

# Sistema de evaluación de habilidades psicomotrices en cirujanos de laparoscopia

An Assessment of Psychomotor Skills in Laparoscopy

Sistema de avaliação de habilidades psicomotoras em cirurgiões de laparoscopia

Camilo Andrés Barrera Álvarez, Ing,<sup>1</sup> Juan David Salamanca Santos, Ing,<sup>1</sup> Eliana María González-Neira, MSc,<sup>1</sup>  
Daniel R. Suárez, PhD,<sup>1</sup> Saúl Rúgeles, MD<sup>2</sup>

Recibido: julio 1 de 2015 • Aprobado: enero 29 de 2016

Doi: <https://dx.doi.org/10.12804/revsalud14.especial.2016.04>

Para citar este artículo: Barrera CA, Salamanca JD, González-Neira EM, Suárez DR, Rúgeles S. Sistema de evaluación de habilidades psicomotrices en cirujanos de laparoscopia. Rev Cienc Salud 2016;14(Especial):57-67. doi: <https://dx.doi.org/10.12804/revsalud14.especial.2016.04>

## Resumen

**Objetivo:** este estudio tiene como propósito diseñar y validar un sistema de evaluación objetivo y práctico de habilidades psicomotrices en laparoscopia que permita discriminar la competencia de los cirujanos. **Materiales y métodos:** fue llevado a cabo en el Centro de Simulación Clínica de la Universidad Javeriana. Voluntariamente participaron seis cirujanos expertos, seis residentes de cirugía y seis estudiantes de Medicina. Cada participante hizo una evaluación compuesta por ocho actividades en simuladores de caja con instrumental regular de laparoscopia. Los resultados obtenidos en los tres grupos fueron analizados por medio de un Anova y un test de Tukey. Además, se hizo un análisis de validez por fuentes de evidencia y la fiabilidad del sistema de evaluación se valoró con alfa de Cronbach. **Resultados:** tres actividades evaluativas discriminaron los tres grupos de estudio o grados de competencia (novato, intermedio y experto) y otras dos actividades revisaron dos de los grupos. La validez del sistema de evaluación fue comprobada al considerar el contenido, la respuesta del proceso, la estructura interna y las consecuencias de la evaluación. La fiabilidad del sistema de evaluación fue sobresaliente (alfa de Cronbach=0,90). **Conclusión:** el sistema de evaluación propuesto es una herramienta viable en procesos de formación de cirujanos que permite complementar el método tradicional de evaluación.

**Palabras clave:** análisis y desempeño de tareas, laparoscopia, cirugía mínimamente invasiva, educación quirúrgica.

1 Departamento de Ingeniería Industrial, Facultad de Ingeniería, Pontificia Universidad Javeriana. Correo electrónico: d-suarez@javeriana.edu.co

2 Departamento de Cirugía General, Facultad de Medicina, Pontificia Universidad Javeriana.

### *Abstract*

*Objective:* this study aims at designing and validating an objective and practical assessment of psychomotor skills in laparoscopy. This assessment would allow a classification of the surgeons and residents by their psychomotor competence. *Materials and methods:* this study evaluated the psychomotor competence of three groups of voluntaries: experts (6 senior surgeons), residents (6 surgery residents), and novices (6 students of medical school). The assessment was performed at the Clinical Simulation Center in the Pontificia Universidad Javeriana. The assessment was designed as a composition of eight different activities to be performed by each of the participants in a box trainer and using minimally invasive surgical instruments. The results were analyzed using an Anova and a Tukey's test. Additionally, a validity test by evidence source and the assessment reliability was evaluated using a Cronbach's alpha analysis. *Results:* three of the eight assessment's activities correctly discriminated three study groups, and other two activities only discriminate two of the three groups. The validity of the assessment was positively evaluated using a content analysis, the response to the assessment, the internal structure and the possible future consequences of the assessment. The reliability of the assessment was considered excellent (Cronbach's alpha=0,90). *Conclusion:* the propose assessment is a good alternative in formative processes of surgeons and allows to complement the traditional assessment method in surgery education.

*Keywords:* task performance and analysis, laparoscopy, minimally invasive surgery, surgical education.

### *Resumo*

*Objetivo:* este estudo tem como objetivo desenhar e validar um sistema de avaliação objetiva e prática de habilidade psicomotoras em laparoscopia que permita discriminar o nível de competência dos cirurgiões. *Metodologia:* O estudo foi levado a cabo no Centro de Simulação Clínica da Universidad Javeriana onde voluntariamente participaram 6 cirurgiões expertos, 6 residentes de cirurgia e 6 estudantes de medicina. Cada participante realizou uma avaliação composta por oito atividades em simuladores de caixa com instrumental regular de laparoscopia. Os resultados obtidos pelos três grupos de participantes foram analisados através de um ANOVA e um teste de Tukey. Adicionalmente, realizou-se uma análise de validez por fontes de evidência e a fiabilidade do sistema de avaliação valorou-se empregando uma análise de Alpha de Cronbach. *Resultados:* três atividades avaliativas discriminaram os três grupos de estudo ou níveis de competência (novato, intermédio e experto), e outras duas atividades discriminaram dois dos grupos. A validez do sistema de avaliação foi comprovada ao considerar o conteúdo, a resposta do processo, a estrutura interna e as consequências da avaliação. A fiabilidade do sistema de avaliação foi sobressalente (Alpha de Cronbach=0,90). *Conclusão:* O sistema de avaliação proposto é uma ferramenta viável em processos de formação de cirurgiões que permite complementar o método tradicional de avaliação de médicos cirurgiões.

*Palavras-chave:* Laparoscopia, Cirurgia laparoscópica, Cirurgia minimamente invasiva, Avaliação, Treinamento, Simulação, Educação cirúrgica.

## Introducción

La cirugía mínimamente invasiva (CMI) ha tenido un gran auge debido a los avances tecnológicos de las últimas décadas (1, 2). La CMI brinda más beneficio al paciente por la reducción de tiempos de cirugía y de recuperación, cortes mínimos en los tejidos y un menor dolor, todo lo cual representa una disminución de riesgo y una mayor satisfacción del paciente (3, 4). Sin embargo, estos beneficios están ligados a la habilidad del médico, pues cuanta mayor competencia o habilidad del cirujano, menor riesgo para el paciente (5, 6).

La CMI, en especial la laparoscopia, requiere habilidades que no son comunes en la mayoría de las cirugías abiertas tradicionales, como la percepción de la profundidad a partir de una imagen plana, la destreza bimanual y el manejo de instrumentos y materiales en un espacio restringido (7, 8). Las condiciones en las que se realiza una CMI hacen que tareas habituales en cirugía como el desplazamiento del instrumento, la disección, el corte y la sutura de tejidos se vean muy afectadas por las habilidades mencionadas.

La falta de métodos de evaluación objetivos y estandarizados del desempeño de residentes de cirugía durante una práctica de CMI es un problema en la Medicina en general, pues no se puede determinar con claridad cuál es su grado de competencia y, en consecuencia, si son aptos para hacer los procedimientos en un paciente (9-11). Esto deja una brecha entre la academia y la práctica que puede afectar el resultado clínico de una intervención quirúrgica real, con graves consecuencias para la integridad del paciente (12, 13). Para disminuirla, se ha introducido el uso de nuevas tecnologías en el proceso de aprendizaje de residentes de cirugía y estudiantes de Medicina (*e-learning*, simuladores, tecnologías emergentes), que son empleadas para llevar a los estudiantes de novatos a expertos, ya sea que estas se usen como método de entrenamiento o de evaluación (14).

Los simuladores se dividen en aquellos que usan *software* computacional (simuladores de realidad virtual) y los que utilizan materiales de fácil acceso (simuladores físicos) (7). Los simuladores de realidad virtual ofrecen la posibilidad de evaluar tareas y parámetros con el uso de *software* sofisticado. Sin embargo, su costo puede oscilar entre US\$ 5.000 y US\$ 200.000, lo que puede ocasionar que las instituciones de educación no tengan el presupuesto para adquirir los equipos necesarios (15).

Se ha identificado la necesidad de un sistema de evaluación objetivo, confiable, de bajo costo y aplicable a la educación médica que permita determinar si un residente tiene o no las habilidades requeridas para la realización de una CMI (9, 10, 12). Este estudio tiene como fin diseñar y validar un sistema de evaluación objetivo, de bajo costo y aplicable en la formación de cirujanos para clasificar la competencia de los residentes de cirugía (novato, intermedio y experto). Se asume que una buena evaluación redundará, a largo plazo, en una reducción del riesgo que corren los pacientes que van a ser sometidos a este tipo de procedimientos.

## 1. Materiales y métodos

### 1.1. Participantes

En este estudio se contó con la participación de seis cirujanos expertos en Cirugía General del Hospital Universitario San Ignacio, seis residentes de Cirugía General que estuvieran entre segundo y quinto año de residencia y seis estudiantes de Medicina de octavo semestre de la Pontificia Universidad Javeriana (PUJ) sin entrenamiento previo en CMI. Esto equivale a un total de dieciocho voluntarios que aceptaron participar en el estudio, tras firmar un consentimiento informado como parte del protocolo experimental avalado por el Comité Institucional de Investigación y Ética.

### 1.2. Equipo e instrumental

El estudio fue llevado a cabo en los simuladores físicos de laparoscopia o *pelvitainer* que se encuentran en el Centro de Simulación Clínica de la Pontificia Universidad Javeriana (Colombia). Cada simulador contó con un set de instrumentos laparoscópicos reales (disectores Maryland, tijeras curvas, portaguñas y *grasper*).

Además, se emplearon dos cámaras GoPro Hero 3 (California, Estados Unidos) para grabar las actividades evaluativas desarrolladas por los participantes y así hacer varias evaluaciones posteriores para estimar la confiabilidad entre calificadores.

### 1.3. Actividades evaluativas

Las actividades analizadas durante el estudio fueron diseñadas y refinadas por un equipo de

ingenieros y cirujanos expertos, el cual garantizó que las mismas cumplieran los requerimientos técnicos y reprodujeran de buena manera la experiencia en CMI (incluso la percepción de la profundidad por medio de pantalla, la destreza bimanual, la autonomía y el manejo de instrumental en un espacio reducido). También se tuvo en cuenta la disponibilidad de los elementos, su facilidad de construcción y su replicabilidad. Las ocho actividades elegidas para ser parte del estudio son explicadas en la tabla 1.

Como parte del sistema de evaluación se diseñó un formato que abarca los elementos a evaluar de cada actividad (figura 1), con el fin de disminuir el sesgo en el momento de evaluar participantes diferentes o de revisar evaluaciones realizadas por examinadores diferentes.

Tabla 1. Actividades evaluativas diseñadas en este estudio

Habilidad	Actividad	Descripción
Desplazamiento	Desplazar cilindros	Desplazar con la mano dominante tres puntillas ubicadas en un tablero y acomodarlas en las marcas establecidas en el otro tablero. Seguir el mismo procedimiento con la mano no dominante.
	Cilindros y <i>stacks</i>	Desplazar los cilindros sin dejarlos caer desde el palo en el que se encuentran ubicados hasta el palo de la mitad del tablero con la mano dominante. Luego, desplazarlos al palo del otro extremo con la otra mano.
Corte	Malla	Cortar los cuatro hilos de color rojo que están en la malla sin cortar o reventar los hilos de los demás colores.
	Cortar látex	Cortar la figura demarcada por la línea guía en la lámina de látex.
Sutura	Ramo	Suturar los tres dedos de guante de cirugía, atravesándolos por la marca indicada. No es necesario anudar la sutura.
	Suturar látex	Suturar las dos láminas de látex por las marcas indicadas. No es necesario anudar la sutura.
Disección	Operar un globo	Cortar el globo por la línea guía y extraer las tres bolas de poliestireno expandido que están dentro de él.
	Mandarina	Separar un gajo de mandarina del resto sin causar daño al mismo o al resto de la mandarina.

Fuente: elaboración propia

Nombre			
Nivel de competencia		Fecha	
<b>Desplazar cilindros</b>			
Duración (minutos)	Puntillas y cilindros soltados	Puntillas mal ubicadas	
<b>Cilindros y stacks</b>			
Duración (minutos)	Cilindros soltados	Cilindros mal ubicados	
<b>Cortar látex</b>			
Duración (minutos)	Número de veces cortado por fuera de la guía		
<b>Malla</b>			
Duración (minutos)	Número de hilos cortados de color incorrecto		
<b>Ramo</b>			
Duración (minutos)	Precisión	Penalización	
<b>Suturar látex</b>			
Duración (minutos)	Precisión	Penalización	
<b>Operar un globo</b>			
Duración (minutos)	Precisión	Penalización	
<b>Mandarina</b>			
Duración (minutos)	Penalización		
<b>Observaciones</b>			

Figura 1. Formato de calificación usado en el estudio

Fuente: elaboración propia

#### 1.4. Análisis estadístico

Puesto que cada una de las actividades tiene elementos de calificación diferentes, fue necesario establecer una variable de respuesta estándar que fuera representativa del desempeño de cada uno de los participantes en cada una de

las actividades. Esta variable de respuesta es el tiempo de ejecución de la actividad más unas penalizaciones dadas por el número de errores o imprecisiones. La forma para calcular el puntaje final de cada actividad es la siguiente:

Una vez calculados los puntajes finales por actividad, se empleó un Anova (con una significancia de 95%) para determinar si existen diferencias significativas en el desempeño obtenido en cada actividad por cada grupo de interés. Adicionalmente, se empleó una prueba de Tukey para determinar los grados de experticia que se diferencian en cada actividad.

### 1.5. Análisis de fuentes de validez

El análisis se hizo con base en las fuentes de variación propuestas por Ghaderi y otros autores (16) que son el contenido, la respuesta del proceso, la estructura interna y las consecuencias de la evaluación de cada una de las pruebas que conforman el sistema de evaluación.

Para determinar la confiabilidad del sistema se revisó la equivalencia y la consistencia interna de las actividades que conforman todo el sistema de evaluación. Para la equivalencia, se empleó un test de confiabilidad entre calificadores. Para esto, se designaron dos calificadores que evaluaron al mismo grupo de ocho voluntarios durante la misma sesión. La calificación de los evaluadores se hizo en el mismo momento de aplicar la prueba o mediante una videograbación. Cada evaluador calificó de forma independiente sin conocer la otorgada por su par y así evitar sesgos en el estudio. La independencia en la evaluación es necesaria, ya que el objetivo de este test es determinar la variación de los desempeños obtenidos al ser calificados por diferentes examinadores. Los resultados fueron analizados por medio de coeficientes de correlación entre los dos puntajes obtenidos. La estructura interna del sistema de evaluación, que determina la reproductibilidad y la generalización de los resultados, fue calificada por medio del alfa de Cronbach.

## 2. Resultados

El Anova de los puntajes obtenidos por los participantes en cada prueba evidenció que de las ocho actividades propuestas, cinco son capaces de diferenciar al menos dos grados de competencia (tabla 2). A partir de la prueba de Tukey se clasificaron los grados de competencia en subgrupos homogéneos. De las cinco actividades que resultaron significativas con el análisis de varianza, tres de ellas (Cortar látex, Ramo y Operar un globo) permiten discriminar los tres grados de competencia deseados (novato, intermedio y experto) y dos, diferenciar solo dos grados de competencia (Desplazar cilindros y Suturar látex) (tabla 3).

### 2.1. Análisis de fuentes de validez

El marco actual del estado del arte define un estudio de validez como el proceso de recolección de evidencia que apoya la interpretación de los resultados obtenidos en una evaluación. Es por esto que el marco tradicional para tipos de validez (de contenido, criterio y construcción) ha sido reemplazado por un nuevo enfoque en el que la validez es un proceso de construcción con múltiples facetas y el proceso de validación de una herramienta requiere la identificación de "fuentes de validez" relevantes para las mismas facetas (16) (tabla 4).

## 3. Discusión

Se puede concluir que de las ocho actividades analizadas durante el estudio, tres permitieron diferenciar los tres grupos de estudio o los grados de competencia (novato, intermedio y experto), dos condujeron a diferenciar dos grupos o grados de competencia y las tres restantes no discriminaron grados de competencia.

Tabla 2. Resumen Anova obtenido en cada actividad evaluativa

		Suma de cuadrados	gl*	Media cuadrática	F†	Sig‡
Puntaje obtenido en la actividad Desplazar cilindros	Entre grupos	23,391	2	11,696	12,281	0,001
	Dentro de grupos	14,285	15	0,952		
	Total	37,676	17			
Puntaje obtenido en la actividad Cilindros y stacks	Entre grupos	16,586	2	8,293	0,806	0,465
	Dentro de grupos	154,426	15	10,295		
	Total	171,012	17			
Puntaje obtenido en la actividad Cortar látex	Entre grupos	58,257	2	29,128	14,092	0,000
	Dentro de grupos	31,005	15	2,067		
	Total	89,262	17			
Puntaje obtenido en la actividad Malla	Entre grupos	2,228	2	1,114	2,695	0,100
	Dentro de grupos	6,199	15	0,413		
	Total	8,427	17			
Puntaje obtenido en la actividad Ramo	Entre grupos	77,379	2	38,689	14,332	0,000
	Dentro de grupos	40,493	15	2,700		
	Total	117,872	17			
Puntaje obtenido en la actividad Suturar látex	Entre grupos	21,661	2	10,830	5,322	0,018
	Dentro de grupos	30,523	15	2,035		
	Total	52,184	17			
Puntaje obtenido en la actividad Operar un globo	Entre grupos	67,456	2	33,728	14,145	0,000
	Dentro de grupos	35,766	15	2,384		
	Total	103,222	17			
Puntaje obtenido en la actividad Mandarina	Entre grupos	,736	2	0,368	1,049	0,375
	Dentro de grupos	5,265	15	0,351		
	Total	6,001	17			

Fuente: elaboración propia

\* gl (grados de libertad): parámetro que define la distribución F, el cual se determina a partir de los tamaños muestrales involucrados (19).

† Valor del estadístico que sigue una distribución F de Fisher con los grados de libertad asociados.

‡ Significancia observada (también conocida como valor-p): es el área bajo la distribución de referencia más allá del valor estadístico de prueba (19).

Tabla 3. Resumen subconjuntos homogéneos obtenidos en la prueba Tukey

Puntaje obtenido en la actividad Desplazar cilindros				
Indica el grado de competencia del participante	N	Subconjunto para alfa=0,05		
		1	2	
HSD Tukey <sup>a</sup>	Experto	6	2,1883	
	Intermedio	6	2,2883	
	Novato	6		4,6550
	Sig.		0,983	1,000
Puntaje obtenido en la actividad Cortar látex				
Indica el grado de competencia del participante	N	Subconjunto para alfa=0,05		
		1	2	3
HSD Tukey <sup>a</sup>	Experto	6	3,4550	
	Intermedio	6		5,6683
	Novato	6		7,8617
	Sig.		1,000	1,000
Puntaje obtenido en la actividad Ramo				
Indica el grado de competencia del participante	N	Subconjunto para alfa=0,05		
		1	2	3
HSD Tukey <sup>a</sup>	Experto	6	2,9300	
	Intermedio	6		5,5200
	Novato	6		8,0083
	Sig.		1,000	1,000
Puntaje obtenido en la actividad Suturar látex				
Indica el grado de competencia del participante	N	Subconjunto para alfa=0,05		
		1	2	
HSD Tukey <sup>a</sup>	Experto	6	7,2467	
	Intermedio	6	8,3200	8,3200
	Novato	6		9,9167
	Sig.		0,415	0,162
Puntaje obtenido en la actividad Operar un globo				
Indica el grado de competencia del participante	N	Subconjunto para alfa=0,05		
		1	2	3
HSD Tukey <sup>a</sup>	Experto	6	2,2367	
	Intermedio	6		4,6450
	Novato	6		6,9783
	Sig.		1,000	1,000



Tabla 4. Análisis de fuentes de validez del estudio

Fuente de validez	Evidencia
Contenido	Para evidenciar que en esta fuente de validez se logra una buena relación entre el contenido de la prueba y lo que se busca medir, se tiene en cuenta que en el desarrollo de las actividades se siguió un proceso de diseño en conjunto con un grupo de investigación compuesto por médicos cirujanos expertos en laparoscopia e ingenieros conocedores del tema. Se logró una buena simulación física de lo que un cirujano puede encontrar mientras hace una laparoscopia.
Proceso de respuesta	Para asegurar la integridad de los datos obtenidos en el estudio y evidenciar que todas las fuentes de error asociadas con la aplicación de las pruebas fueron controladas o eliminadas en la mayor medida posible, se hizo una prueba piloto. Esta permitió determinar la aplicabilidad de las actividades del sistema y entrenar a los calificadores que iban a ser parte del estudio oficial para que conocieran y se familiarizaran con el formato único de evaluación. Además, algunos participantes (seleccionados aleatoriamente) fueron evaluados por dos calificadores diferentes (la calificación podía hacerse en vivo o por medio de videograbación).
Estructura interna	Esta es la fuente de evidencia más demostrada en estudios anteriores. Para mostrar esta fuente, se emplea la relación estadística de los datos obtenidos en el estudio. Es conocida como "confiabilidad" y alude a la reproducibilidad y generalización de los resultados (16). La mayoría de los estudios evalúa la estabilidad y la equivalencia del estudio con el uso de coeficientes de correlación intraclase y la consistencia interna, del alfa de Cronbach. El valor alfa obtenido en el estudio para las actividades que lograron diferenciar los tres grados de competencia deseados es de 0,905. Esto demuestra que la creación del sistema de evaluación con estas actividades presenta una muy buena consistencia interna. El coeficiente de correlación intraclase para medidas únicas alcanzó un valor de 0,761 que, a pesar de no ser muy bueno, está muy cercano al umbral de 0,8 para ser considerado como aceptable. En cuanto al test entre calificadores para las actividades Cortar látex, Ramo y Operar un globo fue bueno, ya que se obtuvieron valores alfa de 0,994, 0,997 y 0,982, respectivamente.
Consecuencias de la evaluación	El sistema de evaluación presenta una oportunidad de mejora a cualquier persona que sea evaluada, pues facilita una retroalimentación inmediata y acertada y da a conocer los puntos débiles y las habilidades técnicas en las cuales tiene que trabajar.

Fuente: elaboración propia

La alta variabilidad en el puntaje final que presentan los residentes en la actividad Suturar látex es la causa por la cual el grado intermedio no se puede diferenciar de los grados experto y novato, por lo que en esta actividad solo se aprecian dos grados de competencia: competente y no competente; los cirujanos expertos son "competentes" y los novatos, "no competentes".

En cuanto a la actividad Desplazar cilindros, solo se distinguen los dos grados de competencia extremos. En el estudio realizado por Chmarra, Klein, De Winter, Jansen y Dankelman, se hizo una actividad similar a Desplazar cilindros, la

cual discriminó los tres grados de competencia y clasificó correctamente a 16 de 31 participantes en el estudio (12). La diferencia entre ambos estudios es explicada por las variables de calificación tenidas en cuenta. En el estudio de Chmarra y otros autores, la evaluación de cada participante es realizada con el sistema TrEndo tracking system, que evalúa los movimientos de cada participante y no considera el tiempo ni la precisión obtenida en la actividad (12). El estudio de Cotin y otros autores también evaluó la actividad Desplazar cilindros (5). Los autores concluyeron que con ella se logra distinguir los grados de competencia novato y experto. Sin

embargo, al igual que en el estudio de Chmarrá y otros autores, se examinó el movimiento espacial que hace la punta de los instrumentos, la suavidad del movimiento, la percepción de la profundidad y la destreza bimanual.

Por otro lado, en las actividades Cilindros y *Stacks*, Malla y Mandarina no se discernió ningún grado de competencia por dos razones: a) el instrumento seleccionado por algunos de los voluntarios para la actividad, independientemente de su grupo, fue inapropiado, y b) estas actividades presentaron una dificultad muy alta respecto al tiempo previsto para su culminación, lo cual implicó que varios de los participantes, independientemente de su grupo, completaran las actividades.

Las actividades Desplazar cilindros y Cortar látex también se encuentran en el sistema de evaluación Mistels (17). En ese sistema, dichas actividades solo clasifican a los participantes como competentes y no competentes. El método de evaluación del presente estudio permitió la discriminación de tres grados de competencia en la actividad Cortar látex.

Por otra parte, se puede afirmar que se satisficieron las fuentes de validez de contenido, proceso de respuesta, consistencia interna y consecuencias de la evaluación al llevar a cabo procesos de validación por parte de expertos, capacitación de evaluadores, prueba piloto, video instructivo, formato de calificación, evaluación por medio de videograbación y análisis estadístico. El sistema de evaluación presentado obtuvo un alfa de Cronbach de 0,905 y coeficientes de correlación intraclase por encima de 0,9 y es comparable con otros métodos de evaluación más sofisticados. Estudios como el de Vassiliou y otros autores evalúan la estructura interna del sistema de evaluación Mistels, que alcanzó un buen alfa de Cronbach (0,86) y una excelente correlación intraclase para el test entre calificadores (0,998) (18). El sistema de

evaluación Osats fue examinado por Martin y Regehr y tuvo valores alfa de Cronbach entre 0,61 y 0,74 y valores de correlación intraclase entre 0,64 y 0,72, lo que indica que el Osats tiene una validez apenas aceptable (19). El sistema de evaluación propuesto es una herramienta viable en procesos de formación de cirujanos que permite complementar el método tradicional de evaluación.

Con base en estudios de diversos autores en la última década, se puede determinar que el desempeño de los estudiantes de CMI se incrementa cuando, en el proceso de aprendizaje, se incluyen la práctica y evaluación en simuladores con respecto a los estudiantes que siguen el método tradicional de aprendizaje (15). Si bien es cierto que los simuladores de realidad virtual (como LapSim, MIST-VR, GI Mentor II y URO Mentor, entre otros) ofrecen grandes beneficios por la capacidad y la complejidad del *software* que manejan, también es cierto que los simuladores físicos representan una alternativa viable para las instituciones de educación que no poseen un poder financiero fuerte y que, por lo tanto, no pueden adquirir equipos que oscilan entre los US\$ 5.000 y US\$ 200.000 (7, 15). Por lo tanto, se invita a desarrollar estudios posteriores para determinar si es equiparable el desempeño cuando se entrena y evalúa en simuladores físicos al obtenido con el uso de simuladores virtuales.

### *Agradecimientos*

Agradecemos especialmente al doctor Adalberto Amaya, director del Centro de Simulación Clínica, al igual que al personal que trabaja en el mismo.

### *Descargos de responsabilidad*

Los autores declaran que no tienen ningún conflicto de intereses y que la única fuente de financiación fue en especie por la Pontificia Universidad Javeriana y el Hospital Universitario San Ignacio.

## Referencias

1. Jiménez CP. Propuesta para el diseño de un sistema práctico de evaluación en laparoscopia [tesis de pregrado]. Bogotá: Repositorio Institucional, Pontificia Universidad Javeriana; 2013.
2. Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM. Fundamentals of Laparoscopic Surgery Simulator Training to Proficiency Improves Laparoscopic Performance in the Operating Room -A Randomized Controlled Trial. *Am J Surg* 2010 enero; 199 (1): 115-20.
3. Gómez-Ramírez MF, Gómez JC, González-Neira EM, Rúgeles S, Suárez DR, Torregrosa L. Evaluación del mejoramiento de habilidades básicas para cirugía laparoscópica por medio del entrenamiento con un videojuego. *Rev Cienc Salud* 2014; 12 (Especial): 9-20.
4. Gaar E. Errors in Laparoscopic Surgery. *J Surg Oncol* 2004 diciembre; 88 (3): 153-60.
5. Cotin S, Stylopoulos N, Ottensmeyer M, Neumann P, Rattner D, Dawson S. Metrics for Laparoscopic Skills Trainers: The Weakest Link! En: Takeyoshi D, Kikins R, editores. *Medical Image Computing and Computer-Assisted Intervention —MICCAI 2002*. Tokio: Springer Science & Business Media; 2002. 35-43.
6. Gumbs AA, Hogle NJ, Fowler DL. Evaluation of Resident Laparoscopic Performance Using Global Operative Assessment of Laparoscopic Skills. *J Am Coll Surg* 2007 febrero; 204 (2): 308-13.
7. Gravante G, Venditti D. A Systematic Review on Low-cost Box Models to Achieve Basic and Advanced Laparoscopic Skills During Modern Surgical Training. *Surg Laparosc Endosc Percutan Tech* 2013; 23 (2): 109-20.
8. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D et al. A Global Assessment Tool for Evaluation of Intraoperative Laparoscopic Skills. *Am J Surg* 2005 julio; 190 (1): 107-13.
9. Martin J, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C et al. Objective Structured Assessment of Technical Skill (OSATS) for Surgical Residents. *Br J Surg* 1997; 84 (2): 273-8.
10. Van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective Assessment of Technical Surgical Skills. *Br J Surg* 2010 julio; 97 (7): 972-87.
11. Zendejas B, Brydges R, Hamstra SJ, Cook DA. State of the Evidence on Simulation-based Training for Laparoscopic Surgery: A Systematic Review. *Ann Surg* 2013 abril; 257 (4): 586-93.
12. Chmarra MK, Klein S, De Winter JCF, Jansen F-W, Dankelman J. Objective Classification of Residents Based on their Psychomotor Laparoscopic Skills. *Surg Endosc* 2010 mayo; 24 (5): 1031-9.
13. Kneebone R. Simulation in Surgical Training: Educational Issues and Practical Implications. *Med Educ* 2003; 37 (3): 267-77.
14. Department of Health. A Framework for Technology Enhanced Learning. *J Surg Simul* 2014 abril; 1-21.
15. Sutherland LM, Middleton PE, Anthony A, Hamdorf J, Cregan P, Scott D et al. Surgical Simulation. *Ann Surg* 2006; 243 (3): 291-300.
16. Ghaderi I, Manji F, Park YS, Juul D, Ott M, Harris I et al. Technical Skills Assessment Toolbox: A Review Using the Unitary Framework of Validity. *Ann Surg* 2015 febrero; 261 (2): 251-62.
17. Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM. Evaluating Laparoscopic Skills: Setting The Pass/Fail Score for the Mistels System. *Surg Endosc* 2003 junio; 17 (6): 964-7.
18. Vassiliou MC, Ghitulescu GA, Feldman LS, Stanbridge D, Leffondré K, Sigman HH et al. The Mistels Program to Measure Technical Skill in Laparoscopic Surgery: Evidence for Reliability. *Surg Endosc* 2006 mayo; 20 (5): 744-7.
19. Gutiérrez H, De la Vara R. Análisis y diseño de experimentos. México D. E.: McGraw Hill; 2003.