Objective and Automated Quantification of Instrument Handling for Open Surgical Suturing Skill Assessment: A Simulation-Based Study

Simar P. Singh, Amir Mehdi Shayan, Jianxin Gao, Joseph Bible, Richard E. Groff, *Senior Member, IEEE*, and Ravikiran Singapogu, *Senior Member, IEEE*

Abstract*—*Goal: **Vascular surgical procedures are challenging and require proficient suturing skills. To develop these skills, medical training simulators with objective feedback for formative assessment are gaining popularity. As hardware advancements offer more complex, unique sensors, determining effective task performance measures becomes imperative for efficient suturing training. *Methods:* 97 subjects of varying clinical expertise completed four trials on a suturing skills measurement and feedback platform (SutureCoach). Instrument handling metrics were calculated from electromagnetic motion trackers affixed to the needle driver. *Results:* The results of the study showed that all metrics significantly differentiated between novices (no medical experience) from both experts (attending surgeons/fellows) and intermediates (residents). Rotational motion metrics were more consistent in differentiating experts and intermediates over traditionally used tooltip motion metrics. *Conclusions:* Our work emphasizes the importance of tool motion metrics for open suturing skills assessment and establishes groundwork to explore rotational motion for quantifying a critical facet of surgical performance.**

***Index Terms*—Motion smoothness, skill training, surgical simulation, surgical skill assessment.**

***Impact Statement*— This study aims to determine the effectiveness of metrics derived from needle driver rotational and tooltip motion tracking to determine differences in clinical expertise in open needle driving.**

Manuscript received 12 October 2023; revised 16 March 2024 and 6 May 2024; accepted 6 May 2024. Date of publication 17 May 2024; date of current version 20 June 2024. This work was supported by the National Institutes of Health under R01 Grant 5R01HL146843-03. The review of this article was arranged by Editor Jean-Francois Daneault. *(Corresponding author: Ravikiran Singapogu.)*

Simar P. Singh, Amir Mehdi Shayan, and Ravikiran Singapogu are with the Department of Bioengineering, Clemson University, Clemson, SC 29634 USA.

Jianxin Gao and Richard E. Groff are with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634 USA.

Joseph Bible is with the Department of Mathematical and Statistical Sciences, Clemson University, Clemson, SC 29634 USA (e-mail: joseph@clemson.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/OJEMB.2024.3402393, provided by the authors.

Digital Object Identifier 10.1109/OJEMB.2024.3402393

I. Introduction

Vascular surgery encompasses a wide range of intricate procedures, including surgeries performed both traditionally (“open” surgery) as well as endoscopically (endovascular surgery). Due to the benefits of endovascular procedures for increasing patient comfort and reducing hospital stay, there is a demand for today’s vascular surgery trainees to learn endovascular techniques. Consequently, surgeon educators recognize the need for trainee development in open surgical techniques, as procedures unable to be done endovascularly are relatively more challenging [1]. To determine educational priorities for technical skill learning, a study exploring necessary procedures to include in a vascular surgery curriculum deemed open surgical techniques to comprise two-thirds of the required procedures for a proper curriculum [2]. Among these, anastomotic technique ranked the highest priority. In line with this, vascular surgeon educators have stressed the importance of learning fundamental vascular skills, such as suturing, as foundational for learning advanced surgical techniques [1]. Unskilled suturing can lead to bleeding and tearing, potentially leading to adverse patient morbidity and mortality [3]. Since vascular procedures can be high-risk, proficient suturing is crucial for well-prepared surgeons.

Given the critical role of suturing in open vascular surgery procedures, there is a demand for practical and widespread objective training methods for effective and efficient skills training. Task trainers provide a relatively affordable and reusable training method as a viable alternative to cadaver training, despite diminished anatomical realism [4], [5]. Such trainers excel in facilitating measurable performance in a focused, simulated anatomical environment that typically allows for sensor metrics to provide objective assessment and targeted feedback on specific skills. The appeal of these trainers is evident in their increasing adoption by surgery boards for performance training and assessment tailored to surgical specialties [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. These trainers utilize game-like training modules to assess performance through expert ratings, time-based metrics, and error-based metrics. As such, task trainers can be valuable pedagogical tools, particularly in instilling fundamental surgical skills.

For task trainers to excel as learning tools, they must incorporate effective metrics that score a trainee on the various characteristics of a surgical procedure. Needle driver motion is frequently evaluated by surgical experts in standardized rating sheets, as a surgeon's ability to manipulate instruments efficiently denotes surgical skill [1], [11]. Thus, this study's primary focus is quantifying the needle driver's distinct motion characteristics during suturing on a simulator. While conventionally used motion metrics, such as path length (*PL*), average velocity, or the number of peaks in the velocity profile (*Pks*), offer a foundational approach to motion analysis, these measures are limited in their formulation and can yield varied results in distinguishing clinical expertise [5], [6], [7], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]. Accordingly, the surgical education community requires more sophisticated metrics to provide trainees with robust skill measurements and meaningful feedback.

Precise instrument handling with minimum hesitation is hypothesized to be associated with smooth motion of the tool, and capturing this behavior holds the potential for effective surgical skills training. While initially used for tracking stroke recovery [26], [27], motion smoothness is increasingly used as a robust tool for measuring surgical proficiency [7], [9], [10], [11], [12], [14], [16], [17], [18], [19], [20], [22], [23]. Among the various motion smoothness metrics defined in previous studies, log dimensionless jerk (*LDLJ*) and spectral arc length (*SPARC*) are considered state-of-the-art in measuring smoothness of motion. To our knowledge, there is no application of motion smoothness for open suturing skills assessment.

Surgical skills assessment on open surgery is limited due to the high demand for minimally invasive surgical skills training [12], [17], [24]. This claim is substantiated by a systematic review by Mitchell et al. [28] that found twenty-nine studies on open vascular skills assessment, reporting eight studies on dexterity analysis of hand motions. Although the studies reported positive results in surgical skills assessment and correlation with expert ratings, there is an additional need to quantitatively assess instrument handling motion as these motions are directly related to suturing quality and provide complexity and depth unique from hand tracking. To our knowledge, only a few studies have done so, likely due to the difficulty of instrumentation without interfering with the subject's needle driver maneuverability [24]. Suturing skills assessment for open surgery requires further research, and evaluating tool motions can provide valuable quantification of some of the various characteristics of skilled suturing.

Studies that use tool motion analysis for surgical skills assessment generally evaluate tooltip motion, but the needle driver's rotational motion is integral to open suturing. Recognizing this, Sharon et al. [24] propose analyzing rotational motion for efficient suturing skill quantification and introduce a novel metric, the orientation rate of change (*RoC*). The study demonstrated potential in their measure by distinguishing expert and novice performance, but the researchers note that their small sample size may affect the generalizability of their results. Similarly, a previous study on the SutureCoach found metrics applied to rotational hand motions were better suited to differentiate clinical expertise than metrics applied to transitional hand motions [29]. We can expect that applying complex metrics to rotational motion may better assess instrument handling motion quality pertinent to skilled suturing.

To compare the importance of rotational vs. translational needle driver motions in open suturing, our study applies equivalent metrics to both domains of motion. Additionally, studies on surgical skills assessment have generally succeeded in differentiating between experts (clinicians, surgeon educators) and novices (medical students, subjects with no experience). However, surgical skill assessment has encountered difficulties in determining differences between experts and intermediates (residents), often attributing this to a small sample size [6], [7], [8], [9], [10], [13], [14], [15]. To mitigate this, our dataset consists of 97 subjects with a vast range of experience, ranging from students with no experience to expert vascular attendings with several decades of experience. In this study, we aim to answer the following questions:

1) Can metrics that quantify tool motion indicate skilled instrument handling?
2) Are metrics that quantify tool rotational motion better suited to assess suturing skill over metrics that quantify tool translational motion?

II. Materials and Methods

A. Simulator Design and Development

The SutureCoach is a custom-built simulator that renders radial suturing employed in vascular surgery and measures suturing skill comprehensively using multi-modal sensors [21]. The current design features a hollow cylinder with a simulated membrane material attached to the surface. Twelve suture locations are marked in a radial suturing pattern modeled from the Fundamentals of Vascular Surgery simulator [1]. The SutureCoach simulates suturing at surface and depth conditions simulated through raised barriers. The depth condition represents vascular suturing in an anatomical cavity. Subjects were instructed to complete four trials on the SutureCoach: one at surface, one at depth, and two more trials of the same sequence. Fig. 1 demonstrates both simulator conditions. The SutureCoach platform interfaces various sensors for a multi-modal, comprehensive assessment of suturing skill. An internal camera (Intel RealSense D435) automatically performs subcutaneous, vision-based needle tracking, recorded at 60 fps, to synchronize sensor data for suture-specific analysis. A force/torque sensor measures membrane forces during suturing, and two inertial measurement units (IMUs) are placed on a subject's hand and wrist. A previous study on our simulator analyzed hand motions obtained from the IMU on the same dataset presented in this study [29]. Specifically, this study examines data obtained from two electromagnetic position and orientation sensors (Ascension trakSTAR Model 180, Northern Digital Inc.), recording x, y, and z Cartesian sensor-frame coordinates and azimuth, elevation, roll, and quaternion orientation at a rate of 100 Hz, are attached to both handles of a needle driver (Mayo-Hegar, 8"). Tooltip location is calculated through rotation calibration.

The electromagnetic sensor attachment to the needle driver was designed in a non-intrusive manner for the surgeon to feel

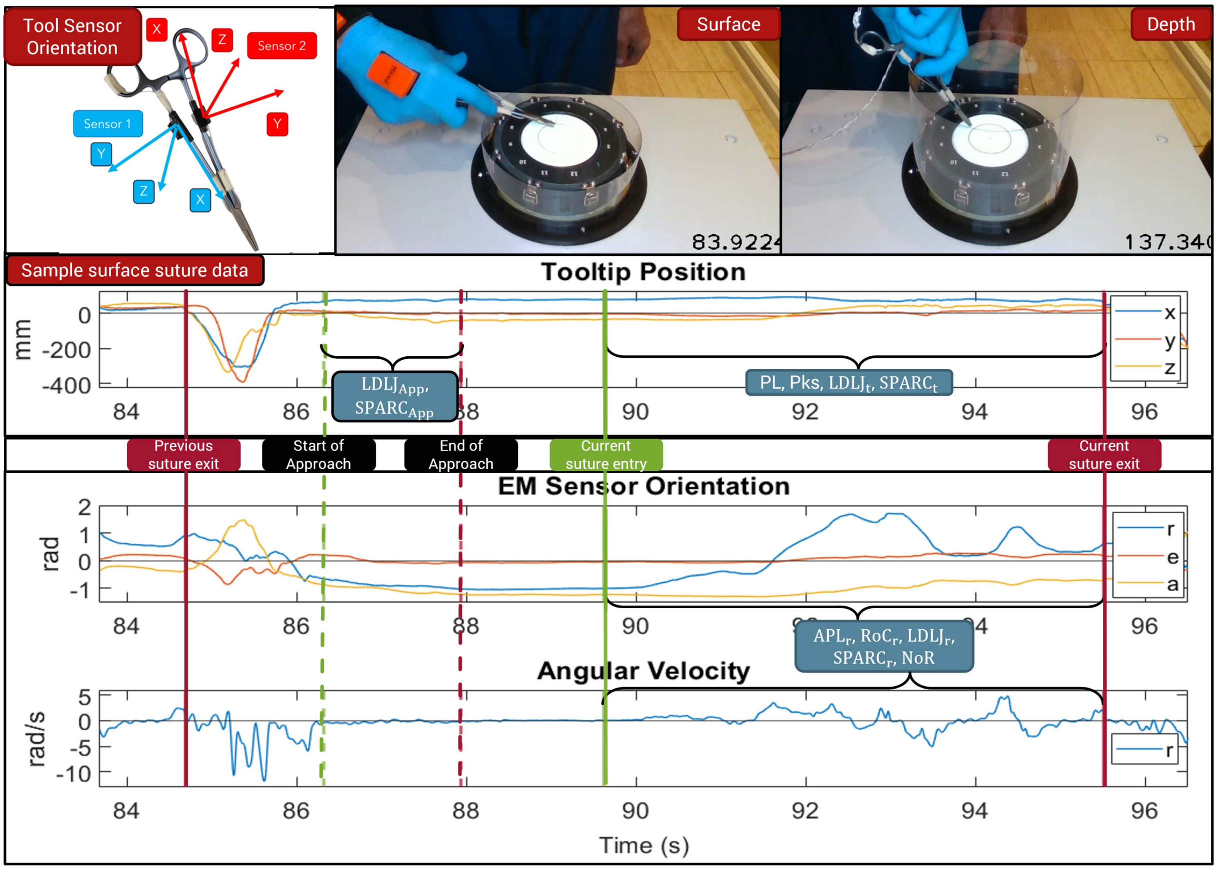

Fig. 1. Figure depicting the titanium needle driver with the electromagnetic motion sensor attachment and orientations on the top left. The top right figures demonstrate an expert suturing at the surface and depth conditions, demonstrating the markedly more constrained motions at depth. The bottom figure depicts sample tooltip position, orientation, and angular velocity profiles across a suture alongside sections of data that metrics are calculated on.

comfortable using instrument handling techniques identical to those in the operating room. The sensor is secured through a molded 3D-printed casing and lid, which is then affixed to each handle of the needle driver. One sensor is flipped so the cable can be wrapped around the needle driver and braided with the other sensor to mitigate interference during the suturing procedure. A representation of the orientation of both sensors is seen in the top left image of Fig. 1. The sensors are rotated 180 degrees in the x and y axis post-processing to align both sensor's movement profiles.

B. Data Processing

Sensor data for metrics that do not require derivatives were processed with a 20 Hz, 2nd order low-pass Butterworth filter applied to the profile. *LDLJ* heavily relies on accurate derivative estimations, as our previous study on cannulation needle motions found that noise increases exponentially per derivative calculation [25]. We found a window length of 25 best suited rotation calibrated tip motion data, as calibrated tooltip values are noisier than raw data. Thus, we used a Savitzky-Golay filter of order three and a window span of 25 for translational tooltip motion derivatives. For rotational motion derivatives, we compared the low-pass Butterworth filtered data to several Savitzky-Golay parameters and found a window span of 13 best matched the filtered data. For further validation, we compared this parameter with x-angular velocity obtained from an IMU placed in parallel with the EM sensor.

C. Metrics

We aimed to establish metrics to pinpoint characteristics of skilled instrument handling needle driver motion. The following section will present the metrics and corresponding formulations used in this paper. Performance is evaluated by suture, and metrics are calculated from the start of needle contact to the surface of the membrane until the tracked needle swage exits the membrane. A flowchart depicting our data processing methods and metric calculations is seen in Fig. 2. Metrics with physical properties deemed applicable to provide feedback were chosen. General formulations will be presented since metrics will be applied to both rotational and translational motion.

To evaluate the needle driver's rotational motion, we applied motion metrics to the x-axis/roll (denoted with $_r$), as the primary angular motions in needle driving encapsulate rotations about this axis. To evaluate the needle driver's translational motion,

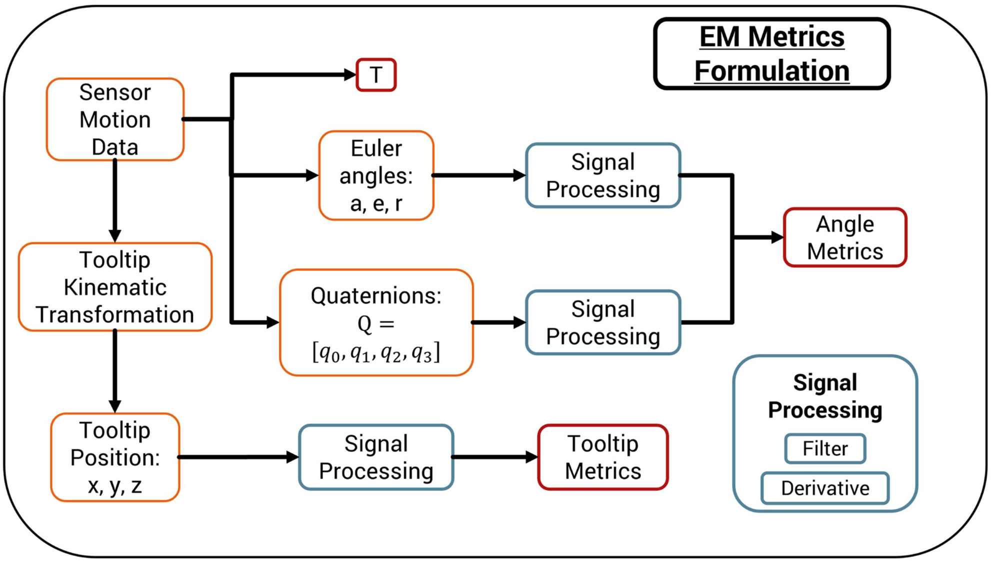

Fig. 2. Flowchart describing sensor data processing and metrics calculations. Derivatives are approximated with a Savitzky-Golay filter. Otherwise, data is filtered with a Butterworth filter, as detailed in the Data Processing section. The formula function lists sample function calculations for the more advanced, repeated metrics for different calculations.

TABLE I
TABLE LISTING MOTION METRICS AND THEIR APPLICATIONS TO TRANSLATIONAL TOOLTIP AND ROLL MOTIONS.

Metric	Tooltip	Roll	Suture Approach Confidence
PL/APL	✓	✓	
Pks	✓		
LDLJ	✓	✓	✓
SPARC	✓	✓	✓
RoC		✓	
NoR		✓	

motion metrics are applied to the calibrated tooltip location (denoted with$_t$). Additionally, our surgeon collaborators frequently stated that a trainee's confidence in their motions when approaching a suture could indicate their suturing skill, leading to the formulation of Suture Approach Confidence metrics. This measure uses a third of the time from the last suture end time to the current suture start time and computes *LDLJ* and *SPARC* (denoted with $_{App}$) on that time window. A list of the metrics used in this study and their application is seen in Table I.

1) *Time (T):* The time from needle contact to needle exit. *T* is the most common measure of surgical skill.

$$T = t_{exit} - t_{entry} \tag{1}$$

2) *Path Length (PL):* Total distance traversed by the tooltip. *PL* is a common measure to assess the economy of motion. Theoretically, it can be surmised that a skilled clinician follows minimal displacement to accomplish their surgical procedure. Thus, total tooltip distance and degree of rotation correlate with clinical expertise.

$$\int_{t_{entry}}^{t_{end}} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt \tag{2}$$

3) *Number of Peaks in the Velocity Profile (Pks):* A precursory motion metric that measures the number of peaks present in the velocity profile with a minimum prominence of 0.05 mm/s with the *findpeaks* MATLAB function. The more unsmooth the velocity profile, the greater amount of peaks.

4) *Log Dimensionless Jerk (LDLJ):* The natural log of jerk integrated and squared, where *T* and *PL* are defined above.

$$LDLJ = \ln \left| \frac{T^5}{PL^2} \int_{t_{entry}}^{t_{exit}} \left(\frac{d^3x}{dt^3}\right)^2 + \left(\frac{d^3y}{dt^3}\right)^2 + \left(\frac{d^3z}{dt^3}\right)^2 dt \right| \tag{3}$$

5) *Spectral Arc Length (*SPARC*):* The arc length of the Fourier transform of the velocity profile.

$$SPARC = \int_0^{\omega_c} \left[\left(\frac{1}{\omega_c}\right)^2 + \left(\frac{d\hat{V}(\omega)}{d\omega}\right)\right]^{\frac{1}{2}} d\omega; \tag{4}$$

$$\hat{V}(\omega) = \frac{V(\omega)}{V(0)}, V(\omega) = \mathscr{F}\{v(t)\}$$

6) *Angular Path Length (APL):* Total angular distance observed by the sensors.

$$APL = 2\sum_{i=1}^{N-1} \left(Q(i+1) \cdot Q(i)^{-1}\right) \tag{5}$$

7) *Rate of Angular Change (RoC):* The rate of change in rotation as defined in [24].

$$RoC = \frac{2}{N-1} \sum_{i=1}^{N-1} \frac{Q(i+1) \cdot Q(i)^{-1}}{t(i+1) - t(i)} \tag{6}$$

TABLE II
A TABLE DESCRIBING THE DISTRIBUTION OF SUBJECTS AND THEIR HANDEDNESS PER LEVEL OF CLINICAL EXPERTISE ALONGSIDE CORRESPONDING DEPTH AND SURFACE TRIALS AND SUTURES

		Surface		Depth		Total
Subject Handedness	**Subjects**	**Trials**	**Sutures**	**Trials**	**Sutures**	**Total Sutures**
Right	89	182	2177	180	2152	4329
Left	8	16	192	16	190	382
Experts	**35**	**71**	**852**	**71**	**847**	**1699**
Fellow	10	19	228	19	226	454
Attending (<= 10 years)	13	26	312	26	310	622
Attending (>10 years)	12	26	312	26	311	623
Intermediates	**32**	**63**	**756**	**63**	**755**	**1511**
PGY1	3	6	72	6	71	143
PGY2	5	10	120	10	120	240
PGY3	5	10	120	10	120	240
PGY4	8	15	180	15	180	360
PGY5	11	22	264	22	264	528
Novices	**30**	**64**	**761**	**62**	**740**	**1501**

As noted in the table, some subjects were unable to complete all four trials or all twelve sutures.

8) *Number of Rotations (NoR):* Building upon the number of rolls metric previously applied to measure hand movements on the SutureCoach [29], *NoR* sums the number of changes in the direction of rotation through roll-axis angular velocity to measure deliberate rotations during suturing performance. A deliberate change in rotation requires an *APL* greater than 3.6 degrees — 1% of the full 360-degree rotation — until the next rotation occurs.

D. Subject Demographics and Statistical Methods

This study evaluated suturing performance of 97 subjects with varying levels of expertise. Ethics approval for this study was provided by Clemson University (IRB number: IRB2020-0387; Date of Approval: May 4th, 2021). Participants included attendings with varying degrees of clinical experience, fellows, residents (Post Graduate Year (PGY) 1-5), and novices (medical students and others with no medical experience) on the SutureCoach suturing simulator. A distribution of subject demographics is seen in Table II.

Statistical analyses were computed using R (version 4.2.2). Tukey HSD tests for multiple pair-wise comparisons between the three levels of clinical standing are computed for each metric. The model uses Tukey's multiplicity adjustment to account for the family-wise error rate based on the number of comparisons made, similar to ANOVA. The linear model fits each metric with the level of clinical standing while controlling for suture location, subject handedness, and the interaction between the two variables, as suturing technique and the needle driver's position and orientation depend on such factors. To analyze the relationships between the metrics, we generated correlation matrices for surface and depth trials.

III. RESULTS

The pair-wise Tukey HSD comparison results calculated from sensor 1 are seen in Fig. 3. All metrics observe similar results between sensor 1 and sensor 2. As such, results from sensor 2 are only presented in the supplementary materials. Additionally, pair-wise Tukey HSD p-value results are recorded in Table III in the supplementary materials.

Among the tooltip metrics, $LDLJ_{App}$ emerged as the sole metric capable of distinguishing between expert and intermediate performance in both superficial and depth conditions. In contrast, angular-based metrics showed improved performance compared to the tooltip metrics, with 4/5 metrics successfully distinguishing between experts and intermediates in the surface condition and 3/5 in the depth condition. $SPARC_r$ failed to differentiate between expert and intermediate performance in both conditions, and NoR could not do so in the depth condition.

Correlation matrices from surface and depth trial metrics revealed high correlations across translational and rotational metrics, particularly with *T* (Fig. 5). Among the translational metrics, Pks and $LDLJ_t$ demonstrated the highest correlations with T (surface: Pks $= LDLJ_t = 0.9$, depth: Pks $= 0.94$ and $LDLJ_t = 0.87$). Likewise, for rotational metrics, $LDLJ_r$ and NoR observed the highest correlation with T (surface: $LDLJ_r$ $= 0.8$ and NoR $= 0.79$, depth: $LDLJ_r = 0.8$ and NoR $= 0.89$).

IV. DISCUSSION

The importance of skilled instrument handling in suturing signifies the potential for tool motion metrics for suturing skills assessment. In support of this claim, all metrics analyzed in this study effectively differentiated novice from intermediate or expert suturing performance. We begin by focusing on the results of the traditional translational tooltip metrics and *T*. *T*, *PL*, and *Pks* were unable to differentiate between expert and intermediate skill levels. Studies incorporating these metrics have found similar difficulties in separating expert clinician and intermediate resident groups, noting their limitation for more nuanced levels of skill assessment [7], [14], [15], [19]. Thus, determining the fine-grain differences between the expert and intermediate skill levels will allow for better assessment of skilled instrument handling characteristics for targeted skills analysis and improved feedback.

More advanced motion smoothness metrics have been found to be effective in surgical skills analysis, although only a few studies have incorporated them thus far [5], [9], [10], [25].

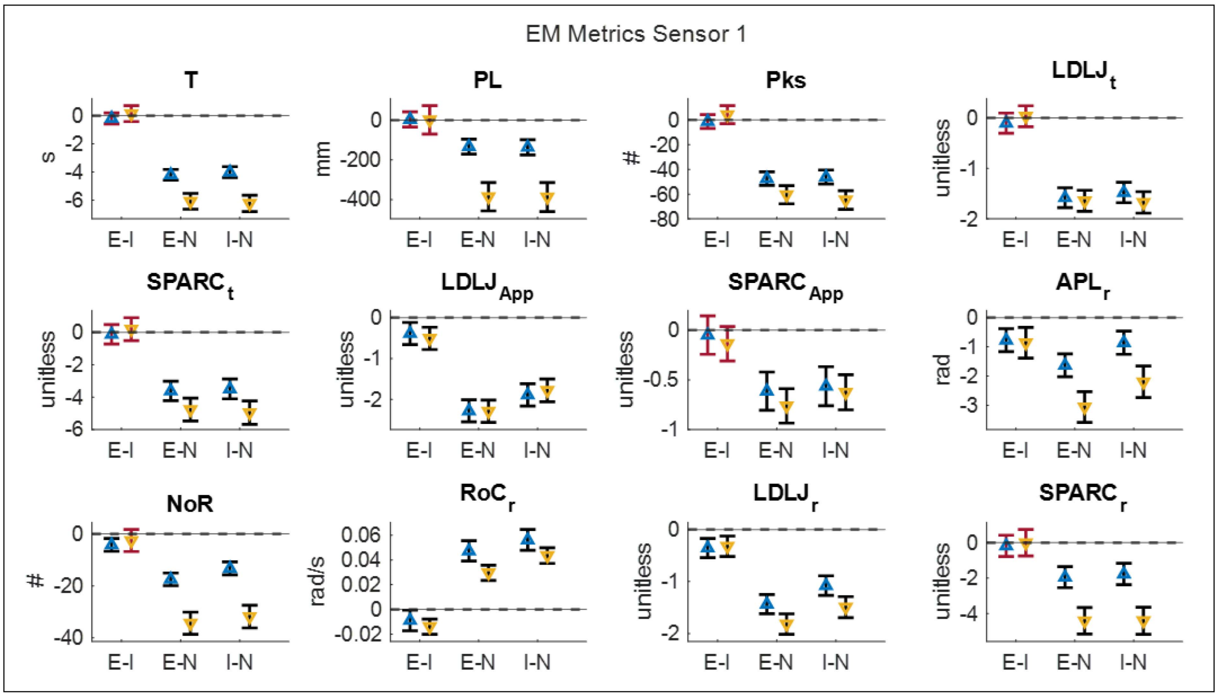

Fig. 3. Confidence intervals of Tukey HSD pair-wise comparisons for EM metrics calculated from sensor 1. Surface and depth intervals are plotted side-by-side and are denoted with a blue △ and a yellow ▽, respectively. E, I, and N are shorthand for expert, intermediate, and novice, with corresponding comparisons marked with a -. If the comparison between groups is not statistically significant, the confidence interval contains zero and is colored red, whereas significantly different comparisons are colored black. To reduce the number of figures in the results, pair-wise comparisons of time to complete a suture are also included.

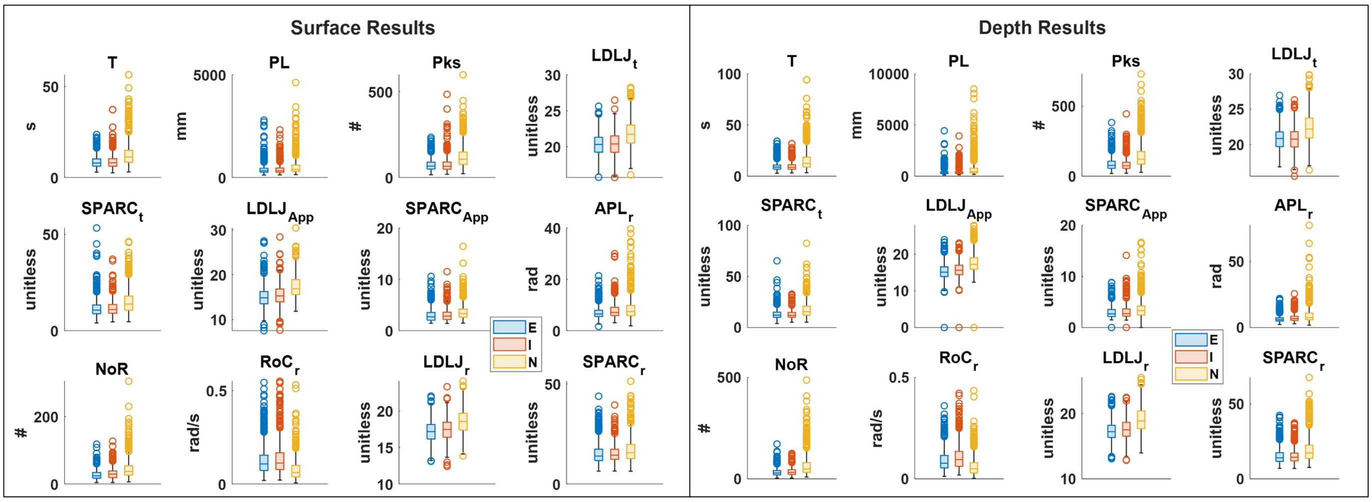

Fig. 4. Score distributions of surface and depth metrics that observed significant differences between experts and intermediates from sensor 1.

Results from Fig. 3 demonstrate that the sole tooltip measure to differentiate between experts and intermediates was found from one of the new applications of smoothness defined in this study, $LDLJ_{App}$, whereas $SPARC_{App}$, $SPARC_t$ and $LDLJ_t$ could not. Despite *SPARC* being developed to address *LDLJ*'s limitations of sensitivity to noise [27], this study evidenced that *SPARC* and *LDLJ* have varying results in terms of superiority. For example, $LDLJ_{App}$ was designed to preemptively assess a surgeon's confidence before suturing, quantifying movements that are highly related to the original purpose of *LDLJ* to assess stroke rehabilitation in hand motion tasks [26], [30]. In summary, while tooltip metrics effectively distinguish novice from intermediate and expert skill levels, they appear to have limited value in differentiating experts and intermediates.

We hypothesized that the most potent suite of metrics for measuring adept suturing skills from instrument handling could be derived from its rotational time series data. Recent seminal research by Sharon et al. [24] has offered insights into the

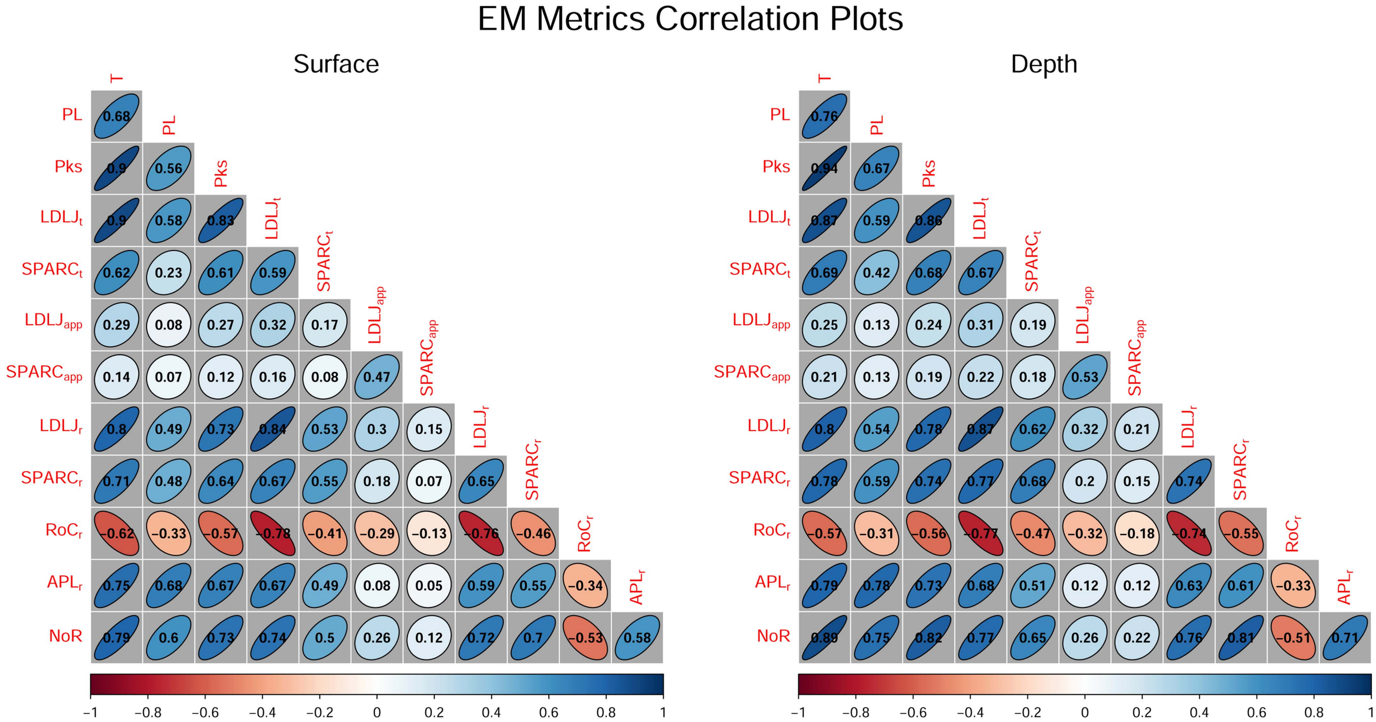

Fig. 5. These correlation matrices visualize the relationships between all metrics defined in the study in the surface condition (left) and the depth condition (right). Metrics are ordered from *T*, translational metrics (*Pks* to $SPARC_{App}$) and rotational metrics ($LDLJ_r$ to *NoR*).

potential of rotational motion metrics. In their study, they found that, while angular displacement (*APL*) could not differentiate expert surgeon performance from non-medical graduate students in the open needle driving trial, it could do so in the teleoperated condition. Conversely, we found that APL_r was among the most effective metrics, observing significant differences between all groups for both conditions. It is important to note that Sharon et al. highlight limitations in their work deriving from a small sample size and the lack of subjects with more varied levels of clinical expertise, likely leading to the differences in observed results. We can surmise that the more experienced the user, the fewer needle-driving rotations are necessary to complete the suture. This conjecture is further validated through one of the new metrics introduced in this study, *NoR*, which was able to differentiate between all three groups in the superficial condition. The measure captures the amount of discrete, intentionally made rotations about the needle driver's roll axis. Our initial assessment of rotational-based position metrics demonstrates substantial improvements over their translational counterparts, further evidenced by a previous study on the SutureCoach platform analyzing IMU hand motions on the same dataset used in this study [29].

Towards the more computationally advanced rotational metrics, RoC_r differentiated between expert and intermediate scores in both conditions. However, unlike the other metrics in this study, the intermediate group observed increased *RoC* scores over experts and novices in Fig. 4 instead of the gradual trend towards lower values from novices to experts. Sharon et al. [24] did not have an intermediate resident population in their study, thus not experiencing these results. Experts are likely effectively utilizing the speed-accuracy trade-off, driving their needle at a slower, constant rotational speed to minimize the errors present during suturing. Overall, metric shows potential for suturing skills assessment by allowing objective feedback on adjustments in rotational speed for experts and intermediates.

The final set of metrics, rotational motion smoothness metrics, has demonstrated remarkable efficacy in distinguishing between the various levels of clinical experience. The $LDLJ_r$ metric successfully distinguished between expert and intermediate groups in both conditions, whereas its *SPARC* counterpart did not. This is in line with two studies that applied *LDLJ* and *SPARC* for rotational motion IMU motion [29], [31]. These results suggest that the differences in metrics' formulation sensitize them to unique task-specific features of motion, as evidenced by our results. Nonetheless, $LDLJ_r$ continues to provide robust results for surgical skills assessment, observing much lower variability than the APL_r metric as seen in Fig. 4. The use of smoothness measures as a robust measure of surgical skill continues to show promising results with budding potential for further exploration in various applications.

High correlations were observed between many of the evaluated metrics. In particular, metrics were highly correlated with *T*, including Pks, $LDLJ_t$, $LDLJ_r$, and NoR. These findings align with the metrics' quantification of smoothness: a subject with a longer suture completion time (*T*) would likely exhibit shakier movements reflected by their velocity (Pks), further amplified in jerk (LDLJ). LDLJ's independence of time (due to its dimensionless nature) strengthens this idea. Studies by Hogan and Sternad [26] and Balasubramanian et al. [32] demonstrated that the same movement characteristics result in the same LDLJ or SPARC value regardless of movement duration.

We would like to highlight a limitation of our study: we do not have a measure of task completion or expert ratings. Ultimately, we aim to develop the SutureCoach into an assessment tool that follows Messick's framework of validity: content validity, response process, internal structure, relationship to other variables, and consequences [33]. The current work, however, provides evidence for content validity by modeling a procedure deemed relevant to vascular surgery skills training [1]. In addition, we further demonstrated both content validity and relationship to other variables by establishing mean score differences between population groups of known levels of clinical expertise. Previous studies have noted the difficulties of defining an expert and argued that expertise is not solely defined by the amount of experience [19]. However, the large number of participants analyzed in this dataset may help to mitigate the effects of noise introduced with these categorizations of skill. In the future, we plan to collect expert ratings of suturing performance on the simulator (from video recordings) and relate them with metrics from the SutureCoach. Another aim of our team is the long-term study of training effects of using the SutureCoach with expert ratings. Through this future study, we aim to further validate the response process, internal structure, and consequences by measuring performance with expert ratings and whether feedback and practice on the SutureCoach improve suturing skills.

A. Conclusion

In summary, our study has demonstrated the efficacy of the metrics we established in distinguishing among the three groups of study participants: (1) attending surgeons and fellows, (2) residents, and (3) novices, including both medical and non-medical students. We observed remarkable success with rotational *LDLJ* and *APL* metrics, effectively differentiating between all groups in both conditions. In contrast, only one tooltip measure achieved similar success in differentiation. Our analysis of open needle driving motion characteristics, coupled with insights from previous studies [24], [29], reveal that rotational motion metrics were more consistent in assessing open suturing skill, highlighting the vast potential for assessing rotational motion for specific surgical skills. Motion smoothness metrics, particularly *LDLJ*, have exhibited substantial promise in this regard. To our knowledge, no prior studies have specifically assessed open suturing skills using rotational motion smoothness metrics. To further enhance our understanding and strengthen the validity of these metrics, future work will involve trial-based ratings, facilitating direct metric comparisons and increasing confidence in their reliability.

Supplementary Materials

The supplementary materials provide additional information to enhance the understanding of this study. The introduction section details a background on the development of motion smoothness metrics leading to *LDLJ* and *SPARC*. The materials and methods section contains details on calculating angular velocity from quaternions, along with clarifications on the formulation of *LDLJ* and *SPARC*. Finally, the results sections provides figures of the confidence interval plots from sensor 2 and tables of p-value results from the metric pair-wise comparisons for data from both sensors.

Author Contributions

S.P.S. wrote the initial draft of the manuscript. S.P.S., A.M.S., and J.G. contributed to simulator development and data collection. S.P.S. performed data processing and metric calculation. S.P.S. and J.B. worked on the statistical analysis of the data. A.M.S., J.G., J.B., R.E.G., and R.S. contributed to data interpretation and feedback for improving the research and analysis. All authors reviewed and edited the manuscript. R.S secured and managed funding for the project as well as conceived the research study.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] M. G. Sheahan, C. Duran, and J. Bismuth, "National simulation-based training of fellows," *Surg. Clin. North Amer.*, vol. 95, no. 4, pp. 781–790, Aug. 2015.
[2] L. Nayahangan et al., "A national needs assessment to identify technical procedures in vascular surgery for simulation based training," *Eur. J. Vasc. Endovascular Surg.*, vol. 53, no. 4, pp. 591–599, Apr. 2017.
[3] K. M. Bennett, K. C. Kent, J. Schumacher, C. C. Greenberg, and J. E. Scarborough, "Targeting the most important complications in vascular surgery," *J. Vasc. Surg.*, vol. 65, no. 3, pp. 793–803, Mar. 2017.
[4] H. B. Colaco, K. Hughes, E. Pearse, M. Arnander, and D. Tennent, "Construct validity, assessment of the learning curve, and experience of using a low-cost arthroscopic surgical simulator," *J. Surg. Educ.*, vol. 74, no. 1, pp. 47–54, Jan. 2017.
[5] S. Singh, C. Vanatter, and R. Singapogu, "Phase-based assessment of arthroscopic skill using motion smoothness metrics: A simulator-based proof-of-concept study," in *Proc. 13th Int. Conf. Appl. Hum. Factors Ergonom.*, 2022, pp. 58–66.
[6] V. Chandra et al., "A comparison of laparoscopic and robotic assisted suturing performance by experts and novices," *Surgery*, vol. 147, no. 6, pp. 830–839, Jun. 2010.
[7] J. B. Pagador et al., "Decomposition and analysis of laparoscopic suturing task using tool-motion analysis (TMA): Improving the objective assessment," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 7, no. 2, pp. 305–313, Mar. 2012.
[8] A. M. Derossis, G. M. Fried, M. Abrahamowicz, H. H. Sigman, J. S. Barkun, and J. L. Meakins, "Development of a model for training and evaluation of laparoscopic skills," *Amer. J. Surg.*, vol. 175, no. 6, pp. 482–487, Jun. 1998.
[9] A. Ghasemloonia, Y. Maddahi, K. Zareinia, S. Lama, J. C. Dort, and G. R. Sutherland, "Surgical skill assessment using motion quality and smoothness," *J. Surg. Educ.*, vol. 74, no. 2, pp. 295–305, Mar. 2017.
[10] S. Estrada, C. Duran, D. Schulz, J. Bismuth, M. D. Byrne, and M. K. O'Malley, "Smoothness of surgical tool tip motion correlates to skill in endovascular tasks," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 5, pp. 647–659, Oct. 2016.
[11] C. Duran et al., "The model for fundamentals of endovascular surgery (FEVS) successfully defines the competent endovascular surgeon," *J. Vasc. Surg.*, vol. 62, no. 6, pp. 1660–1666.e3, Dec. 2015.

[12] R. S. Sidhu, J. Park, R. Brydges, H. M. MacRae, and A. Dubrowski, "Laboratory-based vascular anastomosis training: A randomized controlled trial evaluating the effects of bench model fidelity and level of training on skill acquisition," *J. Vasc. Surg.*, vol. 45, no. 2, pp. 343–349, Feb. 2007.

[13] M. G. C. Pellen, L. F. Horgan, J. R. Barton, and S. E. Attwood, "Construct validity of the promis laparoscopic simulator," *Surg. Endoscopy*, vol. 23, no. 1, pp. 130–139, Jan. 2009.

[14] E. Hiemstra, M. K. Chmarra, J. Dankelman, and F. W. Jansen, "Intracorporeal suturing: Economy of instrument movements using a box trainer model," *J. Minimally Invasive Gynecol.*, vol. 18, no. 4, pp. 494–499, Jul. 2011.

[15] E. F. Hofstad, C. Våpenstad, M. K. Chmarra, T. Langø, E. Kuhry, and R. Mårvik, "A study of psychomotor skills in minimally invasive surgery: What differentiates expert and nonexpert performance," *Surg. Endoscopy*, vol. 27, no. 3, pp. 854–863, Mar. 2013.

[16] R. W. Partridge, M. A. Hughes, P. M. Brennan, and I. A. Hennessey, "Accessible laparoscopic instrument tracking ("InsTrac"): Construct validity in a take-home box simulator," *J. Laparoendoscopic Adv. Surg. Techn.*, vol. 24, no. 8, pp. 578–583, Aug. 2014.

[17] V. Datta, S. Mackay, M. Mandalia, and A. Darzi, "The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model," *J. Amer. College Surgeons*, vol. 193, no. 5, pp. 479–485, Nov. 2001.

[18] A. L. Trejos, R. V. Patel, R. A. Malthaner, and C. M. Schlachta, "Development of force-based metrics for skills assessment in minimally invasive surgery," *Surg. Endoscopy*, vol. 28, no. 7, pp. 2106–2119, Jul. 2014.

[19] G. Retrosi, T. Cundy, M. Haddad, and S. Clarke, "Motion analysis and assessment in pediatric laparoscopy: Construct, concurrent, and content validity for the eoSim simulator," *J. Laparoendoscopic Adv. Surg. Techn.*, vol. 25, no. 11, pp. 944–950, Nov. 2015.

[20] J. A. Sánchez-Margallo, F. M. Sánchez-Margallo, I. Oropesa, S. Enciso, and E. J. Gómez, "Objective assessment based on motion-related metrics and technical performance in laparoscopic suturing," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 2, pp. 307–314, Feb. 2017.

[21] I. Kil, J. F. Eidt, R. E. Groff, R. B. Singapogu, "Assessment of open surgery suturing skill: Simulator platform, force-based, and motion-based metrics,," *Front. Med.*, vol. 9, 2022.

[22] V. M. Belvroy, B. Murali, M. G. Sheahan, M. K. O'Malley, and J. Bismuth, "Motion metrics reliably differentiate competency: Fundamentals of endovascular and vascular surgery," *J. Vasc. Surg.*, p. S0741521420304754, Apr. 2020.

[23] H. Mohamadipanah et al., "Sensors and psychomotor metrics: A unique opportunity to close the gap on surgical processes and outcomes," *ACS Biomaterials Sci. Eng.*, vol. 6, no. 5, pp. 2630–2640, May 2020.

[24] Y. Sharon, A. M. Jarc, T. S. Lendvay, and I. Nisky, "Rate of orientation change as a new metric for robot-assisted and open surgical skill evaluation," *IEEE Trans. Med. Robot. Bionics*, vol. 3, no. 2, pp. 414–425, May 2021.

[25] S. Singh, J. Bible, Z. Liu, Z. Zhang, and R. Singapogu, "Motion smoothness metrics for cannulation skill assessment: What factors matter?," *Front. Robot. AI*, vol. 8, 2021, Art. no. 625003.

[26] N. Hogan and D. Sternad, "Sensitivity of smoothness measures to movement duration, amplitude and arrests," *J. Motor Behav.*, vol. 41, no. 6, pp. 529–534, Nov. 2009.

[27] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, and E. Burdet, "On the analysis of movement smoothness," *J. NeuroEngineering Rehabil.*, vol. 12, no. 1, Dec. 2015, Art. no. 112.

[28] E. L. Mitchell et al., "A systematic review of assessment of skill acquisition and operative competency in vascular surgical training," *J. Vasc. Surg.*, vol. 59, no. 5, pp. 1440–1455, May 2014.

[29] A. M. Shayan et al., "Measuring hand movement for suturing skill assessment: A simulation-based study," *Surgery*, vol. 174, pp. 1184–1192, Aug. 2023.

[30] T. Flash and N. Hogan, "The coordination of arm movements: An experimentally confirmed mathematical model," *J. Neurosci.*, vol. 5, no. 7, pp. 1688–1703, Jul. 1985.

[31] A. Melendez-Calderon, C. Shirota, and S. Balasubramanian, "Estimating movement smoothness from inertial measurement units," *Front. Bioeng. Biotechnol.*, vol. 8, 2021, Art. no. 558771. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fbioe.2020.558771/full

[32] S. Balasubramanian, A. Melendez-Calderon, and E. Burdet, "A robust and sensitive metric for quantifying movement smoothness," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2126–2136, Aug. 2012.

[33] S. Messick, "Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning," *Amer. Psychol.*, vol. 50, no. 9, pp. 741–749, 1995.