



Recomendaciones generales para elaborar un programa de entrenamiento basado en simulación para desarrollar competencias en pregrado y postgrado

General recommendations for the development of a simulation-based training program for undergraduate and graduate competencies

Javier Vela,* Caterina Contreras,*‡ Cristián Jarry,*‡ Julián Varas,*‡ Marcia Corvetto*,§

Palabras clave:

Simulación, modelo, entrenamiento, validez, transferencia, programa.

Keywords:
Simulation, model, training, validity, transference, program.

RESUMEN

La simulación se define como una técnica que recrea elementos de eventos reales de forma controlada con un propósito educativo, evaluativo o de investigación. Su uso es cada vez más frecuente en el Área de Ciencias de la Salud, y se utiliza en los procesos de enseñanza-aprendizaje o en la evaluación o acreditación de los profesionales. Sin embargo, existe una gran heterogeneidad respecto al diseño y reporte de los resultados de diversos programas de simulación, lo que dificulta la comparación y recolección de evidencia científica. El presente artículo de reflexión busca revisar los principales aspectos que se deben considerar a la hora de desarrollar un programa de simulación con la esperanza de ayudar a mejorar la calidad de la evidencia y de la educación mediante la simulación clínica.

ABSTRACT

Simulation is defined as a technique that emulates real-life events in a controlled manner for educational, evaluation or research purposes. It has become increasingly common in health sciences, being used in teaching-learning processes or for evaluation or certification of health professionals. Nonetheless, design and data report quality vary greatly among simulation programs, making it difficult to establish comparisons and collect scientific evidence. This article reflects on the main aspects to consider when creating a simulation program, in the hope of helping to improve the quality of the evidence and education by means of clinical simulation.

INTRODUCCIÓN

La simulación fue definida por la Society for Simulation in Healthcare (SSH) como una técnica que crea una situación o un entorno que permite a las personas experimentar una representación de un evento real con el propósito de practicar, aprender, evaluar, probar o aumentar el conocimiento de sistemas o acciones humanas.¹ Por años se ha utilizado en el entrenamiento militar y en la aviación, lo cual ha logrado disminuir significativamente la tasa de error humano en estas áreas.^{2,3} El elevado número de errores evitables que están asociados con la atención en salud, junto con una menor exposición a los pacientes por parte de los profesionales de salud en formación y la creciente preocupación por la seguridad de los pacientes han fundado la necesidad de entrenar a futuros profesionales en un ambiente seguro y controlado, sin riesgos para los pacientes ni para los profesionales mismos.^{4,5}

En este contexto se han creado numerosos programas de simulación clínica que buscan enseñar y/o evaluar competencias en habilidades no técnicas o procedimentales de los estudiantes del área de la salud, tanto en actividades de pregrado como de postgrado.⁶⁻¹¹ Estos programas varían ampliamente en cuanto a su enfoque, a la competencia que buscan cubrir, al diseño y a la metodología de evaluación. A pesar de que, en forma global, la simulación pareciera ser útil para la educación en las ciencias de la salud, esta heterogeneidad hace que sea sumamente difícil comparar estudios y programas.¹²⁻¹⁴

El presente escrito es un artículo de reflexión con base en la literatura actual. El objetivo es revisar los elementos generales que los autores consideran fundamentales a la hora de crear un programa de simulación, y sirve como guía a la comunidad para ayudar a crear nuevos programas tanto en pregrado como en postgrado. El enfoque está puesto prin-

* Centro de Simulación y Cirugía Experimental.

† Departamento de Cirugía Digestiva.

‡ Departamento de Anestesiología.

Pontificia Universidad Católica de Chile.

Recibido: 11/11/2019

Aceptado: 17/03/2020

doi: 10.35366/92936



cipalmente en el entrenamiento de competencias en habilidades motoras; sin embargo, también se podría aplicar a escenarios que permitan evaluar o entrenar habilidades no técnicas y *debriefing*.

Revisaremos seis puntos a considerar en el diseño de un nuevo programa de simulación.

1. LEVANTAMIENTO DE LA NECESIDAD

La primera etapa consiste en la identificación de una habilidad, un set de competencias o los procedimientos que requieren ser entrenados. Este reconocimiento implica, en ocasiones, una situación clínica en la que se identifica por parte del equipo de profesionales una falencia en el desempeño de alguna tarea asistencial o docente.¹⁵ En otras ocasiones, la literatura reporta los beneficios asociados a algún entrenamiento, el cual puede resultar de interés para un determinado centro de atención en salud. Por ejemplo, el entrenamiento simulado de catéteres venosos centrales se asocia con una mejoría en los outcomes de estos pacientes.¹⁶ El *debriefing* posterior a algún evento crítico, o bien las reuniones clínicas de rutina pueden ser de utilidad para exponer aquellas competencias deficitarias que serían susceptibles de ser optimizadas mediante el entrenamiento simulado.¹⁷

No obstante, dado el impacto del entrenamiento simulado en múltiples currículos, la identificación de necesidades se ha convertido en un proceso

deliberado y cuidadosamente planeado. La aplicación de encuestas, entrevistas semiestructuradas, paneles Delphi o grupos focales es de gran utilidad para buscar dirigidamente aquellas competencias susceptibles de entrenar mediante simulación y que debieran ser comprendidas en un programa de formación.¹⁸ La participación conjunta de autoridades académicas, docentes, residentes, internos y profesionales clínicos debe ser considerada en este proceso, porque¹⁹ la percepción de necesidad varía según el rol de quien la refiere y porque los motivos que la justifiquen también serán variables y podrán abarcar fines académicos, asistenciales, de acreditación o incluso económicos.

Existen distintos tipos de competencias susceptibles de ser entrenadas a través de simulación; éstas se pueden clasificar según el modelo descrito por Miller.²⁰ Para entrenar los diferentes tipos de competencias se han descrito distintos tipos de simulación que se ajustan mejor para lograr cada objetivo. Lo anterior puede verse resumido en la *Tabla 1*.

2. MODELO DE SIMULACIÓN

Modelo de simulación y programa de simulación

Al seleccionar una competencia a entrenar se debe determinar el constructo subyacente. Se

Tabla 1: Tipos de metodologías de simulación.

Niveles	0	1	2	3	4	5
Técnica de Simulación	Simulaciones escritas	Simuladores de baja fidelidad, <i>part task trainers</i> y maniquíes básicos	Simuladores de pantallas computacionales, simulaciones virtuales y simuladores quirúrgicos	Pacientes estandarizados	Simuladores de fidelidad intermedia y maniquíes de tamaño real no totalmente interactivos	Simuladores de alta fidelidad y maniquíes de tamaño real totalmente interactivos
Habilidades que se logran	Cognitivas pasivas	Psicomotoras	Cognitivas interactivas	Psicomotoras, cognitivas e interpersonales	Parcialmente interactivas, psicomotoras, cognitivas e interpersonales	Interactivas, psicomotoras, cognitivas e interpersonales
Uso habitual	Manejo y diagnóstico de pacientes, evaluación	Práctica de habilidades	Manejo clínico de habilidades cognitivas	Igual que nivel 2; realización de examen físico, diagnóstico y manejo de pacientes	Igual que nivel 3; habilidades en procedimientos, entrenamiento de simulación <i>full-scale</i>	Igual que el nivel 4

Reproducido de: Corvetto M, Bravo MP, Montaña R, Utili F, Escudero E, Boza C, et al. Simulación en educación médica: una sinopsis. Rev Med Chil. 2013; 141 (1): 70-79. doi: 10.4067/S0034-98872013000100010.

Tabla 2: Conceptos (parte 1).	
Constructo	Colección intangible de conceptos y principios abstractos. ²¹ En simulación son elementos propios e inherentes de una determinada competencia que deben ser representados en el modelo de simulación o simulador
Modelo de simulación o simulador	Aparato físico o escenario en el que se lleva a cabo el entrenamiento
Programa de entrenamiento	Conjunto de ejercicios que componen el entrenamiento que se realiza en el simulador
Fidelidad funcional	Grado en que las funciones o características del modelo de simulación se asemejan a la realidad
Fidelidad estructural	Grado en que la apariencia estética o física del modelo de simulación se asemeja a la realidad

define el constructo desde el enfoque educacional como “una colección intangible de conceptos y principios abstractos”.²¹ En términos de la simulación, éstos vienen a ser aquellos elementos propios e inherentes de una determinada competencia, los cuales deben ser representados en el modelo de simulación o simulador.

Modelo de simulación y programa de simulación son términos distintos; el primero alude al aparato tangible o al escenario en el que se lleva a cabo el entrenamiento, y el segundo al conjunto de ejercicios que componen el entrenamiento que se realiza en el simulador. Como se verá más adelante, un simulador no tiene validez para entrenar una competencia si no se asocia a un programa de entrenamiento validado.²² Los conceptos importantes de este apartado se sintetizan en la *Tabla 2*.

Elección del modelo de simulación

Una vez identificada la competencia de interés y su constructo, el siguiente paso consiste en buscar un simulador o modelo de simulación y su respectivo programa de entrenamiento.

Se debe revisar en la literatura o el mercado que está disponible para posteriormente realizar una evaluación crítica. Esto conlleva, por ejemplo, considerar su costo o aplicabilidad en el medio local. Se debe evaluar si cuenta con algún tipo de evidencia o validez. Una situación frecuente es encontrar un simulador en el mercado sin un programa de entrenamiento validado y asociado. Este simulador aún puede ser útil, pues, mediante distintas mediciones, y como se describe más adelante, puede desarrollarse y validarse un programa.

En el caso de no existir un modelo o de tener un costo muy elevado, se puede pensar en innovar y desarrollar un simulador y su programa. En primer lugar, se hablará de cómo construir un modelo.

Construcción del modelo: ¿en qué enfocarse?

Como se expresó en el apartado anterior, el desarrollo de un simulador y su programa de simulación nace de una necesidad. Ésta debe asociarse a una competencia y esta competencia debe dividirse en los componentes que definen el constructo. Es útil plantear la siguiente pregunta: “¿qué es lo que se quiere enseñar en forma concreta?”. La respuesta es: los elementos claves de la tarea clínica o técnica que se busca enseñar a los alumnos.

La respuesta a esta pregunta se encuentra en distintas fuentes de información: se puede revisar la literatura publicada y estudiar el paso a paso de una tarea clínica o buscar experiencias previas publicadas sobre simulación en el área de interés. Otra fuente proviene de entrevistar a grupos de expertos que den su opinión sobre cuáles serían los pasos claves de un determinado procedimiento. En general, éstos se identifican como aquellos puntos en los que los sujetos en entrenamiento encuentran dificultades.¹⁵

Una vez identificado los pasos claves se puede acudir a un modelo constructivista para elaborar el simulador y su programa. En este cuadro de trabajo se elaboran modelos intermedios de dificultad progresiva o se asignan tareas de dificultad progresiva en un mismo modelo. Así, el alumno debe ir abordando las distintas etapas del procedimiento de manera progresiva y acumulativa, integrando el conocimiento adquirido en las etapas previas. En este modelo pedagógico se concibe el proceso de aprendizaje como aproximaciones sucesivas al “saber”, siendo el error algo inevitable y útil, pues es la base sobre la que se mejorará.^{23,24} Por ejemplo, si se plantea enseñar a un alumno a realizar una anastomosis intestinal, el modelo constructivista dividiría el procedimiento en sus etapas fundamentales: primero la práctica de puntos de afrontamiento, luego la enterotomía, el cierre de pared posterior y finalmente la pared anterior.⁷ Las etapas se van sumando de forma que, al realizar el cierre de la pared anterior, el alumno tendrá que haber realizado en una misma instancia todos los pasos previos. Así se logra realizar el procedimiento completo.

Como se mencionó, en ocasiones el modelo constructivista obliga a crear modelos de simula-

ción intermedios para exponer al alumno a etapas de dificultades crecientes. Se puede aprovechar esto para crear modelos costo-efectivos, por ejemplo, anticipando qué etapas requerirán modelos de alto recambio para fabricarlos con materiales de menor costo. Se reservan así los modelos de mayor costo para las etapas finales del entrenamiento. Un ejemplo: para practicar anastomosis vasculares, se puede primero practicar sobre modelos sintéticos (de bajo costo y alto recambio) y, en las etapas finales, sobre un vaso de tejido *ex vivo* o *in vivo* (de alto costo y bajo recambio).

Fidelidad del modelo

Se entiende generalmente el concepto de fidelidad de un simulador como el grado en que éste se ve, se siente y actúa como un escenario real (por ejemplo, un paciente humano). Sin embargo, este concepto enfatiza los avances tecnológicos y la similitud física con la que sólo se construye más que la capacidad de lograr un aprendizaje efectivo.²⁵

Un concepto más útil a la hora de elaborar el modelo es dividir a la fidelidad en “fidelidad estructural”, la cual hace referencia a la *apariencia física del simulador* y en “fidelidad funcional”, la cual se refiere a *qué hace el simulador*.²⁶ La fidelidad funcional debe enfocarse en representar de la mejor manera posible aquellos elementos claves que están asociados con el constructo que se identificó al analizar la tarea a simular (ver apartado anterior).

En la elaboración del simulador se debe privilegiar su fidelidad funcional; un ejemplo de esto es el de Matsumoto y colaboradores (2002), quienes usaron un vaso y dos bombillas para entrenar las habilidades de ureteroscopia y lo compararon con un simulador de alta fidelidad estética disponible en el mercado sin encontrar diferencias significativas en el aprendizaje de dos grupos de estudiantes que usaron los distintos simuladores.²⁷

Por el contrario, la priorización de la fidelidad estructural redundó en un aumento de los costos del modelo y limita su replicabilidad sin tener mejores resultados en el aprendizaje.^{27,28}

Es importante enfatizar que lo anterior no quiere decir que no deben valorarse los aspectos físicos de los simuladores, pues existen situaciones en las que éstos son fundamentales para representar determinada labor clínica, por ejemplo, es fundamental la buena representación de los tejidos cuando se quiere enseñar un procedi-

miento quirúrgico. Los aspectos físicos cobran relevancia cuando éstos están en línea con la fidelidad funcional.

Materiales y costos

En este acápite juega un rol fundamental la creatividad del investigador y su accesibilidad a actividades multidisciplinarias con equipos de diseño, ingeniería, etcétera. Con el desarrollo de la tecnología, la oferta de materiales para construir un simulador es amplia y el investigador debe realizar un catastro sobre qué elementos tiene a su disposición en su institución. Al momento de elegir se deben privilegiar aquellos materiales que contribuyan a una mayor fidelidad funcional y se debe considerar tanto la durabilidad (cantidad de usos) como el costo de éstos. Los materiales de baja durabilidad y de alto costo encarecen el modelo y limitan su replicabilidad; por el contrario, los materiales de alta durabilidad y bajos costos generan un simulador replicable, que, de ser efectivo para entrenar determinada labor clínica, puede generar impacto en los programas curriculares.

Los costos relacionados con el desarrollo de un modelo simulado pueden dividirse en distintas categorías: materiales y equipamiento, costos de personal capacitados, costos de sedes, entre otros.²⁹ Ésos deben ser considerados, pues tanto los costos como los materiales deben ser reportados al momento de publicar un modelo simulado. Esto garantiza su comparabilidad con otros simuladores existentes y evalúa su costo-efectividad en el aprendizaje.³⁰

Iteración y evaluación del simulador

El proceso de elaboración de un simulador y su programa de entrenamiento es dinámico e iterativo. Se debe pilotear por los expertos, entrenadores y sujetos a entrenar con el fin de obtener datos para mejorar la fidelidad funcional. Se espera que, a mayor fidelidad funcional, mejore la transferencia del aprendizaje adquirido al ambiente clínico en el modelo.

3. HERRAMIENTAS DE EVALUACIÓN

Una vez obtenido el modelo debe contarse con herramientas de evaluación para medir el aprendizaje durante el programa de entrenamiento. A continuación nos referiremos a los aspectos

esenciales de estas herramientas. Los conceptos esenciales de este apartado se encuentran resumidos en la *Tabla 3*.

Validez y confiabilidad

Establecer la confiabilidad y validez de las herramientas de evaluación es un paso fundamental antes de empezar a considerar la efectividad de un programa de entrenamiento con un modelo dado.³¹ Lo contrario puede inducir a errores de interpretación de los datos obtenidos con las herramientas. Con frecuencia los métodos utilizados para medir confiabilidad y validez son heterogéneos. Esto dificulta la comparación entre los estudios realizados y va en detrimento del desarrollo de evidencia científica de los programas de simulación y su impacto a nivel individual, a nivel del paciente y a nivel institucional.

Confiabilidad y validez son conceptos que se encuentran frecuentemente en la literatura de educación en simulación (*reliability* y *validity* en inglés, respectivamente). Sin embargo, su definición ha sido inconsistente en el tiempo.³¹ Beckman y su equipo (2004) realizaron una revisión de publicaciones relacionadas con la evaluación de educación clínica y encontraron que confiabilidad y validez eran mal interpretados con frecuencia, siendo usados en contextos inadecuados.³²

Validez

Estos conceptos tienen su origen en teorías de psicología educacional.^{33,34} Validez hace referencia

al grado en el que las conclusiones o interpretaciones derivadas de los resultados de una evaluación son plausibles o justificables.²² En otras palabras, es la capacidad de medir lo que se quiere evaluar con cierto grado de confianza.³¹ Los resultados de cualquier evaluación psicométrica sólo tienen validez en el contexto y para el constructo que han sido diseñados.²² Dado esto, la definición del constructo y el contexto en que se empleará la evaluación debe ser el primer paso de cualquier proceso de validación.

Nótese que, de momento, sólo se ha mencionado el concepto de validez en torno a herramientas de evaluación. Tanto validez como confiabilidad —a la que se hará referencia más adelante— son conceptos propios de los instrumentos de evaluación en su contexto (no como herramienta aislada), de los puntajes obtenidos y de su interpretación.^{22,31} Ni los modelos para simulación ni una intervención educacional aislada pueden ser validados, y el uso de estos términos en ese contexto corresponde a un mal uso de los conceptos.

En la literatura se hace referencia a múltiples tipos de validez.^{22,31} Clásicamente se ha clasificado en validez de contenido, de constructo y de criterio.^{22,35} Esta última engloba la validez correlacional, concurrente y predictiva. En este artículo se hará referencia a tres: de constructo, de concurrencia y predictiva por considerarse las más comúnmente usadas. Sin embargo, en el último tiempo ha habido una tendencia a entender la validez como un concepto más holístico y práctico, uniendo toda la evidencia posible en pos de argumentar a favor o en contra de una hipótesis de validez.^{21,35,36}

La validez de constructo se define como el grado en el cual los procedimientos evaluativos son capaces de identificar las cualidades, habilidades o rasgos para los que fueron diseñados.³⁷

Se evalúa frecuentemente al comparar el desempeño de un experto (por ejemplo, médico especialista) con un novato (por ejemplo, el residente de primer año o los alumnos de pregrado). Lo esperable es que antes de aplicar el programa de entrenamiento los expertos tengan un desempeño significativamente superior al de los novatos;³¹ sin embargo, ha de tenerse cuidado de no usar esta evaluación como única forma de validez.³⁸ Existen múltiples factores de confusión que pueden explicar las diferencias observadas entre grupos, sobre todo cuando estos grupos son sistemáticamente distintos a aquéllos en los que se realizará la intervención formal (novatos versus

Tabla 3: Conceptos (parte 2).

Validez	Grado en el que las conclusiones o interpretaciones derivadas de los resultados de una evaluación son plausibles o justificables ²²
Validez de constructo	Grado en el cual los procedimientos evaluativos son capaces de identificar las cualidades, habilidades o rasgos para los que fueron diseñados ³⁷
Validez concurrente	Existencia de correlación entre distintas formas de evaluar un mismo constructo ³¹
Confiabilidad	Posibilidad de confirmar un comportamiento predicho. En un programa de entrenamiento, hace referencia a la transferencia de las competencias desde el modelo inicial hacia uno más complejo o una situación real
Confiabilidad	Reproducibilidad o consistencia de los datos o mediciones obtenidas al utilizar repetidamente una misma forma de evaluación en un sujeto ³⁹

expertos en vez de un grupo de similar nivel de entrenamiento, pero con habilidades variables y desconocidas).³⁸ Que un instrumento sea capaz de detectar diferencias entre grupos no significa que necesariamente esté midiendo el parámetro deseado. Por este motivo, la capacidad de discriminación es necesaria, pero no suficiente para determinar la existencia de validez, ni debiese ocuparse como argumento aislado.³⁸

Messick hace referencia a cinco fuentes de evidencia que apoyan la presencia de validez de constructo:³³

- Contenido: la relación entre el contenido de una evaluación y el constructo para el que fue diseñado.³⁹ Generalmente se hace una descripción detallada de los métodos utilizados para garantizar que los elementos representen adecuadamente al constructo.
- Proceso de respuesta: revisión de las acciones y razonamiento de los evaluados. Medible mediante solicitud de “pensar en voz alta” o mediante métodos que avalen la asignación de puntajes específicos.^{21,39}
- Estructura interna: relacionado con confiabilidad y análisis de los datos.⁴⁰ La confirmación de diferencias entre grupos apoya este punto, mientras que lo contrario refleja una falla estructural del método evaluativo.^{22,39}
- Relación con otras variables: correlación con otros instrumentos diseñados para evaluar el mismo constructo y/o la ausencia de correlación en elementos no asociados.^{34,41}
- Consecuencias: presencia de consecuencias deseadas y ausencia de aquéllas no deseadas. Los métodos utilizados para determinar los puntos de corte que diseñen los intervalos o el puntaje de aprobación/reprobación también caen dentro de esta categoría.²²

La evaluación de una o más de estas fuentes dará peso a la validación de constructo.

Validez concurrente se refiere a la existencia de correlación entre distintas formas de evaluar un mismo constructo.³¹ Esto se puede realizar comparando una nueva metodología de evaluación con un estándar de oro. Por ejemplo, para las habilidades técnicas suele hacerse la comparación contra la escala OSATS (*Objective Structured Assessment of Technical Skills*).⁴² También se pueden comparar distintos modelos de simulación que buscan evaluar una misma destreza, utilizando

una herramienta de evaluación común. En 2001, Grantcharov y otros realizaron un estudio de validez concurrente, evidenciando que 14 residentes de cirugía obtuvieron puntajes similares en cuanto a desempeño, errores y economía de movimiento al realizar tareas de laparoscopia avanzada en un modelo animal porcino y en un modelo de realidad virtual.⁴³

La validez predictiva se relaciona con la posibilidad de confirmar un comportamiento predicho. Esto se refiere particularmente a la transferencia de las competencias desde el modelo inicial hacia uno más complejo o una situación real. Por ejemplo, un estudio realizado por Boza y colaboradores sometió a residentes de cirugía general de primer año a un programa de entrenamiento en laparoscopia avanzada de diecisésis sesiones, el cual estaba basado en simulación.⁴⁴ Finalizado este entrenamiento, se evaluó el desempeño en pabellón donde debían realizar una yeyuno-yeyunostomía mecánica. Estos residentes no sólo demostraron transferencia de habilidades a una situación real, sino que además su desempeño, medido por medio de escala de evaluación global (OSATS) y específica, fue superior al de cirujanos generales no entrenados con simulación.⁴⁴

Adicionalmente, a menudo en la literatura se hace referencia al concepto de validez de aspecto, más conocido por su nombre en inglés: *face validity*. Expertos evalúan los contenidos de una herramienta de evaluación o el aspecto de un nuevo modelo para simulación y emiten un juicio sobre si les parece o no apropiado. Cook y Beckman (2006) se refieren a este concepto como la apariencia de validez en ausencia de evaluación empírica;²² sin embargo, existe una brecha entre apariencia y realidad. Por tanto, se desaconseja su uso.^{22,45}

Confiabilidad

La confiabilidad se refiere a la reproducibilidad o consistencia de los datos o mediciones obtenidas al utilizar repetidamente una misma forma de evaluación en un sujeto.³⁹ Se puede evaluar de distintas formas, tales como:^{22,46,47}

- Consistencia interna. Cuando todos los elementos de la evaluación miden un mismo constructo. El desempeño del evaluado es relativamente consistente a lo largo de la evaluación.

- Estabilidad temporal “test-retest”. Consistencia de los datos obtenidos cuando se aplica la misma evaluación una segunda vez en el mismo sujeto.
- Consistencia interobservador. Correlación existente entre los puntajes obtenidos cuando distintos evaluadores emplean una misma pauta.
- Generabilidad. Medible mediante un coeficiente de generabilidad, el cual excede los objetivos de este artículo. Busca estimar las fuentes de posible error y su peso relativo.

Es necesaria que exista para determinar validez, pero no es suficiente por sí misma.²²

Con frecuencia las publicaciones omiten información sobre la confiabilidad de los instrumentos usados y la validez que se busca evaluar. Esto dificulta enormemente la comprensión del trabajo como tal y la comparación con otros programas nacionales o internacionales. Además, hace dudar al lector acerca del rigor del trabajo e idoneidad de la metodología utilizada. Así, la determinación de estos elementos constituye una fase crítica en el diseño de un programa de simulación.

Selección de las herramientas de evaluación

La gama de herramientas de evaluación es amplia. Se encuentra desde la observación directa hasta las mediciones objetivas mediante instrumentos electrónicos. La elección de la herramienta debe estar en directa relación con el constructo que se planea medir (*Tabla 1*).

Al momento de elegir un instrumento de evaluación se presentan dos posibilidades: crear una *de novo* o utilizar una previamente publicada. Estas últimas suelen estar parcialmente validadas o no suelen ajustarse perfectamente a nuestras necesidades. La opción preferible suele ser una mezcla de ambas alternativas: basarse en una herramienta previamente desarrollada y parcialmente validada, realizarle modificaciones y llevar una investigación que aporte nuevos elementos a su argumento de validez.³⁵

Un error común al momento de seleccionar y validar una herramienta de evaluación es “reinventar la rueda”. No es aconsejable siempre crear un instrumento nuevo cada vez que se pretenda medir algo. Usar o mejorar uno previamente existente tiene ventajas: evita el problema de crear algo *de novo* y permite comparar los resultados con otros trabajos que usen la misma escala.

Además, permite que la evidencia de validez de la escala que se obtenga de la investigación que se esté desarrollando se agregue a la existente para ese instrumento.

Al decidir modificar una herramienta de evaluación preexistente se debe considerar lo siguiente: la validez de un instrumento no es una característica propia de él, sino más bien de las interpretaciones que se obtienen de los resultados.⁴⁸ Esto quiere decir que un instrumento está validado para su uso en cierto contexto y, cuando éste se cambia, se debe valorar la validez de las interpretaciones.

Al modificar un instrumento de evaluación se debe analizar si la evidencia disponible sustenta los cambios, pues la modificación puede invalidar los supuestos en que se apoyaba.

Para evaluar competencias técnicas suele usarse las escalas Likert, escalas dicotómicas o binarias, evaluaciones contra un referente (*Benchmark Assessment*), entre otras. Las escalas Likert se emplean como una escala de evaluación global (GRS o *Global Rating Scale*, por sus siglas en inglés; éstas evalúan el desempeño general durante el entrenamiento) o como una escala específica (en la que los ítems son específicos de la competencia a entrenar). Las escalas dicotómicas también pueden evaluar aspectos generales o específicos, asumiendo forma de lista de chequeo o *checklist*.⁴⁹ Existe debate sobre las características comparadas entre las pautas tipo *checklist* y las Likert. Una revisión sistemática encontró que el *checklist* es en apariencia más objetivo y tiene una confiabilidad interevaluador mayor, pero debe reelaborarse para distintas situaciones. En cambio las GRS pueden ser usadas en distintos escenarios (como el OSATS) y captan de mejor manera distintos niveles de habilidades.⁵⁰ Herramientas más específicas se han desarrollado caso a caso y debe revisarse con cuidado la evidencia que sustenta su uso.⁵¹

En paralelo con la creación de escalas, se han desarrollado herramientas basadas en tecnología que pueden apoyar al argumento de validez. A continuación se mencionan algunos ejemplos. El *Imperial College Surgical Assessment Device* (ICSAD) es un aparato capaz de captar el movimiento de las manos durante algún procedimiento. Sirve para cuantificar de manera objetiva la economía de los movimientos y está validado para cirugía laparoscópica y otros procedimientos.⁵² Otros ejemplos son los dispositivos capaces de rastrear el movimiento ocular y medir la fuerza

ejercida en una superficie (usado por ejemplo en simulación de examen físico mamario).⁵³ En síntesis, las nuevas tecnologías permiten obtener mediciones más objetivas y reproducibles, por lo que el investigador debe estar dispuesto a usarlas, de estar éstas disponibles.

Frecuentemente la evidencia disponible no sustentará a la perfección la validez del uso del instrumento, por lo que se debe llevar a cabo un juicio si la evidencia es al menos suficiente y adecuada. De existir vacíos en la evidencia, éstos pueden llevarse a una investigación asociada al uso de la herramienta.

4. PROGRAMA DE ENTRENAMIENTO

Elaboración de un programa de entrenamiento

Un programa de entrenamiento es mucho más que simplemente contar con el espacio, infraestructura y el equipamiento necesario para realizar una simulación (fantoma, simulador de laparoscopia, etcétera).

Un programa de entrenamiento simulado ideal debería:

- Ser accesible por el aprendiz en cualquier parte, en cualquier momento y no sólo ciertas fechas al año (ser como un “gimnasio” de entrenamiento).⁵⁴
- Idealmente haberse demostrado su efectividad de forma científica (transferencia de habilidades a escenario real, sala operatoria).^{7,55}
- Ser costo-efectivo.
- Tener instructores siempre disponibles para dar retroalimentación (*feedback*).
- Ser fácil de replicar en cualquier lugar del mundo.
- Permitir entrenar técnicas avanzadas y no sólo básicas.

Para efectos prácticos se aplicará un nuevo concepto para entender “entrenamiento básico”; si las habilidades obtenidas con un programa de entrenamiento son consideradas “la introducción a un paciente”, es decir, si en la simulación, la dificultad técnica presentada resulta más fácil que la intervención real en paciente, entonces se hablará de un entrenamiento básico. Este concepto sólo puede ser determinado por varios expertos clínicos probando la simulación. Esto no significa que la habilidad que se intenta aprender

sea básica desde un punto de vista psicomotor (por ejemplo, la laparoscopia básica es compleja, pero comparada con operar a un paciente es técnicamente menos demandante).

En otras palabras, se recomienda no sólo contar con niveles básicos de entrenamientos, sino también avanzados, con el objetivo de optimizar la transferencia de habilidades a un escenario real. Resulta lógico pensar que, si un entrenamiento presenta cierto grado de fidelidad y es avanzado con un nivel de dificultad técnicamente superior a un escenario clínico determinado, el alumno presentará mayor facilidad cuando haga la transferencia de habilidades. El entrenamiento avanzado de laparoscopia del grupo de Simulación de la Universidad Católica presenta esta lógica, y así el alumno (residente de primer año de cirugía general) termina realizando un procedimiento sumamente avanzado (anastomosis intestinal) para la etapa clínica en la que se encuentre, en la que probablemente sólo deba operar casos de baja complejidad como apendicectomía o colecistectomía.⁴⁴

Es así que se sugiere contar con los siguientes tres pilares para lograr obtener un programa de entrenamiento efectivo:

1. Espacio, infraestructura y hardware.
2. Programa de entrenamiento válido.
3. Instructores para la entrega de *feedback*.

Espacio, infraestructura y hardware

Al plantear un programa de entrenamiento para una habilidad específica es necesario contar con un espacio que asegure un entrenamiento de calidad, idealmente aislado, con cámaras que permitan grabar los entrenamientos, o salas espejo, etcétera. Además, se debe intentar lograr la similitud en el entorno o ambiente de la actividad que se quiere replicar (por ejemplo, si es un pabellón quirúrgico, la máquina de anestesia o la mesa quirúrgica).

Programa de entrenamiento válido

Como se dijo en la sección tres “Establecimiento de los objetivos y validación”, es de suma importancia poder contar con un programa de entrenamiento válido, no sólo poder determinar un constructo (diferenciar novatos de expertos), sino poder llevar estos novatos tras entrenamiento

al nivel deseado de competencias en el programa de simulación (muchas veces el nivel del propio experto) al obtener curvas de aprendizaje. Si se demuestra que un residente es competente tras el entrenamiento, ¿será necesariamente hábil en el escenario real? Esto muchas veces no es así, y para estar seguros se debe demostrar validez predictiva: que las habilidades obtenidas con el programa de simulación se transfieran a la sala operatoria y acorten la curva de aprendizaje en el escenario real.

INSTRUCTORES PARA LA ENTREGA DE FEEDBACK

Sin duda, el pilar más difícil de conseguir. Cuando se tiene un curso, ¿quiénes son los instructores? Habitualmente solicitamos a clínicos que han sido expuestos a la situación real para que enseñen la situación en un ambiente simulado. Mientras más exposición al escenario real, se suele considerar que la persona es más experta y, por lo tanto, debería ser un buen profesor en el ambiente simulado. ¿O no?

La afirmación previa es tan absurda como decir que un jugador profesional de fútbol como Messi es el idóneo para entrenar a las generaciones infantiles del Club Barcelona. Primero, Messi tiene un elevado costo de oportunidad; dejar de jugar fútbol profesional (atender la clínica en el caso de nuestros expertos) es caro y, por lo general, los presidentes del club (jefes de departamentos) prefieren que esté ganando dinero al jugar en la liga profesional adulta. Segundo, ¿es Messi un buen profesor? ¿Aprendió a dar feedback? ¿Sabe detectar el error cuando lo ve y decir de buena forma cómo corregirlo? En el fútbol, en la música y en diferentes otras áreas con necesidad de aprendizaje psicomotor, la solución ha sido formar "entrenadores": profesionales que se dedican a enseñar cómo optimizar el entrenamiento (por medio de práctica deliberada) de un alumno. Pero en la práctica de la simulación clínica lo que habitualmente hacemos es pedirle a residentes o expertos clínicos que ayuden con los pasos prácticos. Por lo anterior se debe profesionalizar el sistema de instructores en simulación. Enseguida, estos instructores deberán disponer del tiempo necesario para suplir las demandas de nuestros alumnos.⁵⁴

Con estos tres pilares se puede entrenar prácticamente lo que se quiera. Los pilares habitualmente se dan en orden. Primero se consigue el pilar 1, luego, qué entrenar en el pilar 1. Si tras la revisión de la literatura no se encuentra una

simulación adecuada que esté orientada a cumplir los objetivos sobre qué se quiere entrenar, viene una fase de investigación para obtener una simulación con validez y que permita la transferencia de habilidades. Mientras se desarrolla la validación del programa de entrenamiento es necesario entrenar a quienes serán los tutores del curso, y ver la manera para que siempre estén disponibles y para que los alumnos practiquen cada sesión con máxima calidad, recibiendo feedback efectivo (en su defecto un *debriefing* en el caso de un entrenamiento de habilidades comunicacionales u otras de escenarios). El grupo de simulación UC ha podido llevar estos tres pilares por medio de la tecnología a otros lugares fuera de su propia institución. Este es el caso del entrenamiento de laparoscopia por vía telesimulación y uso de feedback diferido.⁵⁶

5. EVALUACIÓN DE TRANSFERENCIA

Cuando un buen desempeño en un programa de simulación se traduce en un mejor desempeño al realizar una tarea equivalente, pero en un escenario más complejo (idealmente en situación real con pacientes), hablamos de transferencia de las competencias. Esto ayudará a definir qué tan bien funciona un programa de simulación, pues determinará su utilidad real.^{57,58}

Un aspecto importante para lograr la transferencia es la definición de un nivel mínimo de competencia para aprobar el módulo de entrenamiento que está basado en la simulación. Esto generalmente se relaciona con el desempeño que fue previamente medido en los expertos.^{59,60} Una vez obtenido este nivel se puede proceder a realizar el procedimiento con pacientes bajo supervisión.

La determinación de si las destrezas adquiridas mediante simulación se transfieren a un escenario real debe basarse en el uso de pautas de evaluación validadas y en los estándares de atención médica. La pauta de evaluación global de la OSATS ya mencionada es útil para este propósito.⁵⁸ Adicionalmente, se pueden medir otros elementos como el tiempo requerido para completar el procedimiento, las habilidades blandas en el trato con el paciente, la confianza del alumno mientras realiza el procedimiento y la tasa de complicaciones.

Por ejemplo, Dawoud y colaboradores reportaron, en el año 2012, el impacto generado por un programa de entrenamiento en biopsia renal

ecoguiada utilizando material ex vivo. Luego de esta práctica, los residentes de nefrología, que fueron entrenados, reportaron mayor comodidad al realizar el procedimiento en pacientes reales, además de una mayor tasa de éxito en la obtención de la muestra y menor sangrado asociado con el procedimiento.⁶¹

Por otro lado, un estudio publicado este año relata transferencia de habilidades blandas y cognitivas en un grupo de estudiantes de enfermería que fueron sometidas a un entrenamiento simulado de manejo de paciente en cuidados paliativos. Los estudiantes presentaron mejor manejo emocional, trato con los familiares y detección precoz de signos de alarma en pacientes terminales.⁶²

Adicionalmente, luego de un programa de entrenamiento simulado en habilidades laparoscópicas avanzadas utilizando material ex vivo, un grupo de residentes de primer año presentó mejor desempeño que los cirujanos generales formados con programas tradicionales sin simulación al realizar una yeyuno-yejunostomía mecánica en pabellón.⁶³

Cuando se desean comparar dos intervenciones de forma más objetiva, se puede calcular la razón de efectividad de la transferencia (TER o *transfer-effectiveness ratio*, por sus siglas en inglés) descrita inicialmente por la industria militar y aviación.⁶³ Este cálculo le permitía a la industria aeronáutica saber cuánto tiempo y dinero se ahorraba mediante el entrenamiento de un piloto en los simuladores de alta fidelidad.

La fórmula es la siguiente:

$$\text{TER} = (Y_0 - Y_x)/X$$

Y_0 : mediana del tiempo requerido por el grupo control para alcanzar el resultado esperado.

Y_x : mediana del tiempo requerido por el grupo intervenido para alcanzar el resultado esperado, luego de un programa de entrenamiento simulado.

X: mediana del tiempo de entrenamiento en el simulador.

Por ejemplo, un estudio comparó un entrenamiento en habilidades de laparoscopia avanzada al utilizar un simulador de realidad virtual con una caja de entrenamiento. Se randomizó un grupo de 24 residentes de cirugía general a entrenamiento convencional, entrenamiento con simulador de realidad virtual o entrenamiento en

caja.⁶⁴ Los residentes entrenados en forma convencional demoraron 2,627 segundos en lograr el *outcome* primario (Y_0) en pabellón, mientras que aquéllos entrenados mediante realidad virtual demoraron 269 segundos (Y_x) luego de una mediana de 1,020 segundos de entrenamiento (X).⁶⁴ Así, el TER de entrenamiento con realidad virtual correspondió a $(2,627-269)/1,020 = 2.31$. Es decir, cada hora de entrenamiento en el simulador corresponde a 2.31 horas de entrenamiento convencional.

La transferibilidad a situaciones reales con pacientes es una condición necesaria para justificar la inclusión de un programa de simulación a un currículo educativo y los extensos gastos asociados con el desarrollo del modelo, infraestructura y recurso humano.⁶⁴

6. EVALUACIÓN DE IMPACTO CLÍNICO E INSTITUCIONAL

El desarrollo de un programa de simulación clínica con todos los elementos mencionados anteriormente no tiene sentido si no se hace con miras a mejorar la calidad de la atención de los pacientes o a generar un cambio institucional. Sin embargo, existen escasos estudios que evalúen el resultado a este nivel de las intervenciones realizadas.

En 2011, Andreatta y su equipo reportaron los resultados de un estudio realizado a lo largo de cuatro años en el que, aleatoriamente y al menos en forma mensual, se realizaba un entrenamiento de reanimación cardiopulmonar avanzada en una Unidad de Cuidados Intensivos Pediátricos.⁶⁵ Se involucró a la totalidad del equipo: especialistas, residentes en formación, enfermeras, alumnos de medicina y farmaceúticas, entre otros. Luego de cuatro años, la tasa de supervivencia general ascendió de 33 a 56%. Además, la tasa de supervivencia a actividad eléctrica sin pulso aumentó de 15 a 56%, la cual fue significativamente mayor a 27% de supervivencia reportado a nivel nacional.⁶⁵

El trabajo clínico está inmerso en un contexto institucional y rara vez son los mismos profesionales los que administran los recursos económicos y estructurales. Por esto, es importante comunicar a los administradores el impacto clínico y económico de programas de simulación potencialmente costosos para justificar su existencia y racionalizar los costos.⁶⁶

Una forma de hacerlo es mediante el cálculo de costos asociados a error y/o complicaciones potencialmente mejorables mediante entrena-

miento, en comparación con el costo de implementación de un programa de simulación. El método exacto mediante el que se realiza un análisis de costos se encuentra en la literatura y excede los objetivos del presente artículo.⁶⁶ A modo de ejemplo, un entrenamiento para inserción de catéter venoso central basado en simulación realizado en una Unidad de Cuidados Intensivos previno alrededor de 10 casos de infección asociada a catéter vascular (IACV) el primer año. El gasto promedio asociado con cada evento de IACV se evaluó en aproximadamente 82,000 USD. El costo de implementación del programa de entrenamiento fue de 112,000 USD, permitiendo ahorrar cerca de 700,000 USD anuales a la institución.⁶⁷

CONCLUSIÓN

El desarrollo de un programa de entrenamiento basado en simulación es un proceso que requiere el cumplimiento riguroso de múltiples etapas. Uno de los pasos iniciales, y muy relevantes, es realizar una búsqueda en la literatura de los programas y simuladores descritos previamente en entrenamientos similares, ya que muchas veces no es necesario “reinventar la rueda”. En segundo lugar, quisiéramos recalcar la importancia del establecimiento del constructo y de la correcta elección de la herramienta de evaluación. La validación como proceso en simulación es relevante y corresponde a las interpretaciones derivadas del modelo en el contexto particular al usar una determinada herramienta.

Finalmente, todo el esfuerzo de crear un entrenamiento validado debe ir de la mano con una integración formal del programa en el currículo de los estudiantes, ya sean de pregrado o postgrado. El objetivo es que el programa esté cuidadosamente integrado con las otras actividades educativas de los alumnos. Esto significa que tanto el programa como su evaluación deben ser planificados, programados y realizados cuidadosamente en el contexto del plan de estudio de los alumnos. Además, esto agrega el valor de que, si las personas a cargo del programa simulado cambian, el programa tendrá continuidad.

En un mundo ideal, la etapa final de un programa siempre debiera ser la medición de la transferencia de las habilidades logradas por los alumnos al paciente real. No debemos olvidar que la educación mediante simulación clínica tiene por finalidad mejorar los resultados de nuestros pacientes.

REFERENCIAS

1. Lopreiato JO, Downing D, Gammon W, Lioce L, Sittner B, Slot V, et al. Healthcare simulation dictionary. Healthcare. 2a. Ed. Rockville, MD: Agency for Healthcare Research and Quality; 2020.
2. Bradley P. The history of simulation in medical education and possible future directions. Med Educ. 2006; 40 (3): 254-262. doi: 10.1111/j.1365-2929.2006.02394.x.
3. Shy KS, Hageman JJ, Le J. The Role of Aircraft Simulation in Improving Flight Safety Through Control Training. California: Dryden Flight Research Center; 2012.
4. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. N Engl J Med. 1991; 324 (6): 377-384. doi: 10.1056/NEJM199102073240605.
5. Hutter MM, Kellogg KC, Ferguson CM, Abbott WM, Warshaw AL. The impact of the 80-hour resident workweek on surgical residents and attending surgeons. Ann Surg. 2006; 243 (6): 864-871. doi: 10.1097/01.sla.0000220042.48310.66.
6. Corvetto M, Bravo MP, Montaña R, Utili F, Escudero E, Boza C, et al. Simulación en educación médica: una sinopsis. Rev Med Chil. 2013; 141 (1): 70-79. doi: 10.4067/S0034-98872013000100010.
7. Varas J, Mejía R, Riquelme A, Maluenda F, Buckel E, Salinas J, et al. Significant transfer of surgical skills obtained with an advanced laparoscopic training program to a laparoscopic jejunomejunostomy in a live porcine model: Feasibility of learning advanced laparoscopy in a general surgery residency. Surg Endosc. 2012; 26 (12): 3486-3494. doi: 10.1007/s00464-012-2391-4.
8. Varas J, Achurra P, León F, Castillo R, De La Fuente N, Aggarwal R, et al. Assessment of central venous catheterization in a simulated model using a motion-tracking device: an experimental validation study. Ann Surg Innov Res. 2016; 10 (1): 2. doi: 10.1186/s13022-016-0025-6.
9. McCaughey CS, Traynor MK. The role of simulation in nurse education. Nurse Educ Today. 2010; 30 (8): 827-832. doi: 10.1016/j.nedt.2010.03.005.
10. Berragan L. Learning nursing through simulation: a case study approach towards an expansive model of learning. Nurse Educ Today. 2014; 34 (8): 1143-1148. doi: 10.1016/j.nedt.2014.03.005.
11. Perry S, Bridges SM, Burrow MF. A review of the use of simulation in dental education. Simul Healthc. 2015; 10 (1): 31-37.
12. Khan R, Plahouras J, Johnston BC, Scaffidi MA, Grover SC, Walsh CM. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. Cochrane Database Syst Rev. 2018; 8: CD008237. doi: 10.1002/14651858.CD008237.pub3.
13. Elliott S, Murrell K, Harper P, Stephens T, Pellowe C. A comprehensive systematic review of the use of simulation in the continuing education and training of qualified medical, nursing and midwifery staff. JBI Libr Syst Rev. 2009; 9 (17): 538-587. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/27819938>.
14. Norman J. Systematic review of the literature on simulation in nursing education. ABNF J. 2012; 23

- (2): 24-28. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/22774355>.
15. Bjerrum F, Thomsen ASS, Nayahangan LJ, Konge L. Surgical simulation: current practices and future perspectives for technical skills training. *Med Teach.* 2018; 40 (7): 668-675. doi: 10.1080/0142159X.2018.1472754.
 16. Barsuk JH, Cohen ER, Feinglass J, McGaghie WC, Wayne DB. Use of simulation-based education to reduce catheter-related bloodstream infections. *Arch Intern Med.* 2009; 169 (15): 1420-1423. doi: 10.1001/archinternmed.2009.215.
 17. Wehbi NK, Wani R, Yang Y, Wilson F, Medcalf S, Monaghan B, et al. A needs assessment for simulation-based training of emergency medical providers in Nebraska, USA. *Adv Simul (Lond).* 2018; 3 (1): 22. doi: 10.1186/s41077-018-0081-6.
 18. Humphrey-Murto S, Varpio L, Wood TJ, Gonsalves C, Ufholz LA, Mascioli K, et al. The use of the Delphi and other consensus group methods in medical education research: a review. *Acad Med.* 2017; 92 (10): 1491-1498. doi: 10.1097/ACM.00000000000001812.
 19. Glass CC, Acton RD, Blair PG, Campbell AR, Deutsch ES, Jones DB, et al. American College of Surgeons/Association for Surgical Education medical student simulation-based surgical skills curriculum needs assessment. *Am J Surg.* 2014; 207 (2): 165-169. doi: 10.1016/j.amjsurg.2013.07.032.
 20. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990; 65 (9 Suppl): S63-7. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/2400509>
 21. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ.* 2003; 37 (9): 830-837. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/14506816>.
 22. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med.* 2006; 119 (2): 166.e7-166.e16. doi: 10.1016/j.amjmed.2005.10.036.
 23. Hirsch E. The core knowledge curriculum-what's behind its success? *Educational Leadership.* Educ Leadersh. 1993; 50 (8): 23-30. <http://www.ascd.org/publications/educational-leadership/may93/vol50/num08/The-Core-Knowledge-Curriculum-What's-Behind-Its-Success#.aspx>.
 24. Ferrero F. ¿Puede la simulación clínica contribuir al aprendizaje significativo de competencias educativas? Una aproximación constructivista. *Rev Fac Med UNAM.* 2017; 60 (S1): 49-59. Disponible en: <https://www.medigraphic.com/pdfs/facmed/un-2017/uns171f.pdf>.
 25. Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA. Reconsidering fidelity in simulation-based training. *Acad Med.* 2014; 89 (3): 387-392. doi: 10.1097/ACM.0000000000000130.
 26. Allen J, Buffardi L, Hays R, Drillings M, Kaplan M. The relationship of simulator fidelity to task and performance variables. USA; United States Army Research Institute for the Behavioral and Social Sciences: 1991. Disponible en: <https://apps.dtic.mil/dtic/tr/fulltext/u2/a238941.pdf>.
 27. Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD. The effect of bench model fidelity on endourological skills: a randomized controlled study. *J Urol.* 2002; 167 (3): 1243-1247. doi: 10.1016/S0022-5347(05)65274-3.
 28. Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, et al. The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. *Ann Surg.* 2004; 240 (2): 374-381. doi: 10.1097/01.SLA.0000133346.07434.30.
 29. Zendejas B, Wang AT, Brydges R, Hamstra SJ, Cook DA. Cost: the missing outcome in simulation-based medical education research: A systematic review. *Surgery.* 2013; 153 (2): 160-176. doi: 10.1016/j.surg.2012.06.025.
 30. Isaranuwatchai W, Brydges R, Carnahan H, Backstein D, Dubrowski A. Comparing the cost-effectiveness of simulation modalities: a case study of peripheral intravenous catheterization training. *Adv Heal Sci Educ.* 2014; 19 (2): 219-232. doi: 10.1007/s10459-013-9464-6.
 31. Van Nortwick SS, Lendvay TS, Jensen AR, Wright AS, Horvath KD, Kim S. Methodologies for establishing validity in surgical simulation studies. *Surgery.* 2010; 147 (5): 622-630. doi: 10.1016/j.surg.2009.10.068.
 32. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? *J Gen Intern Med.* 2004; 19 (9): 971-977. doi: 10.1111/j.1525-1497.2004.40066.x.
 33. Messick S. Validity BT. Educational measurement. In: Linn RL. Educational measurement. 3rd ed. New York: Macmillan; 1989.
 34. Foster SL, Cone JD. Validity issues in clinical assessment. *Psychol Assess.* 1995; 7 (3): 248-260. doi: 10.1037/1040-3590.7.3.248.
 35. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul.* 2016; 1 (1): 31. doi: 10.1186/s41077-016-0033-y.
 36. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ.* 2015; 49 (6): 560-575. doi: 10.1111/medu.12678.
 37. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc.* 2003; 17 (10): 1525-1529. doi: 10.1007/s00464-003-0035-4.
 38. Cook DA. Much ado about differences: why expert-novice comparisons add little to the validity argument. *Adv Heal Sci Educ.* 2015; 20 (3): 829-834. doi: 10.1007/s10459-014-9551-3.
 39. Eignor DR. The standards for educational and psychological testing. In: Geisinger KF, Bracken BA, Carlson JF, Hansen JL, Kuncel NR, Reise SP. *APA handbook of testing and assessment in psychology, Test Theory and Testing and Assessment in Industrial and Organizational Psychology.* American Psychological Association; 2013. pp 245-250. doi: 10.1037/14047-013.
 40. Clark LA, Watson D. Constructing validity: Basic issues in objective scale development. *Psychol Assess.* 1995; 7 (3): 309-319. doi: 10.1037/1040-3590.7.3.309.

41. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull.* 1959 [Accessed 29 June 2019]; 56 (2): 81-105. <http://www.ncbi.nlm.nih.gov/pubmed/13634291>.
42. Hatala R, Cook DA, Brydges R, Hawkins R. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. *Adv Heal Sci Educ.* 2015; 20 (5): 1149-1175. doi: 10.1007/s10459-015-9593-1.
43. Grantcharov TP, Rosenberg J, Pahle E, Fench-Jensen P. Virtual reality computer simulation. *Surg Endosc.* 2001; 15 (3): 242-244. doi: 10.1007/s004640090008.
44. Boza C, León F, Buckel E, Riquelme A, Crovari F, Martínez J, et al. Simulation-trained junior residents perform better than general surgeons on advanced laparoscopic cases. *Surg Endosc.* 2017; 31 (1): 135-141. doi: 10.1007/s00464-016-4942-6.
45. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ.* 2004 [Accessed 29 June 2019]; 38 (3): 327-333. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/14996342>.
46. Traub RE, Rowley GL. An NCME Instructional module on. Understanding reliability. *Educ Meas Issues Pract.* 1991; 10 (1): 37-45. doi: 10.1111/j.1745-3992.1991.tb00183.x.
47. Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ.* 2004; 38 (9): 1006-1012. doi: 10.1111/j.1365-2929.2004.01932.x.
48. Cook DA, Kuper A, Hatala R, Ginsburg S. When assessment data are words. *Acad Med.* 2016; 91 (10): 1359-1369. doi: 10.1097/ACM.0000000000001175.
49. Szasz P, Louridas M, Harris KA, Aggarwal R, Grantcharov TP. Assessing technical competence in surgical trainees. *Ann Surg.* 2015; 261 (6): 1046-1055. doi: 10.1097/SLA.0000000000000866.
50. Ilgen JS, Ma IWY, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ.* 2015; 49 (2): 161-173. doi: 10.1111/medu.12621.
51. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA.* 2009; 302 (12): 1316-1326. doi: 10.1001/jama.2009.1365.
52. Aggarwal R, Dosis A, Bello F, Darzi A. Motion tracking systems for assessment of surgical skill. *Surg Endosc.* 2007; 21 (2): 339-339. doi: 10.1007/s00464-005-0561-3.
53. Rutherford DN, D'Angelo A-LD, Law KE, Pugh CM. Advanced engineering technology for measuring performance. *Surg Clin North Am.* 2015; 95 (4): 813-826. doi: 10.1016/j.suc.2015.04.005.
54. Jarry C, Inzunza M, Quezada J, Marino C, Zamorano E, Alvarado V, et al. Gimnasio de simulación quirúrgica: una herramienta educacional factible de entrenamiento continuo. Experiencia de un centro universitario. *Rev Latinoam Simulación Clínica.* 2019; 1 (1): 18-24. Disponible en: <https://www.medicgraphic.com/pdfs/simulacion/rsc-2019/rsc191d.pdf>.
55. Castillo R, Buckel E, León F, Varas J, Alvarado J, Achurra P, et al. Effectiveness of learning advanced laparoscopic skills in a brief intensive laparoscopy training program. *J Surg Educ.* 2015; 72 (4): 648-653. doi: 10.1016/j.jsurg.2015.01.016.
56. Quezada J, Achurra P, Jarry C, Asbun D, Tejos R, Inzunza M, et al. Minimally invasive tele-mentoring opportunity-the mito project. *Surg Endosc.* 2019. doi: 10.1007/s00464-019-07024-1.
57. Pugh C, Plachta S, Auyang E, Pryor A, Hungness E. Outcome measures for surgical simulators: is the focus on technical skills the best approach? *Surgery.* 2010; 147 (5): 646-654. doi: 10.1016/j.surg.2010.01.011.
58. Dawe SR, Pena GN, Windsor JA, Broeders JA, Cregan PC, Hewett PJ, et al. Systematic review of skills transfer after surgical simulation-based training. *Br J Surg.* 2014; 101 (9): 1063-1076. doi: 10.1002/bjs.9482.
59. Reznick RK, MacRae H. Teaching surgical skills-changes in the wind. *N Engl J Med.* 2006; 355 (25): 2664-2669. doi: 10.1056/NEJMra054785.
60. Scott DJ, Dunnington GL. The New ACS/APDS Skills curriculum: moving the learning curve out of the operating room. *J Gastrointest Surg.* 2008; 12 (2): 213-221. doi: 10.1007/s11605-007-0357-y.
61. Dawoud D, Lyndon W, Mrug S, Bissler JJ, Mrug M. Impact of ultrasound-guided kidney biopsy simulation on trainee confidence and biopsy outcomes. *Am J Nephrol.* 2012; 36 (6): 570-574. doi: 10.1159/000345305.
62. Valen K, Holm AL, Jensen KT, Grov EK. Nursing students' perception on transferring experiences in palliative care simulation to practice. *Nurse Educ Today.* 2019; 77: 53-58. doi: 10.1016/j.nedt.2019.03.007.
63. Morrison JE, Hammon C. On measuring the effectiveness of large-scale training simulations. Alexandria Virginia: IDA; 2000 [Accessed 8 September 2019]; <https://apps.dtic.mil/docs/citations/ADA394491>.
64. Orzech N, Palter VN, Reznick RK, Aggarwal R, Grantcharov TP. A comparison of 2 ex vivo training curricula for advanced laparoscopic skills. *Ann Surg.* 2012; 255 (5): 833-839. doi: 10.1097/SLA.0b013e31824aca09.
65. Andreatta P, Saxton E, Thompson M, Annich G. Simulation-based mock codes significantly correlate with improved pediatric patient cardiopulmonary arrest survival rates. *Pediatr Crit Care Med.* 2011; 12 (1): 33-38. doi: 10.1097/PCC.0b013e3181e89270.
66. Asche CV, Kim M, Brown A, Golden A, Laack TA, Rosario J, et al. Communicating value in simulation: cost benefit analysis and return on investment. *Acad Emerg Med.* 2018; 25 (2): 230-237. doi: 10.1111/acem.13336.
67. Cohen ER, Feinglass J, Barsuk JH, Barnard C, O'Donnell A, McGaghie WC, et al. Cost savings from reduced catheter-related bloodstream infection after simulation-based education for residents in a medical intensive care unit. *Simul Healthc J Soc Simul Healthc.* 2010; 5 (2): 98-102. doi: 10.1097/SIH.0b013e3181bc8304.

Correspondencia:**Dra. Marcia Corvetto**Maroleta 377, segundo piso, Santiago,
Región Metropolitana, Chile.**E-mail:** marciacorvetto@gmail.com