

Title	A computerised test of perceptual ability for learning endoscopic and laparoscopic surgery and other image guided procedures: Score norms for PicSOr
Authors	Henn, Patrick;Gallagher, Anthony G.;Nugent, Emmeline;Cowie, Roddy;Seymour, Neal E.;Haluck, Randy S.;Hseino, Hazem;Traynor, Oscar;Neary, Paul C.
Publication date	2017-02-14
Original Citation	Henn, P., Gallagher, A.G., Nugent, E., Cowie, R., Seymour, N.E., Haluck, R.S., Hseino, H., Traynor, O. and Neary, P.C. (2017) 'A computerised test of perceptual ability for learning endoscopic and laparoscopic surgery and other image guided procedures: Score norms for PicSOr', American Journal of Surgery, 214(5), pp. 969-973. doi:10.1016/j.amjsurg.2017.01.025
Type of publication	Article (peer-reviewed)
Link to publisher's version	http://go.qub.ac.uk/PICSOR - 10.1016/j.amjsurg.2017.01.025
Rights	© 2017 Elsevier Inc. All rights reserved. - https://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2024-04-25 17:26:35
Item downloaded from	https://hdl.handle.net/10468/3790

A computerised test of perceptual ability for learning endoscopic and laparoscopic surgery and other image guided procedures: Score norms for PicSOr

Patrick Henn MB¹, Anthony G Gallagher DSc^{1,2}, Emmeline Nugent MD,² Roddy Cowie PhD³, Neal E Seymour MD⁴, Randy S Haluck MD⁵, Hazem Hseino MD², Oscar Traynor MCh², Paul C Neary MD⁶.

¹ ASSERT Centre, Brookfield Health Sciences Complex, University College Cork, Cork, Ireland.

p.henn@ucc.ie, ag.gallagher@ucc.ie

² National Surgical Training Centre, Royal College of Surgeons in Ireland, Dublin 2. Ireland.

emmelenugent@rcsi.ie

³ School of Psychology, Queens University Belfast, Belfast, BT7 1NN, UK. R.Cowie@qub.ac.uk

⁴ Tufts University School of Medicine, Dept. of Surgery, 759 Chestnut Street, Springfield, MA 01199, USA. Neal.Seymour@baystatehealth.org

⁵ Dept. of Surgery, Pennsylvania State University, C4628, 500 University Drive, Hershey, PA 17033, USA. rhaluck@hmc.psu.edu

² National Surgical Training Centre, Royal College of Surgeons in Ireland, Dublin 2. Ireland.

hazemhseino@rcsi.ie

² National Surgical Training Centre, Royal College of Surgeons in Ireland, Dublin 2. Ireland.

o.traynor@rcsi.ie

⁶ Dept. of Surgery, Tallaght Hospital and University of Dublin, Trinity College, Ireland.

Paul.Neary@amrch.ie

Correspondence

Dr Patrick Henn, Advanced Southern Simulation, Education, Research and Training (ASSERT) Centre, School of Medicine, Brookfield Health Sciences Complex, College Road, University College Cork, Cork, Ireland. Tel: +353214205605, FAX: +3524901594 Email: p.henn@ucc.ie

Abstract

Background: The aptitude to infer the shape of 3-D structures, such as internal organs from 2-D monitor displays, in image guided endoscopic and laparoscopic procedures varies. We sought both to validate a computer-generated task Pictorial Surface Orientation (PicSOr), which assesses this aptitude, and to identify norm referenced scores.

Methods: 400 subjects (339 surgeons and 61 controls) completed the PicSOr test. 50 subjects completed it again one year afterwards.

Results: Complete data was available on 396 of 400 subjects (99 %). PicSOr demonstrated high test and re-test reliability ($r = 0.807$, $p < 0.000$). Surgeons performed better than controls' (surgeons = 0.874 V controls = 0.747 , $p < 0.000$). Some surgeons ($n = 22 \sim 5.5\%$) performed atypically on the test.

Conclusions: PicSOr has population distribution scores that are negatively skewed. PicSOr quantitatively characterises an aptitude strongly correlated to the learning and performance of image guided medical tasks. Most can do the PicSOr task almost perfectly, but a substantial minority do so atypically, and this is probably relevant to learning and performing endoscopic tasks.

Introduction

Interventional and diagnostic medicine has experienced considerable advances in the use of image guided procedures in areas such as arthroscopy, cystoscopy, endoscopy, gynaecology, otolaryngology, laparoscopic surgery and surgical robotics. As a result of these less invasive procedures, patients experience less trauma, pain, reduced scarring and a faster recovery to normal activities of daily living.^[1] However, image guided procedures impose considerable ergonomic and human factor related demands on the operator.^[2] The operator must perform the procedure with instruments that fulcrum against the body wall.^[3,4] The fulcrum effect is defined as the perceived inversion of movements. Consequentially, an internal movement to the right is displayed as a movement to the left on the monitor.^[3] The perceived inversion also affects horizontal and vertical movements. The operator must also infer the shape of 3-D structures, such as the internal organs of patients, from 2-D displays on a video monitor, an issue that has not been resolved with 3-D camera systems.^[5] This problem is often described as loss of binocularity, but it is simpler and more accurate to call it pictorial perception. So-called primary cues, binocular disparity, convergence, and accommodation are present in abundance. The difficulty is that primary cues and other cues related to lighting and texture present an image with reduced depth cues. The operator has to reconstruct from the image on the monitor screen a 3-D representation of the anatomical environment to perform the procedure effectively, efficiently, and safely. These skills of selection and inference are the basis of pictorial perception. The ability to reconstruct 3-D scenes from pictures varies with age,^[6] and cultural background.^[7] Such variability could clearly contribute to differences in aptitude for pictorially guided diagnostic and interventional medicine.

Pictorial Surface Orientation (PicSO_r) was developed to address this issue of perceptual contributions.^[8,9] Performance on PicSO_r has been found to be reliably associated with learning and performing laparoscopic,^[8,10] endoscopic,^[11-13] and otolaryngology tasks^[14] by novices and procedure experts.^[8,13] The aims of the study reported here were to investigate the reliability of PicSO_r as an instrument for the evaluation of an operator's ability to recover 3-D information from 2-D displays, and to report norms and reference scores with which trainees and educators may wish to use for comparisons.

Methods

Subjects

Between 2001 and 2011, 400 subjects were recruited, 300 in Ireland and 100 in the USA. In the Irish group 239 were surgeons in training (mean age = 30 years, SD = 1.6 years) and 61 controls, who were medical students or medical interns (first year medical resident) with no prior surgical experience, these (mean age = 24.35 years, SD = 5.09). All of the American subjects were surgical residents (mean age = 29.11 years, SD = 2.75 years).

Materials

PicSOr consists of a series of 35 pictures displayed sequentially on a computer monitor. Each picture displays a spinning arrowhead with its point touching the surface of a geometric object, namely a cube, as shown in **Figure 1**. The cube and arrow are presented at random orientations for the 35 picture test trials.

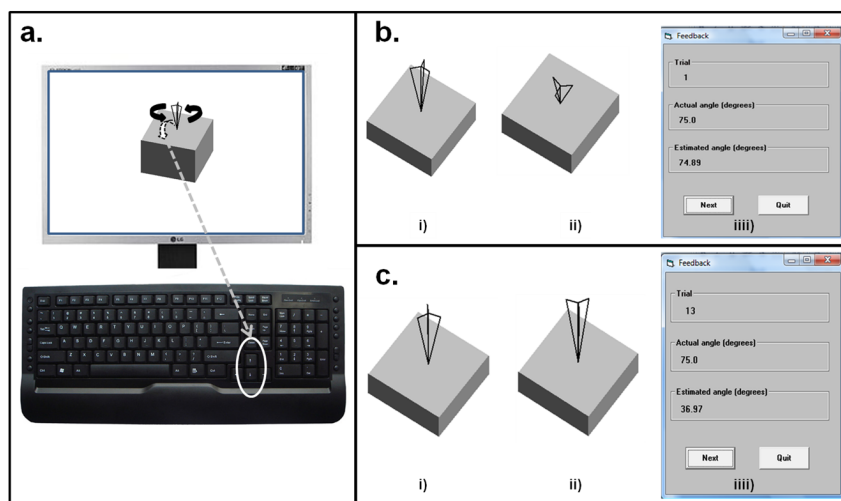


Figure 1a-c. The test set-up for the PicSOr computerised assessment.

Procedure

Subjects performed the task seated at a computer console. During training trials subjects were instructed to “adjust the arrow (spinning on the top of the cube) so that the central shaft of the arrow was perpendicular to the surface of the cube using the forward and backward arrows on their keyboard

(shown in Figure a.)”. During training trials subjects were given feedback on their performance. No performance feedback was given during the test proper.

The subject’s task is to manoeuvre the arrowhead until its shaft is perpendicular to the object surface at the point where they touch (**Figure 1a-c**). The motor element of the task is deliberately trivial. Hence the task is a relatively pure test of the subject’s ability to recover the pictorial cues that specify how structures are oriented in virtual pictorial space, and to compare the implied orientations. The cube and arrow presents the perceptual problem in its simplest form. The most important measures of performance are firstly the correlation (R) between the theoretically correct arrowhead orientation and the settings chosen by the subject and secondly the slope β of the fitted line.

Prior to testing subjects and controls were allowed practice trials including coaching on performing the PicSOr task with no time limits place on completion. Subjects were also given feedback on their angle estimation in comparison to the actual angle of arrow orientation as shown in **Figure 1b** and **1c**.

During testing subjects were instructed to proceed through the 35 test trials as quickly and accurately as possible, however, the subjects were not given feedback on their performance. When the subject had completed their arrow adjustment for each picture trial, they pressed the ‘return’ key and the test automatically advanced. On completion of the 35 test trials, the subjects’ responses were imported into an Excel spread-sheet and Pearson’s Product Moment Correlation Coefficient was used to calculate scores by correlating the settings chosen by the subject with the theoretically correct arrowhead orientation. The PicSOr score ranges from 0.0 to 1.0.

Results

Complete data was available on 396 of 400 subjects (99 %). The mean time taken to perform the 35 PicSOr test trials was 438 seconds ($SD = 152$; range 143 -716). **Figure 2** shows the frequency distribution of PicSOr scores plotted against a normal distribution curve. The median score was 0.9 and the mode was higher at 0.94. PicSOr percentile scores were 0.83 (25th percentile), 0.9 (50th percentile) and 0.95 (75th percentile). The lowest score was 0.09 and the highest was 0.99. The mean PicSOr score was 0.855 ($SD 0.15$). The test results showed that scores were negatively skewed with the

majority of subjects tending to score high on the PicSOr assessment. There was however, a long tail of subjects who performed less well.

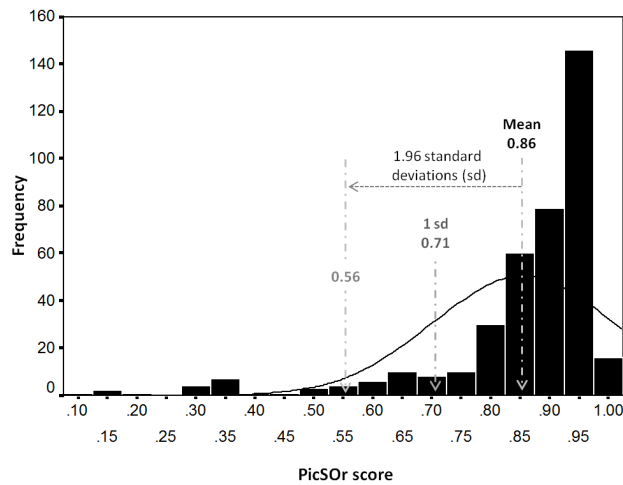


Figure 2. A frequency distribution of the PicSOr performance scores.

Fifty subjects were re-tested on PicSOr one year after the first test, results of which are shown in **Figure 3**. Overall, PicSOr scores showed a small mean improvement on the re-test (0.0243), which was found to be statistically significant when compared with Analysis of Variance (ANOVA) for Repeated Measures ($F(1, 49) = 5.21, p = 0.027$; test CI = 0.838 – 0.960 and re-test CI = 0.871 – 0.925). The test re-test PicSOr scores for males and females are shown in **Figure 3**. The means scores for the male subjects were unchanged (mean difference = 0.0003) at re-testing ($F(1, 29) = 0.001, p = 0.979$; test CI = 0.869 – 0.938 and re-test CI = 0.874 – 0.933). In contrast the female subjects showed a small overall improvement in their scores on re-testing (mean difference = 0.057) which was statistically significant ($F(1, 20) = 8.46, p = 0.009$; test CI = 0.763 – 0.905 and re-test CI = 0.838 – 0.943). The concordance between subjects scores at test and re-test were assessed with Pearson's Product Moment Correlation Coefficient. These scores at test and re-test were highly correlated for the group as a whole ($r = 0.807, p < 0.000$), for the male subjects ($r = 0.831, p < 0.000$) and for the female subjects ($r = 0.822, p < 0.000$). Internal reliability of PicSOr scores showed that coefficient alpha was high ($\alpha = 0.893$).

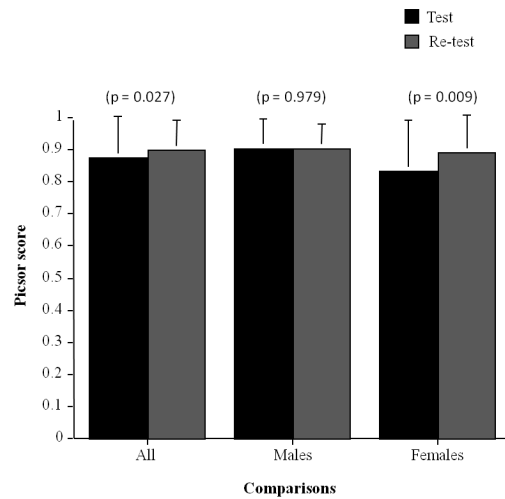


Figure 3. Test re-test reliability score comparisons.

Figure 4a shows the median and inter-quartile range of the PicSO_r scores for surgeons (n = 335) and control subjects (n = 61). The mean PicSO_r scores of the control subjects were considerably lower than the surgeons' scores (surgeons = 0.874 (SD = 0.12) V controls = 0.747 (SD = 0.23)) and showed greater performance variability. When compared with AONVA the mean difference was found to be statistically significant ($F(1, 394) = 39.64, p < 0.000$; Surgeons CI = 0.861 – 0.888 and Controls CI = 0.688 – 0.805).

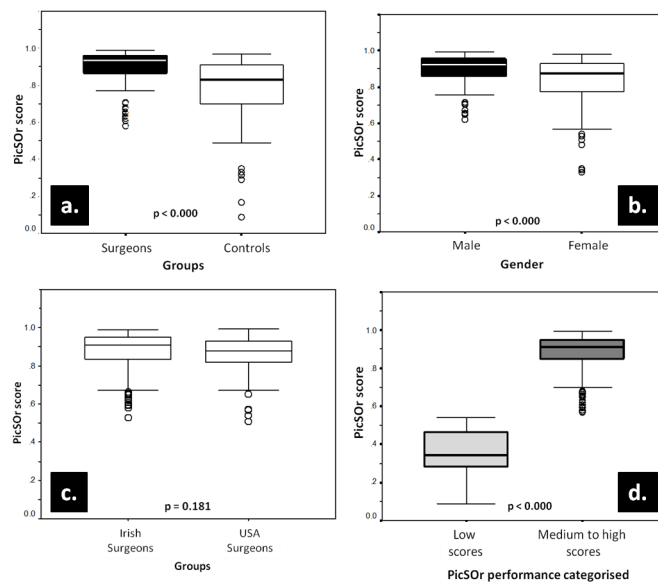


Figure 4. a - d. The median, inter-quartile range and outlier score comparisons group, gender, nationality and PicSO_r test performance.

Significant differences between male and female scores were observed in the test re-test sub-analysis (**Figure 3**). Male (n = 249) and female (n = 147) PicSO_r median and inter-quartile range scores for all of the subjects tested are shown in **Figure 4b**. On average, females' performance scores were lower than males (male mean = 0.887 (SD = 0.12), female mean = 0.8 (SD = 0.19) and showed greater score variability. These differences were compared with one-factor ANOVA, the difference was statistically significant (F (1, 394) = 31.85, p < 0.000; males CI = 0.872 – 0.901 and females CI = 0.77 – 0.831).

100 of the surgeons assessed were in surgical training programs in the USA. **Figure 4c** shows there was no difference in performance patterns between the Irish (mean score = 0.861 (SD = 0.15), and American (mean score = 0.84 (SD = 0.15)) surgeons.

The frequency distribution of all subjects' scores in **Figure 2** indicates the threshold score for subjects who were performing more than 1.96 standard deviations (SDs) from the mean and were thus performing significantly differently from the sample mean. There were 22 subjects with low scores and their median and inter-quartile scores are plotted in **Figure 4d** along with the 374 medium to high score subjects who scored better than the 1.96 SD cut-off. The mean score for the low score subjects was 0.346 (SD = 0.123), considerably lower than the mean for the medium to high score subjects, which was 0.885 (SD = 0.087). When compared with one-factor ANOVA this difference was found to be statistically significant (F (1, 394) = 754.42, p < 0.000; Low scores CI = 0.292 – 0.401 and for Medium to High scores CI = 0.876 – 0.894).

The PicSO_r boundary score that falls ≥ 1.96 standard deviations from the mean of the 400 subjects assessed in this study was 0.556.

Discussion

PicSO_r is reliably and statistically significantly associated with performance on a variety of endoscopic tasks by both novice and experienced surgeons and physicians.^[8,10-14] In this study the performance distribution of the 396 subjects assessed was negatively skewed with 94.5% of subjects' scores falling between 0.99 and 0.56. Only 22 subjects scored lower than this. PicSO_r performance also demonstrated high test re-test reliability, and internal validity. In general, females scored lower on perceptual and

visual spatial tasks than males.^[15,16] In this study we also found that females scored lower than males (e.g., 0.087 lower) on their first testing but improved their scores by the time they were re-tested. Although statistically significant this difference may have only a modest impact on real-world functioning.

No significant differences were observed between the PicSO_r scores of USA surgeons and Irish surgeons but large and statistically significant scores were observed between surgeons and control subjects; (0.87 V 0.75). This is probably due to a training effect where surgeons in training have learned to attend to, select and then draw inferences from subtle but important depth cues, which are the basis of pictorial perception. Training has the effect of increasing performance test scores but weakening correlations with performance on other tasks because it reduces individual score variability.^[17] This robust observation helps to explain the inconsistent correlations observed between aptitude assessment and objectively assessed performance with very experienced trainees or practicing doctors. Westman et al, observed that the correlations with aptitude tests scores were strongest for the difficult colonoscopy case.^[13] Another potential explanation of a failure to find a robust correlation between aptitude scores and image guided task performance is the use of inappropriate study tasks by researchers, that is the task was too easy or difficult, thus depressing score variability because of the ceiling or floor effect.

Several studies using PicSO_r in conjunction with an image guided task in a variety of surgical and medical contexts, demonstrated reliable correlation with task performance.^[8,10-14] Image guided medical tasks impose greater cognitive processing demand on the brain when performing the procedure.^[18] Although the human brain represents only 2% of the body weight, it uses substantive energy resources with active regions of the cortex consuming more energy than inactive regions.^[19]20,21] Individuals who are better at a task or score higher on aptitudes relevant to the task performance have lower rates of brain metabolism, i.e. cognitive capacity is utilized more efficiently.^[22,23]

Cortical regions function collaboratively to perform tasks, and differences between individuals have been found to be reliably associated with scores on aptitude tests.^[24] The implication of this finding is that individuals who score low on an aptitude test can be trained to perform a given task, e.g. an image

guided interventional medical or surgical procedure. However, performance of the task will make greater metabolic cerebral demands in comparison to an individual who scores considerably higher on the test. This means that there will be less spare resource capacity to respond to unanticipated events such as intra-operative complications, which could have profound implications for intra-operative performance^[25,26] and recovery from error.^[27]

The question about what aptitude score may be too low is important but difficult to answer precisely although it is a question that medicine has asked and is asking.^[28] While a score of 0.56, which is ≥ 1.96 standard deviations from the mean almost certainly is important. What is less clear is whether a PicSOOr score of 0.71 one standard deviation below the mean is critical. As doctors become more familiar with the performance of an image guided procedure, they become better at developing coping strategies for dealing with their aptitude deficit under normal operating conditions. Furthermore, the requirement to deal with unanticipated events is uncommon and during training there is the safety net of a more senior supervising colleague present. More worrying is the finding by Bell et al,^[29] that the exposure residents get to procedures during training is nothing like as broad or in-depth as previously thought. They reported that of the 121 procedures program directors believed residents should be at least competent in at graduation, only 18 were performed > 10 times, 83 of the 121 were performed < 5 times and for 63 of the procedures the mode frequency of resident experience was '0'. Of note 33% of experienced and practicing laparoscopic surgeons failed a test of their laparoscopic skills (Fundamentals of Laparoscopic Surgery Assessment,) which trainees are required to pass in order to become board certified in the USA.^[30] An important question, which has not been addressed by previous research, is how perceptual ability might be related to the rate of skills extinction. One potential solution to these problems would be to identify and select candidates with at least average aptitude capacity for their chosen medical discipline.

PicSOOr is the only reported instrument for the assessment of an individual's ability to interpret 3-D from 2-D visual cues.^[8] It has been derived from research on the psychophysics of perceptual performance,^[9] which has a heritage spanning more than a century.^[31] As a computer generated and scored test it is objective, transparent and fair in performance assessment. It is easy to use and runs on a standard PC

without requiring specialist statistical software packages, results can be calculated using Excel. It is also available at no charge to non-commercial users (at <http://go.qub.ac.uk/PICSOR>).

Conclusions

PicSOr has population distribution scores that are negatively skewed with 94.5% of subjects' scores falling between 0.99 and 0.56. It quantitatively characterises an aptitude that appears to be strongly correlated to the learning and performance of image guided medical tasks. Most people can do the PicSOr task almost perfectly, but a substantial minority perform atypically. It is worth knowing that such ability exists and is probably relevant to learning and performing endoscopic tasks.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Ethical approval. This study was approved by the higher surgical training committee of the Royal College of Surgeons in Ireland

References

1. Editorial. Cholecystectomy practice transformed. *Lancet* 1991;338:789-790
2. Gallagher AG, Smith CD. From the operating room of the present to the operating room of the future. Human-factors lessons learned from the minimally invasive surgery revolution. *Semin Laparosc Surg* 2003;10:127-139
3. Gallagher AG, McClure N, McGuigan J, et al. An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. *Endoscopy* 1998;30:617-620
4. Crothers IR, Gallagher AG, McClure N, et al. Experienced laparoscopic surgeons are automated to the "fulcrum effect": an ergonomic demonstration. *Endoscopy* 1999;31:365-369
5. Hanna GB, Shimi SM, Cuschieri A. Randomised study of influence of two-dimensional versus three-dimensional imaging on performance of laparoscopic cholecystectomy. *Lancet* 1998;351:248-251
6. Hagen MA. *The Perception of Pictures: Durer's devices, beyond the projective model of pictures*. Waltham, MA.: Academic Press; 1980
7. Miller RJ. Cross-cultural research in the perception of pictorial materials. *Psychol Bull* 1973;80:135-150
8. Gallagher AG, Cowie R, Crothers I, et al. PicSOR: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surg Endosc* 2003;17:1468-1471
9. Cowie R. Measurement and modelling of perceived slant in surfaces represented by freely viewed line drawings. *Perception* 1998;27:505-540
10. Kolozsvari NO, Andalib A, Kaneva P, et al. Sex is not everything: the role of gender in early performance of a fundamental laparoscopic skill. *Surg Endosc* 2011;25:1037-1042
11. Haluck RS, Gallagher AG, Satava RM, et al. Reliability and validity of Endotower, a virtual reality trainer for angled endoscope navigation. *Stud Health Technol Inform* 2002;85:179-184

12. Ritter EM, McClusky DA, 3rd, Gallagher AG, et al.. Perceptual, visuospatial, and psychomotor abilities correlate with duration of training required on a virtual-reality flexible endoscopy simulator. *The American Journal of Surgery* 2006;192:379-384
13. Westman B, Ritter EM, Kjellin A, et al. Visuospatial abilities correlate with performance of senior endoscopy specialist in simulated colonoscopy. *J Gastrointest Surg* 2006;10:593-599
14. Arora H, Uribe J, Ralph W, et al. Assessment of construct validity of the endoscopic sinus surgery simulator. *Archives of Otolaryngology- Head and Neck Surgery* 2005;131:217
15. Voyer D, Voyer S, Bryden MP. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychol Bull* 1995;117:250-270
16. Linn MC, Petersen AC eds. A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. Baltimore: Johns Hopkins University Press 1986:67-101
17. Cooper C. Individual differences: Arnold London; 2002
18. Gallagher AG, Ritter EM, Lederman AB, et al. Video-assisted surgery represents more than a loss of three-dimensional vision. *The American Journal of Surgery* 2005;189:76-80
19. Siegel G, Agranoff B, Albers R, et al. eds. Basic Neurochemistry: Molecular, Cellular and Medical Aspects. . Philadelphia: Lippincott; 1999:637-670
20. Gusnard DA, Raichle ME. Searching for a baseline: functional imaging and the resting human brain. *Nature Reviews Neuroscience* 2001;2:685-694
21. Parks RW, Loewenstein DA, Dodrill KL, et al. Cerebral metabolic effects of a verbal fluency test: A PET scan study. *J Clin Exp Neuropsychol* 1988;10:565-575
22. Boivin MJ, Giordani B, Berent S, et al. Verbal fluency and positron emission tomographic mapping of regional cerebral glucose metabolism. *Cortex* 1992;28:231–239
23. Charlot V, Tzourio N, Zilbovicius M, et al. Different mental imagery abilities result in different regional cerebral blood flow activation patterns during cognitive tasks. *Neuropsychologia* 1992;30:565-569
24. Newman SD, Just MA. The neural bases of intelligence. In: Sternberg RJ, Pretz J eds, *Cognition and intelligence*. Cambridge: Cambridge University Press; 2005:88-103

25. Gallagher AG, Ritter EM, Champion H, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 2005;241:364-372
26. Gallagher AG, O'Sullivan GC. *Fundamentals of Surgical Simulation; Principles & Practices* London: Springer Verlag; 2011
27. Andersen DK. How Can Educators Use Simulation Applications to Teach and Assess Surgical Judgment? *Acad Med* 2012;87:934-941
28. Stefanidis D, Korndorffer JR, Black FW, et al. Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. *Surgery* 2006;140:252-262
29. Bell Jr RH, Biester TW, Tabuenca A, et al. Operative experience of residents in US general surgery programs: a gap between expectation and experience. *Ann Surg* 2009;249:719 - 724
30. Hafford ML, Van Sickle KR, Willis RE, et al. Ensuring competency: Are fundamentals of laparoscopic surgery training and certification necessary for practicing surgeons and operating room personnel? *Surg Endosc* 2012:on-line first; DOI: 10.1007/s00464-00012-02437-00467
31. Bruce V, Green PR, Georgeson MA. *Visual perception: Physiology, psychology, & ecology.* Hove,; Psychology Press,; 2003