

Validación psicométrica de la Escala Ottawa de Gestión de Recursos en Crisis en simulación clínica: evidencia de fiabilidad interevaluadores y validez convergente, en Argentina

Carolina Iris Pereyra-Girardi¹, Norma Beatriz Raúl¹, Gustavo Alberto Costa¹, Silvia Liliana García¹, Silvia Deborah Ofman²

INTRODUCCIÓN. La simulación clínica (SC) es una herramienta clave para la formación de profesionales de la salud, permitiendo entrenar habilidades técnicas y no técnicas (HNT) en entornos seguros. La evaluación de las HNT constituye un desafío por la falta de instrumentos validados. Este estudio tiene como objetivo analizar las propiedades psicométricas de la Escala Ottawa para la Gestión de Recursos en Crisis, adaptada al contexto argentino.

MATERIAL Y MÉTODOS. Se realizó un estudio cuantitativo no experimental de tipo psicométrico instrumental. La muestra estuvo conformada por 95 casos simulados grabados en video en el Hospital Universitario de Simulación Clínica de la Universidad Nacional Arturo Jauretche. Tres evaluadores expertos evaluaron el desempeño de equipos de salud utilizando la escala. Se calcularon coeficientes de correlación intraclase (ICC) para estimar la fiabilidad interobservador, alfa de Cronbach para consistencia interna y correlaciones de Pearson para validez concurrente.

RESULTADOS. Los valores de consistencia interna fueron excelentes ($\alpha = 0,932-0,977$). La fiabilidad interevaluadores fue moderada para medidas individuales (ICC = 0,455-0,808) y excelente para medidas promedio (ICC = 0,914-0,944). Se observaron correlaciones positivas y significativas entre ambas versiones de la escala ($r = 0,749-0,920$; $p < 0,01$), confirmando su validez concurrente.

CONCLUSIONES. La versión argentina de la Escala Ottawa para la Gestión Global de Crisis cuenta con evidencia de fiabilidad y validez robusta, constituyéndose en una herramienta útil, confiable y culturalmente pertinente, para la evaluación de habilidades no técnicas en simulación clínica en equipos de salud.

Palabras clave: Entrenamiento simulado. Psicometría. Reproducibilidad de los resultados. Variaciones dependientes del observador. Gestión de recursos de personal en Salud. Habilidades no técnicas.

Psychometric validation of the Ottawa Global Rating Scale for Crisis Resource Management in clinical simulation: evidence of inter-rater reliability and convergent validity in Argentina

INTRODUCTION. Clinical simulation (CS) has become a key tool for the training of healthcare professionals, allowing the development of both technical and non-technical skills (NTS) in safe environments. The assessment of NTS remains challenging due to the lack of locally validated instruments. The aim of this study was to analyze the psychometric properties of the Ottawa Global Rating Scale for Crisis Resource Management adapted to the Argentine context.

MATERIAL AND METHODS. We conducted a quantitative, nonexperimental psychometric instrumental study. The sample consisted of 95 simulated cases recorded on video at *Hospital Universitario de Simulación Clínica de la Universidad Nacional Arturo Jauretche*. Three expert judges evaluated the performance of health care teams using the scale. Intraclass correlation coefficients (ICC) were calculated to estimate interrater reliability, Cronbach's alpha for internal consistency, and Pearson correlations for concurrent validity.

RESULTS. Internal consistency values were excellent ($\alpha = 0.932-0.977$). Interrater reliability was moderate for individual measures (ICC, 0.455-0.808) and excellent for average measures (ICC, 0.914-0.944). Positive and significant correlations were observed between both versions of the scale ($r = 0.749-0.920$; $P < .01$), confirming concurrent validity.

CONCLUSIONS. The Argentine version of the Ottawa Global Rating Scale for Crisis Management shows strong evidence of reliability and validity and represents a useful, reliable, and culturally appropriate tool for assessing non-technical skills in clinical simulation among health care teams.

Keywords: Simulation training. Psychometrics. Reproducibility of results. Observer variation. Health workforce resource management. Non-technical skills.

Filiación de los autores: ¹Universidad Nacional Arturo Jauretche, Hospital Universitario de Simulación, Buenos Aires, Argentina.

²Instituto de Neurociencias Buenos Aires (INEBA) - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Ciudad Autónoma de Buenos Aires, Argentina.

Correspondencia: Carolina Iris Pereyra-Girardi. Avenida Calchaquí 6200, Florencio Varela, (CP 1888), Buenos Aires, Argentina.

E-mail: cpereyragirardi@unaj.edu.ar

Información del artículo: Recibido: 11-12-2025. Aceptado: 27-2-2026. Online: 27-3-2026.

Editor responsable: Rafael Castro Delgado.

DOI: 10.55633/s3me/REUE021.2026

Introducción

La simulación clínica (SC) se ha consolidado como estrategia pedagógica central en la formación de profesionales de la salud, permitiendo la adquisición de habilidades técnicas y no técnicas en entornos seguros y controlados¹⁻³. Su implementación mejora la calidad asistencial y contribuye significativamente a la seguridad del paciente, lo que explica su creciente incorporación en programas académicos y su promoción institucional en diversos sistemas sanitarios^{4,5}.

A pesar de su expansión, el desarrollo de la SC continúa siendo desigual: el 37 % de las facultades de medicina en Estados Unidos y Canadá la integran formalmente en sus currículos, mientras que en Latinoamérica esta cifra desciende al 17 %³. Entre las principales barreras se identifican los costos de implementación, el mantenimiento tecnológico y la formación docente especializada^{1,2,5}. En Argentina, diversas iniciativas han fortalecido la SC en ámbitos universitarios y públicos. A nivel regional, Armijo-Rivera *et al.*⁶ registraron 408 centros de simulación en Latinoamérica, con crecimiento sostenido e impacto positivo en la formación clínica.

En este marco, la SC se ha posicionado como una herramienta clave para entrenar habilidades no técnicas (HNT), particularmente bajo el modelo Crisis Resource Management (CRM), que integra competencias de liderazgo, comunicación, conciencia situacional, toma de decisiones y resolución de problemas^{7,8}, consideradas determinantes para el desempeño seguro y eficaz de los equipos de salud en contextos de emergencia. No obstante, la evaluación de las HNT representa un desafío metodológico por su dificultad conceptual y la escasez de instrumentos estandarizados con respaldo psicométrico sólido⁹. Disponer de instrumentos validados localmente constituye un requisito metodológico indispensable, ya que las propiedades psicométricas pueden variar según el contexto cultural. Zamudio Burbano *et al.*¹⁰ validaron la Escala Ottawa en Colombia (ICC = 0,42-0,79), mientras que Sánchez-Marco *et al.*¹¹ confirmaron su validez en España, reforzando la necesidad de una versión argentina.

La relevancia de evaluar y entrenar HNT se vincula directamente con la seguridad del paciente. La Organización Mundial de la Salud (OMS) estima que uno de cada diez pacientes hospitalizados experimenta algún daño prevenible², y más del 50 % de estos eventos podrían evitarse¹³. Los fallos de comunicación, el liderazgo deficiente y la inadecuada gestión de recursos constituyen los principales factores desencadenantes^{14,15}. En este escenario, el entrenamiento sistemático mediante SC emerge como estrategia eficaz¹², requiriendo instrumentos válidos y confiables para la evaluación.

El presente estudio analiza las propiedades psicométricas de la Escala Ottawa de Gestión Global en Crisis para equipos de salud, versión Argentina¹⁵, evaluando consistencia interna, confiabilidad interevaluador y validez concurrente mediante comparación con la versión española¹¹.

Materiales y métodos

Diseño

Se realizó un estudio cuantitativo no experimental de tipo psicométrico instrumental¹⁶. La muestra estuvo conformada por 95 casos con escenarios de SC de mediana fidelidad. El tamaño muestral se consideró adecuado para estudios de validación instrumental. En investigaciones psicométricas se recomiendan muestras entre 50 y 100 sujetos para estimaciones estables de consistencia interna¹⁷ y coeficientes de correlación intraclase¹⁸. Las guías metodológicas enfatizan considerar el número de sujetos y evaluadores para garantizar precisión en las estimaciones del coeficiente de correlación intraclase (ICC)¹⁹. Se incluyeron situaciones de reanimación cardiopulmonar, politraumatismo, *shock* y emergencias respiratorias, desarrollados con simuladores de mediana y alta tecnología en el Hospital Universitario de Simulación Clínica de la Universidad Nacional Arturo Jauretche (Florencio Varela, Buenos Aires, Argentina).

Cada caso correspondió a equipos integrados entre tres y cinco estudiantes de las asignaturas Prácticas Profesionales III y IV del último año de la Tecnicatura en Emergencias Sanitarias y Desastres. En todos los escenarios se designó formalmente un líder responsable de la coordinación del equipo, en concordancia con la estructura operativa del modelo CRM.

Instrumentos

Se utilizó la Escala Ottawa de Gestión Global en Crisis para equipos de salud, versión adaptada culturalmente para Argentina¹⁵. El proceso de adaptación incluyó traducción directa, retrotraducción, evaluación por panel de expertos y estudio piloto previo, siguiendo estándares internacionales de adaptación transcultural. La escala evalúa cinco dimensiones del modelo CRM: liderazgo, resolución de problemas, conciencia situacional, comunicación y utilización de recursos. Está compuesta por 20 ítems con escala ordinal de 1 a 4 puntos, e incluye además una puntuación global de desempeño, que sintetiza la competencia integral del equipo.

Como criterio externo se empleó la versión española de la Ottawa Crisis Resource Management Global Rating Scale, validada por Sánchez-Marco *et al.*¹¹. Este instrumento conserva las mismas dimensiones conceptuales y estructura general del modelo original y ha demostrado consistencia interna elevada ($\alpha \approx 0,98$), así como correlaciones significativas con otras medidas que evalúan el mismo constructo, aportando evidencia de validez basada en la relación con otras variables.

Procedimiento

Todas las SC fueron registradas en video para su posterior análisis. Un panel de tres expertos evaluó el desempeño de los equipos en términos de HNT. La selección de evaluadores expertos constituyó un componente central del proceso de validación, dado que la confiabilidad interevaluador depende en gran medida de la idoneidad y experiencia de quienes realizan las valoraciones. Siguiendo

Tabla 1. Coeficiente de correlación intraclass para la dimensión Liderazgo

Medida	Correlación Intraclass	Límite inferior IC 95 %	Límite superior IC 95 %	Valor F	gl1	gl2	Sig.
Medidas únicas	0,482	0,407	0,566	15,032	94	1.316	0,000
Medidas promedio	0,933	0,912	0,951	15,032	94	1.316	0,000

Nota: Coeficiente de correlación intraclass de tipo A (acuerdo absoluto).

los criterios propuestos por Sk-jong y Wentworth²¹, se priorizó la inclusión de profesionales con experiencia acreditada en SC y evaluación de desempeño, reconocimiento en la comunidad académica, disponibilidad para participar y criterios de imparcialidad. Asimismo, conforme a las recomendaciones metodológicas sobre validación instrumental^{19,21}, el número de observadores puede oscilar entre dos y veinte según la naturaleza del estudio. En el presente trabajo participaron tres evaluadores con trayectoria comprobada en el área: un licenciado en enfermería, una médica y una técnica en emergencias sanitarias y desastres. Esta composición permitió incorporar una perspectiva multiprofesional en la evaluación de las habilidades no técnicas, favoreciendo la representatividad ecológica del proceso evaluativo.

Para la evaluación de desempeño de HNT cada juez utilizó de manera independiente, la versión argentina de la escala¹⁵ y la versión española utilizada como criterio externo¹¹. Esto permitió obtener evidencia de validez concurrente mediante la comparación directa de ambas medidas en un mismo momento temporal.

Análisis de datos

La confiabilidad interevaluador se estimó mediante el ICC, utilizando un modelo de dos factores de efectos mixtos con efectos aleatorios para los sujetos y efectos fijos para los evaluadores, apropiado cuando un conjunto específico de observadores evalúa los mismos casos. La confiabilidad interevaluador se estimó mediante el ICC, considerado una medida robusta para evaluar el acuerdo absoluto cuando todos los evaluadores califican los mismos sujetos¹⁸. Se aplicó un modelo de dos factores de efectos mixtos con acuerdo absoluto, calculando ICC tanto para medidas únicas (desempeño de un evaluador individual) como para medidas promedio (promedio de los tres evaluadores). Se calcularon ICC para medidas únicas (acuerdo de un evaluador individual) y medidas promedio (acuerdo promedio entre observadores). La consistencia interna se estimó mediante el coeficiente alfa de Cronbach por dimensión y por juez. La validez concurrente se evaluó mediante correlaciones de Pearson entre las puntuaciones de la versión argentina y la versión española. Se estableció un nivel de significación estadística de $p \leq 0,05$. Los análisis se realizaron utilizando el software R y SPSS versión 24.

Aspectos éticos

El estudio se desarrolló conforme a los principios éticos establecidos en la Declaración de Helsinki de la Asociación Médica Mundial²⁰ y a la normativa argentina vigente para investigaciones en salud. Todos los participantes firmaron un consentimiento informado previo a su inclusión en el estudio. El protocolo de investigación titulado "Escala Global Ottawa para la gestión de recursos en crisis destinada a equipos de salud versión local. Validación interjueces y evidencias de constructo" fue evaluado y aprobado por el Comité de Ética en Investigación del Hospital de Alta Complejidad en Red El Cruce (Registro N° 023/2010, Folio 062 del Libro de Actas N°1).

Resultados

Confiabilidad interevaluador

En la dimensión Liderazgo, el ICC para medidas únicas fue de 0,482 (IC 95 %: 0,407-0,566), lo que indica un nivel de acuerdo moderado. No obstante, al considerar el promedio de los tres evaluadores, el ICC ascendió a 0,933 (IC 95 %: 0,912-0,951), reflejando un acuerdo excelente (Tabla 1).

En Resolución de Problemas, el ICC individual alcanzó 0,808 (IC 95 %: 0,773-0,841), evidenciando un acuerdo alto incluso a nivel individual. El ICC promedio fue de 0,944 (IC 95 %: 0,932-0,955), consolidando una fiabilidad excelente (Tabla 2).

Para Conciencia de Situación, el ICC de medidas únicas fue 0,471 (IC 95 %: 0,395-0,557), mientras que el ICC promedio alcanzó 0,914 (IC 95 %: 0,867-0,938), lo que indica que la agregación de evaluadores incrementa sustancialmente la estabilidad de la medición (Tabla 3).

En la dimensión Comunicación, el ICC individual fue 0,503 (IC 95 %: 0,426-0,587), compatible con acuerdo moderado, y el ICC promedio fue 0,924 (IC 95 %: 0,899-0,945), evidenciando excelente consistencia grupal (Tabla 4).

Respecto de Utilización de Recursos, el ICC individual fue 0,540 (IC 95 %: 0,465-0,629), mientras que el ICC promedio alcanzó 0,915 (IC 95 %: 0,887-0,938), manteniendo el patrón de mejora sustancial al considerar la puntuación agregada (Tabla 5).

Tabla 2. Coeficiente de correlación intraclass para la dimensión Resolución de Problemas

Medida	Correlación Intraclass	Límite inferior IC 95 %	Límite superior IC 95 %	Valor F	gl1	gl2	Sig.
Medidas únicas	0,808	0,773	0,841	18,179	245	735	0,000
Medidas promedio	0,944	0,932	0,955	18,179	245	735	0,000

Nota: Coeficiente de correlación intraclass de tipo A (acuerdo absoluto).

Tabla 3. Coeficiente de correlación intraclase para la dimensión Conciencia de Situación

Medida	Correlación Intraclase	Límite inferior IC 95 %	Límite superior IC 95 %	Valor F	gl1	gl2	Sig.
Medidas únicas	0,471	0,395	0,557	11,797	94	1.034	0,000
Medidas promedio	0,914	0,867	0,938	11,797	94	1.034	0,000

Nota: Coeficiente de correlación intraclase de tipo A (acuerdo absoluto).

Finalmente, la puntuación global presentó un ICC individual de 0,455 (IC 95 %: 0,365-0,551) y un ICC promedio de 0,930 (IC 95 %: 0,902-0,952), indicando excelente fiabilidad cuando se utiliza la puntuación promedio del panel (Tabla 6).

Consistencia interna

En la consistencia interna, los valores oscilaron entre 0,932 y 0,977, situándose en el rango de excelencia.

La dimensión Utilización de Recursos presentó los coeficientes más elevados (hasta 0,977), seguida por Comunicación, Resolución de Problemas, Conciencia de Situación y Liderazgo, todas con valores superiores a 0,93 (Tabla 7). Estos resultados indican una alta homogeneidad interna de los ítems dentro de cada dimensión y sugieren que cada conjunto de ítems evalúa de manera coherente el constructo específico que pretende evaluar.

Dado que la puntuación global correspondía a un único ítem de valoración integral, no se estimó alfa de Cronbach para esta puntuación.

Validez concurrente

Los resultados mostraron correlaciones positivas, elevadas y estadísticamente significativas ($p < 0,01$) en todas las dimensiones y en la puntuación global para los tres evaluadores. Para el evaluador 1, los coeficientes oscilaron entre $r = 0,865$ y $r = 0,920$, indicando asociaciones fuertes a muy fuertes en todas las dimensiones, incluido el puntaje global ($r = 0,920$). El evaluador 2 presentó correlaciones entre $r = 0,749$ y $r = 0,889$, todas significativas, reflejando alta correspondencia entre ambas versiones culturales del instrumento. Por su parte, el evaluador 3 obtuvo coeficientes comprendidos entre $r = 0,818$ y $r = 0,884$, igualmente elevados y significativos.

Discusión

La formación de profesionales en ciencias de la salud se desarrolla actualmente en un contexto caracterizado por una elevada complejidad asistencial, innovación tecnológica constante y crecientes exigencias en materia de seguridad del paciente²³. Las HNT constituyen competencias críticas para el desempeño clínico seguro^{24,25}, y los errores de comunicación y de coordinación representan factores principales en eventos adversos prevenibles^{13,14}. En consecuencia, su evaluación rigurosa adquiere relevancia tanto en el

ámbito académico, como en el plano clínico y organizacional.

En este marco, los resultados del presente estudio aportan evidencia psicométrica sólida sobre la versión argentina de la Escala Ottawa de Gestión Global en Crisis para equipos de salud¹⁵. En relación con la consistencia interna, los coeficientes alfa obtenidos ($\alpha = 0,93-0,97$ por dimensión) se sitúan en el rango de excelencia y resultan consistentes con los publicados en otras adaptaciones internacionales del instrumento. La validación española realizada por Sánchez-Marco et al.¹¹ informó coeficientes alfa superiores a 0,90 en la mayoría de las dimensiones, mientras que la adaptación colombiana de Zamudio Burbano et al.¹⁰ evidenció adecuados niveles de consistencia interna en población hispanohablante. De manera concordante, la versión italiana publicada por Franc et al.²⁶ comunicó indicadores elevados de confiabilidad y una estructura dimensional coherente con el modelo original. En Latinoamérica se ha comenzado a consolidar la evaluación estructurada de HNT mediante herramientas estandarizadas, como la aplicación de la escala Ottawa CRM GRS en médicos residentes peruanos²⁷. No obstante, aún son necesarios estudios con mayor profundidad psicométrica que aporten evidencia robusta de validez y confiabilidad en contextos locales. Esta convergencia de resultados en estudios psicométricos de la escala respalda la estabilidad conceptual del constructo bajo el modelo de CRM y sugiere consistencia métrica a través de distintos contextos culturales y lingüísticos.

En cuanto a la confiabilidad interevaluadores, los coeficientes de correlación intraclase mostraron un patrón metodológicamente esperable en estudios de evaluación conductual en SC. Las medidas únicas evidenciaron niveles de acuerdo moderados (ICC = 0,455-0,808), lo que refleja la variabilidad inherente a la evaluación de HNT en escenarios simulados. Esta variabilidad puede atribuirse tanto a la naturaleza interpretativa del constructo –que implica la valoración de conductas complejas en tiempo real– como a las diferencias formativas entre los evaluadores. El panel estuvo conformado por profesionales de disciplinas distintas (medicina, enfermería y tecnicatura en emergencias sanitarias y desastres), lo que introduce marcos conceptuales, experiencias clínicas y criterios de valoración potencialmente diversos. Diversos trabajos han destacado que la pericia del evaluador, el entrenamiento específico en el uso de escalas conductuales y la cali-

Tabla 4. Coeficiente de correlación intraclase para la dimensión Comunicación

Medida	Correlación Intraclase	Límite inferior IC 95 %	Límite superior IC 95 %	Valor F	gl1	gl2	Sig.
Medidas únicas	0,503	0,426	0,587	13,242	94	1.034	0,000
Medidas promedio	0,924	0,899	0,945	13,242	94	1.034	0,000

Nota: Coeficiente de correlación intraclase de tipo A (acuerdo absoluto).

Tabla 5. Coeficiente de correlación intraclass para la dimensión Utilización de Recursos

Medida	Correlación Intraclass	Límite inferior IC 95 %	Límite superior IC 95 %	Valor F	gl1	gl2	Sig.
Medidas únicas	0,540	0,465	0,629	11,702	94	752	0,000
Medidas promedio	0,915	0,887	0,938	11,702	94	752	0,000

Nota: Coeficiente de correlación intraclass de tipo A (acuerdo absoluto).

Tabla 6. Coeficiente de correlación intraclass en la Puntuación Global

Medida	Correlación Intraclass	Límite inferior IC 95 %	Límite superior IC 95 %	Valor F	gl1	gl2	Sig.
Medidas únicas	0,455	0,365	0,551	20,230	94	1.410	0,000
Medidas promedio	0,930	0,902	0,952	20,230	94	1.410	0,000

Nota: Coeficiente de correlación intraclass de tipo A (acuerdo absoluto).

bración previa influyen significativamente en los niveles de acuerdo interevaluadores^{28,29}.

Si bien la heterogeneidad disciplinar del panel de evaluadores, puede introducir variabilidad en las medidas individuales, constituye también una fortaleza desde el punto de vista ecológico, ya que evidencia la realidad multiprofesional de los equipos que intervienen en situaciones de crisis.

Al considerar las medidas promedio, los ICC alcanzan valores excelentes (ICC = 0,914-0,944 por dimensión y 0,930 para el puntaje global), lo que indica que la agregación de evaluadores proporciona estimaciones altamente estables. Este patrón (acuerdo moderado en medidas individuales y excelente en medidas promedio), coincide con lo publicado en la validación colombiana de Zamudio Burbano et al.¹⁰, quienes informaron ICC entre 0,42 y 0,79 en medidas individuales. En este sentido, los valores obtenidos en el presente estudio se sitúan dentro del mismo rango de magnitud observado en otros contextos latinoamericanos, lo que refuerza la comparabilidad internacional del instrumento y la coherencia de los hallazgos.

En el contexto hispanohablante, Escibano et al. utilizaron la Ottawa Global Rating Scale CRM para evaluar un programa de simulación de alta fidelidad en estudiantes de ciencias de la salud, evidenciando mejoras significativas en el desempeño observado de las HNT⁸. Aunque su estudio no se centró específicamente en la estimación de confiabilidad interevaluadores, sus resultados respaldan la sensibilidad y aplicabilidad del instrumento en contextos formativos. En comparación, los coeficientes de correlación intraclass obtenidos en el presente trabajo muestran un patrón coherente con la evaluación de conductas complejas: niveles moderados en medidas individuales (ICC = 0,455-0,808) y niveles excelentes en medidas promedio (ICC = 0,914-0,944; 0,930 global).

Así, los presentes hallazgos complementan la evidencia formativa previa, aportando sustento psicométrico específico sobre la confiabilidad interevaluador de la escala en el contexto argentino.

Respecto de la validez concurrente, se observaron correlaciones positivas, elevadas y estadísticamente significativas entre la versión argentina y la versión española del instrumento en todas las dimensiones y en el puntaje global. La decisión de utilizar la versión española como crite-

rio externo se fundamentó en los modelos contemporáneos de validación, los cuales sostienen que la evidencia basada en la relación con otras variables constituye una fuente central de respaldo interpretativo.

Ambas versiones comparten el mismo marco conceptual (el modelo CRM) y mantienen equivalencia estructural en sus dimensiones (liderazgo, resolución de problemas, conciencia situacional, comunicación y utilización de recursos).

Tal como señalan Arias y Sireci³⁰, la evidencia basada en relaciones con otras variables se fortalece cuando las medidas comparadas presentan coherencia conceptual y operacionalización equivalente del constructo. Estos resultados fortalecen la argumentación de validez de la adaptación argentina, al demostrar que ambas versiones producen resultados consistentes al evaluar el mismo constructo teórico en un mismo momento temporal.

Desde una perspectiva formativa, estos hallazgos adquieren especial relevancia. El entrenamiento en HNT mediante simulación de alta fidelidad mejora tanto el desempeño técnico como el no técnico en escenarios de emergencia², generando mejoras significativas en liderazgo, comunicación y toma de decisiones²⁴. En este contexto, disponer de instrumentos válidos y confiables resulta indispensable para sostener procesos formativos orientados a la seguridad del paciente.

Entre las limitaciones de este estudio figura el hecho de no incluir un análisis factorial exploratorio ni confirmatorio, lo cual debería abordarse en futuras investigaciones para examinar la estructura interna del instrumento y la invarianza factorial entre diferentes profesiones y tipos de simulación. Asimismo, el muestreo no probabilístico limita la generalización de los resultados.

Tabla 7. Consistencia interna (α de Cronbach) por dimensión y por evaluador

Dimensión	Ítems	Evaluador 1	Evaluador 2	Evaluador 3
Liderazgo	5	0,949	0,941	0,933
Resolución de Problemas	4	0,960	0,940	0,938
Conciencia de Situación	4	0,946	0,932	0,947
Comunicación	4	0,974	0,957	0,952
Utilización de Recursos	3	0,977	0,960	0,971

Nota: Alfa de Cronbach calculado para cada juez considerando los 95 casos evaluados.

En síntesis, la versión a partir de la Escala Ottawa de Gestión Global en Crisis¹⁵ evidenció adecuados indicadores de consistencia interna, confiabilidad interevaluadores y validez convergente, en consonancia con otras valida-

ciones internacionales.^{16,26} Estos hallazgos corroboran la escala como herramienta robusta para la evaluación de HNT en SC y contribuyen al fortalecimiento de procesos formativos orientados a la seguridad del paciente.

INFORMACIÓN DEL ARTÍCULO

Conflicto de intereses: Los autores declaran no tener conflicto de interés en relación con el presente artículo.

Financiación: El presente trabajo fue financiado por el subsidio UNAJ-Investiga 2023 80020230300001UJ titulado "Escala Global Ottawa para la gestión de recursos en crisis destinada a equipos de salud versión local. Validación Interjueces y evidencias de constructo".

Responsabilidades éticas: Todos los autores han confirmado el mantenimiento de la confidencialidad y respeto de los derechos de los pacientes, acuerdo de publicación y cesión de derechos de los datos a la Revista Española de Urgencias y Emergencias.

Disponibilidad de datos en abierto: Los datos están disponibles bajo solicitud al autor asignado para la correspondencia.

Contribuciones a la autoría del artículo (CRediT): CIPG: Participó en la conceptualización – Ideas; formulación o evolución de los objetivos y metas generales de la investigación, en el Análisis formal – Aplicación de técnicas estadísticas, matemáticas, u otras técnicas formales para analizar o sintetizar datos de estudio. Participó en cada etapa de la Investigación, en la recolección de datos. Metodología, tanto en diseño de la metodología; creación de modelos. Respecto a la Redacción – desarrolló el borrador original – Preparación, creación y/o presentación del trabajo publicado, específicamente la redacción del borrador inicial y finalmente realizó las modificaciones ajustes y redacción – revisión y edición, específicamente revisión crítica, comentario o revisión – incluyendo las etapas previas o posteriores a la publicación. NBR: Participó en la conceptualización del estudio, contribuyendo al desarrollo del enfoque general de la investigación y tuvo un rol activo en la recolección de datos. Participó en la elaboración del manuscrito, redactando secciones del texto e incorporando aportes relevantes para su desarrollo. Intervino en la revisión crítica y edición del documento en sus distintas versiones, realizando correcciones, ampliaciones y mejoras sustanciales del contenido intelectual. Aprobó la versión definitiva del artículo y asumió la responsabilidad por su contenido. GAC: Participó en el Análisis formal: análisis e interpretación de los datos del trabajo. Tareas: Recolección y análisis de datos. Redacción - borrador original: redacción del borrador inicial. Redacción - revisión y edición: revisión crítica, comentarios y aprobación de la versión final a ser publicada. SLG: Participó en el Análisis formal: análisis e interpretación de los datos del trabajo. Tareas: Recolección y análisis de datos. Redacción - borrador original: redacción del borrador inicial. Redacción - revisión y edición: revisión crítica, comentarios y aprobación de la versión final a ser publicada. SDO: Participó en el Análisis formal: análisis e interpretación de los datos del trabajo. Redacción - borrador original: redacción del borrador inicial. Redacción - revisión y edición: revisión crítica, comentarios y aprobación de la versión final a ser publicada.

Uso de herramientas de inteligencia artificial generativa: Los autores declaran que se utilizó para la elaboración del manuscrito las herramientas de IA ChatGPT (OpenAI) para la corrección estilística del texto y la optimización de tablas. Tras utilizar las herramientas los autores llevaron adelante una

revisión y edición del texto obtenido asumiendo la responsabilidad del contenido de la publicación.

Artículo no encargado por el Comité Editorial y con revisión externa por pares.

Agradecimientos: Los autores desean agradecer especialmente al Dr. José L. de Echave y al Dr. Marcelo Muro por su acompañamiento y compromiso académico en la elaboración del presente trabajo.

BIBLIOGRAFÍA

1. Elendu C, Amaechi DC, Okatta AU, Amaechi EC, Elendu TC, Ezech CP, et al. The impact of simulation-based training in medical education: A review. *Medicine (Baltimore)*. 2024;103:e38813.
2. Muñoz-Gualán GG, Sierra RE. La simulación clínica en la educación médica moderna: revisión de revisiones. *Rev Eugenio Espejo* 2025;19:102-16.
3. Thurman W, Nsengi C, Steindal SA, Albrecht S, Gaber SN, Mattsson E, et al. Simulation-based learning in nursing education to counter stigma and discrimination: A scoping review. *Nursing Outlook*. 2025;73:102478.
4. Monteverde E, Klappenbach R, Baliña J. Manual de seguridad del paciente [Internet]. (Consultado 30 Octubre 2025). Disponible en: https://www.argentina.gob.ar/sites/default/files/2022/11/manual_de_seguridad_del_paciente.pdf
5. Park YS, Lee SJ, Hur Y. Facilitators, barriers, and future direction of high-fidelity simulation in nursing education: a qualitative descriptive study. *BMC Nurs*. 2025;24:881.
6. Armijo-Rivera S, Machuca-Contreras F, Raul N, de Oliveira SN, Mendoza IB, Miyasato HS, et al. Characterization of simulation centers and programs in Latin America according to the ASPIRE and SSH quality criteria. *Adv Simul (Lond)*. 2021;6:41.
7. Gaba DM, Howard SK, Fish KJ, Smith BE, Sowa YA. Simulation-based training in anesthesia crisis resource management (ACRM): a decade of experience. *Simul Gaming*. 2001;32:175-93.
8. Escribano S, Sánchez-Marco M, Espinosa-Ramírez S, Mateos-Rodríguez A, Fernández-Lebrusán L, Cabañero-Martínez M-J. Emergency crisis resource management: a simulation-based course developed by the Spanish Society of Emergency Medicine (SEMES) for health sciences students. *Emergencias*. 2024;36:41-7.
9. Khaw SC, Chung L, Fancourt N. Integrating non-technical skills into undergraduate medical simulation: a scoping review and thematic analysis of current practices. *Adv Simul (Lond)*. 2025;10:49.
10. Zamudio Burbano MA, González Giraldo D, López Agudelo LD, Casas Arroyave FD. Validación en castellano de la escala de Ottawa para habilidades no técnicas en personal de salud en situación de crisis. *Rev Esp Anestesiol Reanim*. 2021;68:523-30.
11. Sánchez-Marco M, Escribano S, Cabañero-Martínez MJ, Espinosa-Ramírez S, Muñoz-Reig MJ, Juliá-Sanchis R. Cross-cultural adaptation and validation of two crisis resource management scales. *Int Emerg Nurs*. 2021;57:101016.
12. World Health Organization. Global patient

safety action plan 2021-2030: towards eliminating avoidable harm in health care. Geneva: WHO; 2023.

13. Sauro KM, Machan M, Whalen-Browne L, Owen V, Wu G, Stelfox HT. Evolving factors in hospital safety: a systematic review and meta-analysis of hospital adverse events. *J Patient Saf*. 2021;17:e1285-95.
14. Guttman OT, Lazzara EH, Keebler JR, Webster KLW, Gisick LM, Baker AL. Dissecting communication barriers in healthcare: a path to enhancing communication resiliency, reliability, and patient safety. *J Patient Saf*. 2021;17:e1465-71.
15. Pereyra-Girardi CI, Raúl NB, Costa G, García S, de Echave JL, Muro M, et al. Escala Ottawa de Gestión Global en Crisis para equipos de salud. *Adaptación cultural argentina. Rev Educ Investig Emerg*. 2023;5(4). doi:10.24875/reie.22000278
16. Montero I, León OG. A guide for naming research studies in Psychology. *Int J Clin Health Psychol*. 2007;7:847-62.
17. Nunnally JC, Bernstein I. *Psychometric Theory*. 3a ed. Maidenhead, Inglaterra: McGraw Hill Higher Education; 1993.
18. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med*. 2002;21:1331-5.
19. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155-63.
20. Asociación Médica Mundial. Declaración de Helsinki de la AMM – Principios éticos para las investigaciones médicas en seres humanos [Internet]. 2024. (Consultado 30 Octubre 2025). Disponible en: <https://www.wma.net/es/politicas-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos/>
21. Skjong R, Wentworth B. Expert Judgement and risk perception. En: *In The Eleventh International Offshore and Polar Engineering Conference International Society of Offshore and Polar Engineers, Stavanger, Norway*. 2000.
22. Mccartland D, Berg M, Tebb SS, Lee ES, Rauch S. Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*. 2003;27:94-104.
23. Hughes AM, Smejdir K, Al-Abed S, Zaidan Z, Abualola F, Al-Hamad H. Human Factors in Healthcare: A Systematic Review of the Literature. *J of Patient Saf*. 2020;16:125-31.
24. Sánchez-Marco M, Escribano S, Rubio-Aparicio M, Juliá-Sanchis R, Cabañero-Martínez M-J. Effectiveness of nontechnical skills educational interventions in the context of emergencies: A systematic review and meta-analysis. *Aust Crit Care*. 2023;36:1159-71.
25. O'Connor P, O'Dea A. When I say ... non-technical skills. *Med Educ*. 2024;58:1439-41.
26. Franc JM, Noël A, Laberge M. Validation of the Italian version of the Ottawa Crisis Resource Management Global Rating Scale. *Simul Healthc*. 2016;11:249-55.
27. Ramirez-Roldan RR. Evaluación de habilidades no técnicas mediante la escala de Ottawa CRM GRS en escenarios de alta fidelidad en médicos residentes de áreas críticas [Internet]. [Perú]: Universidad Cayetano Heredia; 2025. (Consultado 30 Octubre 2025).

Disponible en: https://repositorio.uco.edu.pe/bitstream/handle/205001216/17638/Evaluacion_RamirezRoldan_Ruben.pdf?sequence=1

28. Prydz K, Dieckmann P, Fagertun H, Musson D, Wisborg T. Collecting evidence of validity for an assessment tool for Norwegian medi-

cal tudent non-technical skills (NOTS-NTS): usability and reliability when used by novice raters. BMC Med Educ. 2023;23:865.

29. Smith SE, McColgan-Smith S, Stewart F, Mar-don J, Tallentire VR. Beyond reliability: asses-sing rater competence when using a beha-

vioral metrics system. Adv Simul. 2022;9:55.

30. Arias A, Sireci S. Validez y Validación para Pruebas Educativas y Psicológicas: Teoría y Recomendaciones. Rev Iberoam Psicol. 2021;14:11-22.

Anexo 1. Escala Ottawa de Gestión Global en Crisis. Versión argentina adaptada por Pereyra-Girardi et al.

Dimensión	Ítems	Mal/nulo desempeño	Regular desempeño	Buen desempeño	Muy buen desempeño
Liderazgo	Mantiene la calma durante la crisis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Mantiene el control durante la crisis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Toma decisiones a tiempo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Toma decisiones seguras	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Mantiene una perspectiva global (visión amplia y en contexto)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Resolución de problemas	Resolución de problemas organizada y eficiente (secuencia ABCDE)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Resuelve los problemas a tiempo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Resuelve los problemas de modo eficiente	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Considera las diferentes alternativas durante la crisis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conciencia de situación	Evita el error de fijación	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Reevalúa la situación constantemente	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Replantea la situación	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Anticipa eventos probables	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Comunicación	Se comunica de forma clara	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Se comunica de forma concisa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Utiliza comunicación verbal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Utiliza comunicación no verbal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Utilización de recursos	Pide ayuda de forma adecuada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Utiliza los recursos disponibles adecuadamente	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Prioriza las tareas adecuadamente	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>