Validation of a Visual-Spatial Secondary Task to Assess Automaticity in Laparoscopic Skills

Richard Castillo, MD,* Juan Alvarado, MD,* Pablo Moreno, MD,* Pablo Billeke, MD, PhD,† Carlos Martínez, MD,* Julián Varas, MD,* and Nicolás Jarufe, MD*

*School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile; †División de Neurociencia, Centro de Investigacion en Complejidad Social, Universidad del Desarrollo, Concepción, Chile

INTRODUCTION: Our objective was to assess reliability and validity of a visual-spatial secondary task (VSST) as a method to measure automaticity on a basic simulated laparoscopic skill model. In motor skill acquisition, expertise is defined by automaticity. The highest level of performance with less cognitive and attentional resources characterizes this stage, allowing experts to perform multiple tasks. Conventional validated parameters as operative time, objective assessment skills scales (OSATS), and movement economy, are insufficient to distinguish if an individual has reached the more advanced learning phases, such as automaticity. There is literature about using a VSST as an attention indicator that correlates with the automaticity level.

METHODS: Novices with completed and approved Fundamentals of Laparoscopic Surgery course, and laparoscopy experts were enrolled for an experimental study and measured under dual tasks conditions. Each participant performed the test giving priority to the primary task while at the same time they responded to a VSST. The primary task consisted of 4 interrupted laparoscopic stitches (ILS) on a bench-model. The VSST was a screen that showed different patterns that the surgeon had to recognize and press a pedal while doing the stitches (PsychoPsy software, Python, MacOS). Novices were overtrained on ILS until they reach at least 100 repetitions and then were retested. Participants were video recorded and then assessed by 2 blinded evaluators who measured operative time and OSATS. These scores were considered indicators of quality for the primary task. The VSST performance was measured by the detectability index (DI), which is a ratio between correct and wrong detections. A reliable evaluation was defined as two measures of DI with less than 10% of difference, maintaining the cutoff scores for performance on the primary task (operative time <110 seg and OSATS >17 points).

RESULTS: Novices ($n = 11$) achieved reliable measure of the test after 2 (2-5) repetitions on the preassessment and 3.75 (2-5) on the postassessment ($p = 0.04$); whereas laparoscopy experts ($n = 4$) did it after 3.5 (3-4) repetitions. Proficiency cutoff scores for the primary task were achieved on every measure for novices (prepost overtraining) and experts. Expert performance on VSST was DI 0.78 (0.69-0.87). Novice performance was significantly better on postassessment (DI-pre 0.48 [0.06-0.71] vs DI-post 0.78 [0.48-0.95], $p = 0.003$). Overtraining consisted in 140 (100-210) repetitions of ILS for all novices, made in 8 hours (3-15). By categorizing DI based on expert performance, novices with DI-post >0.65 achieved better OSATS score and less operative time than novices with DI-post <0.65 ($p = 0.007$ y, $p = 0.089$, respectively).

CONCLUSION: Measuring automaticity is feasible using a VSST. This instrument is reliable and has a face, content and construct validity. A DI over 0.65 may be a cutoff point correlated with high standard performance on the primary task. This instrument measures performance on laparoscopic skills, and along with conventional indicators, would better define advance levels of expertise. More studies are required applying this VSST to achieve external validity by reproducing our results. (J Surg Ed 75:1001-1005. © 2017 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: automaticity, laparoscopy, simulation training, surgical education

This study was supported by Chilean Research Grant FONDECYT REG. 1141196 and by the Department of Digestive Surgery, School of Medicine, Pontificia Universidad Católica de Chile.

Correspondence: Inquiries to Richard Castillo, MD, UDA Hospital Dr. Sótero del Río. Avenida Concha y Toro 459, Puente Alto, Chile; e-mail: richardcr@me.com

1931-7204/$30.00 https://doi.org/10.1016/j.jsurg.2017.11.007

COMPETENCIES: Practice-Based Learning and Improvement, Professionalism

INTRODUCTION

Currently, laparoscopic techniques are a skill that must be considered in all general surgeon formation.[1,2] For their optimal training, it has been necessary to understand the process behind the learning of these abilities as well as to incorporate new teaching tools such as simulation.[3-7]

Regarding abilities acquisition, Ericsson proposes that expertise is specific to a particular task or domain. This could be achieved through the deliberated practice of a task associated with immediate feedback, which allows individuals to focus on training their weaknesses and, at the same time, improve other aspects of their performance.[4,5]

Meanwhile, Fitts and Posner's learning theory postulates that the acquisition of complex motor skills could be divided into 3 phases: cognitive, associative, and autonomous.[8,9] In the cognitive phase, there is development of basic movement patterns associated with a high attentional demand. In the associative phase, individuals link cognitive elements with motor maneuvers and sensory responses, so the movements become smoother, more efficient, and coordinated. Finally, in autonomous phase psychomotor movements become automatic with minimal demand of attentional resources, allowing performing additional tasks. Expert surgeons in laparoscopy complete their learning curve and reach this phase through exposition, constant training, and perfection of the technique, which can mean a considerable period of their career.[10]

In order to assess the performance and learning level of laparoscopic surgical skills, various indicators have been employed, such as time, error, global rating scales ("objective structured assessment of technical skills": OSATS),[11] kinetic metric (movement economy), psychophysiological (electromyography, electroencephalography), and mental workload measures. Combining these indicators has allowed to determinate the skill acquisition achieved in the earlier stages of laparoscopic surgical training, fostering the onset and development of simulated training models. However, these parameters are insufficient to distinguish if an individual has reached the more advanced learning phases, such as automaticity.[7]

One strategy used to assess automaticity in previous publications is through the performance in a secondary task.[7,12] To do this, the laparoscopic procedure is measured as the primary task with simultaneous measurement of a secondary task, which may consist of a cognitive, visual-spatial, or auditory stimulus. Expert or automatized individuals are able to respond to both tasks with a good performance, because they require less attentional resources to carry out the primary task, which allow them to redistribute the resources in order to accomplish the secondary task, as stated by the theory.[13,14]

The results of these experiments are promising, since through automaticity assessment it is possible to distinguish the level of skills between experts and already trained individuals, which should allow learning optimization and simulated training, and the transfer of skills to the operating room.[14] Nevertheless, the use of a visual-spatial secondary task (VSST) to measure automaticity has only been evaluated in a small group of individuals and its replication has not been reported in other settings yet.[15]

The aim of this study is to assess the reliability and validity of a VSST as a method to measure automaticity in a simulated laparoscopic suturing model in our simulation center.

METHODS

Study Design

An experimental study was conducted, for which novices and experts in laparoscopy were recruited. They participated in the experimental protocol at the Center for Simulation and Experimental Surgery of the Pontificia Universidad Católica de Chile. A sample size calculation was not performed, and the participant's number was defined according to their availability and to the simulation laboratory capacity. Novices were general surgery residents and medicine interns who have already completed and approved the FLS (Fundamentals of Laparoscopic Surgery)[16] course. Experts were surgeons who have performed more than 100 advanced laparoscopic procedures in the last 2 years prior this study.

Tasks Definition

Both novices and experts underwent an initial assessment of their performance on dual tasks. Each participant performed the test giving priority to the primary task while simultaneously responded to a VSST. The primary task, derived from an FLS exercise, consisted in 4 interrupted laparoscopic stitches (ILS) performed on a simulated synthetic box model (penrose, Fig). The VSST consisted in a visual-spatial stimulus presented through a computer screen located under the laparoscopic training box monitor. The stimulus consisted in a 4×4 cm gray square, which appeared over a blue background in randomized series at the center, left, or right of the screen. The stimulus duration was programmed in 300 milliseconds and the interstimulus interval in 1000 milliseconds. Participants had to respond to a defined pattern of successive appearance of the squares at the right side of the screen by pressing a pedal connected to the stimuli generator computer. Targets frequency corresponded to 10% of the generated stimuli, given in randomized intervals. For programming the VSST, the software PsychoPsy (Python, MacOS) was used.

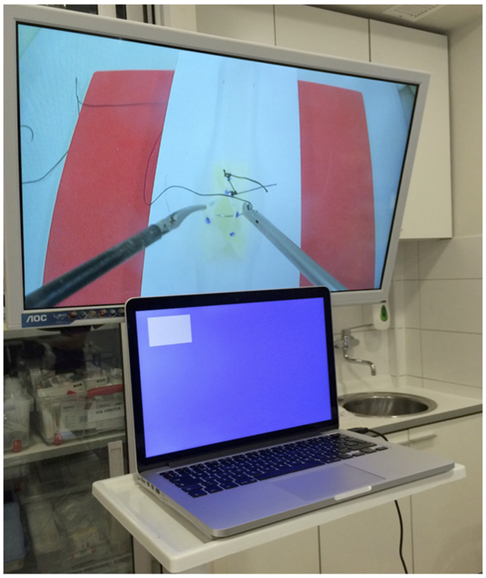

FIGURE. Synthetic box model consisted in 4 interrupted laparoscopic stitches (ILS) and visual-spatial secondary task (VSST).

Working Method

All novices were overtrained on ILS, by perfoming at least 100 ILS on the simulated synthetic box model (penrose). The number of repetitions of the ILS was considered by each novice until he felt able to take the test again, fulfilling a minimum of 3 sessions and 100 repetitions of the task. After completing the overtraining repetitions, they proceeded to the final assessment including the VSST.

Before the evaluation with dual tasks, participants were requested to perform the VSST alone, in order to ensure the proper understanding of the task.

Evaluations

Initial and final assessments were made for each participant, which were videotaped and analyzed by blinded evaluators. Time, OSATS, and detectability index (DI) were measured. Time and OSATS (a higher score, better performance) were considered as quality indicators of the primary task. The DI, used to assess the secondary task, corresponds to a ratio between the correct and wrong answers to the visual stimulus ([correctly answered targets/total targets]/[wrong answers/total of nontargets]).

A reliable measurement was defined at 2 evaluations with a DI difference of less than 10% between them, maintaining the cutoff scores performance in the primary task (time <110 s and OSATS $>$ 17 points). The 10% cutoff for a reliable measurement was determined according to neuroscience studies about significant changes.[17]

Primary outcomes were variations in time, OSATS and DI between the initial and final assessment. The number of ILS repetitions, number of sessions, and overtraining required hours were also considered. Finally, novice's subgroups were defined to further analysis, based on the final-DI obtained. The cutoff point was determined by the minimum DI values obtained by experts on the VSST.

Statistical Analysis

A SPSS 20.0 software (IBM) was performed for statistical analysis. Considering a nonparametric distribution of the data, they were expressed as median (minimum-maximum). Wilcoxon test for paired samples was used to compare scores before and after training, and the Mann Whitney test, to compare scores between different groups. A $p < 0.05$ was stablished as statistically significant.

RESULTS

In total, 11 novices and 4 experts completed their evaluations and overtraining between June and December 2016. Basal characteristics of the participants are summarized in Table 1.

To achieve a reliable measure in the initial assessment, novices required up to 5 measurements and experts 4. Moreover, in the novices group, a significant difference was observed in the number of measurements required to obtain a reliable DI between the initial and final assessment. Experts required a number of measurements comparable to the novices on their final assessment in order to achieve a reliable DI (Table 2).

The performance obtained by novices and experts in the primary task (time and OSATS) and in the VSST is summarized in Table 2. A significant change in performance of the VSST was detected in novices after overtraining, reaching a performance similar to the experts. The minimum performance required in primary task corresponded to FLS approval (defined as time <110 s and OSATS$>$ 17 points). This performance was met in all measurements of novices (initial and final) and experts. Novices met the minimum performance indicators in the primary task and initial assessment and then after overtraining, they improved up to the level of experts.

TABLE 1. Basal Characteristics of Participants

	Novices $n = 11$	**Experts** $n = 4$
Age (y)	28 (25-33)	39 (34-44)
Sex (male)	9	4
Forming stage		
Surgery resident	8	-
Medicine intern	3	-
Digestive surgeon	-	4
Bariatric surgeries performed	-	>100

Age expressed as median and range.

TABLE 2. Results of the Assessment of Novices and Experts Under Dual Tasks Conditions, Using a Visual-Spatial Secondary Task as Automaticity Indicator

Variables	A Novices (n = 11) Initial Evaluation	B Novices (n = 11) Final Evaluation	C Experts (n = 4)	AB p Value (*)	BC p Value (†)
Primary task	4 ILS	4 ILS	4 ILS		
Medium time ILS (s)	87.3 (80-102)	76.3 (55-99)	74.6 (71-81)	0.016	0.896
ILS OSATS	18.5 (18-23)	24 (19-25)	24.5 (23-25)	0.005	0.695
Detectability index	0.48 (0.06-0.71)	0.78 (0.48-0.95)	0.78 (0.69-0.87)	0.003	0.352
Assessments required‡	2 (2-5)	3.75 (2-5)	3.5 (3-4)	0.04	-

Values expressed as median and ranges.
*Wilcoxon rank-sum test.
†Mann-Whitney U test.
‡To obtain a reliable measure.

Novices overtraining was variable, with a median of 140 (100-210) ILS repetitions. These were performed in 4 (3-7) sessions of 8 (3-15) training hours.

In the subgroup analysis of novices categorized by final-DI, it was observed that novices with a final-DI ≥0.65 obtained a better OSATS score and required less operative time in their final assessment than novices with final-DI <0.65 (Table 3). Residents and interns distribution was comparable in both groups.

DISCUSSION

Using a VSST as indirect measure of automaticity is feasible. In our experience, this instrument proved to be reproducible and to have face, content, and construct validity. Its application allows distinguishing the performance of individuals on advanced training phases from experts, which was not possible with the preexisting indicators. Introducing the concept of automaticity in simulation laboratories would help to optimize the motor skills learning, maximizing the potential of basic simulated models before facing more complex ones.[7,15,18] In our view, automaticity must be done once the FLS is approved. The individual has the skills, but it is not automated. Overtraining leads to the automaticity of skills acquired in the FLS.

This is a quasiexperimental study and is not free of limitations. First, it has a small sample size, given the time restrictions of trainees and the resources needed to complete the training. Secondly, although the heterogeneity of the sample (interns and surgical residents) search to improve the applicability of the results, as interns are not exposed to surgical practice like the residents, it could also imply a source of bias that we believe was not detected in the subgroup analysis.

The number of performed repetitions of ILS for overtraining was variable (100-210). As part of the design of the study a minimum and maximum of repetitions were required in order to achieve an adequate performance in the VSST. The variability in the amount of overtraining could be explained by the different skills level of the participants at the beginning of the automation training.

From obtained results, variability was highlighted in obtaining a reliable DI measure between the novices' initial and final assessments, being this last one comparable to the one of the experts. We hypothesize that the number of times (the median) to obtain a reliable measurement in the group of experts and trained novices is greater given the requirement to exploit the technical skills. This situation demonstrates that using this instrument implies certain complexity and that strategies should be developed to deal with DI measurement variability.

This experience demonstrated that overtraining improves the performance in a VSST, when comparing the initial and final assessments. Those participants whose DI approached to the one of the experts (at or above the minimum DI of experts: 0.65), were considered to be at the automaticity phase. Moreover, obtaining a DI at or above 0.65 demonstrated to be a cutoff point correlated with a high performance standard in the primary task (ILS).

Although novices were already trained on ILS, with overtraining they were able to improve even more their

TABLE 3. Subgroup Analysis of Performance in the Final Evaluation of the Primary Task According to Detectability Index (DI)

Variables	Novices Final-DI <0.65 (n = 4)	Novices Final-DI ≥0.65 (n = 7)	p Value*
Medium ILS time (s)	83.4 (76-89)	70.4 (55-99)	0.089
ILS OSATS	21.75 (19-23)	24.25 (24-25)	0.007
Interns number	1	2	-
Residents number	3	5	-
Overtraining (ILS repetitions)	150 (100-160)	120 (100-210)	0.501

Values expressed as median and ranges.
*Mann-Whitney U test.

primary task indicators, getting closer to the experts' skills level.[19] This suggests that the skills level required for approving FLS might not be the optimal (preautomaticity phase). If so, it would be necessary to optimize skills before starting automaticity training. It is possible that this study has mixed optimization and automaticity phases, both improved with overtraining.

Since in the learning curve of an individual time and OSATS indicators improve before automaticity parameters, we believe it is necessary to define optimal time and OSATS standards in order to start automaticity training.[7]

In conclusion, this instrument is useful for evaluating laparoscopic skills, and together with conventional indicators,[20] could better define advanced expertise levels. Recognizing the automaticity phase in learning curves poses a new focus of simulated training, allowing an optimization of it, of the commonly used indicators and of the stablished goals for each stage of instruction. More studies are required to determine the best conditions to reach the automaticity phase and its limitations.

REFERENCES

1. Buia A, Stockhausen F, Hanisch E. Laparoscopic surgery: a qualified systematic review. *World J Methodol*. 2015;5(4):238-254.
2. Bashankaev B, Baido S, Wexner S. Review of available methods of simulation training to facilitate surgical education. *Surg Endosc*. 2011;25(1):28-35.
3. de Montbrun SL, Macrae H. Simulation in surgical education. *Clin Colon Rectal Surg*. 2012;25(3):156-165.
4. Ericsson KA, Prietula MJ, Cokely ET. The making of an expert. *Harv Bus Rev*. 2007;85(7-8):193 114–21.
5. Hashimoto DA, Sirimanna P, Gomez ED, et al. Deliberate practice enhances quality of laparoscopic surgical performance in a randomized controlled trial: from arrested development to expert performance. *Surg Endosc*. 2015;29(11):3154-3162.
6. van Merrienboer J, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ*. 2010;44(1):85-93.
7. Stefanidis D, Scerbo MW, Korndorffer JR, Scott DJ. Redefining simulator proficiency using automaticity theory. *Am J Surg*. 2007;193(4):502-506.
8. Fitts PM, Posner MI. Human Performance. Belmont, CA: Brooks/Cole; 1967.
9. Reznick RK, MacRae H. Teaching surgical skills changes in the wind. *N Engl J Med*. 2006;355(25):2664-2669.
10. Hopper AN, Jamison MH, Lewis WG. Learning curves in surgical practice. *Postgrad Med J*. 2007;83(986):777-779.
11. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273-278.
12. Stefanidis D, Scerbo MW, Montero PN, Acker CE, Smith WD. Simulator training to automaticity leads to improved skill transfer compared with traditional proficiency-based training: a randomized controlled trial. *Ann Surg*. 2012;255(1):30-37.
13. Logan GD. Automaticity, resources, and memory: theoretical controversies and practical implications. *Hum Factors*. 1988;30(5):583-598.
14. Shiffrin RM, Schneider W. Automatic and controlled processing revisited. *Psychol Rev*. 1984;91(2):269-276.
15. Poolton JM, Zhu FF, Malhotra N, Leung GK, Fan JK, Masters RS. Multitask training promotes automaticity of a fundamental laparoscopic skill without compromising the rate of skill learning. *Surg Endosc*. 2016;30(9):4011-4018.
16. Peters JH, Fried GM, Swanstrom LL, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery*. 2004;135(1):21-27.
17. Lachin JM. The role of measurement reliability in clinical trials. *Clin Trials*. 2004;1(6):553-566.
18. Zheng B, Cassera MA, Martinec DV, Spaun GO, Swanström LL. Measuring mental workload during the performance of advanced laparoscopic tasks. *Surg Endosc*. 2010;24(1):45-50.
19. Stefanidis D, Scerbo MW, Sechrist C, Mostafavi A, Heniford BT. Do novices display automaticity during simulator training? *Am J Surg*. 2008;195(2):210-213.
20. Schlachta CM, Mamazza J, Poulin EC. Evaluating laparoscopic skills. *Can J Surg*. 1999;42(6):470.