Supplemental Materials: Table S1. Comparing laparoscopic simulation performance to VSA:

Simulation Task	VSA Test (VSA component)	Reference	Key Points of Authors' Interpretations
MIST-VR	Purdue Spatial Visualization Test (visualization)	Birbas et al.,[54] 2006	VSA predicted performance on MIST-VR for novice laparoscopic surgeons.
	HFT (flexibility of closure)	Gallagher et al.,[59] 2001	Urologists' performance on the VSA task did not significantly differ from the general population.
	MRT-A, MRT-C, SDT, PFT (visualization); CC (spatial orientation)	Hedman et al.,[22] 2006	VSA scores for novice surgery trainees were significantly correlated to performance on the instrument navigation, suggesting VSA scores may predict early performance on the key surgical task for novices.
	MRT-A (visualization)	Hedman et al.,[47] 2007	VSA scores were significantly correlated with several performance scores on the MIST manipulative diathermy and GI Mentor II tasks (such as efficiency of screening and time to complete task).
	CC, CR (spatial orientation); PicSOr (depth perception); MPT (spatial scanning)	McClusky et al.,[56] 2005	No VSA tests were significantly correlated to duration of training, however the number of trials to reach criterion-based performance goals on the MIST-VR task was significantly correlated to perceptual and psychomotor ability.
	MRT-A (visualization)	Schlickum et al.,[113] 2016	Suggest VSA is the most important component, followed by intrinsic motivation, for MIST-VR performance in male medical students.
	CR (spatial orientation); PIcSOr (depth perception); FBT (visualization); MPT (spatial scanning); RFT (visual-motor/memory/VSA process); Matrix Reasoning Test (visual problem solving)	Stefanidis et al.,[24] 2006	CR test correlated with baseline simulation performance on video trainer and LCN/manipulative diathermy tasks.
	MRT-A (visualization)	Schlickum et al.,[61] 2009	There were improvements in performance on the manipulative diathermy task after training on the Half Life and Chessmaster videogames, but this was not mediated by MRT-A score.
LapSimGyn VR	MRT-A (visualization)	Ahlborg et al.,[57] 2011	Suggest VSA in OBGYN consultants predicts performance on LapSimGyn VR.
	MRT-A (visualization)	Ahlborg et al.,[73] 2012	VSA correlated with simulation performance, however after two days of training, VSA was no longer correlated with performance, suggesting training/experience may compensate for lower VSA in novices.
	MRT-A (visualization)	Ahlborg et al.,[62] 2013	There was not a strong relationship between VSA and simulation performance. It is possible that VSA training can identify trainees who need extra support.
	MRT-A (visualization)	Enochsson et al.,[48] 2008	Suggest VSA is important for successful performance of advanced gynecological procedures on the LAPSimGyn VR simulation.
FLS box trainer and/or bench model	MRT-A (visualization)	Abe et al.,[39]2017	MRT helps for initial skill acquisition phase for laparoscopic suturing skills.
	CC, CR (spatial orientation); MPT (spatial scanning)	Andalib et al.,[23] 2006	Spatial orientation, but not spatial scanning, correlated with the rate at which proficiency was achieved on the

			bimanual transferring FLS task using a monocular optical system.
	Identify visual spatial centers of the brain using fMRI during FLS tasks	Bahrami et al.,[14] 2014	Parietal areas of the brain were active during more complex tasks.
	CR, CC (spatial orientation); MRT-A (visualization); OSATS	Burkhardt et al.,.,[42]2019	Lower VSA scorers correlated with OSATS ratings of surgical performance; however, high scoring VSA did not correlate with OSATS ratings.
	MRT-A (visualization); Spatial Orientation Test; OSATS	De Witte et al.,.,[40] 2018	Decreased performance in MRT post laparoscopic suturing and knot tying training as compared to pre-training. One group showed improvement in spatial orientation performance post laparoscopic skills training.
	Effect of cognitive imaging on laparoscopic suturing	Donnon et al.,[66] 2005	Visual-spatial discordance between genders where males tend to outperform females in visual-spatial simulated laparoscopic tasks.
	CR, CC (spatial orientation)	Kolozsvari et al., [117]2011	No correlation between CC and CR with the learning curve for FLS; however CC was correlated with overall virtual reality simulator score while CR correlated with performance on laparoscopic camera navigation task.
	FCT (speed of closure); Orientation Test (spatial orientation); BTT (spatial relation)	Risucci et al.,[27] 2000	Score on the orientation test was inversely correlated with performance speed on the rope pass drill. BTT score was inversely correlated with performance speed on both rope pass and cup drop drills. Suggest performance on mental rotation tasks and prior surgical experience facilitate resident performance on laparoscopy.
	BTT (spatial relation); FCT (speed of closure); Orientation Test (spatial orientation)	Risucci et al.,[25] 2001	Suggest age, surgical experience, and VSA predict performance on laparoscopic suturing skills.
	MRT-A (visualization); 3D Blocks Game; Keyhole test	Stransky et al.,[74] 2010	Suggest that training on a variety of mental rotation tasks improves performance on FLS tasks.
MISTELS	CR (spatial orientation); MPT (Spatial Scanning)	Mistry et al.,[26] 2013	Monoscopic visualization shown to be same or better than stereoscopic visualization at basic training levels utilizing the MISTELS standardized evaluation criteria. VSA was weakly correlated (r=0.012) to MISTELS performance, albeit not significantly so.
	MRT-A (visualization) during mono or stereoscopic viewing	Roach et al.,[53] 2014	No significant difference was observed between those with high VSA and low VSA on the MISTELS tasks in either monoscopic or stereoscopic views.
Angled laparoscope simulation	CR (spatial orientation); PFT (visualization)	Eyal et al.,[28] 2001	VSA was significantly correlated with performance on the angled laparoscope simulation tasks. Participants with low VSA had the lowest performance, implying some degree of VSA is a prerequisite for adequate learning.
	Purdue Spatial Visualization Test, MRT-A (visualization)	Keehner et al.,[11] 2006	VSA correlated with initial and final performance on the angled laparoscope simulation task. Suggest VSA is important throughout training as the correlation between performance and VSA did not diminish over time.
Laparoscopic abdominal cavity	PicSOr (depth perception)	Gallagher et al.,[121] 2003	PicSOr predicted performance on the laparoscopic circle cutting task.

trainer	CR (spatial orientation); MRT-A, PFT (visualization)	Keehner et al.,[58] 2004	Participants with low VSA were more affected by camera offset during the laparoscopic trainer simulation than those with high VSA.
Surgical Science LapSim	MRT-A, PFT (visualization); CC, SDT, Rotating Shapes test (spatial orientation)	Groenier et al.,[10] 2014	Those with high VSA were more efficient in their movements during the laparoscopic simulation. Suggest VSA is essential for early learning of minimally invasive surgery and may predict rate of learning.
	Stumpf-Fay Cube Perspectives Test (spatial orientation)	Hassan et al.,[30] 2007	VSA correlated with laparoscopic performance. Participants with high spatial perception performed the tasks faster.
	CR, CC (spatial orientation); PicSOr (depth perception)	Louridas et al.,[6] 2016	Those with high CC scores had more accurate camera navigation path length and angle path scores for the LapSim. However, VSA was not predictive of baseline surgical skill, there was no significant correlation between PicSOr scores and performance on the LCN task, no significant correlation between CR scores and performance, and no significant correlation between LCC time, error, or total score and any of the VSA tasks. Suggest using VSA as a metric of surgical skill is questionable.
	MRT-A (visualization); CR (spatial orientation); HFT (flexibility of closure); IPT (perceptual speed)	Luursema et al.,[112] 2012	CR score was correlated to performance on the damage index on the LapSim. Perceptual speed was correlated to motion efficiency. Visualization was correlated to all performance variables. Suggest that VSA is the most important component during basic laparoscopic simulation training, and the ability to mentally manipulate complex visual stimuli is key to laparoscopic skills development.
	CC (spatial orientation)	Roch et al.,.,[68] 2018	VSA has impact on LCN performance, particularly for complex tasks.
	MRT-A (visualization); HFT (flexibility of closure); IPT, NCT (perceptual speed)	Sliwinski et al.,[75] 2010	Visualization was significantly correlated with performance on the 3rd and 4th training session. Flexibility of closure and perceptual speed were not significantly correlated with any individual training session. VSA component visualization was found to be correlated to laparoscopic performance.
	MRT-A, PFT, SDT, (visualization); Rotating Shapes Test (spatial orientation); Corsi block tapping test (spatial memory); NCT, IPT (perceptual speed)	Groenier et al.,[41] 2015	Perceptual speed was the only VSA that predicted the rate at which medical students learned the grasping and instrument navigation tasks.
ProMIS Simulator	CC, CR (spatial orientation); SDT (visualization); PicSOr (depth perception); MPT (spatial scanning)	Buckley et al.,[32] 2013	Those with high VSA attained proficiency faster when completing laparoscopic appendectomy compared to those with low VSA.
	SDT (visualization); CC, CR (spatial orientation); MPT (spatial scanning)	Buckley et al.,[33] 2014	Those with high VSA achieved proficiency faster than those with low VSA.

	PicSOr (depth perception); CC, CR (spatial orientation); MPT (spatial scanning)	Nugent[34], 2012	Surgical trainees had higher depth perception when compared to a control group. There was a significant association between VSA, psychomotor aptitude, and MIS performance on a laparoscopic colectomy. VSA was also significantly correlated with the number of trials to reach proficiency (those with high VSA required fewer trials).
	CR (spatial orientation); MPT (spatial scanning)	Nugent et al.,[49] 2012	VSA was significantly correlated with performance on basic laparoscopic tasks such as: locating, coordination, and sharp dissection. VSA was significantly correlated with path smoothness and number of tray errors on the laparoscopic sigmoid colectomy. Those with high VSA and psychomotor aptitude performed better on surgical tasks than those with low VSA.
VR simulator SimSurgery	CC (spatial orientation)	Jungmann et al.,[35] 2011	High VSA correlated with precise movements and faster performance during the SimSurgery tasks. CC also correlated with knot-tying skills in surgical novices.
Virtual Environment for Surgical Training & Augmentation (VESTA)	Virtual environments for training perceptual motor skills, spatial skills, and critical steps of surgical procedures	Tendick et al.,[107] 2000	Suggest surgeons who demonstrate high VSA rely on visualization. The ability to utilize visualization in surgery may be limited for surgeons with low VSA.
Xiact	SRT (spatial relation)	Schijven et al.,[122] 2004	Spatial reasoning task was significantly correlated to Xitact outcome.
LapMentor	CC (spatial orientation)	Rosenthal et al.,[36] 2010	VSA was significantly correlated with LapMentor navigation task, VR tasks (scores, task completion time, and economy of movement), but not endoscope travel speed. There were no differences in VSA between and trainees with previous laparoscopic surgery experience.
	Wechsler Intelligence Scale for Children–III (WISC-III) cubes (spatial ability)	Rosenthal et al.,[126] 2011	Performance on the LapMentor task was correlated with the WISC-III.
Not directly assessed	Meta-analysis on VSA, perceptual ability and psychomotor ability predicting MIS performance	Kramp et al.,[123] 2016	Across all articles, there was a significant correlation between VSA and laparoscopy performance (r=0.32). There was also a significant correlation between PicSOr and laparoscopy performance (r=0.31). Due to the low r value, suggest aptitude tests can be used to predict some of the variance in laparoscopic performance, but there are likely many other factors.
Videoscopic live animal courses for advanced surgeons and laparoscopic urology novices	PFT (visualization)	Keehner et al.,[29] 2004	There was a significant correlation between VSA and novice performance (r=0.393), but no correlation with experienced surgeons' performance (r= 0.020). Surgeons had better operative scores, but not VSA. As no significant correlation was found between videoscopic skill and VSA for experts, the authors suggest the importance of VSA may diminish with experience in terms of laparoscopic performance.
Peg transfer and key threading tasks	Graded Circle Test (depth perception)	Suleman et al.,[63] 2010	Medical students with depth perception defects had significantly lower performance ratings. All students, regardless of whether or not they had depth perception defects, benefited from simulation training.

Esophageal myotomy and Dor fundoplication on a box trainer	Depth perception as measured by the global operative assessment of laparoscopic skills (GOALS) tool	Bellorin et al.,[79] 2016	The experienced trainees were significantly better than novice trainees in all GOALS domains, including the depth perception domain.
Laparoscopic appendectomy, laparoscopic cholecystectomy, and laparoscopic inguinal hernia repair in a modified U.S. Surgical Trainer	Spatial orientation as measured on a five-point scale	Adrales et al.,[78] 2004	Spatial orientation was highly correlated with competence at the tasks.

Table S2. Comparing endoscopy simulation performance to VSA:

Simulation	VSA Test (VSA component)	Reference	Key Points of Authors' Interpretations
GI Mentor II	PicSOr (depth perception)	Enochsson et al.,[124] 2004	Performance on PicSOR correlated well with percentage of time spent with a clear view and efficiency of screening in a gastroscopy task. However, this correlation was only significant in men and not in women.
	CR (spatial orientation)	Enochsson et al.,[69] 2006	There was no correlation between VSA and performance on the gastroscopy task for experts, but there was a correlation between VSA and performance on the task for residents.
	MRT-A (visualization)	Hedman et al.,[47] 2007	There was a significant correlation between VSA and efficiency of screening and total time on a gastroscopy task.
	CR (spatial orientation); MRT-A, Guay's visualization test (visualization); HFT (flexibility of closure); IPT (perceptual speed); Gestalt Completion or Closure test (speed of closure)	Luursema[16], 2010	High VSA scoring group improved faster for measurement variable 'time on task' for both the Endobubble task and VR-Colonoscopy task. For the EndoBubble task on GI Mentor II, only MRT and Guay's visualization tests correlated significantly and negatively with time on task. Individuals who scored high on the MRT-A test required fewer training sessions than those who scored low on MRT-A to achieve proficiency on the colonoscopy task.
	Purdue Spatial Visualization Test, CR, MRT-A (visualization); HFT (flexibility of closure); IPT, NCT (perceptual speed); Gestalt Completion or Closure Test (speed of closure)	Luursema et al.,[46] 2010	Significant correlations were found with MRT and Guay's visualization tasks with the following VR-Colonoscopy performance variables: time, clear view, and lost lumen.
	PicSOr (depth perception); CC, CR (spatial orientation); MPT (spatial scanning); SDT (visualization)	Nugent[34,] 2012	There was a significant negative correlation between the CR and SDT tests and number of endoscopy trials to reach proficiency.
	PicSOr (depth perception); CC (spatial orientation)	Ritter et al.,[45] 2006	There was a strong correlation between PicSOr and CC and duration of training to proficiency for the EndoBubble task.
	MRT-A (visualization)	Schlickum et al.,[44] 2011	Score on the MRT-A test was significantly correlated with total score and time for the GI Mentor II simulation and UroMentor.
	PicSOr (depth perception); CC, CR (spatial orientation)	Westman et al.,[55] 2006	VS tests correlated to GI mentor II endoscopic simulation tasks (percent of time spent with clear view and visualization of colon lumen). No correlations with gastroscopy module and VSA tasks.
	MRT-A (visualization)	Schlickum et al.,[61] 2009	There were improvements in performance on the Gastroscopy module 1 task after training on the Half Life videogame, but this was not mediated by MRT-A score.

Lap Mentor (Simbionix) trainer	RFT (visual-motor/memory & VS process)	Jardine et al.,[106] 2015	Additional testing on VS or psychomotor aptitude on interview day can be done and may provide additional information to selection committee.
Natural orifice surgical simulated phantom (NOTES)	Target navigation assessed spatial awareness as analysis of endoscope tip position	Karimyan et al.,[80] 2012	Poor VSA via path analysis was characterized by erratic maneuvers with the endoscope, often leading to poor performance during simulation.
Hiroshima University endoscopic surgical assessment device (HUESAD)	MRT-A (visualization)	Egi et al.,[50] 2015	The high VSA group was more accurate on the HUESAD assessments. There was no significant difference in VSA groups when compared to smoothness using the HUESAD.
	VSA analyzed by Motion analysis in HUESAD assessment, OSATS checklist and GRS	Egi et al.,[9] 2013	VSA, smoothness and accuracy analyzed by the HUESAD were reliable parameters when assessing the endoscopic surgical skills.

Table S3. Comparing VSA to Other Simulations (i.e., not laparoscopy or endoscopy):

Simulation	VSA Test (VSA component)	Reference	Key Points of Authors' Interpretations
Surgical knot tying	MRT (visualization)	Brandt & Davies[1] 2006	Higher MRT scoring students correlated with performance on total number of surgical knots completed.
Z-plasties on pig thighs	FBT, MRT-A (visualization)	Wanzel et al.,[52] 2002	Residents with high VSA as tested by MRT-A and FBT performed significantly better in the procedure than low-scoring residents, and were able to transfer skills to more complex four-flap Z-plasty more easily. Low-scoring VSA residents improved performance after practice, so residents with lower VSA might need additional practice and feedback for learning new spatially complex procedures.
Rhombic flap procedure and double z-plasty procedure in 2D vs 3D methods	MRT (visualization)	Roach et al.,[53] 2012	MRT-A was assessed in addition to 2D vs. 3D video footage to capture spatially complex surgical translation flaps and assess these videos as a medium to support acquisition of complex surgical skills in novices using a global rating scale (GRS)
Anastomosis on pithed rats	SDT (visualization)	Murdoch et al.,[114] 1994	Trainee performance on microsurgical tasks was significantly correlated with VSA.
Anastomosis on porcine jejunum	HFT (flexibility of closure)	Steele et al.,[125] 1992	Correlations reported between VSA and improvement in performance for anastomosis on porcine jejunum. VSA was more important than pure motor ability for predicting capacity to perform anastomosis.
Vascular Intervention Simulation Trainer (VIST)	MPT (spatial scanning); RFT (visual-motor/memory/organization)	Van Herzeele et al.,[120] 2010	VSA as measured by the RFT and MPT did not correlate with initial performance during virtual renal artery stent procedures. However, over five anastomosis trials, there was a correlation between VSA and performance improvement.
Virtual-Fracture-Carving Simulation	MRT-A (visualization)	Pahuta et al.,[60] 2012	Performance on the MRT-A test was not correlated with learning from a Virtual-Fracture-Carving Simulation.
Relationship between MRT-A, SDT and training on 3D perception of 2D angiographic images of abdominal aortic aneurism (AAA) between novices and vascular surgeons.	MRT-A, SDT (visualization)	Sidhu et al.,[67] 2004	There were no significant correlations between VSA and novice depth map comparisons with expert group. Innate VSA (measured by MRT-A and SDT) was not correlated with accuracy of perceived 3D structure of angiographic images among novices. Perception of 3D structures from 2D images was affected by experience and training, with experts perceiving more elaborate 3D structures than novices prior to training. Although novice and experts had statistically significant differences in perception of 3D from 2D images, the

			novices' perception became more similar to experts after educational intervention.
Da Vinci Robot Simulator	PFT, Keyhole test (visualization); CC (spatial orientation)	Suozzi et al.,[7] 2013	Better VSA was correlated with improved performance on a robotic surgery simulation.
	Not specified	Teishima et al.,[64] 2012	VSA was not correlated to scores on a suturing task.
	Perceptual Ability Test (PAT) from the Dental Aptitude Test (DAT) (includes the Keyhole Test (visualization))	Finnegan et al.,[115] 2013	The group with better spatial ability as measured by the PAT had significantly higher scores on four of the seven metrics (time to completion, economy of motion, excessive instrument force, and work space range).
Human mandible model fixation of 3D plate bending and screwing	MRT (visualization); Gestalt completion test	Wanzel et al.,[70] 2003	VSA correlated with surgical performance scores amongst dental students but not for residents or staff surgeons.
Loading the needle on the driver on a mitral valve model	MRT, Purdue Visualization of Views (visualization)	Sheikh et al.,[65] 2014	There was no significant correlation between the visualization tests and performance on the task.

Table S4. VSA Aptitude Testing:

VSA discussed/tested	Reference	Key Points of Authors' Interpretations
Not directly assessed (narrative review)	Anastakis et al.,[13] 2000	No strong consensus to link surgical ability and VSA and do not recommend assessment of surgical trainees. Need for future research to identify specific VSA relevant to surgical tasks.
Not directly assessed (editorial)	Bishawi & Pryor,[92] 2014	This paper suggests that some skills required for success as a resident are not necessarily associated with spatial aptitude, and that further research is needed to understand the correlation between aptitude testing and resident performance.
MRT-A (visualization); Career path questionnaire	Brandt & Wright,[94] 2005	Medical students initially interested in VS-intense residency programs scored higher on the MRT. However, findings did not persist to time of application/acceptance.
HFT (flexibility of closure); Technical skills ability of trainees measured by staff	Gibbons et al.,[38] 1983	Significant correlations found between HFT and ratings of technical skills of trainees despite current entry into surgical training programs not requiring HFT testing.
SDT (visualization)	Gilligan et al.,[5] 1999	This study found no significant difference for spatial reasoning scores between geriatricians and surgeons at the beginning of training.
Not directly assessed (narrative review)	Graham and Deary,[93] 1991	Spatial visualization is important for surgical performance, but there are too few studies investigating how VSA contributes to performance of surgery. Limitations of studies to date included (1) limited well-designed studies on psychological and psychomotor abilities specific to surgical skill, (2) subjective ratings on surgical ability in most studies, and (3) no definition of superior surgical ability.
Not directly assessed (narrative review)	Hamstra et al.,[96] 2006	Mixed evidence whether innate VSA is predictive of surgical performance, or, if VSA can be acquired through practice and training.
EFT (flexibility of closure)	Harris et al.,[95] 1994	This study compared surgical, medicine, anaesthesia, and psychiatry trainees on aptitude tests, finding no significant differences in VSA by either sex or specialty. Thus, it is unlikely that trainees self-select based on VSA. The authors do not recommend using VSA as a main factor in selecting trainees.
CR, CC (spatial orientation)	Henn et al.,[15] 2018	Assessed surgical trainees and medical students on CR and CC tests, concluding that surgical trainees outperformed the controls on all tests.
Wechsler Adult Intelligence Scale (spatial ability); HFT (flexibility of closure); FBT (visualization); Maze (spatial scanning)	Schueneman et al.,[88] 1984	General surgery residents were tested on visual spatial organization, stress tolerance, and psychomotor ability. Visual spatial organization and stress tolerance explained a significant amount of variance in their performance.
Assessed via MIST-VR	Gallagher et al.,[59] 2001	Urologists did not differ from non-surgeons in their innate spatial awareness skills, as measured by number of errors, economy of movement, and time to complete tasks on the MIST-VR. This suggests that other psychometric tests may be important for acquiring surgical skills other than innate VSA. Further, it suggests urologists are not self-selected due to innate VSA.
Not directly assessed (narrative review)	Gallagher et al.,[4] 2009	This review describes the importance of VSA for learning MIS and endoscopic procedures. The review also explains that the Royal College of Surgeons in Ireland test short-listed candidates for Higher Surgical Training Program on VSA tests (CC, PicSOr and MPT) in addition to other psychomotor tests to select candidates.
MRT-A, MRT-C (visualization)	Langlois et al.,[2] 2015	Medical students entering residency were tested on VSA to see if there was a correlation to specialties. Although lower MRT-A scores were reported in family medicine and internal medicine compared to surgery

		and anesthesia, it was not statistically significant for sex, year, or residency program.
Not directly assessed (systematic review)	Louridas et al.,[31] 2016	Out of 38 studies, 25 different VSA tests were used. Of these 25 tests, two were correlated with technical performance (PicSOr [5/8 studies] and MRT [6/9 studies]). PicSOr scores were correlated with laparoscopic skills learned in a box trainer and a VR simulation. MRT scores were correlated with open surgical skills. The authors concluded that only three VSA tests (CR, MRT, and PicSOr) have consistently demonstrated significant correlations with technical ability and performance.
Not directly assessed (survey study)	Tansley et al.,[87] 2007	The authors conducted a survey with program directors from nine surgical training programs in the London Deanery on methods of surgical assessment and selection. One specialty (general surgery) incorporated VSA in candidate selection. Assessing VSA may aid the surgical selection process by identifying individuals who are not suited for such training. Future residency selection should include assessment of VSA.
Maze Test (spatial scanning); SRT (spatial relation)	Francis et al.,[71] 2001	Master surgeons made significantly fewer errors on the maze test, with similar execution times for dexterity, compared to medical students. However, performance on the SRT was lower for surgeons than for medical students.
Titmus Stereo Fly Test, TNO Stereopsis Test, Frisby Stereotest (depth perception)	Biddle et al.,[97] 2014	The authors measured depth perception in surgeons in a variety of specialties. Most surgeons have high-grade stereoacuity, supporting the notion that it should be assessed prior to surgical training. However, high stereoacuity should not be a requirement.
Not directly assessed (narrative review)	Bann et al.,[98] 2005	No single aptitude test can provide conclusive evidence that surgeons have superior VSA and/or determine the need for any particular attribute for surgeons.
Not directly assessed (letter to the editor)	Buckley et al.,[89] 2014	Discuss Nugent et al.,[(119)], 2012 study, which found evidence to suggest that medical students should undergo VSA testing to advise future paths based on their abilities.
Computerized Pilot Aptitude and Screening System (COMPASS)	Stolk-Vos et al.,[86] 2017	Pilot candidates and medical trainees differ in some aspects of the COMPASS. Pilot candidates have better eye-hand-foot coordination and spatial orientation, while medical trainees have better eye-hand coordination.
Not directly assessed (narrative review)	Torkington et al.,[101] 2000	While measures of depth perception abilities are available, they are not useful for surgical training. Further, it seems that trainee abilities can be improved with training regardless of their inherent abilities.
Perceptual ability	Krespi et al.,[90] 1986	The authors describe a screening test for surgical training applicants and suggest that programs measure perceptual ability of applicants.
Not directly assessed (letter to the editor)	Moglia et al.,[99] 2018	While there is that some VSA components predict performance in simulation, more evidence about how performance in simulation translates to performance in the clinical setting is needed for VSA to be used in selection.
Spatial memory	Roitberg et al.,[91] 2015	The authors demonstrated validity evidence for their test of sensory-motor performance using a virtual reality surgical simulator.

Table S5. 2D (monoscopic) versus 3D (stereoscopic) Training in Surgical Education:

Objective	Reference	Key Points of Authors' Interpretations
Examined performance of surgeons in training on a range of psychometric abilities: FBT, PFT, SDT, MRT (visualization); CR, CC (spatial orientation); MPT (spatial scanning); Stereoscopic vision test; Alice Helm 6 (cognitive ability).	Deary et al.,[104] 1992	Participants with superior stereoscopic discrimination correlated with surgical ability factors. Higher scoring visualization PFT correlated with surgical ability. No significant correlations between MRT and surgical abilities, specifically reaction time and error scores.
Narrative review discussing the benefits and drawbacks of medical applications of stereoscopic displays	Held and Hui[12], 2011	Stereoptic displays can assist with spatial understanding of anatomy for surgical procedures. Students who easily recognize 3D structures are more likely to perform well in surgery, while students with low VSA benefit more than high VSA using stereoptic 3D displays. Stereoptic displays help novice surgeons perform basic laparoscopic tasks with fewer errors. Overall, they reported mixed views on the usefulness of stereo displays: two-thirds of interviewed surgeons prefer non-stereoptic views for laparoscopy. Stereoptic imaging appears to significantly benefit performance of novices on laparoscopic and complex tasks.
Proposed a study to evaluate the effects of stereopsis during a learning phase on VS reasoning in two tasks related to medical diagnosis	Luursema et al.,[17] 2004	MIS has visual feedback through 2D monitors, which causes loss of spatial information. Stereopsis is beneficial for VS task performance in laparoscopy. The authors propose a study to determine the effect of stereopsis during a learning phase on performance of two tasks.
Evaluated differences in laparoscopic skill performance in medical students between monoscopic and stereoscopic on the MISTELS tasks	Mistry et al.,[26] 2013	Monoscopic visualization shown to be the same or better than stereoscopic visualization at basic training levels utilizing the MISTELS standardized evaluation criteria. VSA was reported to be weakly correlated (r=0.012) to MISTELS performance, albeit not significantly so. 3D stereoscopic technology may not provide addition help on surgical skill development in novice trainees.
MRT-A (visualization) was assessed in addition to 2D versus 3D video footage to capture spatially complex surgical translation flaps and assess these videos as a medium to support acquisition of complex surgical skills in novices using a global rating scale (GRS)	Roach et al.,[51] 2012	There were no significant differences in GRS for 2D versus 3D e-learning for either skin flap procedure. Students with higher VSA performed better on each task regardless of 2D or 3D video. Novice participants with low VSA benefited the most from stereoscopic learning environments due to additional depth information.
Tested individuals on MRT-A (visualization) and MISTELS laparoscopic simulator tasks to determine whether there is a difference in scores between high- and low-VSA individuals during monoscopic or stereoscopic viewing	Roach et al.,[53] 2014	No significant differences were observed between high- and low-scoring VSA participants for the MISTELS tasks for either viewing modality. High-VSA individuals on average scored higher on MISTELS tasks. The authors suggest additional practice may enhance performance for individuals with low VSA.
Investigated the influence of MRT-A and SDT (visualization), experience,	Sidhu et al.,[67] 2004	There were no significant correlations between VSA and novice depth map comparisons with expert group.

and training on 3D perception of 2D angiographic images of abdominal aortic aneurism (AAA) between novices and vascular surgeons		Innate VSA (measured by MRT-A and SDT) was not correlated with accuracy of perceived 3D structure of angiographic images among novices. Perception of 3D structures from 2D images was affected by experience and training, with experts perceiving more elaborate 3D structures than novices prior to training. Although novice and experts had statistically significant differences in perception of 3D from 2D images, the novices' perception became more similar to experts after educational intervention.
Monitored function of the superior and inferior parietal lobules during depth perception and hand-eye coordination tasks in 2D and 3D modalities	Paggetti et al.,[84] 2015	Both the superior and inferior parietal lobules were activated in the 2D and 3D conditions. There was significantly greater activation of the lobules in the hand-eye coordination task compared to the depth perception task in both the 2D and 3D conditions.
Explored whether visualization abilities (measured by the MRT) differentially influence performance in 2D and 3D modalities	Egi et al.,[50] 2016	Those who had low visualization ability significantly improved their performance in 3D compared to 2D, while those who had high visualization ability did not.
Examined how depth cues affect performance in a laparoscopic simulator	Shah et al.,[103] 2003	While medical students and practicing surgeons do not differ in how much they rely on depth cues (stereo, texture, and outline), there are some differences in the impact of stereo cues on performances of medical students versus surgeons.