Who Beats the Expert? Building Precision into Simulators for Surgical Skill Assessment

Birgitta Dresp

ISSN: 2574-1241
DOI: 10.26717/BJSTR.2019.13.002429
Birgitta Dresp-Langley. Biomed J Sci & Tech Res

BIOMEDICAL Journal of Scientific & Technical Research

Opinion Open Access

Who Beats the Expert? Building Precision into Simulators for Surgical Skill Assessment

Birgitta Dresp-Langley*

Department of Cognitive Neuroscience, ICube Lab, France

Received: January 12, 2019; **Published:** January 25, 2019

***Corresponding author:** Birgitta Dresp-Langley, Department of Cognitive Neuroscience, ICube Lab, France

Abstract

Simulator training for image-guided surgical interventions allows tracking task performance in terms of speed and precision of task execution. Simulator tasks are more or less realistic with respect to real surgical tasks, and the lack of clear criteria for learning curves and individual skill assessment if more often than not a problem. Recent research has shown that trainees frequently focus on getting faster at the simulator task, and this strategy bias often compromises the evolution of their precision score. As a consequence, and whatever the degree of surgical realism of the simulator task, the first and most critical criterion for skill evolution should be task precision, not the time of task execution. This short opinion paper argues that individual training statistics of novices from a simulator task should therefore always be compared with the statistics of an expert surgeon from the same task. This implies that benchmark statistics from the expert are made available and an objective criterion, i.e. a parameter measure, for task precision is considered for assessing learning curves of novices.

Keywords: Surgical Simulator Training; Individual Performance Trend; Surgical Expertise; Precision; Speed-Precision Trade-Off Functions

Opinion

Objective performance metrics [1-12] form an essential part of surgical simulator systems for optimal independent training. Putting such metrics to the test on users with different levels of expertise appears mandatory for perfecting existing systems. The exploitation of individual performance metrics to establish learning curves that a user can see and understand and that will help him/her elaborate task strategies for measurable skill improvement is the most important aspect of an effective training system [6-12]. Metric-based skill assessment ensures that training sessions are more than simulated clinical procedures, and that trainees are provided with insight about how they are doing in a task, and how they could improve their current scores. Not all simulator tasks are based on surgically realistic physical task models, but at the earliest stages of training surgical task realism is probably not what matters most [13-20]. Whatever the degree of realism of the simulator task, metric based skill assessment gets rid of subjectivity in evaluating skill evolution, and there is no ambiguity about the progress of training.

Moreover, some work has shown that benchmarking individual levels of proficiency against the performance levels of experts on a validated, metric-based simulation system has well-established intrinsic face validity [1,2,10]. It therefore appears the better approach compared with benchmarking on abstract performance concepts or on the basis of expert consensus. Building expert performance in terms of benchmark metrics into simulator training programs would provide an almost ideal basis for automatic skill assessment and ensure that desired levels of skill are defined on the grounds of realistic criteria. Such are, in principle, available in the proficiency levels of individuals who are highly experienced at performing clinical procedures with the highest level of precision [1,11], which is probably the strongest argument for building expert performance data into any simulator system for a direct comparison with novice data at any moment of the training procedure.

Recently, early simulator training models for assessing skill evolution on the basis of individual speed-precision trade-offs, recorded and directly exploited in the light of the benchmark statistics of an expert surgeon in the context of an experimental simulator environment, have been proposed [1,5]. The approach is based on a simple and universal psychophysical human performance model of individual strategies during motor learning [9,21-32]. They focus on individual speed-precision trade-off functions during training at early stages and are based on objective criteria for task precision and task time at any moment in the evolution of individual performance. The individual speed-precision trade-off functions of complete novices having trained for a large number of simulator

sessions reveal, indeed, completely different speed-precision strategies. precision focused strategy most closely matches the benchmark statistics (mean and standard deviations for time and precision, of an expert surgeon, as shown previously [1].

This is illustrated here in Figure 1 by two learning curves of two novice trainees from some of these studies [5], validated further in the light of an expert's benchmark performance statistics. The strategy differences highlighted here in Figure 1 occur spontaneously in novices during early training and are most frequently completely non-conscious. In other words, trainees at these early stages of simulator training do not really know what they are doing, or what to do to improve their skills effectively. Therefore, it is in the very early stages of simulator training where the most effective guidance is required to bring out the full potential in trainees. How individual strategies should be detected, and if necessary, controlled for and modified as early as possible in simulator training, has been discussed on the principle of semi-automatic control procedures in the light of expert benchmark data [1,5]. A training system should be able to know the expert surgeon's statistics relative to task precision and task time, not only time [20,21], and these data should be in-built to the system.

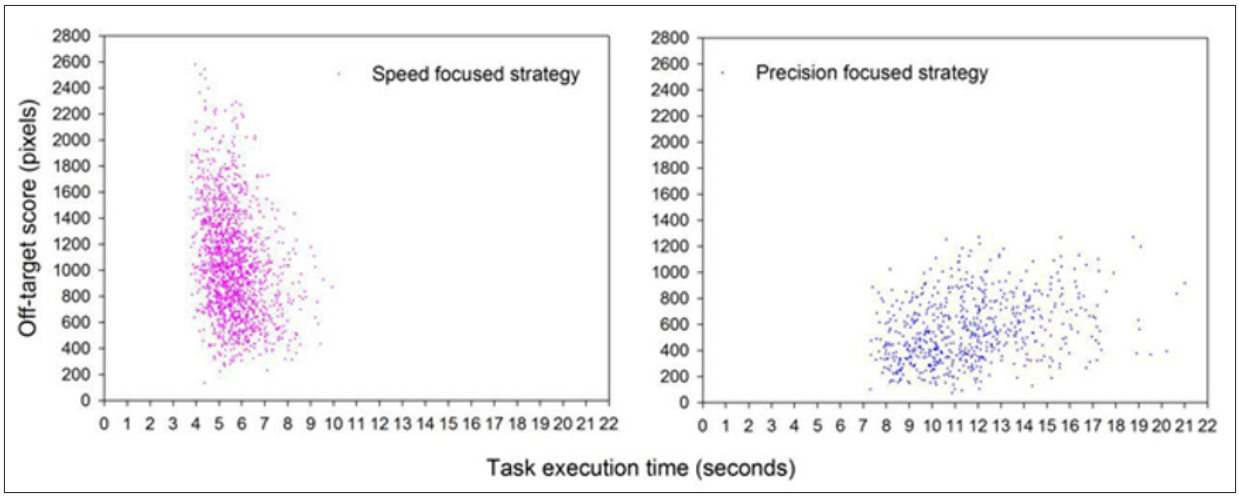

Figure 1: Two different learning curves expressed in terms of the individual speed-precision trade-off functions of two novices having trained in a large number of sessions.

On this basis, it will be possible to detect and if necessary, controls for individual speed-precision strategies by comparing a trainee's performance statistics to the expert's benchmarks. The means and standard deviations from the precision focused strategy closely match the performance benchmark statistics of an expert surgeon highly skilled in clinical precision interventions and performing in the simulator task with no prior training in that specific task, as shown by Batmaz, Dresp-Langley, de Mathelin, and other colleagues in some of their recent work [5-9].

Building reliable precision scores into learning curve approaches in surgical simulator training will ultimately enable the human expert tutors in charge of training programs to provide appropriate user feed-back to the novice when necessary and as early as possible in the program. The goal of early simulator clearly should be to empower, to bring out the full extent of individual potential and give any trainee, without the direct intervention of human tutors who even when they are experts may be biased, the possibility to attain the highest level of skill he/she is capable of on the simulator. This will ultimately result in fostering, and ultimately selecting for, strategy-aware soon-to-be surgeons with optimal precision skills.

How increasingly objective skill assessment will help improve surgical simulator training is still by and large an open question. Although artificial intelligence provides well-suited concepts for knowledge implementation, automatic feed-back procedures and the exploitation of prior (learnt) benchmark knowledge, building such procedures into simulator training is anything but straightforward. Early-stage "dry-lab" training programs are offered to large numbers of individuals, often on experimentally developed simulators, and supervision of the training programs by one or two experts is mandatory. Automatic control procedures [13] that exploit metric-based benchmark criteria and perform statistically driven performance comparisons, with trial-by-trial feed-back at any given moment in time, may prove helpful if they are exploited effectively. The goal of early simulator training should be to help the largest possible number of registered individuals reach their optimal performance levels as swiftly as possible [15,23]. Skill assessment in terms of an end-of-session performance status that highlights differences between trainees is not enough, especially when there is no way of knowing what these differences actually mean, i.e. what they tell us about true surgical talent.

Faced with the problem of defining reliable performance standards, it is important that simulator systems and the metrics they exploit to control performance evolution, including automatic or AI assisted procedures with feedback [22], have been validated by the performance parameters of an expert. This will ensure that the training criteria are likely to match those required for performing

real surgical tasks, and that the learning task measures some of the most relevant characteristics of surgical skill. Since many different physical task models exist, surgical simulator training is permanently confronted with the problem of generalization of the learning curves and, ultimately, uncertainty about skill transfer to real-world surgical interventions.

The task models and control principles reviewed here and more extensively explained and conceptualized in previous work referred to herein should ideally be implemented at the earliest stages of "dry lab" simulator training. Some of them could be adapted to a variety of eye-hand-tool coordination tasks that allow for computer controlled criteria relative to task precision p at any critical task moment in time t. Early simulator training tasks should successfully tell apart the performance levels of a large number of novices from those of a surgical expert, not necessarily trained on the simulator but exhibiting a stable near-optimal performance with respect to task precision. Clearly, if an early training system satisfies this criterion, then it is indeed likely to measure critical aspects of surgical skill that should transfer to real surgical tasks [14-23]. Building precision scores into the learning curves early, should help produce a selection of trainees that will perform better later-on, in more specific tasks on physical models, and in the clinical context. Then, direct supervision by experts will allow taking individual skills to the next level.

Whatever the simulator system and task, be it surgically realistic or less, a single performance metric will inevitably give an incomplete assessment of user performance [1,11,20]. Task completion time in the absence of other, more telling criteria, is a poor and largely misleading measure of surgical skill evolution [19-21]. Some metrics imply that there would be some global optimum performance value, such as a minimal tool path length, a minimal completion time, or other minimal quantities such as forces [7,22] or velocities [24,33-36] reflecting optimal performance. These supposedly optimal values, however, may vary in relation to changes in conditions, which need to be considered. The assumed optimum per se can, in reality, on be known through analysis of expert performance in the same task and on the specific simulator. Such analysis only will give insight into the nature of user-task-condition dependencies and, ultimately, help develop better simulators. Also, some important elements of surgical proficiency have not yet been explored to become part of largely unsupervised simulator training programs and the field is still in need of a large amount of experimental and conceptual work leading in that direction.

Metric-based criteria for task precision in limited task time are sometimes difficult to define. How, for example, does one measure the precision with which a surgical knot [15] is tied in a given time t? In procedures where the camera moves along with the tool [7,10-16], as is the case with most robot-assisted procedures, the frequency and duration of camera movements or camera movement intervals may be highly important indicators of technical skill. The ease with which the trainee controls the tool may be a direct correlate of the precision of tool-target alignment, for example. In combination with other performance metrics such as task completion time, economy of tool motion, or master workspace range used, a variety of precision measures may be exploitable, but have not yet been fully explored or validated. Training control procedures based on device-specific expert performance benchmarks and clear precision metrics will, sooner or later, provide new solutions to the old and still unresolved problem of heuristic validity of learning curves in surgical simulator training.

References

1. Dresp-Langley B (2018) Towards expert-based speed-precision control in early simulator training for novice surgeons. Information 9: 316.
2. Stunt J, Wulms P, Kerkhoffs G, Dankelman J, van Dijk C, et al. (2014) How valid are commercially available medical simulators? Adv Med Educ Pract 5: 385-395.
3. Marcano, L, Komulainen T, Haugen, FA (2017) Implementation of performance indicators for automatic assessment. Computer Aided Chemical Engineering 40: 2971-2976.
4. Marcano L, Yazidi A, Ferati M, Komulainen T (2017) Towards effective automatic feedback for simulator training.
5. Batmaz AU, de Mathelin M, Dresp-Langley B (2016a) Getting nowhere fast: Trade-off between speed and precision in training to execute image-guided hand-tool movements. BMC Psychology 4: 55.
6. Batmaz AU, de Mathelin M, Dresp-Langley B (2016b) Effects of indirect screen vision and tool-use on the time and precision of object positioning on real-world targets. Perception 45: S286.
7. Batmaz AU, Falek M, Nageotte F, Zanne P, Zorn L, et al. (2017) Novice and expert haptic behaviours while using a robot controlled surgery system. 13th IASTED International Conference on Biomedical Engineering (BioMed).
8. Batmaz AU, de Mathelin M, Dresp-Langley B (2017) Seeing virtual while acting real: Visual display and strategy effects on the time and precision of eye-hand coordination. PLoS One 12(8): e0183789.
9. Batmaz AU, de Mathelin M, Dresp-Langley B (2018) Effects of 2D and 3D image views on hand movement trajectories in the surgeon's peri-personal space in a computer controlled simulator environment. Cogent Medecine p. 5.
10. Gallagher AG, O'Sullivan C (2011) Fundamentals in surgical simulation: Principles and practice. Improving Medical Outcome - Zero Tolerance Series, Springer Science & Business Media pp. 374.
11. Gallagher AG (2012) Metric-based simulation training to proficiency in medical education: What it is and how to do it. Ulster Med J 81(3): 107-113.
12. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, et al. (2005) Virtual reality simulation for the operating room: Proficiency-based training as a paradigm shift in surgical skills training. Ann Surg 241(2): 364-372.
13. Dreyfus HL, Dreyfus SE, Athanasiou T (1986) Mind over machine: The power of human intuition and expertise in the era of the computer. Artificial Intelligence 33(1): 135-140.
14. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Andersen DK, et al. (2004) Analysis of errors in laparoscopic surgical procedures. Surg Endosc 18(4): 592-595.
15. Van Sickle K, Smith B, McClusky DA, Baghai M, Smith CD, et al. (2005) Evaluation of a tensiometer to provide objective feedback in knot-tying performance. Am Surg 71(12): 1018-1023.
16. Van Sickle KR, Gallagher AG, Smith CD (2007) The effect of escalating feedback on the acquisition of psychomotor skills for laparoscopy. Surg Endosc 21(2): 220-224.

17. A Reznick RK (1993) Teaching and testing technical skills. Am J Surg 165(3): 358-361.

18. Chen C, White L, Kowalewski T, Aggarwal R, Lintott C, et al. (2014) Crowd-sourced assessment of technical skills: A novel method to evaluate surgical performance. J Surg Res 187(1): 65-71.

19. Moorthy K, Munz Y, Sarker SK, Darzi A (2003) Objective assessment of technical skills in surgery. BMJ 327(7422): 1032-1037.

20. Sewell C, Morris D, Blevins NH, Dutta S, Agrawal S, et al. (2008) Providing metrics and performance feedback in a surgical simulator. Comput Aided Surg 13(2): 63-81.

21. Ritter EM, McClusky DA, Gallagher AG, Smith CK (2005) Real-time objective assessment of knot quality with a portable tensiometer is superior to execution time for assessment of laparoscopic knot-tying performance. Surgical Innovation 12(3): 233-237.

22. Rosen J, Hannaford B, Richards CG, Sinanan MN (2001) Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. IEEE Trans Biomed Eng 48(5): 579-591.

23. Jarc AM, Curet MJ (2017) Viewpoint matters: Objective performance metrics for surgeon endoscope control during robot-assisted surgery. Surg Endosc 31: 1192-1202.

24. Dresp-Langley B (2015) Principles of perceptual grouping: Implications for image-guided surgery. Front Psychol 6: 1565.

25. Fogassi L, Gallese V (2004) Action as a binding key to multisensory integration. In: Calvert G, Spence C, Stein BE (Eds.), Handbook of multisensory processes. MIT Press, Cambridge, pp. 915.

26. Bonnet C, Dresp B (1993) A fast procedure for studying conditional accuracy functions. Behavior Research, Instruments & Computers 25: 2-8.

27. Fitts PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 47(6): 381-391.

28. Meyer DE, Irwin A, Osman, AM, Kounios J (1988) The dynamics of cognition and action: Mental processes inferred from speed-accuracy decomposition. Psychol Rev 95: 183-237.

29. Luce RD (1986) Response times: Their role in inferring elementary mental organization. Oxford University Press, New York, pp. 562.

30. Held R (2009) Visual-haptic mapping and the origin of cross modal identity. Optom Vis Sci 86(6): 595-598.

31. Henriques DY, Cressman EK (2012) Visuo-motor adaptation and proprioceptive recalibration. J Mot Behav 44(6): 435-444.

32. Krakauer JW, Mazzoni P (2011) Human sensorimotor learning: adaptation, skill and beyond. Curr Opin Neurobiol 21(4): 636-644.

33. Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ (2012) Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. J Urol 187(1): 247-252.

34. Smith R, Patel V, Satava R (2014) Fundamentals of robotic surgery: A course of basic robotic surgery skills based upon a society consensus template of outcomes measures and curriculum development. Int J Med Robot Comput Assist Surg 10(3): 379-384.

35. Aiono S, Gilbert JM, Soin B, Finlay PA, Gordan A (2002) Controlled trial of the introduction of a robotic camera assistant (Endo Assist) for laparoscopic cholecystectomy. Surg Endosc Other Interv Tech 16(9): 1267-1270.

36. King BW, Reisner LA, Pandya AK, Composto AM, Ellis RD, et al. (2013) Towards an autonomous robot for camera control during laparoscopic surgery. J Laparoendosc Adv Surg Tech 23(12): 1027-1030.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2019.13.002429

Birgitta Dresp-Langley. Biomed J Sci & Tech Res