Barbara Kim
Group 12

# Critical Review: Objective Evaluation of Expert and Novice Performance during Robotic Surgical Training Tasks

Judkins, Timothy N., Dmitry Oleynikov, and Nick Stergiou. "Objective evaluation of expert and novice performance during robotic surgical training tasks." Surgical endoscopy 23.3 (2009): 590-597.

## Project Review

Our project is concerned with the Robotic Ear, Nose, and Throat Microsurgery System (REMS). We have three subprojects: one where we validate the REMS via an IRB approved sinus study, one where we calibrate the robot to take in account the distortion of the tool, and one in which we focus on measuring free rotation in the tool. This paper is relevant to the first subproject. The REMS validation study will compare novices to experts both via the REMS and freehand. Subjects will navigate the sinus environment in a cadaver head using a tracked instrument either attached to the REMS or freehand. The two groups will then be assessed using certain metrics for their surgical skill. This paper reviews and evaluates a list of quantitative metrics which may be useful for our first subproject.

## Paper Selection

The paper selected is *Objective Evaluation of Expert and Novice Performance during Robotic Surgical Training Tasks*, published in 2009. The study was conducted by three researchers in University of Nebraska Medical Center. The goal of this paper is to identify objective measures that quantitatively and completely describe subjects' performances in robotic surgical tasks. Previously, measures used to compare robot surgical performances have either been subjective or incomplete. Examples of such measures include expert opinion and time to task completion (TTC). In this paper, the authors first describe the experimental setup, and then summarize the different objective metrics they are going to test, and finally summarize the results. In discussion, they further discuss any interesting results and evaluate which of the objective metrics are most successful.

We identified this paper as an excellent resource in relation to our first sub-project: the REMS sinus validation study. The experimental setup is similar to our study in its comparison of novice to expert performance during robotic surgical tasks. Our surgical task will be more complex than the surgical tasks performed in this study. However, the statistical tests and quantitative metrics used for surgical skill assessment are still highly relevant and may be applied in our own analyses. Thus, we shall refer to this paper for its description of its methods, its data analysis, and its suggested measures for assessing surgical performance.

## Experimental Setup

The study recruited five novices and five expert users of the Da Vinci Surgical System (dVSS) to participate in the study. The subjects performed three tasks using the dVSS that varied in difficulty: bimanual carrying (BC), needle passing (NP), and suture tying (ST). The bimanual carrying task involved picking up and relocating two metal caps using both the left and right instruments (Fig. 1A). The needle passing task involved passing a needle through six holes (Fig. 1B). The suture tying task, according to its name, involved tying a surgical suture (Fig. 1C).

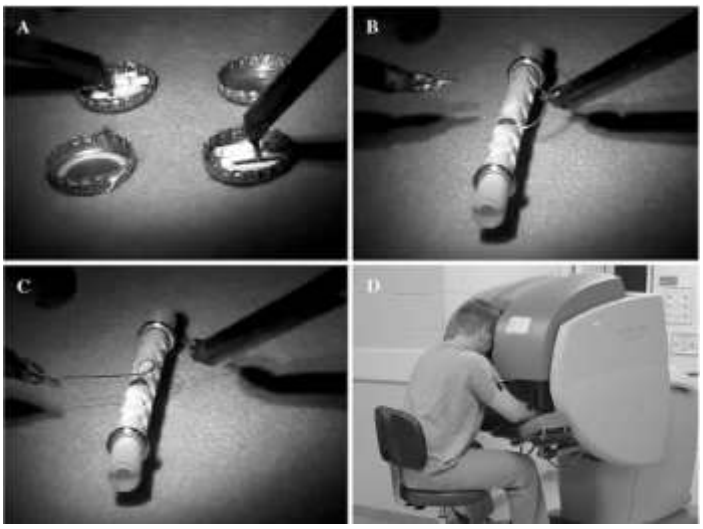


*Figure 1 Depiction of Surgical Tasks*

The novices underwent a training module in which they performed each task 3 times as part of pre-training (N-PRE). They then performed 10 trials of the 3 tasks as part of a training segment. Finally, the trained novices performed each task 3 times again as part of post-training (N-POST). N-PRE and N-POST were recorded and were compared to 3 trials of the experts' performance (EX).

## Objective Measures

The study evaluated 6 different quantitative metrics in their ability to assess surgical skill: time to task completion (TTC), total distance traveled (D), speed (S), curvature (κ), and relative phase (Φ). TTC and D had previously been used as objective measures but were insufficient in describing the subject's

performance. Time to task completion is the time needed to complete the given surgical task. Total distance traveled is the sum of the Euclidean distances between each time sample. Speed (S) is self-explanatory. Curvature (κ) measures the straightness of the subject's path and was calculated using the following equation:

$$\kappa = \frac{|\dot{r} * \ddot{r}|}{\dot{r}^3}$$

$\dot{r}$ is the velocity of a point $r$ on the three dimensional path and $\ddot{r}$ is the acceleration. If the value of κ was closer to 0, this indicated that the subject used relatively smooth and straight movements whereas larger values indicated curved and jerky movements. Relative phase is the difference in the phase angles ($\Phi = \varphi_A - \varphi_B$) of the left and right instruments to indicate whether the direction of the instrument tips are coordinated. Relative phase is calculated by the equation below:

$$\Phi = \tan^{-1}(\frac{\dot{x}_A}{x_A})$$

A value of Φ closer to zero indicates an in-phase relationship while values closer to 180 indicates an anti-phase relationship.

# Results

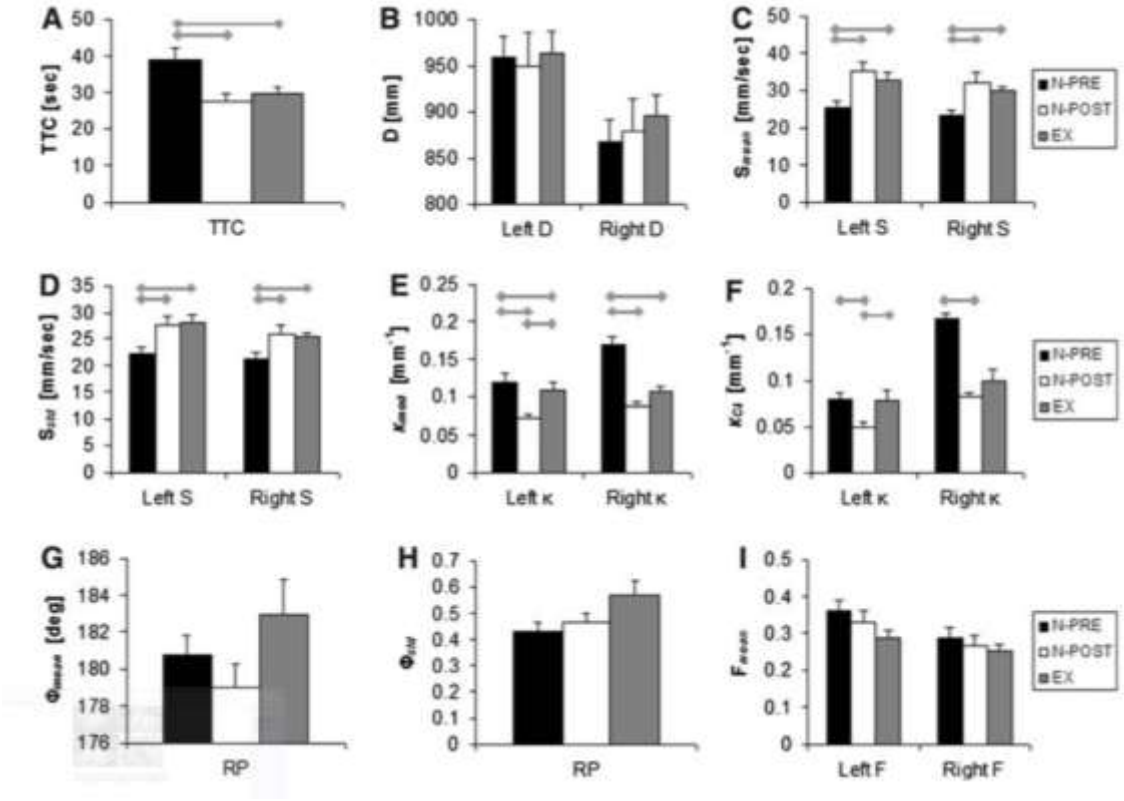


*Figure 2 Group means for pretraining trials (N-PRE), posttraining trials (N-POST), and experts (EX) during the BC task*

Independent t-tests were used to compare group means of the objective measures between N-PRE and EX and between N-POST and EX at $\alpha$ =0.05. Paired t-tests were used to compare group means of the objective measures between N-PRE and N-POST at $\alpha$ = 0.05.

Depicted in Fig. 2 are the results of the bimanual carrying task. In Fig. 2A, N-POST TTC and EX TTC are significantly lesser than the N-PRE TTC. In Fig. 2C, N-POST Smean and EX Smean are significantly higher than the N-PRE Smean. The Sstd was also significantly larger for the novices after training and the experts than the Sstd for the novices before training. Finally, κmean was significantly lesser for the novices after training and the expert than the κmean for the novices before training. Novices after training performed comparably to experts, as can be seen by the TTC, the $S_{mean}$, $S_{std}$, and the $\kappa_{mean}$.

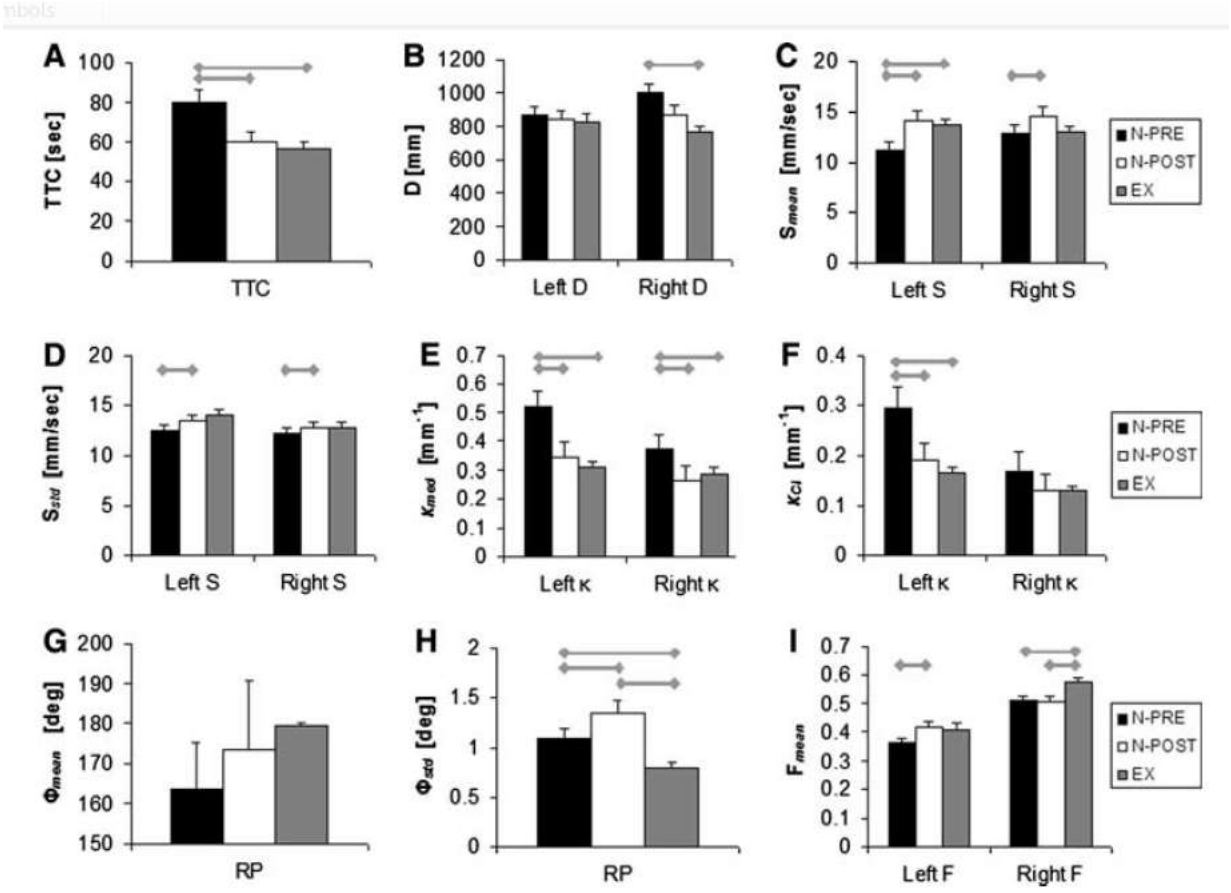


*Figure 3 Group means for pretraining trials (N-PRE), posttraining trials (N-POST), and experts (EX) during the needle-passing (NP) task*

Fig. 3 shows the results of the needle-passing task. General trends remain similar to what was observed in the bimanual carrying task. In Fig. 3C, the $S_{mean}$ for the novices after training was significantly higher than that of the novices before training on both the left and right sides. The $S_{mean}$ of the experts was significantly higher on the left side but not on the right side. In Fig. 3E, $\kappa_{mean}$ was significantly lesser for the novices after training and the expert than the $\kappa_{mean}$ for the novices before training.

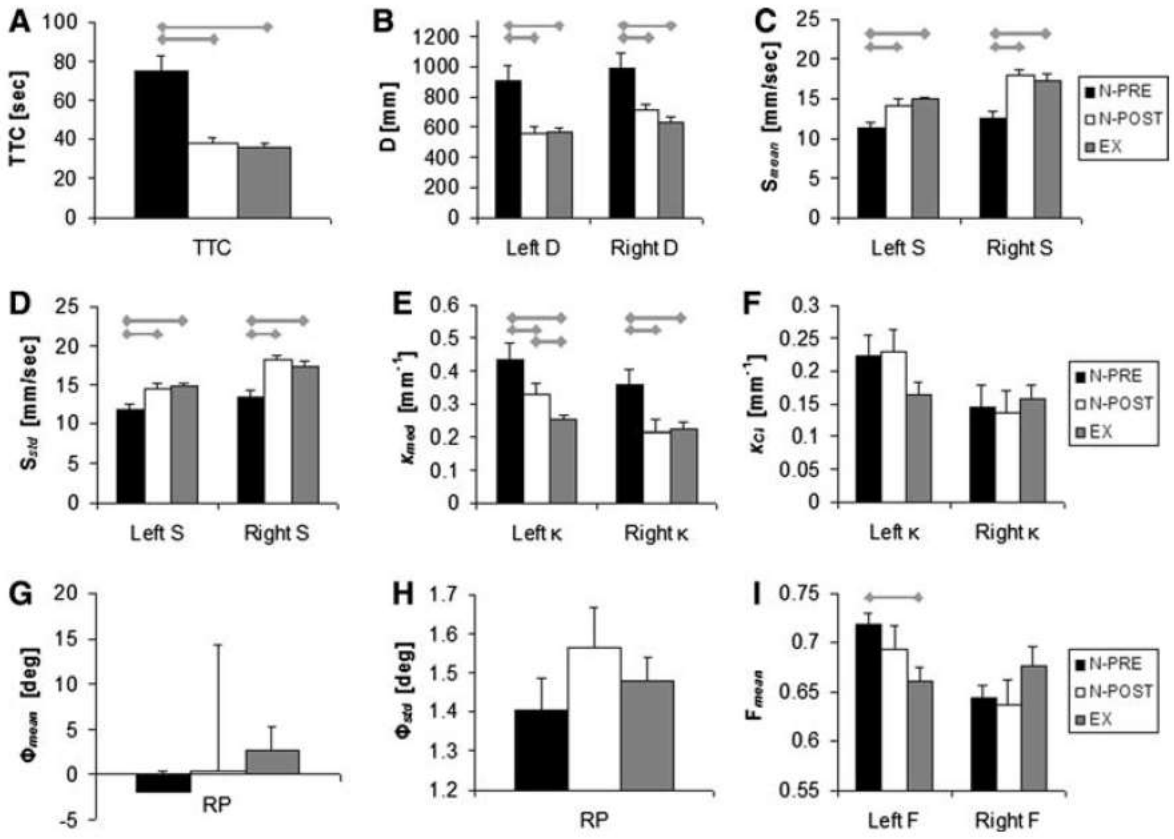


*Figure 4 Group means for pretraining trials (N-PRE), posttraining trials (N-POST), and experts (EX) during the suture-tying task (ST) task*

Fig. 4 shows the results of the suture-tying task. Again, TTC and $\kappa_{mean}$ are significantly shortened with novices after training and experts (Fig. 4A, Fig. 4E). $S_{mean}$ is significantly heightened with novices after training and experts in both the left and right side (Fig. 4C). In the suture tying, the total distance traveled to complete the task (D) was significantly higher for the novices before training than the D for the novices after training and the experts (Fig. 4B).

Overall, the study noted that novices after training improved in TTC, S, and κ. D was not always significant and Φ did not give statistically significant results. The study notes that the number of objective measures with significant differences increased with task complexity. In addition, the novices after training performed comparably to the experts. The study also noted that the novices were able to achieve comparable results with the experts with relatively little training (n = 10 trials).

## Paper Assessment

The paper explained its experimental setup well and clearly explained the training module of the novices. The objective measures were also clearly explained. Their graphs were thorough and they

explained each evaluated objective measure in each task. Their suggested objective measures will prove useful in our own subproject and we will likely use κ and S to evaluate surgical skill assessment. However, there were objective measures that were not measured, such as deviance from the expected path. Jerk could also have been an objective measure that they could have evaluated.