of varying expertise	Experts performed better than novices; Experts had lower CL in single- but not dual-task condition; Moving from single- to dual-task increased CL to a greater degree among novices compared with experts; CL declined with practice but eventually plateaued
Haji 2015 ⁴²	Canada	Quantitative, experimental	28 1 st -year medical students	Surgical knot tying, Simulation	Quantitative surveys, Objective observations or tests	Paas’ Subjective Rating of Mental Effort single item scale, Vibrotactile secondary task	To compare sensitivity of Subjective Rating of Mental Effort and reaction time to a vibrotactile stimulus when varying intrinsic load (simple versus complex task) during simulated surgical knot tying	CL did not differ between simple and complex groups; Subjectively rated CL decreased faster in simple group but no differences in rates of decrease in response time to vibrotactile stimulus
Haji 2016 ⁴³	Canada	Quantitative, experimental	38 2 nd -year medical students	Lumbar puncture, Simulation	Quantitative surveys, Objective observations or tests	Paas’ Subjective Rating of Mental Effort single item scale, Vibrotactile secondary task	To compare effects of simple versus complex lumbar puncture training scenario on CL and performance at acquisition, retention and transfer to	Simple group had lower CL and better performance during acquisition and retention phases; Simple group had slightly better performance at transfer to very complex scenario

Supplemental table B. Included studies.

							a “very complex” lumbar puncture scenario	(fewer sterility breaches), but no differences in CL at transfer
Hautz 2017 ⁴⁴	Germany	Quantitative, experimental	49 3 rd - and 4 th -year medical students	Breast examination, Simulation	Quantitative surveys, Objective observations or tests	CL not measured	To investigate how shame affects CL and learning during simulated breast examination by comparing two groups: high-fidelity simulated breast examination and actual breast examination of a standardized patient; Inferred that shame may impact CL	Students in the standardized patient group had higher shame but also performed better, inferring that shame may not impact CL
Hsieh 2012 ⁴⁵	Taiwan	Quantitative, experimental	32 graduate students studying nuclear power plant operations	Abnormal operating condition management, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether a support system to help nuclear power plant operators manage abnormal operating procedures would reduce CL and improve performance compared with control condition during simulation	With support system, subjects made decisions more quickly (under both simple and complex conditions) and accurately (under complex conditions only), and CL was lower
Hsu 2015 ⁴⁶	Taiwan, United Kingdom	Quantitative, non-experimental	18 in-service F16 pilots	Flight, Simulation	Objective observations or tests	Eye tracking	To determine whether CL was associated with performance in simulated flight	Pupil size differed during different parts of the simulation; Better performance associated with greater gaze fixation; Experience alone was not associated with CL
Hu 2015 ⁴⁷	Singapore	Quantitative, non-experimental	47 medical students	Laparoscopic surgical tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To examine impact of experience level on performance and CL before and after training on simulated laparoscopic tasks, and to determine whether NASA-TLX could be used to assess CL during laparoscopic surgical training	CL decreased and performance improved after training; CL was higher with more complex tasks
Hussein 2016 ⁴⁸	Egypt, United States	Quantitative, non-experimental	Surgical trainee and mentor	Robotic assisted urologic surgery, Workplace	Quantitative surveys, Objective observations or tests	NASA-TLX, EEG	To assess for correlations between CL of a surgical mentor and surgical trainee during robotic-assisted urologic surgery	No correlation between mentor’s EEG and trainee’s NASA-TLX score; Several correlations were identified between NASA-TLX subscores of the mentor and the trainee, most notably that when the trainee had higher CL, the mentor was paying more attention and was less distracted
Jayaprakash 2016 ⁴⁹	United States	Quantitative, non-experimental	11 fellow and 3 attending critical care physicians	Assessment of critically ill patient, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare cognitive load of collecting a clinical history for a critically ill patient following, or in parallel with, the primary survey	No difference in CL for series versus parallel timing for collecting clinical history; CL decreased with subsequent cases

Supplemental table B. Included studies.

Jutric 2017 ⁵⁰	United States	Quantitative, experimental	5 high school students, 5 attending hepatobiliary surgeons	Targeting tumors for laparoscopic ablation, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare CL and performance during four different approaches using a novel 3D guidance system for targeting tumors in simulated laparoscopic ablation: in line with and without 3D guidance, and off-axis with and without 3D guidance	Performance was better and CL lower with 3D guidance; Novices initially performed more poorly but with final trial there was no difference between novices and experts
Kataoka 2011 ⁵¹	Japan	Quantitative, experimental	32 10 nursing students, 9 “inexperienced” nurses, 13 “experienced” nurses	Operating medication infusion pump, Simulation	Objective observations or tests	NASA-TLX, Eye tracking, Heart rate variability, Respiratory frequency	To examine how time pressure and dual tasking impacted CL of nursing students and nurses during simulated operation of a medication infusion pump	CL as measured by NASA-TLX was higher under both time pressure and dual tasking for all subjects; Eye fixation was lower for all subjects under time pressure; LF/HF was higher for students only under dual tasking
Klein 2009 ⁵²	United States	Quantitative, experimental	15 medical students	Robotic assisted surgery, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, Multiple Resources Questionnaire	To compare CL and performance under 2D and 3D vision when performing simulated robotic assisted surgical tasks	In 3D condition (compared with 2D), performance somewhat better (more transfers) and CL lower
Klein 2012 ⁵³	United States	Quantitative, experimental	15 1 st -year medical students	Laparoscopic and robotic-assisted surgical tasks, Simulation	Quantitative surveys, Objective observations or tests	Multiple Resources Questionnaire, Dundee Stress State Questionnaire	To compare CL among novices using the Foundations of Laparoscopic Surgery for performing simulated surgical tasks	CL did not differ between systems; Minor differences in performance
Lee 2014 ⁵⁴	United States	Quantitative, experimental	4 surgical residents, 6 laparoscopic surgery attendings, 3 robotic surgery attendings	Laparoscopic and robotic-assisted surgical tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare CL and performance among laparoscopic and robotic surgery novices and experts performing simulated laparoscopic and robotic surgery	Novices and robotic experts had lower CL and better performance in robotic versus laparoscopic approach; Laparoscopic experts had lower CL with laparoscopic versus robotic
Lee 2017 ⁵⁵	United States	Quantitative, non-experimental	6 neuroradiology fellows, 12 attending neuroradiologists	Interpreting neuroradiology images, Workplace	Quantitative surveys	NASA-TLX	To determine whether redesign of a radiology workspace would reduce CL required to completed daily work and improve overall satisfaction	After redesign participants reported reduced levels of CL, better work quality, and greater satisfaction
Lee 2017 ⁵⁶	United States	Quantitative, experimental	32 surgical residents and fellows	Surgical tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether feedback from an expert mentor during learning of simulated surgical tasks would impact CL and performance, compared with those who completed the learning in a self-directed fashion	Subjects that received feedback performed better and had lower CL than those not receiving feedback
Liang 2010 ⁵⁷	Taiwan, United States	Quantitative, experimental	22 “junior expert” and 20 “senior	Aircraft maintenance, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether work instructions administered to aircraft mechanics via an online maintenance	CL was lower and performance better with online maintenance assistance platform than with paper instructions

Supplemental table B. Included studies.

			expert” aircraft technicians				assistance platform would be associated with better performance and lower CL compared with simple paper instructions	
Lowndes 2015 ⁵⁸	United States	Quantitative, experimental	12 medical students	Laparoscopic single site surgery, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare CL in simulated conventional laparoscopy with two approaches to laparoscopic single site surgery (extracorporeal crossing of hands versus intracorporeal crossing of instruments)	Extracorporeal crossing of hands had higher CL and lower performance than conventional laparoscopy or intracorporeal crossing of instruments
Maxwell 2017 ⁵⁹	United States	Quantitative, experimental	573 Army National Guard soldiers	Infantry soldier training, Simulation	Surveys (open-ended questions), Objective observations or tests	NASA-TLX	To determine whether virtual, live, or combined virtual and live training of infantry soldiers would be associated with different CL and performance	Combined virtual and live training was associated with lower CL compared with virtual training alone; Combined virtual and live training was associated with better performance than either virtual or live training alone
Mazur 2014 ⁶⁰	United States	Quantitative, experimental	5 radiation oncology residents, 4 radiation oncology attendings	Radiation therapy planning, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine how CL affects errors during simulated radiation therapy planning	Higher CL was associated with more frequent errors; Residents made errors more often than faculty
Modi 2016 ⁶¹	UK	Quantitative, non-experimental	27 “higher surgical trainees”	Laparoscopic suturing, Simulation	Quantitative surveys, Objective observations or tests	SURG-TLX, Heart rate	To determine impact of temporal stressor on resident surgeons' CL and technical performance during simulated laparoscopic suturing	Time pressure increased CL and decreased performance
Mohamed 2014 ⁶²	Canada	Quantitative, non-experimental	8 1 st -year gastroenterology residents, 10 practicing gastroenterologists and colorectal surgeons	Endoscopic procedures, Workplace	Quantitative surveys	Modified NASA-TLX	To modify the NASA-TLX for use during endoscopy training and to use the modified tool to map trainees’ workload during early phases of endoscopy training	Gastroenterology residents had higher CL than attendings; CL decreased more quickly for upper endoscopy than colonoscopy procedures
Montero 2011 ⁶³	United States	Quantitative, experimental	14 surgical residents and fellows	Single- and multi-incision laparoscopic tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare CL and performance associated with simulated standard laparoscopic approach versus single incision laparoscopic surgery approach	Single incision approach was associated with higher CL and lower performance
Moore 2015 ⁶⁴	China, New Zealand, United Kingdom	Quantitative, non-experimental	32 “qualified and trainee surgeons”	Robotic and laparoscopic tasks, Simulation	Quantitative surveys, Objective observations or tests	Rating Scale for Mental Effort, SURG-TLX, Heart rate variability	To CL and performance among “qualified and trainee” surgeons performing simulated laparoscopic versus robotic surgical tasks	CL was lower and performance higher on robotic versus laparoscopic platform

Supplemental table B. Included studies.

Moos 2014 ⁶⁵	United States	Mixed methods, non-experimental	26 pre-service teachers	Teaching, Workplace	Quantitative surveys, Interviews	Paas’ Subjective Rating of Mental Effort single item scale, Investigator-developed multi-item instrument	To characterize how types and levels of CL change throughout teacher training	Overall CL decreased over the course of training; Subjects became increasingly aware of their CL throughout training and felt that their CL as teachers limited their ability to make real-time modifications during teaching
Morris 2006 ⁶⁶	Australia, Hong Kong	Quantitative, experimental	42 trainee pilots	Flight, Simulation	Objective observations or tests	CL not measured	To determine whether errors occurred more frequently with higher-CL flight simulations	Higher complexity and CL was associated with increased likelihood of errors
Mouraviev 2016 ⁶⁷	United States	Quantitative, experimental	21 surgical residents	Nephrectomy, Simulation	Quantitative surveys	NASA-TLX	To determine whether simulated nephrectomy using a virtual simulation training environment, or a porcine model, was associated with differing levels of CL	CL was similar for both platforms
Mulcock 2017 ⁶⁸	United States	Quantitative, non-experimental	140 nursing students	Care of inpatients, Workplace	Quantitative surveys	Investigator-developed single item	To determine whether having continuity of instructors during inpatient clinical rotations was associated with lower levels of stress than conventional placements	Working with the same instructor and in the same facility over first two clinical rotations was associated with reduced stress, which the authors suggest relates to reduce extraneous load
Murai 2010 ⁶⁹	Japan	Quantitative, non-experimental	10 ship piloting cadets	Ship piloting, Simulation	Objective observations or tests	Heart rate variability	To evaluate the CL of a ship cadet undergoing simulator training	Various aspects of ship navigation task were associated with varying levels of CL
Muresan 2010 ⁷⁰	United States	Quantitative, experimental	41 medical students	Intracorporeal suturing, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare CL and performance among medical students learning to perform simulated intracorporeal suturing using one of four platforms: laparoscopic suturing, laparoscopic drills, open suturing, and virtual reality drills	Laparoscopic groups had slightly reduced CL compared with other two groups; Performance varied among the four groups but was not clearly linked to level of CL
Naismith 2015 ⁷¹	Canada	Mixed methods, non-experimental	38 medical residents	Chest tube insertion, Simulation	Quantitative surveys, Interviews	Paas’ Subjective Rating of Mental Effort single item scale, NASA-TLX, Investigator-developed multi-item instrument	To compare several measures of CL among medical residents performing simulated chest tube insertion	Total CL scaled to 0-1 scale differed across the different measurement techniques but all were correlated; Qualitative themes included those related to intrinsic load (prior experience, appropriateness for level of training, task complexity), extraneous load (fidelity, anxiety, impact of hidden curriculum), and germane load (promoting effortful processing, opportunities to observe

Supplemental table B. Included studies.

								variations in practice, opportunities for repetitive practice, opportunities for expert feedback)
Rojas 2014 ⁷²	Canada	Quantitative, non-experimental	13 1 st -year medical students	One-handed surgical knot tying, Simulation	Objective observations or tests	Vibrotactile secondary task	To determine whether simple reaction time or recognition reaction time as a secondary task measure would be more sensitive to detect differences in CL as medical students learn to tie one-handed surgical knots in a simulated setting	Simple reaction time (detecting a vibrotactile stimulus) was more sensitive than recognition reaction time (detecting only one of two different types of vibrotactile stimuli)
Saleem 2007 ⁷³	United States	Mixed methods, experimental	16 practicing nurses	Using clinical reminders, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether a redesigned clinical reminder system would be associated with lower CL and improved learnability, usability, and efficiency than the existing system	Redesigned system associated with lower levels of mental demand and frustration (though only statistically significant with one-tailed t-tests); Redesigned system associated with better learnability, usability, and efficiency
Sato 2016 ⁷⁴	Japan	Quantitative, non-experimental	13 anesthesiology residents, 9 anesthesiology attendings	Administering anesthesia, Simulation	Objective observations or tests	Secondary mathematics task	To determine whether anesthesiology attendings or residents have different CL and performance during simple and complex simulated anesthesia scenarios	No difference in CL or performance for simple scenarios; For complex scenarios residents had higher CL and lower performance than residents; CL was higher in both groups for complex versus simple tasks
Saus 2006 ⁷⁵	Norway	Quantitative, experimental	40 1 st -year police academy students	Police firearm use, Simulation	Quantitative surveys, Objective observations or tests	Heart rate variability	To determine whether situational awareness training would reduce CL and improve performance among police academy students undergoing simulated firearm use training, compared with control group	Situational awareness trained group had lower CL but did not perform differently from control group
Saus 2010 ⁷⁶	Norway	Quantitative, non-experimental	32 Navy officer cadets	Ship navigation, Simulation	Quantitative surveys, Objective observations or tests	Heart rate variability	To determine whether a low- or high-CL training condition during simulated ship navigation training would impact contributors to learning	In low CL condition learning was impacted by experience, perceived realism and situational awareness; In high CL condition perceived realism and situational awareness impacted learning
Scerbo 2013 ⁷⁷	United States	Quantitative, non-experimental	7 “expert surgeons,” 11 “intermediate surgeons,” 17 surgical assistants	Laparoscopic surgical tasks, Simulation	Objective observations or tests	Visual secondary task	To determine whether a visual secondary task could distinguish among levels of expertise in simulated laparoscopic tasks	Advancing expertise associated with lower CL, better performance, and faster completion times
Scerbo 2017 ⁷⁸	United States	Quantitative, non-experimental	12 surgical residents and surgical assistants	Single- and multi-incision laparoscopic	Quantitative surveys, Objective observations or tests	Visual secondary task	To examine differences in CL and performance associated with single-incision laparoscopic procedures	CL was higher and performance lower in single-incision versus traditional laparoscopic approach

Supplemental table B. Included studies.

				surgical tasks, Simulation			compared with traditional laparoscopic procedures	
Scerbo 2017 ⁷⁹	United States	Quantitative, experimental	22 surgical assistant students, 5 premedical students	Intracorporeal suturing and knot tying, Simulation	Quantitative surveys, Objective observations or tests	Visual secondary task	To assess CL and performance associated with simulated intracorporeal suturing and knot tying over varying levels of transfer	Performance declined on initial reassessment but then returned to baseline; CL was inversely correlated with performance
Seidelman 2010 ⁸⁰	United States	Quantitative, non- experimental	14 medical students	Laparoscopic ring transfer, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To assess CL and performance associated with single versus tiled surgical displays for a simulated laparoscopic task	Single versus tiled display did not affect CL or accuracy
Sewell 2016 ⁸¹	Netherlands, United States	Quantitative, non- experimental	110 1 st -year, 179 2 nd - year, 174 3 rd -year, 12 4 th -year gastroenterology fellows	Colonoscopy, Workplace	Quantitative surveys	Investigator- developed multi- item instrument	To develop and collect validity for evidence for an instrument to estimate intrinsic, extraneous, and germane load during colonoscopy learning	Intrinsic, extraneous, and germane load were all inversely associated with prior experience
Sewell 2017 ⁸²	Netherlands, United States	Quantitative, non- experimental	110 1 st -year, 179 2 nd - year, 174 3 rd -year, 12 4 th -year gastroenterology fellows	Colonoscopy, Workplace	Quantitative surveys	Investigator- developed multi- item instrument	To identify characteristics of learners, patients/tasks, environments, and teachers that are associated with intrinsic, extraneous, and germane load during colonoscopy learning	Intrinsic load associated with year in training, prior colonoscopy experience, fatigue, patient tolerance, number of maneuvers performed, and supervisor takeover of colonoscopy; Extraneous load associated with fatigue, colonoscopy queue order, supervisor engagement and confidence, and supervisor takeover of colonoscopy; Germane load associated with supervisor engagement
Sexton 2017 ⁸³	Egypt, United States	Quantitative, non- experimental	3 lead surgeons, 3 assistant surgeons, 3 physician assistants, 7 scrub nurses, 11 circulating nurses	Robot-assisted prostatectomy, Workplace	Quantitative surveys, Objective observations or tests	NASA-TLX	To investigate how requests during robotic-assisted prostatectomy impact CL and efficiency in the operating room	Number of requests and request duration increased surgeons' CL
Shachak 2009 ⁸⁴	Canada, Israel	Qualitative	5 residents, 20 practicing primary care physicians	Electronic health record use, Workplace	Interviews	CL not measured, but was estimated qualitatively	To examine the how use of an electronic medical record impacts physicians' CL	Highest CL was perceived with diagnosing, reasoning and treating severe or multiple medical conditions; the electronic medical record was felt to reduce CL, particularly in terms of extraneous load, as results and histories were easy to find
Shewokis 2017 ⁸⁵	United States	Quantitative, experimental	10 3 rd -year medical students	Laparoscopic tasks, Simulation	Quantitative surveys, Objective observations or tests	Functional near infrared spectroscopy	To compare CL and performance in blocked versus random practice during simulated laparoscopic tasks	Compared with blocked practice, random practice was associated with lower CL and better performance

Supplemental table B. Included studies.

Sibbald 2013 ⁸⁶	Canada, Netherlands	Quantitative, experimental	191 internal medicine residents	Cardiac auscultation, Simulation	Objective observations or tests	Paas’ Subjective Rating of Mental Effort single item scale	To determine whether using a checklist during simulated cardiac auscultation learning impacts CL and diagnostic accuracy	Use of checklist was associated with reduced CL and improved accuracy compared with no checklist
Sibbald 2013 ⁸⁷	Canada, Netherlands	Quantitative, non-experimental	15 cardiology fellows	Electrocardiogram interpretation, Simulation	Quantitative surveys, Objective observations or tests	Paas’ Subjective Rating of Mental Effort single item scale	To determine whether use of a checklist during experts’ verification of electrocardiogram interpretation would reduce errors or impact CL	Using the checklist during verification of interpretation was associated with correction of errors; CL was not affected by use of checklist (i.e., expertise reversal effect was not seen)
Stefanidis 2010 ⁸⁸	United States	Quantitative, non-experimental	34 medical students	Laparoscopic and robotic intracorporeal suturing, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare laparoscopic versus robotic approach to simulated intracorporeal suturing	CL was lower and performance better on robotic versus laparoscopic platform
Strang 2014 ⁸⁹	Australia, United States	Quantitative, non-experimental	5 Royal Australian Air Force Air Battle Management Officers	Flight, Simulation	Quantitative surveys, Objective observations or tests	Overall Workload Scale, Heart rate variability	To assess for correlations between a subjective measure of CL and heart rate variability among pilots performing flight simulation	Heart rate variability and subjectively rated CL correlated, but the correlations varied across different simulator tasks
Svensson 1993 ⁹⁰	Sweden	Quantitative, non-experimental	Experiment 1: 21 ground attack pilots Experiment 2: 6 ground attack pilots	Military flight, Simulation	Quantitative surveys	Urinary adrenaline and noradrenalin	To identify contributors to CL during simulated military flight and to correlated CL with performance	Three primary constructs contributed to CL: challenge, effort, and performance; Performance was inversely correlated with CL and challenge positively correlated; CL decreased with training
Szulewski 2014 ⁹¹	Canada	Mixed methods, non-experimental	1 medical student, 3 emergency medicine residents, 1 emergency medicine attending	Resuscitation team leading, Simulation	Interviews, Video and audio recordings	CL not measured	To assess feasibility of using eye tracking technology to monitor CL in simulated resuscitation team leading	Eye tracking device did not interfere with simulation tasks
Tattersall 1995 ⁹²	UK	Quantitative, non-experimental	11 trainee flight engineers	Diagnosing flight faults and incidents, Simulation	Quantitative surveys, Objective observations or tests	Subjective Workload Assessment Technique, Heart rate variability	To monitor “naturally occurring” fluctuations in CL during simulated diagnosis of flight faults and incidents, and to compare subjective ratings of CL with heart rate variability	Heart rate variability lower during more challenging simulation tasks; Subjective assessment of CL correlated with heart rate variability
Teel 2009 ⁹³	Japan, United States	Quantitative, non-experimental	10 Marine cadets	Ship navigation, Simulation	Objective observations or tests	Heart rate variability	To evaluate the CL of a cadet undergoing simulator training using heart rate variability	Heart rate variability varied with different tasks during simulation
Theodoraki 2015 ⁹⁴	Germany	Quantitative, experimental	8 otorhinolaryngology residents, 2 otorhinolaryngology attendings	Functional endoscopic sinus surgery, Workplace	Objective observations or tests	Heart rate variability, Respiratory frequency, Masticator muscle EMG	To compare CL and mental distress of surgeons performing functional endoscopic sinus surgery with and without a navigation system	Navigation system did not impact CL; CL was higher in less experienced group

Supplemental table B. Included studies.

Thomas 2017 ⁹⁵	United States	Quantitative, non-experimental	79 practicing nurses	Administering medications to patients, Workplace	Quantitative surveys, Objective observations or tests	NASA-TLX	To examine how interruptions and distractions impact CL among nurses administering medications to patients in the workplace, and to assess for impact on medication administration errors	CL increased with number of distractions and interruptions; CL did not impact medication administration errors (though the absolute number of errors was very low overall)
Tien 2015 ⁹⁶	UK	Quantitative, non-experimental	7 “expert” surgeons (attendings and senior residents), 6 “junior” surgeons	Open inguinal hernia repair, Workplace	Quantitative surveys, Objective observations or tests	NASA-TLX, Eye tracking	To assess differences in CL and gaze behavior between expert and junior surgeons during open inguinal hernia repair	CL was higher for junior surgeons; Experts had higher fixation frequency and dwell time during particular parts of the procedure
Tomasko 2012 ⁹⁷	United States	Quantitative, experimental	31 3 rd - and 4 th year medical students	Surgical tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, Visual secondary task	To compare CL and performance during simulated surgical tasks in sleep-deprived versus non-sleep-deprived states	CL was higher in sleep-deprived group, but performance did not differ
Tremblay 2017 ⁹⁸	Canada, Netherlands	Mixed methods, non-experimental	143 2 nd -year undergraduate pharmacy students	Medication dispensing and patient counseling, Simulation	Quantitative surveys, Objective observations or tests, Focus groups	Investigator-developed multi-item instrument	To determine whether a more authentic “simulated clinical environment” is associated with different intrinsic and extraneous CL compared with a less authentic environment among undergraduate pharmacy students performing simulated medication dispensing and counseling with a standardized patient	Both intrinsic and extraneous load were higher in simulated clinical immersion versus standard environment; Simulated clinical environment was associated with higher self-perceived learning; Students felt they could learn clinical reasoning better in standard environment but could learn to accomplish technical tasks better in simulated clinical environment
Walker 2015 ⁹⁹	Canada	Qualitative	3 2 nd -year dental students, 1 dental attending	Tooth mold production, Simulation	Interviews, Video recordings	Investigator-developed single item	To identify aspects of simulated tooth mold production that were associated with CL, comparing a dental attending with dental students	Many parts of the simulation were noted by students to have high CL, but only a single part was noted by the attending to have high CL
Weigl 2015 ¹⁰⁰	Germany	Quantitative, non-experimental	63 attending surgeons	General and orthopedic surgical operations, Workplace	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether interruptions during actual surgical operations were associated with higher CL, and to identify other factors impacting CL, among attending surgeons	Situational stress was higher during teaching than non-teaching cases; CL was higher for longer cases; Interruptions were associated with increased CL (distraction component of NASA-TLX)
Weigl 2016 ¹⁰¹	Germany	Quantitative, experimental	19 1 st - and 2 nd -year surgical residents	Vertebroplasty surgery, Simulation	Quantitative surveys, Objective observations or tests	SURG-TLX	To examine the impact of interruptions on resident surgeons’ CL and performance on simulated vertebroplasty	CL was higher and performance lower in disrupted scenario; Patient discomfort as source of disruption caused greater increase to CL than telephone call

Supplemental table B. Included studies.

Weinger 1994 ¹⁰²	United States	Quantitative, non-experimental	11 1 st -year anesthesiology residents, 3 3 rd -year anesthesiology residents, 8 certified registered nurse anesthetists	Administering anesthesia, Workplace	Quantitative surveys, Objective observations or tests	Visual secondary task, Borg’s workload scale	To develop and to validate objective techniques of CL measurement of anesthesiologists in an actual operating room environment, and to compare CL of novices and experienced anesthesiologists	CL was higher in novices than experienced subjects; Subjective instrument correlated with visual secondary task
Weinger 2000 ¹⁰³	United States	Quantitative, non-experimental	33 anesthesiology residents, 10 certified registered nurse anesthetists, 83 attending anesthesiologists	Airway management, Workplace	Quantitative surveys	Investigator-developed multi-item instrument	To measure the CL associated with 10 different airway management tasks among three levels of anesthesia providers	CL varied across the 10 tasks; CL did not differ by type or level of provider
Wilson 2011 ¹⁰⁴	China, United Kingdom	Quantitative, non-experimental	30 medical students	Laparoscopic tasks, Simulation	Quantitative surveys, Objective observations or tests	SURG-TLX	To develop and validate the SURG-TLX as a measure of CL among surgeons	The results supported discriminant sensitivity of the SURG-TLX to different sources of stress
Workman 2007 ¹⁰⁵	United States	Quantitative, experimental	42 3 rd - and 4 th -year medical students, 3 board-certified intensivists	Interpreting data in intensive care unit, Simulation	Objective observations or tests	Visual secondary task	To determine whether a novel graphical format for information display (knowledge-enhanced graphical symbols, KEGS) would reduce CL and improve performance interpreting data in simulated critical care medicine setting, as compared with traditional textual display	CL was lower among medical students using KEGS versus traditional display; among attendings, CL was lower and accuracy higher using KEGS
Wucherer 2015 ¹⁰⁶	Canada, Germany	Quantitative, non-experimental	19 “junior surgeons”	Vertebroplasty surgery, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To estimate CL among “junior surgeons” and assess for relationships with performance, during simulated surgical crisis setting compared with control simulation setting	CL was higher and work slower during crisis compared with control simulation
Young 2010 ¹⁰⁷	United States	Quantitative, experimental	16 hypothetical psychiatry residents	Ambulatory psychiatric care, Workplace	Objective observations or tests	CL not measured	To compare predicted ambulatory caseloads for incoming psychiatry residents a traditional model versus novel CL-based model	CL-based model resulted in more evenly balanced patient panels with less inter-caseload variation
Young 2016 ¹⁰⁸	Netherlands, United States	Quantitative, experimental	23 2 nd -year and 29 6 th -year medical students	Patient handovers, Simulation	Quantitative surveys, Objective observations or tests	Paas’ Subjective Rating of Mental Effort single item scale	To examine impact of simulated handover complexity on medical students’ CL and handover accuracy	CL was negatively correlated with handover accuracy and illness script maturity (measure of experience); Illness script maturity was positively correlated with handover accuracy
Young, 2016 ¹⁰⁹	Netherlands, United States	Quantitative, non-experimental	54 2 nd -year and 33 3 rd -year medical students	Patient handovers, Simulation	Quantitative surveys, Objective observations or tests	Paas’ Subjective Rating of Mental Effort single item scale,	To develop an instrument to measure subtypes of CL among medical students performing simulated handovers, and to assess for	2-factor model for CL developed including intrinsic and germane load; intrinsic load and germane load were higher in less experienced subjects;

Supplemental table B. Included studies.

						Investigator-developed multi-item instrument	relationships between intrinsic and germane load with experience level and performance	No significant associations between CL and performance; Paas’ single item positively correlated with intrinsic but not germane load
Young, 2016 ¹¹⁰	Netherlands, United States	Quantitative, non-experimental	23 2 nd -year and 29 6 th -year medical students	Patient handovers, Simulation	Quantitative surveys	Paas’ Subjective Rating of Mental Effort single item scale, Investigator-developed multi-item instrument	To provide evidence for validity of a revised instrument to measure CL during handovers, and to assess for relationships between CL, case complexity, and experience	Intrinsic and germane load formed a single factor, which was positively correlated with handover complexity and negatively correlated with experience
Yu, 2016 ¹¹¹	United States	Quantitative, non-experimental	4 anesthesiologists, 12 certified registered nurse anesthetists, 38 circulating nurses, 35 surgical technicians, 26 surgical assistants, 45 surgical residents, 32 attending surgeons	Surgical operations, Workplace	Quantitative surveys, Objective observations or tests	SURG-TLX	To quantify and compare CL among surgical team members across different surgical techniques and specialties	CL was highest among surgical residents (most novice team members); Ancillary staff (circulating nurse, surgical technician) had lowest CL; Surgical duration positively associated with overall team CL
Yurko 2010 ¹¹²	United States	Quantitative, non-experimental	28 2 nd -year medical students and senior premedical students	Laparoscopic suturing, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To examine relationships between CL and performance among medical and premedical students performing simulated laparoscopic suturing, first on Fundamentals of Laparoscopic Surgery model, and then on transfer to porcine model	CL and performance were negatively correlated; Higher CL associated with more damage to porcine tissues; Performance increased and CL decreased on post-training testing, but then reversed on transfer to a porcine model
Zheng 2010 ¹¹³	Canada	Quantitative, non-experimental	12 surgical residents, 9 surgical fellows and attendings	Laparoscopic tasks, Simulation	Objective observations or tests	Visual secondary task	To compare CL and performance while performing simulated laparoscopic tasks among experienced versus novice laparoscopic surgeons	Experienced surgeons had lower CL and better performance than novice surgeons
Zheng 2012 ¹¹⁴	Canada	Quantitative, experimental	5 novice and 5 expert surgeons	Laparoscopic and natural orifice transluminal endoscopic surgery tasks, Simulation	Objective observations or tests	NASA-TLX, Visual secondary task	To compare CL and performance among novice and expert surgeons performing simulated laparoscopic and natural orifice transluminal endoscopic surgery (NOTES)	CL was higher in NOTES versus laparoscopic procedures (based on secondary task, but not based on NASA-TLX) but this was attenuated by prior experience; Performance lower in NOTES procedures and higher among those with greater prior experience

Supplemental table B. Included studies.

Zheng 2012 ¹¹⁵	Canada	Quantitative, non-experimental	23 surgical residents, fellows, and attendings	Laparoscopic cholecystectomy, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, Eye tracking	To correlate eye tracking and NASA-TLX as means to estimate CL during simulated laparoscopic cholecystectomy	Higher CL was associated with fewer blinks and shorter blink duration; Performance did not vary with eye tracking data
Zheng 2014 ¹¹⁶	Canada	Quantitative, experimental	12 surgical residents, 3 laparoscopic fellows, 2 surgical attendings	Using surgical instruments, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To examine how complexity (open versus laparoscopic versus endoscopic approach) impacts surgeons' CL and performance during simulated use of surgical tools	CL was higher in most complex approach (endoscopic) versus least complex approach (open); Time to complete task higher in more complex approaches; With practice, performance improved but CL did not change

Abbreviations: CL, cognitive load; EEG, electroencephalogram; EMG, electromyography; LF/HF, low frequency/high frequency (higher ratio suggests higher sympathetic tone and higher CL); NASA-TLX, NASA Task Load Index; SURG-TLX, Surgery Task Load Index

^aDescription of participants is provided using the level of detail included in each study manuscript.

^bFor consistency and ease of table interpretation, we use the term CL (cognitive load) when referring to the constructs of cognitive load, mental effort, and mental workload.

REFERENCES FOR SUPPLEMENTAL TABLE B

1. Andersen SA, Mikkelsen PT, Konge L, Caye-Thomasen P, Sorensen MS. Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. *Laryngoscope*. 2015;126:E74-E79.
2. Andersen SA, Mikkelsen PT, Konge L, Cayé-Thomasen P, Sørensen M. Cognitive Load in Mastoidectomy Skills Training: Virtual Reality Simulation and Traditional Dissection Compared. *J Surg Educ*. 2016;73(1):45-50.
3. Andersen SA, Konge L, Cayé-Thomasen P, Sørensen MS. Retention of mastoidectomy skills after virtual reality simulation training. *JAMA Otolaryngol Head Neck Surg*. 2016;142(7):635-640.
4. Andrade AD, Cifuentes P, Mintzer MJ, Roos BA, Anam R, Ruiz JG. Simulating Geriatric Home Safety Assessments in a Three-Dimensional Virtual World. *Gerontol Geriatr Educ*. 2012;33(3):233-252.
5. Arico P, Borghini G, Di Flumeri G, et al. Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Front Hum Neurosci*. 2016;10:539.
6. Attrill S, Lincoln M, McAllister S. Supervising international students in clinical placements: perceptions of experiences and factors influencing competency development. *BMC Med Educ*. 2016;16:180.
7. Avansino J, Leu MG. Effects of CPOE on provider cognitive workload: A randomized crossover trial. *Pediatrics*. 2012;130(3):e547-e552.
8. Bertram DA, Hershey CO, Opila DA, Quirin O. A measure of physician mental work load in internal medicine ambulatory care clinics. *Med Care*. 1990;28(5):458-467.
9. Bertram DA, Opila DA, Brown JL, et al. Measuring physician mental workload: reliability and validity assessment of a brief instrument. *Med Care*. 1992;30(2):95-104.
10. Bharathan R, Vali S, Setchell T, Miskry T, Darzi A, Aggarwal R. Psychomotor skills and cognitive load training on a virtual reality laparoscopic simulator for tubal surgery is effective. *Eur J Obstet Gynecol Reprod Biol*. 2013;169(2):347-352.
11. Blissett S, Morrison D, McCarty D, Sibbald M. Should learners reason one step at a time? A randomised trial of two diagnostic scheme designs. *Med Educ*. 2017;51(4):432-441.
12. Boet S, Sharma B, Pigford AA, Hladkowicz E, Rittenhouse N, Grantcharov T. Debriefing decreases mental workload in surgical crisis: A randomized controlled trial. *Surgery*. 2017;161(5):1215-1220.
13. Britt RC, Scerbo MW, Montano M, Kennedy RA, Prytz E, Stefanidis D. Intracorporeal suturing: Transfer from Fundamentals of Laparoscopic Surgery to cadavers results in substantial increase in mental workload. *Surgery*. 2015;158(5):1428-1433.
14. Broyles TW, Epler CM, Waknine JW. A case study examining the impact of cognitive load on reflection and pre-service teachers' transfer of specific teaching behaviors. *Career and Technical Education Research*. 2011;36(1):49-66.
15. Byrne AJ, Sellen AJ, Jones JG. Errors on anaesthetic record charts as a measure of anaesthetic performance during simulated critical incidents. *Br J Anaesth*. 1998;80(1):58-62.
16. Byrne AJ, Murphy A, McIntyre O, Tweed N. The relationship between experience and mental workload in anaesthetic practice: an observational study. *Anaesthesia*. 2013;68(12):1266-1272.
17. Cao CGL, Zhou M, Jones DB, Schwaitzberg SD. Can surgeons think and operate with haptics at the same time? *J Gastrointest Surg*. 2007;11(11):1564-1569.

Supplemental table B. Included studies.

18. Chen R, Grierson LE, Norman GR. Evaluating the impact of high- and low-fidelity instruction in the development of auscultation skills. *Med Educ*. 2015;49(3):276-285.
19. Chen R, Grierson L, Norman G. Manipulation of cognitive load variables and impact on auscultation test performance. *Adv Health Sci Educ Theory Pract*. 2015;20(4):935-952.
20. Chowriappa A, Raza SJ, Fazili A, et al. Augmented-reality-based skills training for robot-assisted urethrovesical anastomosis: A multi-institutional randomised controlled trial. *BJU Int*. 2015;115(2):336-345.
21. Chui MA, Stone JA. Exploring information chaos in community pharmacy handoffs. *Res Social Adm Pharm*. 2014;10(1):195-203.
22. Crosby JJ, Parkinson SR. A dual task investigation of pilots' skill level. *Ergonomics*. 1979;22(12):1301-1313.
23. Cummings ML, Smith CA. Improving precision vertical flight operations through a direct-perception-action display. *Journal of Aircraft*. 2009;46(4):1270-1279.
24. Dahlstrom N, Nahlinder S. Mental workload in aircraft and simulator during basic civil aviation training. *International Journal of Aviation Psychology*. 2009;19(4):309-325.
25. Dankbaar ME, Alisma J, Jansen EE, van Merrienboer JJ, van Saase JL, Schuit SC. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. *Adv Health Sci Educ Theory Pract*. 2016;21(3):505-521.
26. Davis DHJ, Oliver M, Byrne AJ. A novel method of measuring the mental workload of anaesthetists during simulated practice. *Br J Anaesth*. 2009;103(5):665-669.
27. de Man FR, Erwtaman M, van Groeningen D, et al. The effect of audible alarms on anaesthesiologists' response times to adverse events in a simulated anaesthesia environment: a randomised trial. *Anaesthesia*. 2014;69(6):598-603.
28. Di Stasi LL, Díaz-Piedra C, Ruiz-Rabelo JF, Rieiro H, Sanchez Carrion JM, Catena A. Quantifying the cognitive cost of laparo-endoscopic single-site surgeries: Gaze-based indices. *Appl Ergon*. 2017;65:168-174.
29. Doig AK, Albert RW, Syroid ND, Moon S, Agutter JA. Graphical arterial blood gas visualization tool supports rapid and accurate data interpretation. *Comput Inform Nurs*. 2011;29(4 Suppl):TC53-60.
30. Dominessy ME, Lukas JH, Malkin FJ, Monty RA, Oatman LC. The effect of information display formats on helicopter pilots' target acquisition and flying performance. *Mil Psychol*. 1991;3(3):163-176.
31. Donato AA, Kaliyadan AG, Wasser T. Self-directed study using MP3 players to improve auscultation proficiency of physicians: a randomized, controlled trial. *J Contin Educ Health Prof*. 2014;34(2):131-138.
32. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?). *Adv Health Sci Educ Theory Pract*. 2012;17(1):65-79.
33. Durning S, Artino AR, Jr., Pangaro L, van der Vleuten CP, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ*. 2011;45(9):927-938.
34. Fenik Y, Celebi N, Wagner R, et al. Prepackaged central line kits reduce procedural mistakes during central line insertion: a randomized controlled prospective trial. *BMC Med Educ*. 2013;13:60.
35. Flindall IR, Leff DR, Pucks N, Sugden C, Darzi A. The Preservation of Cued Recall in the Acute Mentally Fatigued State: A Randomised Crossover Study. *World J Surg*. 2016;40:56-65.
36. Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K. Emotion, cognitive load and learning outcomes during simulation training. *Med Educ*. 2012;46(11):1055-1062.
37. Fraser K, Huffman J, Ma I, et al. The emotional and cognitive impact of unexpected simulated patient death: a randomized controlled trial. *Chest*. 2014;145(5):958-963.
38. Gaba DM, Lee T. Measuring the workload of the anesthesiologist. *Anesth Analg*. 1990;71(4):354-361.
39. Gardner AK, Clanton J, Jabbour, II, Scott L, Scott DJ, Russo MA. Impact of seductive details on the acquisition and transfer of laparoscopic suturing skills: Emotionally interesting or cognitively taxing? *Surgery*. 2016;160(3):580-585.
40. Guru KA, Esfahani ET, Raza SJ, et al. Cognitive skills assessment during robot-assisted surgery: separating the wheat from the chaff. *BJU Int*. 2015;115(1):166-174.
41. Haji FA, Khan R, Regehr G, Drake J, de Ribaupierre S, Dubrowski A. Measuring cognitive load during simulation-based psychomotor skills training: sensitivity of secondary-task performance and subjective ratings. *Adv Health Sci Educ Theory Pract*. 2015;20(5):1237-1253.
42. Haji FA, Rojas D, Childs R, Ribaupierre S, Dubrowski A. Measuring cognitive load: performance, mental effort and simulation task complexity. *Med Educ*. 2015;49(8):815-827.
43. Haji FA, Cheung JJ, Woods N, Regehr G, de Ribaupierre S, Dubrowski A. Thrive or overload? The effect of task complexity on novices' simulation-based learning. *Med Educ*. 2016;50(9):955-968.
44. Hautz WE, Schroder T, Dannenberg KA, et al. Shame in Medical Education: A Randomized Study of the Acquisition of Intimate Examination Skills and Its Effect on Subsequent Performance. *Teach Learn Med*. 2017;29(2):196-206.
45. Hsieh MH, Hwang SL, Liu KH, Liang SFM, Chuang CF. A decision support system for identifying abnormal operating procedures in a nuclear power plant. *Nuclear Engineering and Design*. 2012;249:413-418.
46. Hsu CK, Lin SC, Li WC. Visual movement and mental-workload for pilot performance assessment. *Lecture Notes in Computer Science*. 2015;9174:356-364.

Supplemental table B. Included studies.

47. Hu JS, Lu J, Tan WB, Lomanto D. Training improves laparoscopic tasks performance and decreases operator workload. *Surg Endosc.* 2015;30:1742-1746.
48. Hussein AA, Shafiei SB, Sharif M, et al. Technical Mentorship during Robot-Assisted Surgery: A Cognitive Analysis? *BJU Int.* 2016;118(3):429-436.
49. Jayaprakash N, Ali R, Kashyap R, Bennett C, Kogan A, Gajic O. The incorporation of focused history in checklist for early recognition and treatment of acute illness and injury. *BMC Emerg Med.* 2016;16(1):35.
50. Jutric Z, Grendar J, Brown WL, et al. Novel Simulation Device for Targeting Tumors in Laparoscopic Ablation: A Learning Curve Study. *Surg Innov.* 2017;24(5):492-498.
51. Kataoka J, Sasaki M, Kanda K. Effects of mental workload on nurses' visual behaviors during infusion pump operation. *Jpn J Nurs Sci.* 2011;8(1):47-56.
52. Klein MI, Lio CH, Grant R, Carswell CM, Strup S. A mental workload study on the 2d and 3d viewing conditions of the da Vinci surgical robot. *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting.* 2009;2:1186-1190.
53. Klein MI, Warm JS, Riley MA, et al. Mental workload and stress perceived by novice operators in the laparoscopic and robotic minimally invasive surgical interfaces. *J Endourol.* 2012;26(8):1089-1094.
54. Lee GI, Lee MR, Clanton T, Sutton E, Park AE, Marohn MR. Comparative assessment of physical and cognitive ergonomics associated with robotic and traditional laparoscopic surgeries. *Surg Endosc.* 2014;28(2):456-465.
55. Lee MH, Schemmel AJ, Pooler BD, et al. Radiology Workflow Dynamics: How Workflow Patterns Impact Radiologist Perceptions of Workplace Satisfaction. *Acad Radiol.* 2017;24(4):483-487.
56. Lee GI, Lee MR. Can a virtual reality surgical simulation training provide a self-driven and mentor-free skills learning? Investigation of the practical influence of the performance metrics from the virtual reality robotic surgery simulator on the skill learning and associated cognitive workloads. *Surg Endosc.* 2018;32(1):62-72.
57. Liang GF, Lin JT, Hwang SL, Wang EM, Patterson P. Preventing human errors in aviation maintenance using an on-line maintenance assistance platform. *International Journal of Industrial Ergonomics.* 2010;40(3):356-367.
58. Lowndes B, Abdelrahman A, McCrory B, Hallbeck S. A Preliminary Study of Novice Workload and Performance During Surgical Simulation Tasks for Conventional vs. Single Incision Laparoscopic Techniques. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting.* 2015:498-502.
59. Maxwell DZ, J. Large Scale Testing and Evaluation of Virtual Environments for Infantry Soldier Tasks Comparing Mental Effort for Live Versus Virtual Training Performance Assessments. *MODSIM World 2017*; April 25-27, 2017; Virginia Beach, VA.
60. 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.
61. Modi H, Leff D, Singh H, Darzi A. The influence of cognitive load on technical ability among surgical trainees. *Int J Surg.* 2016;36:S34-S35.
62. Mohamed R, Raman M, Anderson J, McLaughlin K, Rostom A, Coderre S. Validation of the National Aeronautics and Space Administration Task Load Index as a tool to evaluate the learning curve for endoscopy training. *Can J Gastroenterol Hepatol.* 2014;28(3):155-159.
63. Montero PN, Acker CE, Heniford BT, Stefanidis D. Single incision laparoscopic surgery (SILS) is associated with poorer performance and increased surgeon workload compared with standard laparoscopy. *Am Surg.* 2011;77(1):73-77.
64. Moore LJ, Wilson MR, McGrath JS, Waine E, Masters RSW, Vine SJ. Surgeons' display reduced mental effort and workload while performing robotically assisted surgical tasks, when compared to conventional laparoscopy. *Surg Endosc.* 2015;29(9):2553-2560.
65. Moos DC, Pitton D. Student Teacher Challenges: Using the Cognitive Load Theory as an Explanatory Lens. *Teaching Education.* 2014;25(2):127-141.
66. Morris CH, Leung YK. Pilot mental workload: how well do pilots really perform? *Ergonomics.* 2006;49(15):1581-1596.
67. Mouraviev V, Klein M, Schommer E, et al. Urology residents experience comparable workload profiles when performing live porcine nephrectomies and robotic surgery virtual reality training modules. *J Robot Surg.* 2016;10(1):49-56.
68. Mulcock PM, Grassley J, Davis M, White K. Beyond the Dedicated Education Unit: Using Cognitive Load Theory to Guide Clinical Placement. *J Nurs Educ.* 2017;56(2):105-109.
69. Murai K, Ishikura A, Hayashi Y, et al. Toward evaluation of ship bridge teamwork for marine simulator training based on heart rate variability. *World Automation Congress.* 2010:1-6.
70. Muresan III C, Lee TH, Seagull J, Park AE. Transfer of training in the development of intracorporeal suturing skill in medical student novices: A prospective randomized trial. *Am J Surg.* 2010;200(4):537-541.
71. Naismith LM, Cheung JJH, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. *Med Educ.* 2015;48(8):805-814.
72. Rojas D, Haji F, Shewaga R, Kapralos B, Dubrowski A. The impact of secondary-task type on the sensitivity of reaction-time based measurement of cognitive load for novices learning surgical skills using simulation. *Stud Health Technol Inform.* 2014;196:353-359.

Supplemental table B. Included studies.

73. Saleem JJ, Patterson ES, Militello L, et al. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. *J Am Med Inform Assoc.* 2007;14(5):632-640.
74. Sato H, Miyashita T, Kawakami H, Nagamine Y, Takaki S, Goto T. Influence of Mental Workload on the Performance of Anesthesiologists during Induction of General Anesthesia: A Patient Simulator Study. *BioMed Research International.* 2016;2016:1058750.
75. Saus E, Johnsen BH, Eid J, Riisem PK, Andersen R, Thayer JF. The effect of brief situational awareness training in a police shooting simulator: an experimental study. *Mil Psychol.* 2006;18(Suppl):S3-S21.
76. Saus E, Johnsen BH, Eid J. Perceived learning outcome: the relationship between experience, realism and situation awareness during simulator training. *Int Marit Health.* 2010;62(4):258-264.
77. Scerbo MW, Kennedy A, Montano M, Britt RC, Davis SS, Stefanidis D. A Spatial Secondary Task for Measuring Laparoscopic Mental Workload: Differences in Surgical Experience. *Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting.* 2013:728-732.
78. Scerbo MW, Britt RC, Stefanidis D. Differences in mental workload between traditional and single-incision laparoscopic procedures measured with a secondary task. *Am J Surg.* 2017;213(2):244-248.
79. Scerbo MW, Britt RC, Montano M, Kennedy RA, Prytz E, Stefanidis D. Effects of a retention interval and refresher session on intracorporeal suturing and knot tying skill and mental workload. *Surgery.* 2017;161(5):1209-1214.
80. Seidelman W, Carswell CM, Lio CH, et al. Potential performance costs associated with large-format tiled displays for surgical visualization. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting.* 2010;2:1430-1434.
81. 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.
82. 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.
83. Sexton K, Johnson A, Gotsch A, Hussein AA, Cavuoto L, Guru KA. Anticipation, teamwork and cognitive load: chasing efficiency during robot-assisted surgery. *BMJ Qual Saf.* 2018;27(2):148-154.
84. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. *J Gen Intern Med.* 2009;24(3):341-348.
85. Shewokis PA, Shariff FU, Liu Y, Ayaz H, Castellanos A, Lind DS. Acquisition, retention and transfer of simulated laparoscopic tasks using fNIR and a contextual interference paradigm. *Am J Surg.* 2017;213(2):336-345.
86. Sibbald M, de Bruin AB, Cavalcanti RB, van Merrienboer JIG. Do you have to re-examine to reconsider your diagnosis? Checklists and cardiac exam. *BMJ Qual Saf.* 2013;22(4):333-338.
87. Sibbald M, de Bruin ABH, van Merrienboer JIG. Checklists improve experts' diagnostic decisions. *Med Educ.* 2013;47(3):301-308.
88. Stefanidis D, Wang F, Korndorffer JR, Jr., Dunne JB, Scott DJ. Robotic assistance improves intracorporeal suturing performance and safety in the operating room while decreasing operator workload. *Surg Endosc.* 2010;24(2):377-382.
89. Strang AJ, Best C, Funke GJ. Heart rate correlates of mental workload in a large-scale air-combat simulation training exercise. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting.* 2014:2325-2329.
90. Svensson E, Angelborg-Thanderz M, Sjöberg L. Mission challenge, mental workload and performance in military aviation. *Aviat Space Environ Med.* 1993;64(11):985-991.
91. Szulewski A, Howes D. Combining first-person video and gaze-tracking in medical simulation: A technical feasibility study. *The Scientific World Journal.* 2014;2014(975752):eCollection 2014.
92. Tattersall AJ, Hockey GRJ. Level of operator control and changes in heart rate variability during simulated flight maintenance. *Hum Factors.* 1995;37(4):682-698.
93. Teel S, Sanders J, Parrott DS, et al. Evaluation of marine simulator training based on heart rate variability. *Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics.* 2009:233-238.
94. Theodoraki MN, Ledderose GJ, Becker S, et al. Mental distress and effort to engage an image-guided navigation system in the surgical training of endoscopic sinus surgery: a prospective, randomised clinical trial. *Eur Arch Otorhinolaryngol.* 2015;272(4):905-913.
95. Thomas L, Donohue-Porter P, Stein Fishbein J. Impact of Interruptions, Distractions, and Cognitive Load on Procedure Failures and Medication Administration Errors. *J Nurs Care Qual.* 2017;32(4):309-317.
96. Tien T, Pucher PH, Sodergren MH, Sriskandarajah K, Yang GZ, Darzi A. Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. *Surg Endosc.* 2015;29(2):405-413.
97. Tomasko JM, Pauli EM, Kunselman AR, Haluck RS. Sleep deprivation increases cognitive workload during simulated surgical tasks. *Am J Surg.* 2012;203(1):37-43.
98. Tremblay ML, Lafleur A, Leppink J, Dolmans DH. The simulated clinical environment: Cognitive and emotional impact among undergraduates. *Med Teach.* 2017;39(2):181-187.
99. Walker J, von Bergmann H. Lessons from a pilot project in cognitive task analysis: the potential role of intermediates in preclinical teaching in dental education. *J Dent Educ.* 2015;79(3):286-294.

Supplemental table B. Included studies.

100. Weigl M, Antoniadis S, Chiapponi C, Bruns C, Sevdalis N. The impact of intra-operative interruptions on surgeons' perceived workload: an observational study in elective general and orthopedic surgery. *Surg Endosc.* 2015;29(1):145-153.
101. Weigl M, Stefan P, Abhari K, et al. Intra-operative disruptions, surgeon's mental workload, and technical performance in a full-scale simulated procedure. *Surg Endosc.* 2016;30(2):559-566.
102. Weinger MB, Herndon OW, Zornow MH, Paulus MP, Gaba DM, Dallen LT. An objective methodology for task analysis and workload assessment in anesthesia providers. *Anesthesiology.* 1994;80(1):77-92.
103. Weinger MB, Vredenburg AG, Schumann CM, et al. Quantitative description of the workload associated with airway management procedures. *J Clin Anesth.* 2000;12(4):273-282.
104. Wilson MR, Poolton JM, Malhotra N, Ngo K, Bright E, Masters RS. Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). *World J Surg.* 2011;35(9):1961-1969.
105. Workman M, Lesser MF, Kim J. An exploratory study of cognitive load in diagnosing patient conditions. *Int J Qual Health Care.* 2007;19(3):127-133.
106. Wucherer P, Stefan P, Abhari K, et al. Vertebroplasty Performance on Simulator for 19 Surgeons Using Hierarchical Task Analysis. *IEEE Trans Med Imaging.* 2015;34(8):1730-1737.
107. Young JQ, Niehaus B, Lieu SC, O'Sullivan PS. Improving resident education and patient safety: a method to balance initial caseloads at academic year-end transfer. *Acad Med.* 2010;85(9):1418-1424.
108. 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.
109. 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. *Perspect Med Educ.* 2016;5(1):24-32.
110. Young JQ, Boscardin CK, van Dijk SM, et al. Performance of a cognitive load inventory during simulated handoffs: Evidence for validity. *SAGE Open Med.* 2016;4:2050312116682254.
111. 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.
112. Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D. Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simul Healthc.* 2010;5(5):267-271.
113. Zheng B, Cassera MA, Martinec DV, Spaun GO, Swanström LL. Measuring mental workload during the performance of advanced laparoscopic tasks. *Surg Endosc.* 2010;24(1):45-50.
114. Zheng B, Rieder E, Cassera MA, et al. Quantifying mental workloads of surgeons performing natural orifice transluminal endoscopic surgery (NOTES) procedures. *Surg Endosc.* 2012;26(5):1352-1358.
115. Zheng B, Jiang X, Tien G, Meneghetti A, Panton ONM, Atkins MS. Workload assessment of surgeons: Correlation between NASA TLX and blinks. *Surg Endosc.* 2012;26(10):2746-2750.
116. Zheng B, Fu B, Al-Tayeb TA, Hao YF, Qayumi AK. Mastering instruments before operating on a patient: the role of simulation training in tool use skills. *Surg Innov.* 2014;21(6):637-642.