Comparison of the Usability of Robotic Surgery Simulators

Alyssa Tanaka, M.S., Courtney Graddy, M.S., Roger Smith, Ph.D.
Florida Hospital Nicholson Center
Celebration, FL
Alyssa.tanaka@flhosp.org, courtney.graddy@flhosp.org, roger.smith@flhosp.org

Haider M. Abdul-Muhsin, M.D.
Mayo Clinic
Scottsdale, AZ
Hma980@gmail.com

ABSTRACT

The introduction of simulation into minimally invasive robotic surgery is relatively recent and has seen rapid advancement; therefore, a need exists to develop training curriculums and to identify systems that will be most effective at improving surgical skills. Several robotic simulators have been introduced to support these aims, but their effectiveness has yet to be fully evaluated.

Currently, there are three simulators -- the daVinci Skills Simulator, Mimic dV-Trainer, and Surgical Simulated Systems' RoSS. While multiple studies have been conducted to demonstrate the validity of each system, no studies have been conducted which compare the value of these devices as tools for education and skills improvement.

This paper presents the results of an experiment comparing value, usability, and validity of all three systems. Subjects who were qualified as medical students or physicians (n=105) performed one exercise on each of the three simulators and completed two questionnaires, one regarding their experience with each device and a second regarding the comparative effects of the simulators. This data confirmed the face, content, and construct validity for the dV-Trainer and Skills Simulator. Similar validities could not be confirmed for the RoSS. Greater than 80% of the time, participants chose the Skills Simulator in terms of physical comfort, ergonomics, and overall choice. However, only 55% thought the skills simulator was worth the cost of the equipment. The dV-Trainer had the highest cost preference scores with 71% percent of respondents feeling it was worth the investment.

This work is the second component of a three-part analysis. In the previous study, the simulators were objectively reviewed and compared in terms of their system capabilities. The third part will evaluate the transfer of training effect of each simulator. Collectively, this work will offer end users and potential buyers a comparison of the value and preferences of robotic simulators.

ABOUT THE AUTHORS

Alyssa D.S. Tanaka, M.S. is a Systems Engineer at Florida Hospital's Nicholson Center. Her research work focuses on robotic surgery simulation and effective surgeon training. Her current projects include rapid prototyping of surgical education devices, the validation of a robotic surgical curriculum and evaluation of robotic simulation systems. She is a Modeling and Simulation PhD student at the University of Central Florida and previously earned a M.S. in Modeling and Simulation, Graduate Simulation Certificate in Instructional Design, and a B.S. in Psychology and Cognitive Sciences from the University of Central Florida.

Courtney Graddy, M.S. is a Human Studies Research Coordinator at the Celebration Health Research Institute where she manages projects aimed at improving patient health outcomes, employee health, process improvement and simulation research. Her current projects focus on integrating technology into standard of care and evaluating its effects on patient health and patient satisfaction, as well as evaluating teaching modalities used to train surgeons. Her career began at the North Florida South Georgia Veterans Health System where she aided in the development of employee education materials and program planning and evaluation with the Geriatric Research Education and Clinical Center. She holds a Bachelors of Science in Health Education from the University of Florida and a Masters of Health Administration from the University of South Florida.

Haidar Mohamed Abdul-Muhsin, M.D. is an endourology fellow at Mayo Clinic in Scottsdale, AZ. Prior to this role, he was a urology fellow at the Global Robotics Institute at Florida Hospital at Celebration. Dr. Abdul-Muhsin has been practicing medicine for 10 years. His research interests include prostate cancer research with emphasis on robotic prostatectomy outcomes and research related to robotic surgical simulation. Currently, he is researching the contribution of laparoscopic surgical experience in developing robotic proficiency.

Roger Smith, Ph.D. is an expert in the development of simulation devices and training programs. He has spent 25 years creating leading edge simulators for the Department of Defense and Intelligence agencies, as well as accredited methods for training with these devices. He is currently the Chief Technology Officer for the Florida Hospital Nicholson Center where he is responsible for establishing the technology strategy and leading technology implementation. He has served as the CTO for the U.S. Army PEO for Simulation, Training and Instrumentation (PEO-STRI); VP and CTO for training systems at Titan Corp; and Vice President of Technology at BTG Inc. He holds a Ph.D. in Computer Science, a Doctorate in Management, and an M.S. in Statistics. He has published 3 professional textbooks on simulation, 10 book chapters, and over 100 journal and conference papers. His most recent book is *Innovation for Innovators: Leadership in a Changing World*. He has served on the editorial boards of the *Transactions on Modeling and Computer Simulation* and the *Research Technology Management* journals.

Comparison of the Usability of Robotic Surgery Simulators

Alyssa Tanaka, M.S., Courtney Graddy, M.S., Roger Smith, Ph.D.
Florida Hospital Nicholson Center
Celebration, FL
Alyssa.tanaka@flhosp.org, courtney.graddy@flhosp.org, roger.smith@flhosp.org

Haider M. Abdul-Muhsin, M.D.
Mayo Clinic
Scottsdale, AZ
Hma980@gmail.com

INTRODUCTION

Robotic surgery has introduced a new dimension into the surgical field. With the introduction of robotic technology between patient and surgeon, a need to master new skills has emerged. Medicine has come to the conclusion that the Halstedian training model (See one, do one, teach one) is no longer sufficient for teaching complex skills, especially robotic surgical skills (Cameron, 1997). A number of simulators have been developed to support training and skill assessment in robotic surgery. The currently available dedicated robotic simulators include: the da Vinci Skills Simulator (dVSS) by Intuitive Surgical Inc., also known as the "Backpack Simulator"; the dV-Trainer from Mimic Technologies Inc.; and the RoSS by Simulated Surgical Sciences LLC (Figure 1). The purpose of these simulators is to train surgeons prior to using the actual system and to allow them to acquire the necessary robotic skills to perform a safe surgery. All of these da Vinci simulators utilize a visual scene that is presented in a computer generated 3D environment providing challenging tests for practicing dexterity and machine operations. Originally, the simulated exercises trained basic robotic skills; however with advances in technology, surgeons can now train for specific procedures (e.g. nephrectomy and hysterectomy).

Figure 1. Simulators of the da Vinci robotic surgical system

Our hospital research laboratory has purchased each of these three simulators for the purpose of studying their effectiveness and applying them to the education of robotic surgeons, specifically for the Department of Defense (DoD). The DoD is interested in the effectiveness of the simulators to train military surgeons prior to and after returning home from deployments. This research is structured as three distinct stages.

From the first stage of this work, the authors summarized the objective characteristics of the three systems. This included descriptions of the exercises offered in each, metrics used to evaluate students, overview of the system administration functions, physical dimensions and configurations of the equipment, and comparisons of the costs of the devices and their support equipment (Smith & Truong, 2013). In the first simulator, the trainee sits at and operates the simulated environment using the actual da Vinci surgical console. The simulator is a custom computer appended to the surgical console through the actual surgical data port. While the simulator costs approximately

$100,000, the surgical console costs $500,000 incurring an investment of $600,000. Using this simulator, users can train using the actual hardware they would use during surgery; however, this requires the use of the surgical console that may be needed to conduct surgeries. Most hospitals may not have a dedicated training console, meaning that users would not have appropriate access to the simulator. The second is a standalone system that utilizes a graphic/gaming computer, connected to a custom desktop viewing and control device that replicates the hardware of the da Vinci surgeon's console. This system shares similar software with the dVSS, but does not require the use of any actual da Vinci hardware. The cost of this simulator is approximately $100,000. The third is composed of a completely customized replica of the da Vinci surgeon's console. Internally the simulator contains a graphic computer, a 3D monitor, and commercial Omni Phantom haptic controllers. This simulator uses unique software and is a little more than $100,000 (Smith &Truong, 2013).

This paper reports on the second stage of this research, in which the validity and usability of the simulators is examined. The third stage will be a measure of learning effectiveness using the systems.

Validity in Surgical Simulation

The validity of medical and surgical simulators is usually measured by the categories defined by McDougal (2007). This paper defines the most commonly recognized forms of validation as: *face, content, construct, concurrent, and predictive validity*. *Face validity* is typically assessed informally by users and is used to determine whether the simulator is an accurate representation of the actual system (i.e. the realism of the simulator). *Content validity* is the measure of the appropriateness of the system as a teaching modality. Experts who are knowledgeable about the device typically assess this via a formal evaluation. *Construct validity* is the ability of a simulator to differentiate between the performances of experienced users and those who are novices. *Concurrent validity* is the extent to which the simulator correlates with the "gold standard" and *predictive validity* is the extent to which the simulator can predict a user's future performance. Collectively, concurrent and predictive validity are known as criterion validity and are used as measures of the simulator's ability to correlate trainee performance with their real life performance. Face and content validity are most effective in evaluating the ability of a simulator to train a surgeon; however construct, concurrent, and predictive validity are most useful for evaluating the effectiveness of a simulator to assess a trainee.

The validity of all three simulators has been tested and reported separately for the da Vinci skill simulator (Hung, Zehnder, Patil, 2011; Kelly, Margules, Kundavaram, 2012; Liss, Abdelshehid, Quach, 2012), the dV-Trainer (Kenney, Wszolek, Gould, Libertino, Moinzadeh, 2009; Sethi, Peine, Mohammadi, 2009; Lee, Mucksavage, Kerbl, 2012) and the RoSS (Seixas-Mikelus, Kesavadas, Srimathveeravalli, 2010; Stegemann et al., 2013; Colaco, Balica, Su, 2012; Raza et al., 2013). To our knowledge only one publication has compared features of two of the simulators, but no comparative studies have been performed with all three of the systems (Liss MA, Abdelshehid C, Quach S., 2012). Thus, the current study aimed to compare all three commercially available da Vinci simulators and detail the findings for face, content, and construct validity for the three systems.

METHODS

Recruitment

Participants in this study included medical students, residents, fellows, and attending physicians. Participants were recruited from the University of Central Florida Medical School, courses held at the Nicholson Center, and two medical robotic conferences (World Robotics Gynecology Congress and Society of Robotic Surgeons Scientific Meeting). Subjects were excluded from participating if they indicated that they had participated in a formal robotic simulation-training course.

Each participant was categorized into one of three groups (i.e. Expert, Intermediate, or Novice) according to the self-reported number of robotic cases (i.e. procedures) he or she had performed. Individuals performing 0-19 robotic cases in which they had 50% or greater console time were categorized as Novices, individuals with 20-99 robotic cases were considered to be Intermediates, and individuals with 100 or more cases were considered to be Experts.

Materials

After being categorized into an experience level, each participant was assigned a specific order in which they used each of the simulators (Figure 2). This order system was used to identify and potentially eliminate any bias that may exist by using a specific system first. All participants completed one exercise on each of the simulators. The tasks chosen were Peg Board 1 in both the dV-Trainer and the dVSS and Ball Placement 1 in the RoSS. The same task was used for both the dV-Trainer and the dVSS because these systems share similar software and exercises. The RoSS software contains unique exercises and Ball Placement 1 is designed to teach the same skills as Peg Board 1.

Participant	Simulator Order
01	CBA
02	BAC
03	ACB
04	CBA
05	BAC
06	ACB
07	CBA
08	BAC
09	ACB

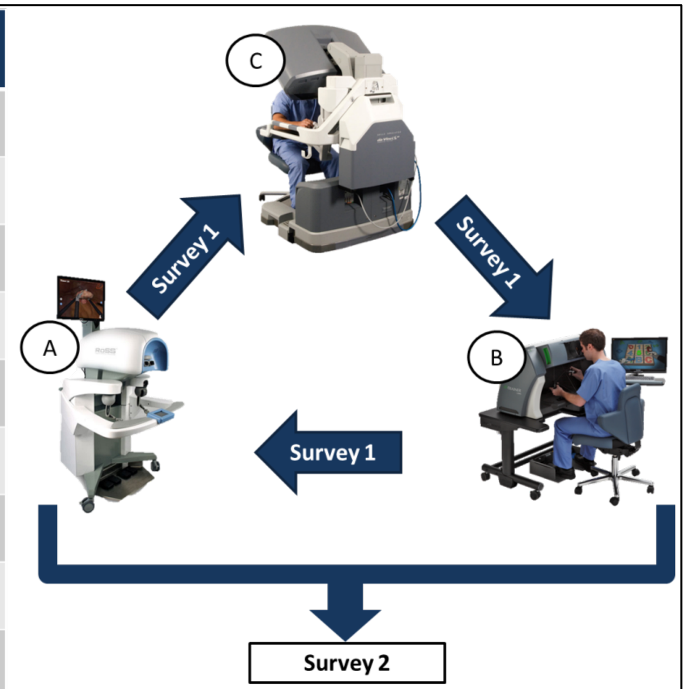

Figure 2. Rotating order of use by subjects, with survey order.

After each exercise on each simulator, participants completed a post questionnaire (Survey 1), which asked for feedback regarding their experience on that specific simulator. After using all three systems, subjects completed a second post questionnaire (Survey 2), which asked them to compare all three systems to each other. The participant's performance metrics were also collected from each of the simulators.

RESULTS

Demographics

Subjects were categorized as Novice (n=37), Intermediate (n=31), or Expert (n=37). Sixty-two percent of subjects were men and 38% were women with an average age of 43. On average, participants had 15 years in practice and 3 years of robotic experience. Seventy-six percent were attending physicians and 73% of participants were currently or had received robotic training, while 41% provided that they train residents and fellows. There were differences in the average age and number of years in practice of participants based on the classification of expert, intermediate or novice (number of robotic procedures). These are to be expected, since higher ages are required to achieve higher number of years of practice and larger numbers of robotic procedures.

Validation

The types of validity evaluated in this experiment were face, content, and construct. To analyze the systems for face validity and content validity, questions from Survey 1 were used. The questions were evaluated on a five point Likert scale (Strongly Disagree, Disagree, Neither Agree or Disagree, Agree, and Strongly Agree). Face validity was analyzed by expert and intermediate feedback as recommended by Van Nortwick et al. (2010) because these are the users most familiar with the robotic system; however, only expert feedback was used for content validity because they have the best ability to judge the appropriateness of the system as a training tool. For construct validity, performance metrics such as Overall Score, Time to Complete, Number of Errors, and Economy of Motion were analyzed (Table 1).

Table 1. Questions and data used for different levels of validity.

Type of Validity	Evaluation	Type of Participant	Question/Metric
Face Validity	Survey 1	Expert and Intermediate	Q1: The hand controllers on this simulator are effective for working in the simulated environment (Likert).
			Q4: The device is a sufficiently accurate representation of the real robotic system (Likert).
Content Validity	Survey 1	Expert	Q2: The 3D graphical exercises in the simulator are effective for teaching robotic skills (Likert).
			Q5: The scoring system effectively communicates my performance on the exercise (Likert).
			Q6: The scoring system effectively guides me to improve performance on the simulator (Likert).
Construct Validity	Simulator	Experts and Novices	Overall Score (points)
			Number of Errors (count)
			Time to Complete (seconds)
			Economy of Motion (centimeters)

Face Validity

The responses of Intermediate and Expert participants (n=68) were used to determine face validity (Table 2). A Chi-square test of independence was used to evaluate the distribution of scores for a specific simulator in relation to the order of the system's presentation to the subject. This analysis indicated that there was no difference in participants' answers according to the order in which the systems were presented; and established that no bias was present due to the presentation order ($p>0.05$). These questions asked participants to evaluate whether the hand controllers on the simulator were effective for working in the simulated environment (Question 1) and if the device is a sufficiently accurate representation of the real robotic system (Question 4). For both questions, the RoSS had the lowest average score, dV-Trainer had the second highest score, and the dVSS had the highest score of the three. A repeated measures ANOVA verified that the systems were scored differently for both questions ($p<0.001$).

Table 2. Average scores from a 5-point Likert scale on face validity.

	DVSS	dV-Trainer	RoSS
Q1: The hand controllers on this simulator are effective for working in the simulated environment.	4.80	3.62	2.17
Q4: The device is a sufficiently accurate representation of the real robotic system.	4.65	3.45	1.82

Content Validity

Expert (n=34) responses were used to determine whether the simulators were appropriate teaching modalities (Table 3). As seen in Table 3, 100% of participants either agreed or strongly agreed that the 3D graphical exercises in the dVSS were effective for teaching robotic skills while 59% disagreed or strongly disagreed that the RoSS' capabilities were effective. When asked if the scoring system effectively communicated their performance, 88% of dVSS users agreed or strongly agreed, while 79% of dV-Trainer users agreed or strongly agreed. Similarly, 91% and

82% of participants agreed or strongly agreed that the dVSS and dV-Trainer, respectively, effectively guided them to improve their performance, while only 36% felt the RoSS provided the same guidance.

Table 3. Scores on a 5 point Likert scale for content validity questions.

Likert Score	Strong Dis	Disagree	Neither	Agree	Strong Agree
Q2: *The 3D graphical exercises in the simulator are effective for teaching robotic skills.*					
DVSS	0%	0%	0%	35.3%	64.7%
dV-Trainer	2.9%	5.9%	11.8%	50.0%	29.4%
RoSS	20.6%	38.2%	17.6%	17.6%	5.9%
Q5: *The scoring system effectively communicates my performance on the exercise.*					
DVSS	2.9%	5.9%	2.9%	38.2%	50.0%
dV-Trainer	2.9%	2.9%	14.7%	55.9%	23.5%
RoSS	17.6%	20.6%	26.5%	29.4%	5.9%
Q6: *The scoring system effectively guides me to improve performance on the simulator.*					
DVSS	0%	0%	8.8%	61.8%	29.4%
dV-Trainer	2.9%	2.9%	11.8%	61.8%	20.6%
RoSS	18.2%	18.2%	27.3%	33.3%	3.0%

Construct Validity

The overall score, number of errors, time to complete, and economy of motion scores collected by the simulators for Experts (n=37) and Novices (n=37) were used to compare construct validity (Table 4). Overall score is a metric synthesized by multiple metrics and is specific to the individual simulator. Intermediate subjects were not included in the construct validity analysis because it was only necessary to look if the simulator could distinguish specifically between novice and expert users.

For the RoSS, the analysis has 23 missing data points because the system does not report scores when a user exceeds a maximum exercise time or chooses to terminate the exercise before completion. This resulted in a sample of 30 experts and 21 novices on that system. A Mann-Whitney U test showed that the distributions of time (p=0.221), number of errors (p=0.644), and economy of motion (p=0.566) were not statistically different for the experts compared to the novice group. The overall score metric is not automatically exported by the simulator and therefore was not analyzed for this sytem.

The dV-Trainer analysis of experts (n=37) and novices (n=37) had three missing values for economy of motion and completion time and five for the overall score metric, thus the analysis contained varying number of subjects. A Mann-Whitney U test showed that the distribution of the overall scores was not significantly different for the expert compared to the novice group (p=0.061). These tests did confirm statistical differences for economy of motion ($p<0.001$) and time to complete ($p<0.001$) for this system with a lower economy of motion value and shorter completion time for expert users compared to novices.

The dVSS analysis included all novice (n=37) and expert (n=37) participants. Using a Mann-Whitney U test, time to complete ($p<0.001$) and overall score (p=0.006) were significantly different for the expert compared to the novice group. The expert group had a higher score and a shorter completion time compared to the novice group. However, economy of motion did not show a statistical difference with this analysis (p=0.216).

Table 4. Mann-Whitney U test level of significance on construct validity measures

	DVSS	dV-Trainer	RoSS
Time to Complete	p<0.001	p<0.001	p=0.221
Overall Score	p<0.01	p=0.061	n/a
Economy of Motion	p=0.216	p<0.001	p=0.566
Number of Errors	n/a	n/a	p=0.644

The construct validity of the simulators was more specifically analyzed in terms of the self-reported number of cases of all participants (n=105) using a non-parametric correlation coefficient (Spearman's). For the RoSS, 30 participants were excluded from the analysis. For the participants that were included in the analysis (n=75), there was not a significant correlation between time to complete (p=0.181), number of errors (p=0.563), or economy of motion (p=0.390) with the total number of robotic cases performed.

For the dV-Trainer, four participants were excluded from the entire analysis and two participants were excluded from the overall score (Overall Score n=99; Economy of Motion and Time to Complete n=101). When analyzing the number of participants' robotic cases, there was a statistically significant correlation between overall score (p=0.03), economy of motion (p<0.01), and time to complete (p<0.01). The correlation value was negative for economy of motion and time to complete, showing that with a greater number of robotic cases, the time taken and distance moved decreased. The correlation was positive for overall score indicating that the participants' score increased with the number of robotic cases performed.

For the dVSS, two participants were excluded from the analysis (n=103). When analyzing the metrics in terms of the total number of robotic cases performed, there was a statistically significant difference between overall score (p =0.01) and time to complete (p <0.01). The correlation value was negative for time and positive for overall score, signifying that with more robotic cases the time taken decreased and the score increased. There was not a statistically significant correlation between economy of motion and the total number of robotic cases performed (p=0.105).

Table 5. Correlation between level of experience and simulator scores

	DVSS	dV-Trainer	RoSS
Overall Score	p=0.001	p=0.031	n/a
Time to Complete	p<0.001	p<0.001	p=0.181
Economy of Motion	p=0.105	p<0.001	p=0.390
Number of Errors	n/a	n/a	p=0.563

Usability (Preference)

The questions from the Survey 2 were used to understand the preference of the subjects when using the simulators. All subjects were included in this analysis except for two participants who were dropped from the analysis because they did not complete the questionnaire. The participant's responses to the usability questions can be seen in Figure 3:

- *If you are (were) a program director, which simulator would you choose for your trainees;*
- *In which simulator were you physically more comfortable;*
- *Which simulator had the best hand controls;*
- *Which simulator had the best foot controls;*
- *Which simulator had the best 3D vision;*
- *Were you feeling stressed or annoyed by any of the simulators?*

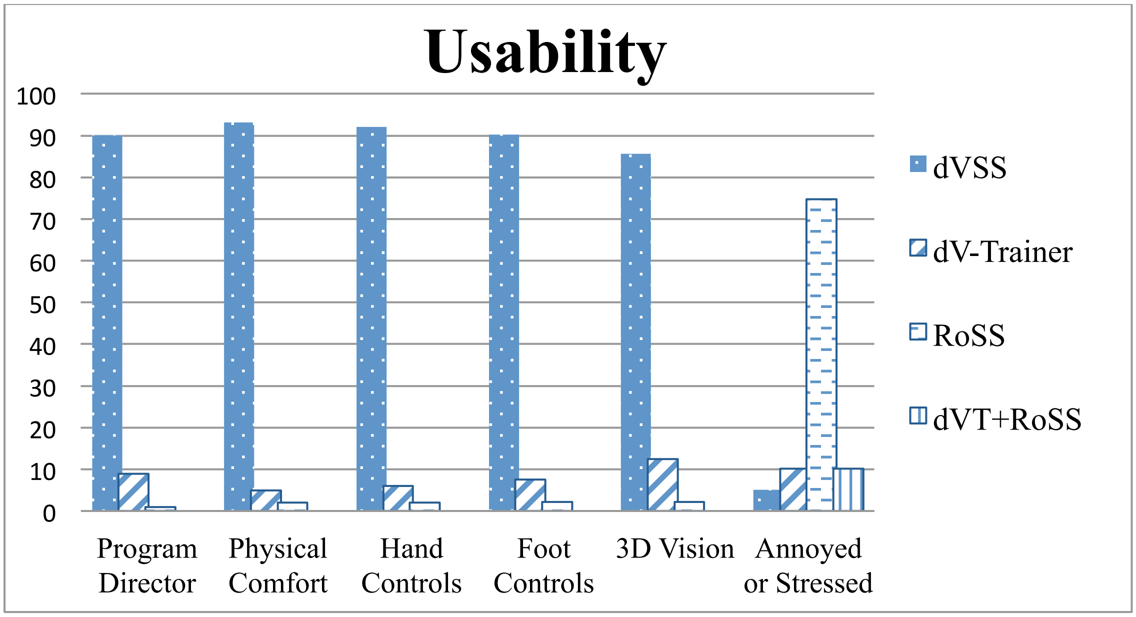

Figure 3. Description of usability responses

Overall, most participants preferred the dVSS and indicated that they would choose this device as a training system if they were a program director. Participants not only felt most comfortable in the dVSS, but also felt that the system had the best control and vision equipment. The least preferred system was the RoSS which most participants also agreed made them feel stressed or annoyed. Ten percent of participants also responded that they felt stressed or annoyed by both the dV-Trainer (dVT) and the RoSS.

Cost

All participants were also asked to provide feedback on their simulator preference in terms of the cost of the system. The responses were analyzed in terms of the frequency of the responses given. Most participants felt that the mimic dV-Trainer was worth the investment; while most felt that the RoSS was not worth the money. When asked about the dVSS, only 56% of participants agreed that it was worth the investment. Figure 4 provides a full description of the responses.

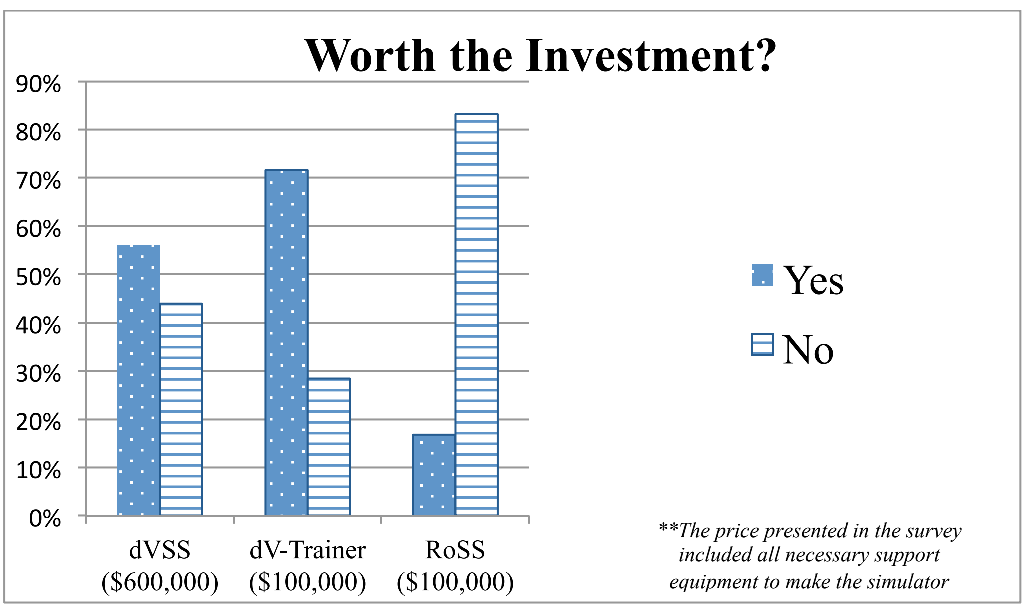

Figure 4. Description of cost preferences

DISCUSSION

The aim of this study was to conduct a comparison of the three commercially available simulators used to train surgeons on the daVinci robotic system. The study was performed for the US Army to assist them in making a purchasing and deployment decision regarding robotic simulators. Their interest is in re-training robotic surgeons who have been deployed to combat zones, where they have served as trauma surgeons for many months. Prior to resuming their robotic specialties, these surgeons need a program to both refresh and re-validate their robotic skills. This study provided information about the face, content, and construct validity as well as usability of the systems. The simulators were perceived to be different in their representation of the real robotic system. The dVSS was most preferred in terms of ergonomics and usability; however, most participants did not feel that this system was worth a $600,000 investment. In terms of cost, most participants agreed that the dV-Trainer had the best cost-effectiveness. The RoSS was the least preferred system for comfort and other usability aspects (i.e., hand controls, foot controls, and 3D interface), with most participants feeling stressed or annoyed when using the system. This study was unable to validate the face, content, or construct validity for this system.

The dVSS leverages the actual hardware used to perform robotic surgeries for use in the simulated environment, which allows for a more realistic experience, but decrease its availability and creates a higher cost for training than other robotic simulators. Economy of motion was not able to differentiate novices from experts in the dVSS, which could be attributed to the ease of use of the controllers allowing novices to move the controls as efficiently as experts. The generous workspace of the dVSS could also have an impact on the lack of difference. In contrast to the dVSS, the dV-Trainer is a standalone simulator and does not require the support of the daVinci hardware to operate. This allows for better accessibility and requires less of an investment for training. The overall score aspect of construct validity may not have shown a difference between novices and experts because of the way that the scoring is developed. The scoring system is constructed with a “ceiling” that prevents users from achieving a high overall score without attaining high scores across multiple metrics.

Currently, there is limited data available that confirms construct validity of the RoSS. Similarly to Raza (2013), this study was unable to confirm a difference between experts and novices in terms of time taken to complete the exercise. Time to complete, as well as economy of motion, is considered a highly relevant measurement of expertise levels for robotic surgeons (Perrenot, Perez, Tran, Jehl, Felblinger, Bresler, & Hubert, 2012). To our knowledge this three-part study is the first to compare all three available systems. This study involved the largest sample size and diversity of participants (i.e., experience levels, number of robotic cases, and subspecialty type) thus far in relevant publications. The lack of consistency in the available exercises and scoring systems across the three systems was a limitation to the study. Considerations for future research would be to use more complex exercises and increase the depth of the face and content validity evaluation.

Current research is focused on the effectiveness of the simulators and objectively measuring the transfer of training to the actual robotic system. All three simulators will be examined in this final stage of the experiment; however, the results of this three-part study will guide the choice of simulators used for future studies at Florida Hospital Nicholson Center and may also influence decisions at other laboratories. Also, this research may impact the purchasing decisions of customers for these devices.

REFERENCES

Cameron, J. L. (1997). William Stewart Halsted. Our surgical heritage. *Annals of surgery*, *225*(5), 445.

Colaco M, Balica A, Su D. (2012) Initial experiences with RoSS surgical simulator in residency training: a validity and model analysis, J Rob Surg, 7(1):71-75

Hung AJ, Patil MB, Zehnder P. (2012). Concurrent and predictive validation of a novel robotic surgery simulator: a prospective, randomized study. J Urology, 187(2):630-7.

Hung AJ, Zehnder P, Patil MB. (2011). Face, content and construct validity of a novel robotic surgery simulator. J Urology, 186(3):1019-24.

Kelly DC, Margules AC, Kundavaram CR. (2012). Face, content, and construct validation of the da Vinci Skills Simulator. J Urology, 79(5):1068-72.

Kenney P, Wszolek M, Gould J, Libertino J, Moinzadeh A. (2009). Face, Content, and Construct Validity of dV-Trainer, a Novel Virtual Reality Simulator for Robotic Surgery. J Urology, 73(6):1288-92.

Lee JY, Mucksavage P, Kerbl DC. (2012) Validation study of a virtual reality robotic simulator--role as an assessment tool? J Urology, 187(3):998-1002.

Liss MA, Abdelshehid C, Quach S. (2012). Validation, Correlation, and Comparison of the da Vinci Trainer(™) and the da Vinci Surgical Skills Simulator(™) Using the Mimic(™) Software for Urologic Robotic Surgical Education. J Endourol, 26(12):1629-34.

McDougall EM. (2007). Validation of surgical simulators. J Endourol, 21(3):244-7.

Sethi AS, Peine WJ, Mohammadi Y. (2009). Validation of a novel virtual reality robotic simulator. J Endourol, 23(3):503-8.

Perrenot, C., Perez, M., Tran, N., Jehl, J. P., Felblinger, J., Bresler, L., & Hubert, J. (2012). The virtual reality simulator dV-Trainer® is a valid assessment tool for robotic surgical skills. *Surgical endoscopy*, *26*(9), 2587-2593.

Raza SJ, Chowriappa A, Stegemann A, Ahmed K, Shi Y, Wilding G, Kaouk J, Peabody J, Menon M, Kesavadas T, Guru KA. (2013). Construct Validation of the Fundamental Skills of Robotic Surgery (FSRS) Curriculum. Podium presentation, AUA Annual Congress, San Diego,

Seixas-Mikelus SA, Kesavadas T, Srimathveeravalli G. (2010). Face validation of a novel robotic surgical simulator. J Urology, 76(2):357-60.

Smith, R., & Truong, M. (2013, January). Robotic Surgical Education with Virtual Simulators. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* (Vol. 2013, No. 1). National Training Systems Association.

Stegemann A, Kesavadas T, Rehman S, Sharif M, Rao M, duPont N, Shi Y, Wilding G, Hassett J, and Guru K. (2013). Development, Implementation, and Validation of a Simulation-Based Curriculum for Robot-Assisted Surgery. J Urology, 81(4):767-74.

Van Nortwick, S. S., Lendvay, T. S., Jensen, A. R., Wright, A. S., Horvath, K. D., & Kim, S. (2010). Methodologies for establishing validity in surgical simulation studies. *Surgery*, *147*(5), 622-630.