Surg Endosc (2013) 27:1451–1455
DOI 10.1007/s00464-013-2786-x

EDITORIAL

Measuring surgical skill: a rapidly evolving scientific methodology

**Anthony G. Gallagher · Richard M. Satava ·
George D. Shorten**

Received: 14 December 2012 / Accepted: 28 December 2012 / Published online: 1 February 2013

Minimally invasive surgery (MIS): understanding human performance

The adoption and spread of MIS has forced the surgical profession to investigate why the learning and practice of MIS challenged otherwise proficient surgeons and posed considerable difficulties for the novice or trainee. It has also forced surgery and medicine to develop a better understanding of the factors that can and do compromise operator learning and performance [1, 2]. These developments in turn fostered better insights into the processes of learning surgical procedures (the learning curve) and stimulated the development of tools [3] and measurement strategies [4] that helped to supplant the Halstedian training paradigm (i.e., that of learning from exposure to a large volume of clinical cases). Much of the knowledge and expertise necessary to facilitate this process already existed in the findings of more than a century of research in psychology and the behavioral sciences. Surprisingly, surgery and medicine seemed unaware of this body of research despite the fact that other high-skills domains had already drawn upon and helped develop this body of knowledge (e.g., aviation, space flight).

Human performance, whether in aviation, space flight, nuclear surety or surgery, is correctly characterized as a complex and dynamic interplay between the related human factors of sensation, perception, cognition, and kinesthetic. Each of these factors impacts on psychomotor performance and learning. The fact that they are not directly observable is a major obstacle to developing a comprehensive understanding of these factors. Usually their effect(s) can only be inferred from observations on performance outcomes. However, the aviation and space communities have developed entire industries around the quantitative measurement and understanding of these human attributes and how they impact on operator performance under different high-risk situations. They even use "near accidents" or "near misses" as occasions for learning rather than as a punishment. In-depth and detailed analysis of such events is used to recreate them and to develop strategies for anticipating, avoiding, or coping with them in the future. Pilots and aviators are then made aware of the findings and solutions.

Measurement and motion analysis: the value of hand tracking

The MIS revolution forced the surgical community to develop a similar scientific understanding of performance through quantitative metrics and analytics; however, by comparison to the aviation and space sectors, this enterprise in surgery (and medicine) is in its infancy. There are, however, encouraging signs that this understanding and use of it are growing, particularly for MIS and image-guided interventions. One of the earliest and most elegant approaches to the measurement of surgical skill/performance derived from the observation that talented surgeons

A. G. Gallagher (✉) · G. D. Shorten
School of Medicine, University College Cork, Brookfield Health Sciences Complex, College Rd., Cork, Ireland
e-mail: anthonyg.gallagher@btinternet.com

G. D. Shorten
e-mail: G.Shorten@ucc.ie

R. M. Satava
University of Washington Medical Center, 1959 NE Pacific St., BB487, Seattle, WA 98195-6410, USA
e-mail: rsatava@u.washington.edu

had “good hands,” a qualitative assessment of performance. Darzi’s group at Imperial College London developed the Imperial College Surgical Assessment Device (ICSAD), which they have used to quantify surgical performance and, by inference, skill [5–7]. In their paper, Mason et al. [8] have reviewed the published evidence as it relates to motion analysis and the assessment of surgical performance. Published evidence indicates that time taken, path length, and number of hand movements are indicative of “better” performance, although lesser known measures such as dwell time and reaction time provide additional information. The importance of hand tracking as a technology for the assessment of surgical skill and performance cannot be underestimated. It quantifies a valid human observation that surgeons who are reputed to be skilled and who have good outcomes also appear to perform their operations smoothly, efficiently, and almost effortlessly. Hand-tracking data appear to confirm that skilled individuals demonstrate a shorter path length, make fewer movements, and take less time to perform an operation, but with the caveat that this improved performance is not accompanied by an increase in errors. Indeed, this is the precise measurement strategy that underpins many virtual reality (VR) simulations.

However, hand-tracking measurement on its own could lead to incorrect conclusions. For example, a surgeon may show a shorter path length, less time to perform the procedure, and make fewer hand movements because they have omitted significant parts of the operation! Good VR simulations have overcome this problem by concurrently reporting the “errors” that the operator has enacted or the steps omitted. Thus, the individual’s performance can be judged in context. Once its limitations are recognized, hand tracking is an elegant component to the quantification of a complex and important human endeavor.

Measurement and eye tracking: value added to inferring cognition

Another and equally refined approach to measurement of human performance is eye tracking. This technique can be used to quantify the point of gaze by following the motion of the eye as it relates to a particular scene. The relationship between eye-tracking patterns and visual activities was first made in the 19th century from direct observations of eye movement and reading. Eye movements were believed to reflect human thought processes. Thus, the observer’s thought(s) may be followed to some extent from records of eye movement (the thought accompanying the eye-tracked visual examination of the particular object or task). From these records it was considered possible to determine which elements attract the observer’s eye (and, consequently, their thought), in what order, and how often. According to Rehder and Hoffman [9], visual attention is always slightly (100–250 ms) ahead of the eye; as soon as attention is attracted to a new position, the eyes will “want” to follow. This makes eye tracking a very appealing methodology for helping to quantify human performance, particularly for image-guided tasks. However, as with hand tracking, the relationships between eye movements and thoughts or perceptions are not straightforward. For example, Johansson et al. [10] have suggested that fixation on a face in a picture may indicate a wide variety and contradictory cognitive phenomena such as recognition, liking, dislike, and puzzlement. This means that we cannot infer *specific* cognitive processes directly from fixation on a particular object in a scene. However, eye tracking coupled with other methodologies such as verbal reports (“thinking aloud”) may provide useful insights into performance or performance deficits. Ahmidi et al. [11] reported that, together, eye tracking and motion analysis better characterized MIS performance than either technique on its own: eye tracking implies the cognitive intention and hand tracking demonstrates the completion of the intent. This and similar studies point the way to how technology can be used to quantify complex human performance. However, they are only tools that may be used as part of a scientific enterprise to better understand both the cognitive and psychomotor components of the complex processes that contribute to surgical performance.

One of the most important and yet elusive human factors that is central to operating room performance is attentional capacity. This refers to the ability to focus mental powers upon an object or activity, e.g., carefully observing an object, listening to instructions, decision making, concentrating. The hypothesis is that the brain has a limited amount of cognitive capacity and that certain tasks or activities can be “automated” (away from focused “higher” cortical cognition toward automatic “reflexive” sublimation to midbrain or lower function) through the use of deliberative practice. This is especially pertinent to psychomotor skills, where skills can become automatic (such as knot tying), freeing up the attention center to concentrate upon cognitive tasks such as perception, judgment, and decision making. Attentional capacity as a mediator of surgical performance, first proposed by Gallagher et al. [4], who inferred that surgical performance is improved with (simulation) training because it afforded the trainee spare attentional capacity by ensuring that some of his/her performance repertoire during a surgical procedure was automated. Thus, performing the procedure consumed less of the fixed attentional resources in comparison to a less well trained or experienced operator (as shown in Fig. 1). Based on this model, it was proposed that a surgeon who trained to proficiency using a simulator in the skills laboratory before

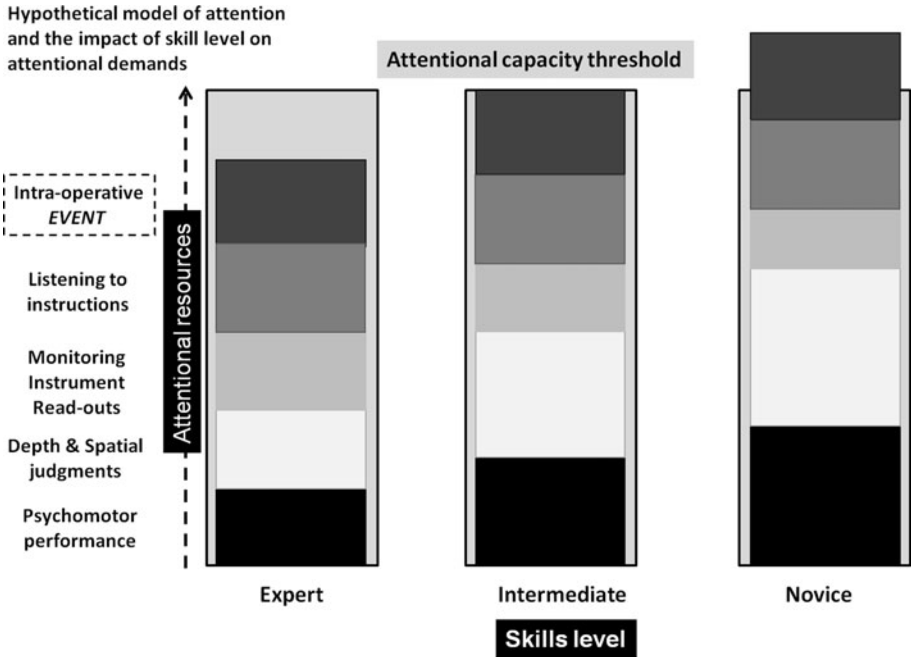

Fig. 1 Hypothetical model of attentional resource allocation of three surgeons with different levels of experience and skill during a minimally invasive surgical procedure

performing surgery would be better able to focus upon instruction and knowledge acquisition during a procedure because of his/her remaining attentional capacity was greater (Fig. 1). Palter et al. [12] have quantitatively validated this effect on attentional resources. They demonstrated that surgical residents who were trained to proficiency prior to performing surgery were better able to recall (and by inference attend to) procedure-relevant information that was presented to them (orally) while they closed the fascia of a patient who had just undergone abdominal surgery. In the study design, the authors' elegantly married experimental methodology with in vivo clinical care to answer an important question that directly impacts on how we configure learning situations for surgical trainees. The results confirm that pretrained novices (i.e., trained to proficiency in the skills lab before performing surgery) were better able to learn during procedural training in the operating room. The study also demonstrated that surgeons have developed the scientific sophistication to address important findings from experimental psychology and evaluate how and if they apply to surgical skill acquisition and practice.

These researchers did not use eye tracking or motion analysis, but their study still provided robust evidence and valid conclusions about the benefits of attentional capacity as a result of training to proficiency. This is no small matter as attentional capacity is one of the most important aspects of human performance that will have a direct and harsh impact on operating surgeons. Surgery differs from some other disciplines in medicine in that the operator must be prepared to make decisions in real time and with incomplete information which will have an immediate and direct impact on patient safety and operative outcome. This means that they must have both the technical capability to perform the surgery and the cognitive resources to manage unanticipated events (such as emergencies or recovery from error). Attentional resource capacity is one of the major rate-limiting performance factors, even for very skilled and experienced surgeons.

Figure 2 is a theoretical construct showing the attentional resources used as a function of the operative activities of two surgeons during a procedure in which two unanticipated events occur, i.e., intraoperative bleeding and an instrument misfire. It is unlikely that a novice would be left alone to deal with this scenario. One or more of these events will probably cause sympathetic nervous system activation. This stimulus as well as the surgeon's appraisal of the urgency of the situation will directly impact on his/her cognitive and technical performance. In this situation, skills and performance elements that the individual previously mastered or "automated" will have become more honed and efficient thus requiring fewer attentional resources devoted to hand–eye coordination, thereby freeing the attention center to focus on the cognitive aspects of the event [13, 14]. However, performance attributes or characteristics that are less well developed or familiar will be negatively impacted by an unanticipated event. This means that for the pretrained novice, because their technical performance, psychomotor skills, and depth and spatial judgments consume less attentional resources, their intraoperative decision making and event response will require more attentional space and make considerable cognitive demands but not necessarily exceed their

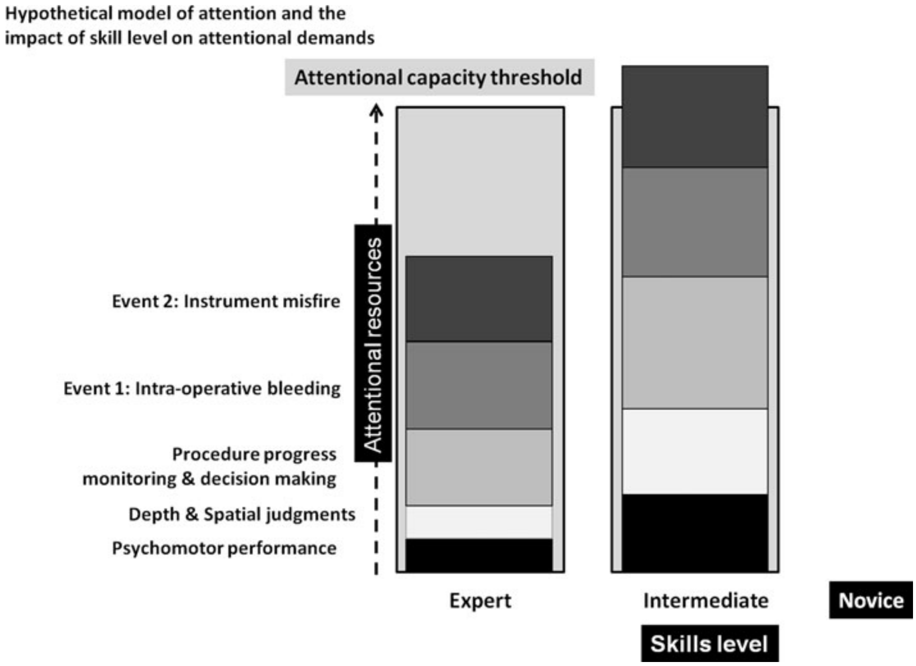

Fig. 2 Hypothetical model of attentional resource allocation of two surgeons with different levels of experience and skill during a minimally invasive surgical procedure where two operative "events" occurred

attentional capacity. The more dominant response for the experienced operator is to deal with an operative situation that is probably familiar to them, hence requiring attentional resources well within their capacity. Even the well-developed attentional capacity of an experienced surgeon can be exceeded: for certain high-risk surgical procedures (e.g., pediatric cardiac surgery, the "switch procedure"), it is clear that the cumulative effect of several unanticipated though minor events can exceed the corrective capacity of an experienced surgeon and increase the likelihood of a poor outcome [15]. It has been found that cognitive performance, e.g., attention, memory, and problem solving, are best when glucocorticoid levels are mildly elevated but poor at high or low levels [16].

Imaging studies

Not surprisingly, the seemingly effortlessness of human sensing, perceiving, remembering, thinking, and problem solving is not without cost. Although the human brain represents only 2 % of the body's weight, it receives 15 % of cardiac output, consumes 20 % of total body oxygen, and uses 25 % of total body glucose [17]. The energy consumption of the brain does not vary greatly over time, but active regions of the cortex consume somewhat more energy than inactive regions; this observation is the basis for the functional brain imaging methods of positron emission tomography and functional magnetic resonance imaging (fMRI) [18]. What neuroscientists had been hoping to find was that the problem-solving workload could be associated with discrete parts of the brain. In contrast, what they have found is that rather than task performance being correlated with a specific brain region, it is distributed across diverse brain regions and both hemispheres [19]. Another consistent finding that has emerged is an inverse relationship between metabolic rates and performance [20]. The results from these brain imaging studies indicate that individuals who are better at a task have lower rates of brain metabolism than those who are not.

Better performance that is resistant to the disruptive effects of stress tends not to happen by accident. Instead, it results from attention to seemingly nontrivial details that impinge on operative performance, with potentially distracting consequences. These can come from unfamiliar/novel or dysfunctional operating equipment (including instruments that misfire), operating in an unfamiliar operating room with unfamiliar staff, or performing an unfamiliar procedure. These apparently inconsequential events may have no obvious or detrimental effects the vast majority of the time; however, they erode important cognitive resources that would be better utilized for solving intraoperative problems when they occur and not used for trying to recall whether a particular piece of equipment was in the operating room, how to tie a knot, or whether the scrub nurse was going to hand over the instrument requested or require prompting again. Small and apparently inconsequential events can have a disproportionate effect, even on apparently well learned behaviors.

Observational assessment: still a fundamental principle

Technologies, such as those that support eye and motion tracking, are attractive to the scientist in the pursuit of objective, verifiable answers. However, one of the most straightforward and robust methodologies for the assessment of surgical skills is direct observation. It is deceptively difficult to do well (i.e., reliably and validly) than perceived by many, is expensive to develop and apply, and is notoriously unreliable (if not well constructed), and yet, for this purpose, it is probably the best methodology available. It should derive from a thorough task analysis by procedure experts/experienced operators and be filtered through an iterative process to a series of performance metrics that accurately characterize optimal and suboptimal procedure performance [21]. These surgeon-derived metrics should shape scientific questions about skilled performance and learning. They may also be the dependent variables that facilitate the measurement of performance with motion analysis, eye movement, or indeed fMRI data. However, these are simply tools for the measurement of aspects of surgical performance. The data derived from these measurements may be used to infer various aspects or attributes of operative performance. The value of these inferences should be measured against rigorous standards, namely, real or potential effects on the outcomes of patients who undergo surgery.

Published evidence indicates that surgeons have become more sophisticated in their use of technology to investigate, quantify, and understand the human attributes that characterize and contribute to operative performance. These developments have almost certainly been driven by the MIS community, which was forced to find answers as to why this type of surgery was more difficult to learn and practice than traditional open surgery. In developing and addressing more of these questions, surgeons have had to become sophisticated in measurement methodologies and related technologies. But no matter how sophisticated or advanced the technology or the methodologies, they still rely on two of the most fundamental aspects of the scientific process: observation and experimental methodology.

Disclosures R. M. Satava has acted as a technical consultant for Karl Storz, GmbH; holds stock in In-Touch Health Technologies, Inc; holds a research grant from Intuitive Surgical, Inc.; and has received honoraria from Ethicon Endosurgery, Inc. and Covidien, Inc. for conference participation. A. G. Gallagher and G. D. Shorten have no conflicts of interest or financial ties to disclose.

References

1. Patkin M, Isabel L (1993) Ergonomics and laparoscopic general surgery. In: Graber JN, Schultz LS, Pietrafitta J, Hickok DF (eds) Laparoscopic abdominal surgery. McGraw-Hill, New York
2. Cuschieri A (1995) Whither minimal access surgery: tribulations and expectations. Am J Surg 169:9–19
3. Satava RM (1993) Virtual reality surgical simulator: the first steps. Surg Endosc 7:203–205
4. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, Smith CD, Satava RM (2005) Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Ann Surg 241:364–372
5. Darzi A, Smith S, Taffinder N (1999) Assessing operative skill. Br Med J 318:887
6. Smith SG, Torkington J, Brown TJ, Taffinder NJ, Darzi A (2002) Motion analysis. Surg Endosc 16:640–645
7. Datta V, Chang A, Mackay S, Darzi A (2002) The relationship between motion analysis and surgical technical assessments. Am J Surg 184:70–73
8. Mason JD, Ansell J, Warren N, Torkington J (2013) Is motion analysis a valid tool for assessing laparoscopic skill? Surg Endosc. doi:10.1007/s00464-012-2631-7
9. Rehder B, Hoffman AB (2005) Eyetracking and selective attention in category learning. Cogn Psychol 51:1–41
10. Johansson R, Holsanova J, Holmqvist K (2010) Eye movements during mental imagery are not reenactments of perception. In: Ohlsson S, Catrambone R (eds) Proceedings of the 32nd annual conference of the Cognitive Science Society, Portland, OR, 11–14 Aug 2010, pp 1968–1973
11. Ahmidi N, Hager G, Ishii L, Fichtinger G, Gallia G, Ishii M (2010) Surgical task and skill classification from eye tracking and tool motion in minimally invasive surgery. Med Image Comput Comput Assist Interv 13(Pt 3):295–302
12. Palter VN, Grantcharov T, Harvey A, MacRae HM (2011) Ex vivo technical skills training transfers to the operating room and enhances cognitive learning: a randomized controlled trial. Ann Surg 253:886–889
13. Zajonc RB (1965) Social facilitation. Science 149:269–274
14. Yerkes RM, Dodson JD (1908) The relationship of stimulus to rapidity of habit formation. J Comp Neurol Psychol 18:459–482
15. de Leval MR, Carthey J, Wright DJ, Farewell VT, Reason JT (2000) Human factors and cardiac surgery: a multicenter study. J Thorac Cardiovasc Surg 119:661–672
16. Lupien SJ, Maheu F, Tu M, Fiocco A, Schramek TE (2007) The effects of stress and stress hormones on human cognition: implications for the field of brain and cognition. Brain Cogn 65:209–237
17. Siegel G, Agranoff B, Albers R, Fisher S, Uhler M (eds) (1999) Basic neurochemistry: molecular, cellular and medical aspects. Lippincott, Philadelphia, pp 637–670
18. Gusnard DA, Raichle ME (2001) Searching for a baseline: functional imaging and the resting human brain. Nat Rev Neurosci 2:685–694
19. Parks RW, Loewenstein DA, Dodrill KL, Barker WW, Yoshii F, Chang JY, Emran A, Apicella A, Sheramata WA, Duara R (1988) Cerebral metabolic effects of a verbal fluency test: a PET scan study. J Clin Exp Neuropsychol 10:565–575
20. Boivin MJ, Giordani B, Berent S, Amato DA, Lehtinen S, Koeppe RA, Buchtel HA, Foster NL, Kuhl DE (1992) Verbal fluency and positron emission tomographic mapping of regional cerebral glucose metabolism. Cortex 28:231–239
21. Gallagher AG, O'Sullivan GC (2011) Fundamentals of surgical simulation; principles & practices. Springer Verlag, London