Time to complete questions and exercises (seconds)			
	Expert	**Novice**	
Module 1 - Colon Mobilization	Mean ± Stdev		p value
Anatomy recognition (overall)	137±55	163±38	0.11
Section one	39±20	71±27	0.002
Section two	64±33	49±11	0.04
Section three	33±11	43±21	0.16
Technical questions	110±70	131±44	0.29
Steps of operation	38±16	51±14	0.02
Module 2 - Kocherization of Duodenum			
Anatomy recognition	51±6	99±40	0.001
Steps of operation	46±11	64±23	0.03
Module 3 - Hilar Dissection			
Anatomy recognition	100±31	156±70	0.04
Technical questions	24±15	25±8	0.85
Steps of operation	8±2	10±5	0.15

Number of incorrect answers			
	Expert	**Novice**	
Module 1 - Colon Mobilization	Mean ± Stdev		p value
Anatomy recognition	1±0.9	7±3.6	<0.001
Technical questions	1.1±1.5	0.6±0.7	0.19
Steps of operation	0.5±1.0	2.5±2.1	0.007
Module 2 - Kocherization of Duodenum			
Anatomy recognition	0.1±0.3	4.6±2.5	<0.001
Steps of operation	0.1±0.3	4.1±2.5	<0.001
Module 3 - Hilar Dissection			
Anatomy recognition	0.4±0.5	7.6±4.6	<0.001
Technical questions	0.1±0.4	1.8±1.2	0.001
Steps of operation	0±0	0.5±0.9	0.11

Figure 1
A – Anatomy identification
B – Technical task (retraction)
C – Steps of surgery

PD6-08

CROWD-SOURCED ASSESSMENT OF TECHNICAL SKILLS (C-SATS™): FAST, ECONOMICAL AND ACCURATE ASSESSMENT OF ROBOTIC SURGERY

Lee W. White*, Stanford, CA; Timothy Kowalewski, Minneapolis, MN; Timothy Brand, Tacoma, WA; Jonathan Harper, Bryan Comstock, Blake Hannaford, Thomas S. Lendvay, Seattle, WA

INTRODUCTION AND OBJECTIVES: Surgeon skill has recently been shown to predict patient outcome when measured using objective structured assessment tools. However, the time-cost to have expert surgeons grade surgical videos is great. We hypothesized that crowd-sourcing of surgical skills assessment would be as accurate as and faster than conventional expert surgeon graders rating the same surgical performances.

METHODS: Fifty-one urology residents and faculty performed two dry-lab surgical training tasks on a da Vinci surgical robot: 1) Fundamentals of Laparoscopic Surgery intracorporeal suturing and 2) a rocking pegboard transfer task. Forty-nine recorded performances from each were available to be uploaded to a website built to facilitate efficient grading using the depth perception, bimanual dexterity and efficiency domains of the validated Global Evaluative Assessment of Robotic Surgery scoring tool. Three surgical faculty graded the tasks after completing a grader training session to maximize agreement. Each performance was then scored by 30 Amazon.com Mechanical Turk crowd workers in return for a payment of $0.25 to $0.50. Mean surgeon and mean crowd scores were computed and compared using correlation coefficient (CC) and Cronbach's Alpha (CA), a measure of crowd-surgeon agreement.

RESULTS: Figure 1. shows the agreement between surgeon scores and C-SATS™ scores. The CC between surgeon grade and C-SATS™ was 0.79 for rocking pegboard and 0.86 for suturing. CA was 0.84 for the rocking pegboard and 0.92 for suturing, indicating 'good' and 'excellent' agreement, respectively. 67% of the C-SATS™ scores for rocking pegboard and 69% of the C-SATS™ scores for suturing fell within 1 point of the surgeon-provided score on a 3-15 point possible score range. The cost to assess these surgical performances was small: $10.07 per rocking pegboard video and $15.67 per suturing video. Furthermore, crowds provided scores in 9 hours for 49 suturing videos compared to over a month for the surgeon panel.

CONCLUSIONS: This is the first demonstration of untrained crowds accurately assessing robotic surgical performances on diverse tasks, representing a wide range of surgeon skill levels. Also, C-SATS™ can be much faster than relying on faculty assessments. Future studies are needed to compare crowd assessment of surgeries on real patients to complications and clinical outcomes. If correlated, this type of assessment could be central to the surgical resident training and credentialing processes.
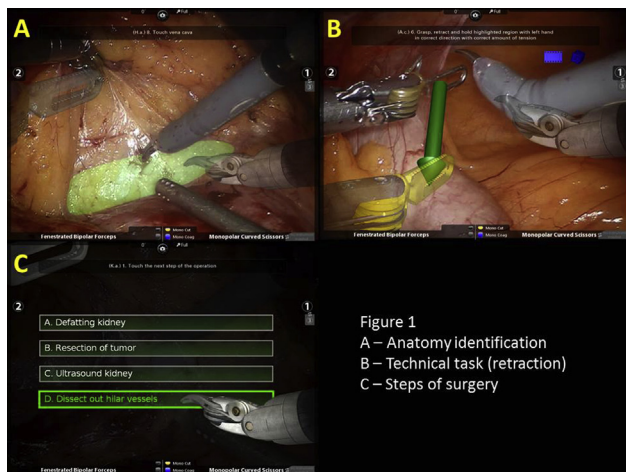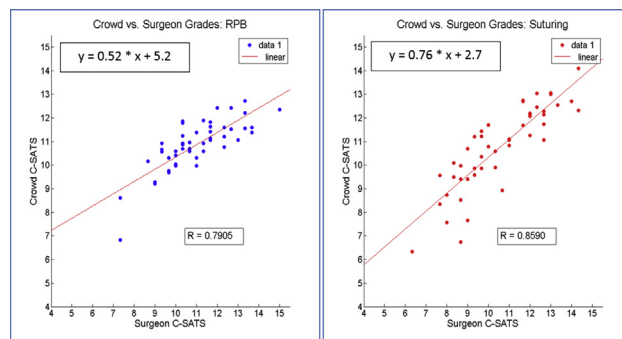

Figure 1: Scores for 49 recorded suturing tasks and 49 recorded rocking pegboard tasks (each point is one performance). Presented with crowd to surgeon correlation coefficients and lines of best fit. LEFT: Rocking Pegboard (RPB). RIGHT: Suturing.

Department of Urology, University of Washington, Seattle, WA

PD6-09

EFFECT OF POST-CALL FATIGUE MEASURED BY ROBOTIC SKILLS SIMULATOR

Aaron Weinberg*, Solomon Woldu, Tammer Yamany, Ruslan Korets, Ketan Badani, New York, NY

INTRODUCTION AND OBJECTIVES: Studies on post-call residents have shown that manual dexterity and surgical skills are worsened by fatigue following a 24-hour call. National work-hour restrictions have been initiated for certain levels of residency training to prevent complications and maximize patient care. The purpose of this