The American Journal of Surgery (2015) 209, 645-651

The American Journal of Surgery®

Society of Black Academic Surgeons

Idle time: an underdeveloped performance metric for assessing surgical skill

Anne-Lise D. D'Angelo, M.D.[a,*], Drew N. Rutherford, M.S.[a,b], Rebecca D. Ray, Ph.D.[a], Shlomi Laufer, Ph.D.[a,c], Calvin Kwan, B.S.[a], Elaine R. Cohen, M.Ed[a], Andrea Mason, Ph.D.[b], Carla M. Pugh, M.D., Ph.D.[a]

[a]*Department of Surgery, School of Medicine and Public Health, University of Wisconsin – Madison, 750 Highland Avenue, Madison, WI 53726, USA;* [b]*Department of Kinesiology, School of Education, University of Wisconsin – Madison, 2000 Observatory Drive, Madison, WI 53706, USA;* [c]*Department of Electrical and Computer Engineering, College of Engineering, University of Wisconsin – Madison, 1415 Engineering Drive, Madison, WI 53706, USA*

KEYWORDS:
Motion tracking;
Surgical skills;
Assessment;
Idle time;
Path length;
Simulation

Abstract

BACKGROUND: The aim of this study was to evaluate validity evidence using idle time as a performance measure in open surgical skills assessment.

METHODS: This pilot study tested psychomotor planning skills of surgical attendings (n = 6), residents (n = 4) and medical students (n = 5) during suturing tasks of varying difficulty. Performance data were collected with a motion tracking system. Participants' hand movements were analyzed for idle time, total operative time, and path length. We hypothesized that there will be shorter idle times for more experienced individuals and on the easier tasks.

RESULTS: A total of 365 idle periods were identified across all participants. Attendings had fewer idle periods during 3 specific procedure steps ($P < .001$). All participants had longer idle time on friable tissue ($P < .005$).

CONCLUSIONS: Using an experimental model, idle time was found to correlate with experience and motor planning when operating on increasingly difficult tissue types. Further work exploring idle time as a valid psychomotor measure is warranted.

Supported by the National Institutes of Health (Number 1F32EB017084-01) entitled "Automated Performance Assessment System: A New Era in Surgical Skills Assessment" and the Department of Defense (Number W81XWH-13-1-0080) entitled "Psycho-Motor and Error Enabled Simulations Modeling Vulnerable Skills in the Pre-Mastery Phase–Medical Practice Initiative Procedural Skill Decay and Maintenance (MPI-PSD)". Dr. Pugh holds a patent entitled "Medical Examination Teaching System" US 6428323 B1 that relates to the use of sensor technology for capturing hands-on performance. Study sponsors were not involved in study design; data collection, analysis and interpretation; writing of the manuscript; or decision to submit the manuscript for publication.

* Corresponding author. Tel.: +1-608-263-5218; fax: 608-263-0454.

E-mail address: amaag@uwhealth.org

Manuscript received August 12, 2014; revised manuscript December 6, 2014

0002-9610/$ - see front matter © 2015 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amjsurg.2014.12.013

Objective measures of technical surgical skills are needed for accurate feedback and competency evaluations.[1] Current assessments of technical skills include observer-generated task-specific checklists and global rating scales,[2–5] and technology-based performance measures.[6–9] Observer-based scoring metrics are readily accessible and inexpensive. Although commonly used to assess trainees, observer-based scoring metrics remain subject to bias[10] and can be time consuming. In contrast, although technology-based performance measures are more expensive, they provide a unique and unparalleled opportunity for automated and objective assessment methods. Further integration of technology into surgical skills assessment is critical for developing objective performance measures and providing an explicit path to surgical mastery.

The movement of motor skills acquisition outside the operating room and into the simulation environment permits the incorporation of technology-based metrics. Specifically, motion tracking technology allows for the objective measurement of motor behavior. Prior research on psychomotor skills assessment in open surgery has employed an electromagnetic tracking device called the Imperial College Surgical Assessment Device.[11–15] This work demonstrated evidence of construct validity with relationships among number of hand movements and experience level,[13] final product scores,[14] and observer-based global rating scores.[12,13]

The majority of studies evaluating motor behavior in surgical skills focus on metrics regarding hand movements including path length and the amount of time to complete the procedure.[11–15] These metrics represent a subset of surgical skills performance and technical skills metrics available with motion tracking technology. Although the prior work on motor movement provides a strong foundation for understanding surgical performance, there is a paucity of research on those instances where there is no movement—idle time. Idle time is characterized by a lack of movement of both hands and may represent periods of motor planning or decision making that can be used to differentiate performance.[16] Additionally, as individuals progress through stages of motor learning, there is a rapid reduction in the cognitive processes associated with difficult motor planning and a convergence on more rapid automaticity.[17] This is expected to translate into less idle time during a task performed by more experienced individuals who have had more practice.

Our work involves development and implementation of decision-based simulators with a variety of anatomical presentations that provide a range of task complexity.[18–21] The variable tissue simulator was developed to present suturing tasks of varying levels of difficulty based on the materials presented. The first aim of this pilot study was to evaluate idle time as a potential surgical performance metric. We hypothesize that there will be shorter idle times for more experienced individuals and on the easier tasks. The second aim of this pilot study was to assess the ability of the variable tissue simulator to differentiate psychomotor performance based on experience and task difficulty. We hypothesize that the time and path length to complete the suturing task will be shorter for more experienced individuals and on less difficult tasks.

Methods

Setting and participants

The study participants (n = 15) were medical students (n = 5), surgery residents (n = 4), and attending surgeons (n = 6) at a Midwestern academic hospital. Participants were recruited through departmental list serves and participation was voluntary and based on availability. Participants performed a simulated suturing task on the variable tissue simulator in a single setting. Performance data were collected with video recordings and an optical motion tracking system. There were no time restrictions on the suturing task and feedback was not provided.

Study approval was granted by the University of Wisconsin Health Sciences Institutional Review Board and written informed consent was obtained from all participants.

Surveys

Before performing the suturing task, participants completed a survey that collected demographic information including the level of training, surgical specialty, and handedness. Following the suturing task, participants responded to a question asking if the motion tracking system interfered with their ability to perform the task (5-point Likert scale: 1 = strongly disagree, 5 = strongly agree).

Variable tissue simulator

The variable tissue simulator was designed to present a simulated suturing task that induced decision making by varying the materials and difficulty of the suturing task. The simulator was composed of a board with simulated materials held in place by clips. In this study, 3 different simulated tissue types were presented: foam (dense connective tissue), rubber balloons (artery), and tissue paper (friable tissue) (Fig. 1). The suturing task was to place 3 interrupted instrument-tied sutures on 2 opposing pieces of material. 3-0 Prolene suture was provided along with a needle driver, surgical forceps and suture scissors. The development of the variable tissue simulator was guided by prior cognitive task analysis with 2 expert surgeons.[22]

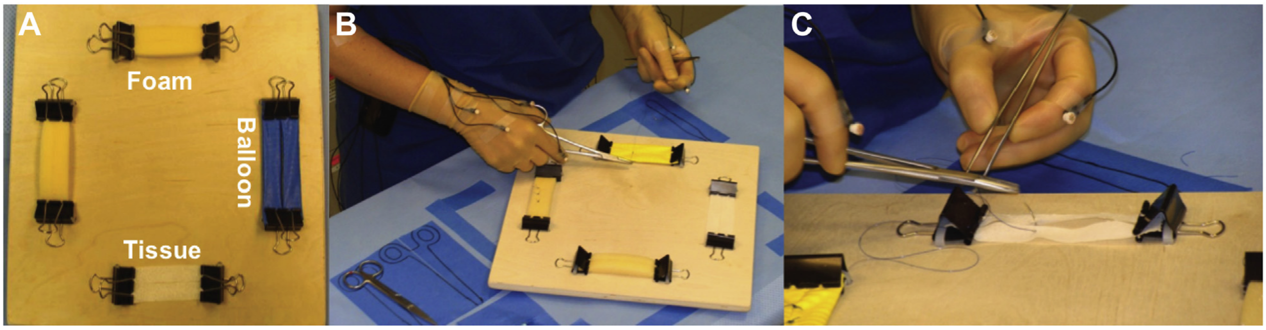

Figure 1 Variable tissue simulator with simulated tissue types (A) on which participants performed the suturing task (B and C) with motion capture from an optical motion tracking system.

Motion tracking system

Hand movements were measured with an optical motion tracking system (Visualeyez 3000; Phoenix Technologies, Inc, Campbell, CA). Four infrared light emitting diode markers were affixed by medical tape (Transpore, 3M Company, St. Paul, MN) to the participant's gloved hands. Locations on both hands were as follows: (1) distal dorsal surface of the second digit phalanx; (2) distal dorsal surface of the first digit metacarpal; (3) mid-dorsal surface of the hand; and (4) dorsal surface of the forearm just proximal to the wrist joint. Data were collected from both of the participant's hand movements. Traces of marker positions in millimeters were sampled at 180 Hz. To remove any high-frequency noise artifacts, data were filtered with a dual-pass second-order Butterworth filter with a low-pass cutoff frequency of 7 Hz.

Motion tracking measures

Motion tracking data of both hands were analyzed for differences in idle time, total operative time, and path length. An idle threshold analysis was performed by calculating the proportions of time the hands spent below velocity values ranging from 1 to 50 mm/seconds. The threshold yielding the greatest differences in proportion among participants, 20 mm/seconds, was the velocity selected to define idle periods. Also, a minimum duration of .5 seconds was chosen to define a deliberate idle period to avoid capturing very brief pauses. Total operative time was calculated from the time participants' hands left the starting position until participants cut the tail of the third suture. Path length is the 3-dimensional distance the hands traveled from the starting position.

Video analysis

Idle periods were identified from the motion tracking data and correlated to specific time points in the video recordings of the participants. The video recordings were then used to code when in the procedure the idle periods were occurring. The procedure steps were broken down into 7 main categories: (1) Entering the tissue with the needle; (2) Driving the needle through the tissue; (3) Pulling the needle out of the tissue; (4) Tightening a knot; (5) Cutting a suture tail; (6) Grasping suture with an instrument; and (7) Managing the suture. Idle periods occurring when the participant was interacting with the experimenter were also noted. For analysis, videos were reviewed by a single blinded rater. To assess inter-rater agreement, a 20% random sample was reviewed and classified by a second blinded rater.

Data analysis

Our hypothesis is that validity evidence exists for the use of 3 different psychomotor measures (idle time, total operative time, and path length) to differentiate performance by experience level and task difficulty. To test this hypothesis, repeated measures analysis of variance (ANOVA) with experience level (medical students, surgery residents, and attending surgeons) as a between-subjects variable and tissue type (foam, rubber, and tissue paper) as a repeated within-subjects variable was used to evaluate main effects of experience level and tissue type. Chi-square analysis was performed to assess for an association between procedure steps when idle periods were occurring and participant experience level. Inter-rater reliability for video analysis of idle periods was assessed using a Cohen's kappa (κ). Pearson's correlation coefficients were calculated to evaluate correlations among idle time, total operative time, and path length. All analyses were performed using IBM SPSS Statistics, version 22 (IBM SPSS Statistics, Armonk, NY).

Results

Surveys

The study participants ($n = 15$) were medical students ($n = 5$), surgery residents ($n = 4$), and attending surgeons ($n = 6$). The medical students were either third or fourth

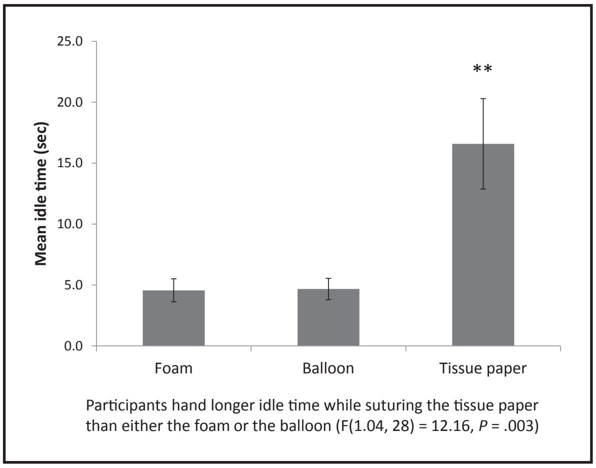

Figure 2 Mean idle time per task by tissue type during the suturing task.

year students who had completed at least one surgical rotation. Surgical residents ranged from postgraduate year 1 to 3. Attending surgeons had an average length of time in practice of 14.2 years (standard deviation [SD] = 12.0). All participants completed all 3 suturing tasks. Overall, participants felt that the motion tracking system did not interfere with their ability to perform the task (M = 1.9/5.0, SD = 1.1).

Idle time

Motion tracking measures. The repeated measures ANOVA of idle time revealed a main effect for tissue type (F(1.04, 28) = 12.16, P = .003) (Fig. 2). All groups paused longer when working on the tissue paper (M = 16.58 seconds, SD = 14.35) compared with the balloon (M = 4.67 seconds, SD = 3.39, $t(14)$ = 3.73, P = .003) or foam (M = 4.56 seconds, SD = 3.67, $t(14)$ = 3.35, P = .005). Cohen's effect size value (tissue vs balloon d = 1.1; tissue vs foam d = 1.2) suggests a high practical significance.

Repeated measures ANOVA of idle time was not significant for experience level ($F(2.00, 12.00)$ = 27.13, P = .742).

Video analysis. A total of 365 idle periods were identified by the motion tracking system across all participants and was further evaluated with video analysis. Video analysis of procedure steps when idle periods occurred revealed high inter-rater reliability (k = .887). Most of the idle periods (99.2%) occurred while participants were performing the task. There was a significant difference in the number of idle periods during specific procedure steps ($\chi^2(6)$ = 238.65, $P < .001$). Idle periods by procedure step and experience level are shown in Fig. 3. Results show that attending surgeons had fewer idle periods while "Entering the tissue with the needle" (P = .001); "Driving the needle through the tissue" ($P < .001$); and "Pulling the needle out of the tissue" (P = .007). Attending surgeons had more idle periods when "Tightening a knot" ($P < .001$).

Total operative time

The repeated measures ANOVA of total operative time revealed a main effect for tissue type (F(2,24) = 60.02, $P < .001$) (Fig. 4A). It took significantly longer for all groups to complete the suturing task on tissue paper (M = 101.73 seconds, SD = 30.89) compared with balloon (M = 57.53 seconds, SD = 13.65, $t(14)$ = 8.67, $P < .001$) or foam (M =

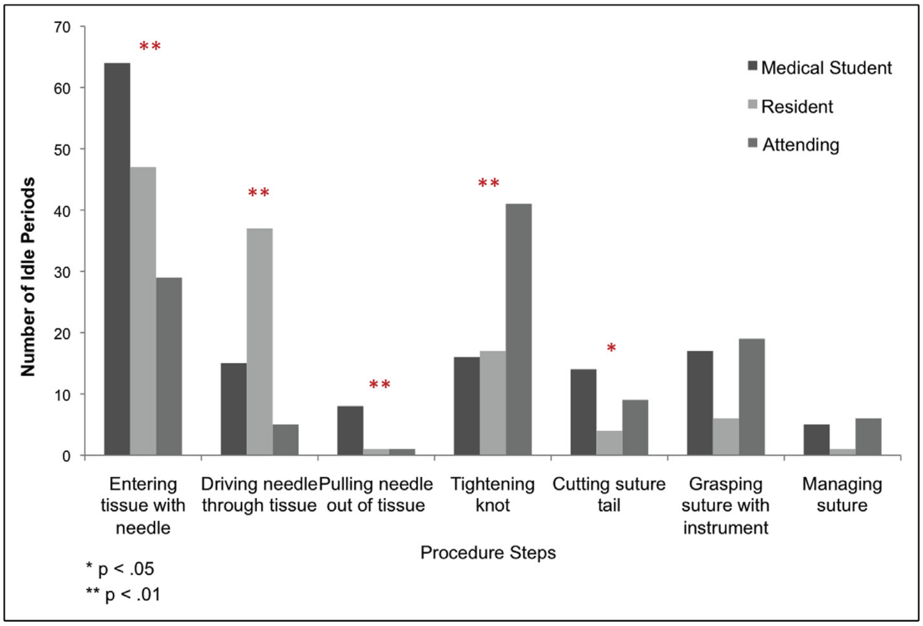

Figure 3 Number of idle periods by experience level during specific procedure steps of the suturing task.

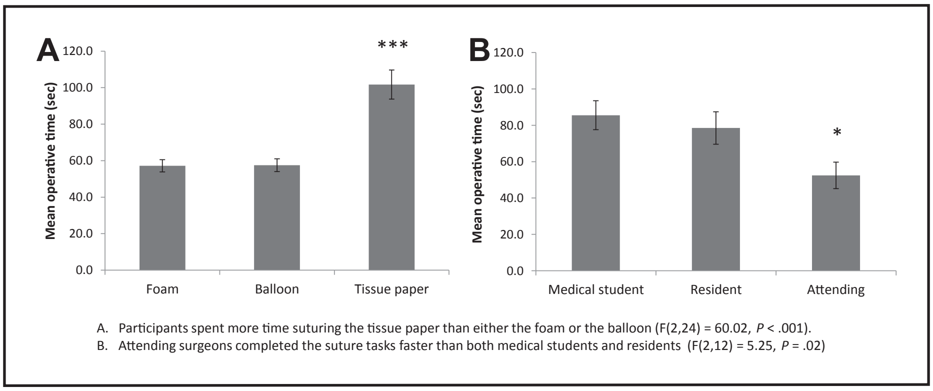

Figure 4 Mean total operative time per stitch by tissue type and experience level during the suturing task.

57.20 seconds, SD = 12.99, $t(14) = 7.46$, $P < .001$). Cohen's effect size value (tissue vs balloon $d = 1.9$; tissue vs foam $d = 1.9$) suggests a high practical significance.

A repeated measures ANOVA of total operative time showed a main effect of experience level (F(2,12) = 5.25, $P = .02$) (Fig. 4B). Attending surgeons ($M = 52.43$ seconds, SD = 17.84) performed the procedure in less time than surgical residents ($M = 78.50$ seconds, SD = 17.84, $t(9) = 2.26$, $P < .05$) and medical students ($M = 85.52$ seconds, SD = 17.83, $t(10) = 3.06$, $P = .01$). Cohen's effect size value (attending vs resident $d = 1.5$; attending vs medical student $d = 1.9$) suggests a high practical significance.

Path length

The repeated measures ANOVA of path length also revealed a main effect of tissue type ($F(2, 24) = 21.84$, $P < .001$) (Fig. 5A). Participant's hands moved farther when suturing tissue paper ($M = 6.64$ m, SD = 2.01) compared with balloon ($M = 4.81$ m, SD = 1.22, $t(14) = 5.77$, $P < .001$) or foam ($M = 4.93$ m, SD = 1.02, $t(14) = 4.64$, $P = .001$). Cohen's effect size value (tissue vs balloon $d = 1.1$; tissue vs foam $d = 1.09$) suggests a high practical significance.

There was a significant main effect of experience level on path length (F(2,12) = 5.38, $P = .02$) (Fig. 5B). Attending surgeons ($M = 4.03$ m, SD = 1.29) had shorter path lengths than surgical residents ($M = 5.85$ m, SD = 1.29, $t(9) = 2.17$, $P < .05$) and medical students ($M = 6.50$ m, SD = 1.29, $t(10) = 3.15$, $P = .008$). Cohen's effect size value (attending vs resident $d = 1.4$; attending vs medical student $d = 1.9$) suggests a high practical significance.

Correlation of motion tracking measures

Idle time was significantly correlated with total operative time for the tissue paper task ($r = .56$, $P = .03$). Idle time was not significantly correlated with total operative time for the foam ($r = .44$, $P = .097$) and balloon ($r = .48$, $P = .07$) tasks. In addition, idle time did not correlate with path length ($P > .05$). Finally, there was a significant correlation between total operative time and path length for all 3 tasks (foam $r = .84$, balloon $r = .80$, tissue $r = .760$; $P < .001$ for all correlations).

Comments

This pilot study sought validity evidence for the use of idle time as a performance metric in an open surgical skills task. Validity evidence was evaluated using known groups and response process with video analysis.[23,24] Additionally, we sought validity evidence for the psychomotor metrics generated when using motion tracking technology with the newly developed variable tissue simulator. Fifteen

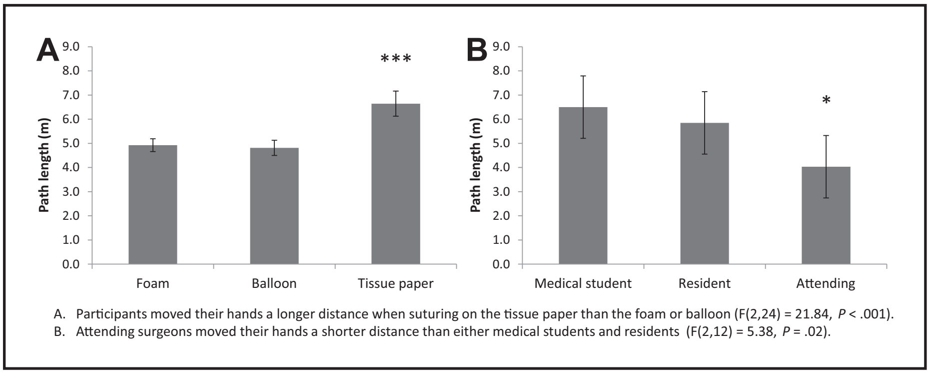

Figure 5 Path length per stitch by tissue type and experience level during the suturing task.

participants with different levels of experience (medical students, surgical residents, and attending surgeons) completed 3 suturing tasks on different materials (foam, balloon, and tissue paper) that were purposefully selected to provide varying degrees of complexity.

This study is the first time that idle time was experimentally investigated as a performance metric in open surgical skills assessment. All participants had greater amounts of idle time while performing the more difficult suturing task. This is consistent with our hypothesis that idle times will be shorter for easier tasks. These findings provide validity evidence for the use of idle time as a psychomotor performance metric. Idle time may represent periods of motor planning or decision making that can be used to differentiate performance.[16] This suggests that evaluating what participants are doing or thinking while not moving their hands may be just as important as while they are moving their hands.

Prior studies assessing laparoscopic skills noted that novices spend a greater amount of time in the idle state compared with experts.[16,25,26] Our study did not find a significant relationship between overall idle time and experience level. However, differences in idle time were noted when evaluating specific procedure steps by experience. Entering the tissue with the needle, driving the needle through the tissue, and tightening a knot had the greatest number of idle periods. Additionally, the idle periods during these steps were not evenly distributed among participants by experience level. Medical students and residents had a greater number of idle periods during entering the tissue with the needle and driving the needle through the tissue, while attending surgeons had a greater number of idle periods during tightening a knot. Medical students and residents may exhibit more idle periods when performing motor planning for contacting the needle with the tissue and moving it through the tissue. In contrast, attendings exhibit idle periods while placing and evaluating tension on the knot. Our findings warrant additional research evaluating the relationship among idle time, experience level, and procedure steps.

Study results also demonstrate that more experienced practitioners took less time and moved their hands a shorter distance to complete the suturing tasks compared with less experienced trainees. This finding is consistent with motor behavior research demonstrating improvements in movement components of motor skills with increased practice.[27] Additionally, participants took more time and moved their hands a longer distance to complete the suturing task on the material (tissue paper) that was designed to be the most friable and complex. Both these findings are consistent with prior work evaluating surgical performance with motion tracking technology,[11–15] thus providing construct validity evidence for the use of psychomotor metrics to differentiate performance on the newly developed variable tissue simulator.

There were significant positive correlations between idle time and total operative time on the more difficult suturing task (tissue paper). Participants who take longer to complete the task have greater idle periods. This may result from those participants requiring larger amounts of psychomotor and procedural planning. The lack of correlation between idle time and path length may result from other factors confounding the relationship or our small sample size and limited power.

Limitations of this study include the small sample size, which decreases statistical power and generalizability of the results. This pilot study was undertaken to evaluate validity evidence of this newly developed performance metric. The preliminary results show great promise with large effect sizes despite low power. Thus we plan to collect data from a larger sample size to address the limitations of statistical power and allow for broader inferences regarding our results. Generalizability is also limited by the sample size and lack of randomization. Participants were included in the study based on availability and willingness to participate, and therefore may be inherently different than those who did not choose to participate. Finally, another limitation may be the relative simplicity of using a suturing task rather than a full procedure. The study was specifically designed with a simple task and complexity was added through the use of different materials. This allowed for more detailed understanding of the motor behavior components that may have not been possible with a longer task involving multiple motor components. We are interested in quantifying periods of motor planning or decision making and our future work is aimed at evaluating these same psychomotor metrics of idle time, total operating time, and path length in more complex tasks and full procedures.

The implications of this work relate to ongoing development and evaluation of performance metrics in the research setting, as well as the need for a paradigm shift in how motion tracking technology is used for surgical skills assessment. The surgical profession is still in its infancy regarding best practice for technology-based performance assessments. Several research groups are using commercial technologies and discovering new metrics. As additional metrics continue to be developed, future directions may call for a meta-analysis to better understand the gaps, strengths, and weaknesses of these new approaches. Idle time represents an evolution in traditional motion analysis research. Our findings imply that this performance metric is meaningful and may add to our current understanding of technical skills research.

References

1. Schmitz CC, Darosa D, Sullivan ME, et al. Development and verification of a taxonomy of assessment metrics for surgical technical skills. Acad Med 2014;89:153–61.
2. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997;84:273–8.

3. Fried GM, Feldman LS. Objective assessment of technical performance. World J Surg 2008;32:156–60.
4. Regehr G, MacRae H, Reznick RK, et al. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998;73:993–7.
5. Yule S, Paterson-Brown S. Surgeons' non-technical skills. Surg Clin North Am 2012;92:37–50.
6. Cano A, Gayá F, Lamata P, et al. Laparoscopic tool tracking method for augmented reality surgical applications. In: Bello F, Edwards PJ, editors. Biomedical Simulation. Berlin: Springer; 2008. p. 191.
7. Dosis A, Aggarwal R, Bello F, et al. Synchronized video and motion analysis for the assessment of procedures in the operating theater. Arch Surg 2005;140:293–9.
8. Chmarra MK, Grimbergen CA, Dankelman J. Systems for tracking minimally invasive surgical instruments. Minim Invasive Ther Allied Technol 2007;16:328–40.
9. Datta V, Mackay S, Mandalia M, et al. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. J Am Coll Surg 2001;193:479–85.
10. Angrosino MV. Observer bias. In: Lewis-Beck MS, Bryman A, Liao TF, editors. Encyclopedia of Social Science Research Methods. Thousand Oaks, CA: SAGE Publications; 2004. p. 758–9.
11. Bann S, Davis IM, Moorthy K, et al. The reliability of multiple objective measures of surgery and the role of human performance. Am J Surg 2005;189:747–52.
12. Datta V, Bann S, Mandalia M, et al. The surgical efficiency score: a feasible, reliable, and valid method of skills assessment. Am J Surg 2006;192:372–8.
13. Datta V, Chang A, Mackay S, et al. The relationship between motion analysis and surgical technical assessments. Am J Surg 2002;184:70–3.
14. Datta V, Mandalia M, Mackay S, et al. Relationship between skill and outcome in the laboratory-based model. Surgery 2002;131:318–23.
15. Mackay S, Datta V, Mandalia M, et al. Electromagnetic motion analysis in the assessment of surgical skill: relationship between time and movement. ANZ J Surg 2002;72:632–4.
16. Oropesa I, Sanchez-Gonzalez P, Lamata P, et al. Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery. J Surg Res 2011;171:e81–95.
17. Fitts PM, Posner MI. Human Performance. Belmont, CA: Brooks/Cole; 1967.
18. Balkissoon R, Blossfield K, Salud L, et al. Lost in translation: unfolding medical students' misconceptions of how to perform a clinical digital rectal examination. Am J Surg 2009;197:525–32.
19. Pugh CM, DaRosa DA, Santacaterina S, et al. Faculty evaluation of simulation-based modules for assessment of intraoperative decision making. Surgery 2011;149:534–42.
20. Pugh CM, Domont ZB, Salud LH. A simulation-based assessment of clinical breast examination technique: do patient and clinician factors affect clinician approach? Am J Surg 2008;195:874–80.
21. Pugh CM, Rosen J. Qualitative and quantitative analysis of pressure sensor data acquired by the E-Pelvis simulator during simulated pelvic examinations. Stud Health Technol Inform 2002;85:376–9.
22. Pugh CM, Santacaterina S, DaRosa DA, et al. Intra-operative decision making: more than meets the eye. J Biomed Inform 2011;44:486–96.
23. Cook DA, Brydges R, Zendejas B, et al. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. Acad Med 2013;88:872–83.
24. Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ 2003;37:830–7.
25. Rosen J, Brown JD, Chang L, et al. Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete Markov model. IEEE Trans Biomed Eng 2006;53:399–413.
26. Rosen J, Solazzo M, Hannaford B, et al. Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. Comput Aided Surg 2002;7:49–61.
27. Magill RA. Motor Learning and Control: Concepts and Applications. 9th ed. New York: McGraw-Hill; 2011.