



**UNIVERSIDAD PERUANA  
CAYETANO HEREDIA**  
FACULTAD DE MEDICINA

Trabajo de investigación para optar por el  
grado académico de Bachiller en Medicina

## **Validación de una herramienta de evaluación de conocimientos y destrezas en el análisis de gases arteriales**

## **Validation of a tool for the evaluation of knowledge and skills in the analysis of arterial blood gases**

**Autores:**

**Herrera Rolla, Daniel Fernando  
Tamashiro Tovar, Julio Eduardo  
Zaldívar Facundo, Juan Fernando**

**Asesor:**

**Dr. Cristian León**

**2018**

## **Tabla de contenidos**

Resumen	2
Summary	3
Introducción	3
Material y métodos	6
Resultados	7
Discusión	9
Conclusiones	12
Declaración de conflictos de interés	13
Referencias bibliográficas	13

## **Resumen**

**Antecedentes:** La revisión de la literatura mostró que los niveles de conocimientos eran bajos en la interpretación gasométrica y las herramientas utilizados en estos estudios no estaban validados. **Objetivo:** El presente estudio busca validar una herramienta de evaluación de los conocimientos y destrezas en la interpretación de análisis de gases arteriales. **Materiales y métodos:** Se evaluó la validez de la herramienta por el método de agregados individuales por 4 expertos . Posterior a la validación se realizó una prueba piloto en 30 alumnos de la Universidad Peruana Cayetano Heredia que actualmente cursan el internado médico, con lo cual se determinó la confiabilidad mediante alfa de Cronbach, los índices de discriminación y dificultad, y la correlación entre la calificación en el test y autopercepción del nivel de conocimiento de los participantes. **Resultados:** Los expertos calificaron los ítems con puntajes de “bueno” y “excelente”, concluyendo en la apreciación general que el test es “bueno”. La concordancia entre los 4 evaluadores tuvo un Kappa de: 0.5309, 0.5294, 0.4008 y 0.3535 para congruencia, redacción, contenido y pertinencia respectivamente, estimándose una confiabilidad de 0.8125. La mayor parte de las preguntas de la herramienta tuvo un nivel de dificultad difícil, y un nivel de discriminación excelente. El coeficiente de correlación entre el desempeño en el test y la seguridad del nivel de conocimiento de los participantes fue de 0.58. **Conclusiones:** La herramienta diseñada es válida parcialmente y confiable, mostrando índices de discriminación adecuados, y de dificultad distintos a la literatura.

*Palabras clave: Estudio de validación, reproducibilidad, gasometría*

## **Summary**

**Background:** The review of the literature showed that the levels of knowledge in the interpretation of arterial blood gas analysis were low and the tools used in these studies were not validated. **Objective:** The present study seeks to validate a tool for evaluating knowledge and skills in the interpretation of arterial blood gas analysis. **Materials and methods:** The validity of the tool was evaluated by the method of individual aggregates by 4 experts. After the validation, a pilot test was carried out on 30 students of the Universidad Peruana Cayetano Heredia who are currently enrolled in the medical internship, which determined the reliability through Cronbach's alpha, the discrimination and difficulty indexes, and the correlation between the qualification in the test and self-perception of the level of knowledge of the participants. **Results:** The experts rated the items with scores of "good" and "excellent", concluding in the general assessment that the test is "good". The agreement between the 4 evaluators had a Kappa of: 0.5309, 0.5294, 0.4008 and 0.3535 for congruence, writing, content and relevance respectively, estimating a reliability of 0.8125. Most of the questions on the tool had a difficult level of difficulty, and an excellent level of discrimination. The correlation coefficient between the performance in the test and the level of knowledge of the participants was 0.58. **Conclusions:** The tool designed is partially valid and reliable, showing adequate indexes of discrimination, and difficulty different from the literature.

*Keywords: Validation studies, reproducibility of results, blood gas analysis.*

## **Introducción**

El análisis de gases arteriales es una herramienta necesaria y de uso cotidiano en diversos servicios de emergencia, hospitalización y cuidados críticos, puesto que es clave para el diagnóstico y el manejo de pacientes que presenten patologías que alteren su estado de oxigenación, hidroelectrolítico o balance ácido base, siendo su correcta interpretación de vital importancia (1). Los estudiantes que actualmente se encuentran en el internado médico han recibido entrenamiento en cuanto a análisis de gases arteriales, por lo cual resulta relevante conocer la habilidad que tienen para interpretar este examen bajo un contexto clínico.

Los trastornos ácido-base son una de las principales alteraciones presentes en diferentes enfermedades tales como la diabetes mellitus, la injuria renal o la insuficiencia respiratoria. Existen 4 tipos de trastornos divididos en acidosis o alcalosis, ya sea metabólica o respiratoria; entre estas la acidosis metabólica, una de las manifestaciones de la cetoacidosis diabética (2).

Además, se pueden clasificar en trastornos ácido-base simples o mixtos considerando la presencia de 2 o más alteraciones, siendo el más común la acidosis metabólica con alcalosis respiratoria.

Según los objetivos planteados dentro del plan curricular de la carrera profesional de medicina, los alumnos deben tener la capacidad de reconocer cuándo indicar e interpretar exámenes auxiliares, entre ellos el análisis de gases arteriales (3). Con el fin de evaluar sus conocimientos y destrezas, resulta beneficioso aplicar en ellos una herramienta de evaluación que se encuentre validada según los criterios establecidos por la comunidad científica.

La validación de una herramienta se basa en tres principios básicos: validez, confiabilidad y utilidad práctica. La validez se define como la capacidad de una herramienta de recolectar información coherente con los objetivos del estudio; y se puede dividir en tres: validez de contenido, de constructo y predictiva. La primera se refiere al dominio del estudio, que es medido por cada ítem de la herramienta para cada parámetro a evaluar. La segunda se refiere a la congruencia de lo recolectado con la idea construida por el autor. La última se refiere al grado de predicción de los resultados de la herramienta con respecto a una variable externa. La confiabilidad se define como la precisión con la que los ítems de la herramienta recolectan la información requerida por el investigador. Dependiendo del tipo de resultados que se esperan, existen diversos métodos para el cálculo de esta. La utilidad práctica abarca distintos factores, tales como: complejidad de la prueba, facilidad para su aplicación e interpretación, recursos económicos, entre otros (4).

Un estudio realizado por Austin y Jones (2010) a 80 médicos de emergencia sobre interpretación de análisis de gases arteriales utilizó una herramienta de evaluación con componentes objetivos, preguntas escritas sobre conocimientos, y subjetivos, una pregunta de autocrítica sobre cuán seguros estaban los participantes de sus respuestas, y los comparó para observar su relación; a su vez, comparó estos resultados con las respuestas de expertos sobre el tema, esto dio como resultado notas inferiores a los niveles esperados para expertos según el artículo, y además se observó que a pesar que los médicos con mayor experiencia tuvieran un mayor grado de confianza con respecto a su interpretación sobre las gasometrías, esto no se correlaciona con la nota obtenida (5). Ertok *et al* (2014) realizó también otro estudio parecido, en el que evaluó los conocimientos de médicos residentes de emergencia sobre interpretación de análisis de gases arteriales considerando factores como su entrenamiento previo y su año de residencia, dentro del cual se halló que a pesar del año de residencia, el número de respuestas

correctas era similar; pero el grupo que presentaba mayor número de respuestas correctas, eran residentes que se habían sometido a un entrenamiento corto para la interpretación de gasometrías durante su residencia (6). Ninguno de ellos, considerando el análisis sistemático hecho por cada uno, utilizó una herramienta validada.

Xafis (2014) realizó otro estudio similar al presente en el que utilizó un cuestionario de interpretación de gases arteriales en 30 médicos residentes de medicina de emergencia y 23 médicos asistentes, relacionando posteriormente la calificación obtenida con el nivel de satisfacción del entrenamiento en la lectura de análisis de gases arteriales y con el nivel de seguridad de cada médico con respecto a su respuesta. En el caso de los residentes, el nivel de seguridad, así como la calificación final, fueron ascendiendo con respecto al año de residencia, siendo ambas menores en los residentes de primer año y mayores en los de cuarto año. En general, el estudio encontró niveles insuficientes de conocimiento básico en lo que respecta al análisis de gases arteriales, así como una pobre destreza en la interpretación de los mismos (7). Asimismo, Sullivan y Jeavons (2005) realizaron una encuesta en 66 médicos; usando cinco ejemplos de análisis de gases arteriales, y encontraron que el 54% de los participantes identificaron correctamente los rangos normales de los valores mostrados en un análisis de gases arteriales, mientras que 71% encontraron la anormalidad mostrada en los ejemplos mencionados (8).

Como se puede observar en todos los estudios anteriormente citados, se encuentran niveles insuficientes en conocimiento e interpretación de los resultados del análisis de gases arteriales, así como también una falta de certeza en las herramientas utilizadas en los mismos.

Es por ello que se efectuará una validación parcial de una herramienta que permita una evaluación real de las habilidades y conocimientos de los estudiantes de medicina, para poder conocer el nivel en el que se encuentran los estudiantes con respecto a este ámbito.

## **Objetivos**

### *Objetivo General:*

1. Validar una herramienta para evaluar las destrezas y conocimientos de internos de medicina sobre el análisis e interpretación de los gases arteriales.

### *Objetivos específicos:*

1. Evaluar la validez de la herramienta diseñada a través del juicio de expertos.
2. Evaluar la confiabilidad de la herramienta en estudiantes de medicina de la Universidad Peruana Cayetano Heredia.

3. Determinar el Índice de discriminación y dificultad de las preguntas planteadas en la herramienta de evaluación.
4. Correlacionar la calificación de los participantes en el test con la autopercepción sobre su nivel de conocimiento en el análisis de gases arteriales.

### **Material y métodos**

El presente estudio, de carácter descriptivo y de corte transversal, se realizó con el fin de validar una herramienta que evalúe conocimientos y destrezas en estudiantes de medicina para la interpretación del análisis de gases arteriales, conocer la autopercepción del nivel de conocimientos de los mismos y realizar una comparación entre estas. Se diseñó una herramienta que busca medir aspectos objetivos con un cuestionario, y aspectos subjetivos con una encuesta con escala tipo Likert acerca de la autopercepción de conocimientos en la lectura e interpretación de los gases arteriales para el diagnóstico de los desórdenes ácido base. Dicho herramienta se basó en la estructura de la herramienta utilizada por Austin y Jones (2010) (5) y se dividió en dos partes: la primera, con 4 preguntas generales sobre las fórmulas a emplearse al interpretar un análisis de gases arteriales, y los rangos de normalidad de los valores obtenidos; y la segunda, con 1 caso ejemplo tomado del estudio ya mencionado y 4 casos similares creados por los autores del estudio del tema en cuestión, los cuales fueron evaluados y aprobados en una primera instancia por un médico nefrólogo del Hospital Cayetano Heredia. En esta segunda parte, posterior a cada caso, se empleó una escala Likert, midiendo qué tan seguros se sentían los participantes con respecto a sus respuestas, siendo 1 la menor seguridad y 5 la mayor seguridad.

Para poder validar una herramienta, posterior a su elaboración, se debe obtener la validez y la confiabilidad de la misma. La estimación de la validez de contenido y constructo se realizó por el método de agregados individuales; se invitó a médicos nefrólogos del hospital Cayetano Heredia con más de cinco años de experiencia en dicha especialidad, para valorar los ítems de la herramienta a través de un formato (Ver ANEXO N°1) en base a cuatro criterios: contenido, congruencia, redacción y pertinencia. Así mismo, se valoró el índice Kappa, dato que sirve para evaluar la concordancia entre evaluadores por medio del programa de análisis estadístico STATA v14.0 (9).

Se realizaron las modificaciones planteadas por los expertos, luego de lo cual se procedió a la aplicación de la herramienta (Ver ANEXO N°2) en un grupo de 30 estudiantes de medicina de

la Universidad Peruana Cayetano Heredia elegidos aleatoriamente que cumplan con los criterios de inclusión del estudio (estar cursando el internado de la carrera de medicina de la Universidad Peruana Cayetano Heredia con matrícula regular). Esta evaluación fue realizada por los autores del estudio, previa firma del consentimiento informado para formar parte del estudio (Ver ANEXO N°3). Luego, se procedió al llenado de una base de datos para evaluar las características de la herramienta a través de un análisis psicométrico por medio de STATA v14.0 y Microsoft Excel v.2016. Se determinaron los índices de discriminación y dificultad de cada uno de los ítems de la herramienta, además de la confiabilidad del mismo (10). Para interpretar dichos índices de discriminación y dificultad, se tomaron los parámetros establecidos por Ortiz *et al.* (2015) (11). Por último, se calculó el coeficiente de correlación de Pearson entre las notas obtenidas y la autopercepción del nivel de conocimientos de cada participante extraída por medio de la escala Likert descrita anteriormente. La herramienta según lo estipulado por Kirkpatrick (1979) nos permite evaluar el primer nivel de aprendizaje, es decir la reacción de los participantes al conocimiento aprendido; y además por medio de los ítems que medían la interpretación de los hallazgos gasométricos y la habilidad al realizar el diagnóstico diferencial con respecto a los casos planteados, se evaluó de manera indirecta el segundo nivel de aprendizaje (12).

## **Resultados**

La validez del estudio se realizó bajo el método de agregados individuales; 4 expertos en el tema evaluaron el test bajo los criterios de redacción, contenido, congruencia y pertinencia, asignándole a cada uno de los 51 ítems valores de excelente, bueno, mejorar, cambiar y eliminar según su juicio. Finalmente, se les pidió dar una apreciación general, señalando si el test les impresionaba ser deficiente, aceptable, bueno o excelente. En primera instancia, se obtuvo el promedio de las puntuaciones para cada criterio por evaluador, obteniéndose “bueno” y “excelente” en la totalidad de los ítems.

Posteriormente, se halló el valor del índice Kappa de concordancia por pares entre los 4 evaluadores obteniéndose un Kappa promedio de: 0.7022 y 0.5309 para congruencia, 0.7234 y 0.5294 para redacción, 0.4754 y 0.4008 para contenido y 0.5552 y 0.3535 para pertinencia respectivamente (Ver tabla n°1). Finalmente, en la apreciación general, todos los expertos participantes estuvieron de acuerdo en que el test era “bueno”.

Luego para evaluar la confiabilidad, se aplicó la herramienta como prueba piloto en 30 estudiantes de medicina de la Universidad Peruana Cayetano Heredia que cursan el internado médico con matrícula regular. Fueron 42 ítems dentro de la herramienta evaluados en total, estimándose una confiabilidad de 0.8125 mediante el uso del coeficiente alfa de Cronbach.

De los 30 estudiantes a los que se les aplicó el test, 16 son varones y 14 mujeres, con un promedio de edad de  $23.8 \pm 1.67$  años. El 80% de los evaluados refiere que la metodología de enseñanza de mayor utilidad en cuanto a interpretación de análisis de gases arteriales es la práctica clínica, el 10% las clases teóricas, el 6.67% la lectura de textos y el 3.33% restante el método de aprendizaje basado en problemas. En cuanto a la calidad de este entrenamiento durante pregrado, el 63.33% considera que fue aceptable, el 26.67% muy aceptable y el 10% restante, deficiente.

De todo el grupo, el 26.67% resultó aprobado, mientras que el 73.33% restante desaprobó. La nota promedio fue de 9.27, siendo la más baja 6.05 y la más alta 15.35. Con respecto a la autopercepción del nivel de conocimientos de los participantes, se calculó el coeficiente de correlación entre las notas obtenidas en el test y los valores de autopercepción extraídos con la escala Likert, resultando en 0.58, que es una correlación positiva entre las dos variables (ver gráfico n°1).

Se realizó la estimación de los índices de discriminación y dificultad para cada ítem de la herramienta. Los índices de dificultad tuvieron un resultado menor de 0.40 para preguntas difíciles; entre 0.41 y 0.50 para preguntas medianamente difíciles; entre 0.51 y 0.80 para preguntas de dificultad media; entre 0.81 y 0.90 para preguntas medianamente fáciles; entre 0.91 y 1.00 para preguntas fáciles según lo propuesto por Ortiz *et al.* (2015) (11). Se analizaron los ítems desde el 3 hasta el 51, exceptuando el 15,16, 24, 32 y 42 los cuales corresponden a enunciados de los casos y a la seguridad del alumno con respecto a sus respuestas. Dentro de este análisis, se encontró que la dificultad de las preguntas ascendía progresivamente dentro de cada caso, hallándose mayores índices de dificultad en las primeras preguntas de cada caso, es decir mayor facilidad, y menores índices en las últimas preguntas de cada caso, es decir mayor dificultad (ver gráfico n°2).

Además, se realizó la distribución de ítems dependiendo del índice de dificultad obtenido como se puede observar en gráfico n°3, la herramienta cuenta con el 48.8% de los ítems con un nivel de dificultad difícil, 7.0% medianamente difícil, 25.6% con dificultad media, 9.3% con dificultad medianamente fácil y el 9.3% con dificultad fácil.



Posteriormente se realizó el análisis del índice de discriminación, el cual permite discernir los ítems que permiten separar a los estudiantes con mejor desempeño de los estudiantes con peor desempeño. Los resultados del índice de discriminación varían desde menor a cero, de 0 hasta 0.14, de 0.15 hasta 0.25, de 0.26 hasta 0.35 y mayor a 0.35, dando resultados de discriminación negativa, pobre discriminación, regular discriminación, buena discriminación y excelente discriminación, tomados de Ortiz *et al* (2015) (11). En el gráfico n°4, se puede observar nuevamente una tendencia en el índice de discriminación mientras se va avanzando en las preguntas de casos incluidas en la herramienta, llegando en algunos casos a tener un índice de discriminación de cero, como los casos del ítem 8, 14, 17, 31, 33, 40, 41, 50 y 51.

En gráfico n°5, se presentan la distribución de ítems dependiendo de su nivel de discriminación, en el cual el 44.2%, 2.3%, 14.0%, 37.2%, 2.3% cuentan con un poder discriminatorio excelente, bueno, regular, pobre y negativo respectivamente. El ítem 11 fue el único que contó con un poder discriminativo bueno, es decir un índice de discriminación entre 0.26 y 0.35. Hubo un ítem que presentó un índice de discriminación negativo, este fue el ítem 43.

## **Discusión**

La validación de una herramienta involucra tres partes, como se mencionó anteriormente: la validez, la confiabilidad y la utilidad clínica; en este trabajo se analizaron las dos primeras. Para considerar válida la herramienta se utilizó el método de agregados individuales, que consta de una revisión de cada ítem del test por parte de los 4 expertos nefrólogos; estos expertos evaluaron cada uno de ellos y se obtuvieron, dentro de los cuatro criterios para el análisis (congruencia, redacción, contenido y pertinencia), puntuaciones de “bueno” y “excelente” con una observación mínima sobre la redacción de los ítems 4 y 6, lo cual nos permite constatar la validez de constructo. Para darle mayor consistencia a este aspecto, se halló el índice kappa de concordancia entre evaluadores.

Según la literatura, puede considerarse como aceptable un índice Kappa mayor a 0.21 y va siendo mayor a medida que se acerca a 1 (9). Se puede observar que hay una diferencia entre los valores del índice Kappa obtenidos, resultando mayor cuando se realiza entre 3 expertos, esto debido a que, siendo el análisis de los criterios en cierta medida subjetivo y al haber mayor cantidad de evaluadores, hay menor probabilidad de una concordancia considerable. A pesar de esto, se puede interpretar que sí existen coincidencias importantes entre los expertos revisores del test, existiendo una concordancia “aceptable” (Kappa entre 0.21 y 0.40) para el

contenido y pertinencia, y una concordancia “moderada” (Kappa entre 0.41 y 0.60) para la congruencia y redacción.

La confiabilidad de una herramienta se define como la exactitud con la que dicha herramienta mide lo que debe medir. Para estimar la confiabilidad, se debe aplicar la herramienta como prueba piloto en un grupo pequeño (14-30 sujetos) no pertenecientes a la muestra de estudio, pero que sí pertenecen a la población o que cumplan características similares a dicha muestra. Posterior a la aplicación de la herramienta, existen diferentes métodos para estimar la confiabilidad: método de test-retest, hemitest, división por mitades de Rulon, coeficiente alfa de Cronbach, entre otros (4). El método utilizado en el presente trabajo fue el coeficiente alfa de Cronbach, con el cual se estimó una confiabilidad de 0.8125.

Según Corral (2009), los valores de confiabilidad se pueden interpretar de la siguiente manera: 0.01 a 0.20 es muy baja, 0.21 a 0.40 es baja, 0.41 a 0.60 es moderada, 0.61 a 0.80 es alta, y 0.81 a 1.00 es muy alta (4). Como antes ya se mencionó, la confiabilidad de la herramienta designada en este trabajo fue estimada en 0.8125, con lo cual se afirma que la herramienta es muy altamente confiable.

La utilidad práctica de la herramienta, para otorgarle una validez completa a la herramienta, podrá ser evaluada en un futuro al aplicarla en diversos contextos académicos y con una población adecuada, como por ejemplo al finalizar el curso de clínica médica 1, previa al externado o internado médico, entre otras.

Una vez realizada la validez y confiabilidad de la herramienta, se procedió a la evaluación psicométrica de la herramienta como parte de un objetivo secundario, la cual permitirá una futura valoración del puntaje para cada ítem; cabe resaltar que ésta no significó ningún cambio en la estructura ni contenido de la herramienta. Se realizó por medio del cálculo de los índices de dificultad y discriminación de cada ítem con preguntas objetivas del test, siendo 43 ítems los evaluados por los dos índices mencionados previamente. Inicialmente se realizó el cálculo del índice de dificultad de cada ítem como lo plantea Ortiz *et al.* (2015) (11), dando como resultado un 48.8% de preguntas difíciles, 7.0% de preguntas medianamente difíciles, 25.6% de preguntas de dificultad media, 9.3% de preguntas medianamente fáciles y 9.3% de preguntas fáciles; dentro de las preguntas difíciles se pudieron observar seis preguntas que no fueron respondidas por ningún participante, las cuales fueron en su mayor parte, ítems que evaluaban la habilidad de presunción diagnóstica del mismo; en los ítems con nivel de dificultad fácil, se encontró que solo hubieron 3 ítems en los cuales el 100% de los alumnos respondió

correctamente. Lo resaltante de este resultado es que se esperaba que los estudiantes pudieran reconocer el trastorno ácido base en los 5 casos planteados, pero se contaron con estudiantes que respondieron incorrectamente en los ítems 25 y 43 (Ver ANEXO N°2).

Según Ortiz *et al.* (2015), los ítems deben estar distribuidos en 5%, 20%, 50%, 20% y 5% con un nivel de dificultad fácil, medianamente fácil, media, medianamente difícil y difícil respectivamente (11). Por ejemplo, Backhoff *et al.* (2000) realizó la evaluación psicométrica de un test, donde evidenció que la media del nivel de dificultad de su herramienta era 56% (10). Estos resultados difieren con los del estudio en donde se encontró un porcentaje del 25.6% y 48.8% para ítems con dificultad media y difícil respectivamente; esto puede deberse a que 6 ítems no fueron contestados por ninguna persona, es decir el 14.0% del total de ítems, hecho que afecta la distribución de la dificultad de las preguntas debido al bajo desempeño que presentaron los estudiantes en la prueba piloto. Un punto final a tomar en cuenta acerca de los índices de dificultad encontrados en el estudio es la tendencia vista en el gráfico n°2, la cual sufre una disminución drástica al llegar a las preguntas de determinación del trastorno ácido base y presunción diagnóstica, hasta en algunos ítems como se ha planteado anteriormente llegando a 0%, lo cual podría denotar habilidades diagnósticas deficientes de los participantes de medicina; sin embargo, sería necesario realizar un estudio con mayor número de participantes.

El índice de discriminación se utiliza para evaluar la capacidad de un ítem de identificar a los evaluados con mejor desempeño de los evaluados con peor desempeño (11). En la herramienta utilizada en el estudio, se obtuvieron el 44.2% de ítems con un poder discriminatorio excelente, 2.3% con buen poder discriminatorio, 14% con regular poder discriminatorio, 37.2% con pobre poder discriminatorio y 2.3% con poder discriminatorio negativo. Se presenta un número elevado de ítems con poder discriminatorio pobre, esto debido a la presencia de ítems que evalúan conocimiento que se esperaba que el estudiante recordara por su naturaleza memorística. En la mayoría de los ítems de la herramienta, se cuenta con un poder discriminativo excelente, dentro de los cuales el de mayor valor es 0.81. El índice discriminatorio promedio del test es de 0.28 lo cual le da un nivel de discriminación bueno, comparable con lo encontrado por Backhoff *et al.* (2000), quien presentó un índice de discriminación de 0.34, el cual también es de poder discriminativo bueno (10). De la totalidad de los ítems, sólo uno presentó un índice de discriminación negativa; esto fue debido a que solamente un participante no respondió a la pregunta del ítem 43, por ello se obtuvo un

resultado negativo, ya que la nota promedio de ese participante era mayor que la nota promedio de los 29 participantes restantes.

Por último, con el promedio de nota obtenido por los participantes de esta prueba piloto, si bien no representa al universo de los estudiantes de último año de medicina, se podría inferir que la preparación y entrenamiento recibido en años previos no resulta suficiente para una correcta interpretación del análisis de gases arteriales, ya que más del 70% resultó desaprobado. Gracias al coeficiente de correlación se observa una tendencia por parte de los participantes, si bien no es marcada, a obtener mayor puntaje al presentar mayor seguridad de sus respuestas.

Dentro de las limitaciones del estudio, no se pudo realizar el cálculo de la utilidad práctica, lo cual le da una validez parcial a la herramienta; así mismo no se pudo realizar el método de test-retest para la confirmación de los resultados obtenidos en la prueba piloto. Otra limitación de la herramienta fue que no permite evaluar los niveles tres y cuatro estipulados por Kirkpatrick, es decir el cambio en el comportamiento y los resultados de los conocimientos adquiridos.

Esta herramienta podría resultar beneficiosa al ser incorporada en diversos cursos de clínica médica para evaluar el nivel de los alumnos en distintas etapas de su formación y así propiciar un mejor desempeño en las prácticas preprofesionales.

### **Conclusiones**

La herramienta fue validada parcialmente por medio de juicio de expertos y resultó muy altamente confiable. La mayor proporción de ítems tuvo un poder discriminatorio excelente, característica ideal en una herramienta de evaluación según lo encontrado en la literatura. En cuanto al índice de dificultad, la mayoría de ítems fueron difíciles, hallazgos que difieren a lo encontrado en la literatura, en donde se evidencia que la mayor proporción de ítems sea de dificultad media. Esto último puede deberse al bajo nivel de habilidad diagnóstica de los participantes. El coeficiente de correlación de Pearson entre la calificación de los participantes en la herramienta y su autopercepción del nivel de conocimientos obtuvo una correlación positiva leve.

### **Recomendaciones**

Se encontró cierto nivel de dificultad en los ítems con interpretación de los hallazgos gasométricos y de diagnóstico diferencial, por lo cual se recomienda la realización de un estudio con una muestra representativa para poder determinar un estimado real del nivel de

conocimiento que poseen los estudiantes de medicina próximos a graduarse acerca de la interpretación del análisis de gases arteriales.

### **Declaración de conflictos de interés**

Los autores del estudio declaran que no existen conflictos de intereses.

### **Referencias bibliográficas**

1. Sood P, Paul G, Puri S. Interpretation of arterial blood gas. *Indian Journal of Critical Care Medicine*. 2010; 14(2):57.
2. Kose A, Armagan E, Oner N, Koksall O, Kostak Mert D, Ozdemir F et al. Acid-Base Disorders in the Emergency Department: Incidence, Etiologies and Outcomes. *Journal of Academic Emergency Medicine*. 2014; 13(1):4-9.
3. Universidad Peruana Cayetano Heredia. Plan curricular carrera profesional de medicina. Lima: Vicerrectorado Académico de la Universidad Peruana Cayetano Heredia. 2016; p. 21-22.
4. Corral Y. Validez y confiabilidad de los instrumentos de investigación para la recolección de datos. *Revista Ciencias de la Educación*. 2009; 19: 228-247.
5. Austin K, Jones P. Accuracy of interpretation of arterial blood gases by emergency medicine doctors. *Emergency Medicine Australasia*. 2010;22(2):159-165.
6. Ertok I, Kurtoglu Celik G, Sahin Kavakli H, Dogan N, Icme F, Becel S et al. Evaluation of Emergency Medicine Residents' Level of Knowledge of Arterial Blood Gases. *Journal of Academic Emergency Medicine*. 2014; 13:100-103.
7. Xafis P. Arterial Blood Gases in Emergency Medicine: how well do our registrars and consultants currently enrolled in the Western Cape Division of Emergency Medicine interpret them [MMed Degree]. Stellenbosch University; 2014.
8. O'Sullivan I. Survey of blood gas interpretation. *Emergency Medicine Journal*. 2005; 22(5):391-392.
9. Cerda J, Villarroel L. Evaluación de la concordancia inter-observador en investigación pediátrica: Coeficiente de Kappa. *Revista chilena de pediatría*. 2008; 79(1).
10. Backhoff E, Larrazolo N, Rosas M. Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *REDIE Revista*

Electronica de Investigación Educativa [Internet]. 2000 [cited 30 March 2018];2(1). Available from: <http://www.redalyc.org/articulo.oa?id=15502102>

11. Ortiz G, Díaz P, Llanos O, Pérez S, González K. Difficulty and discrimination of the items of the exams of Research Methodology and Statistics. Rev EDUMECENTRO [Internet]. 2015 [cited 27 March 2018];7(2):19-35. Available from: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2077-28742015000200003&lng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2077-28742015000200003&lng=es).
12. Kirkpatrick DL. Techniques for evaluating training programs. Training and development journal. June 1979: 178-192

## Tablas y gráficos

Tabla n°1. Índice de Kappa de cada criterio por 3 y 4 evaluadores.

	Congruencia		Redacción		Contenido		Pertinencia	
	3 Evaluadores	4 Evaluadores	3 Evaluadores	4 Evaluadores	3 Evaluadores	4 Evaluadores	3 Evaluadores	4 Evaluadores
Kappa	0.7022	0.5309	0.7234	0.5294	0.4754	0.4008	0.5552	0.3535
Z	8.69	9.29	8.95	9.26	5.88	7.01	9.87	6.18

Gráfico n°1. Correlación entre notas y autopercepción del nivel de conocimiento de los participantes.

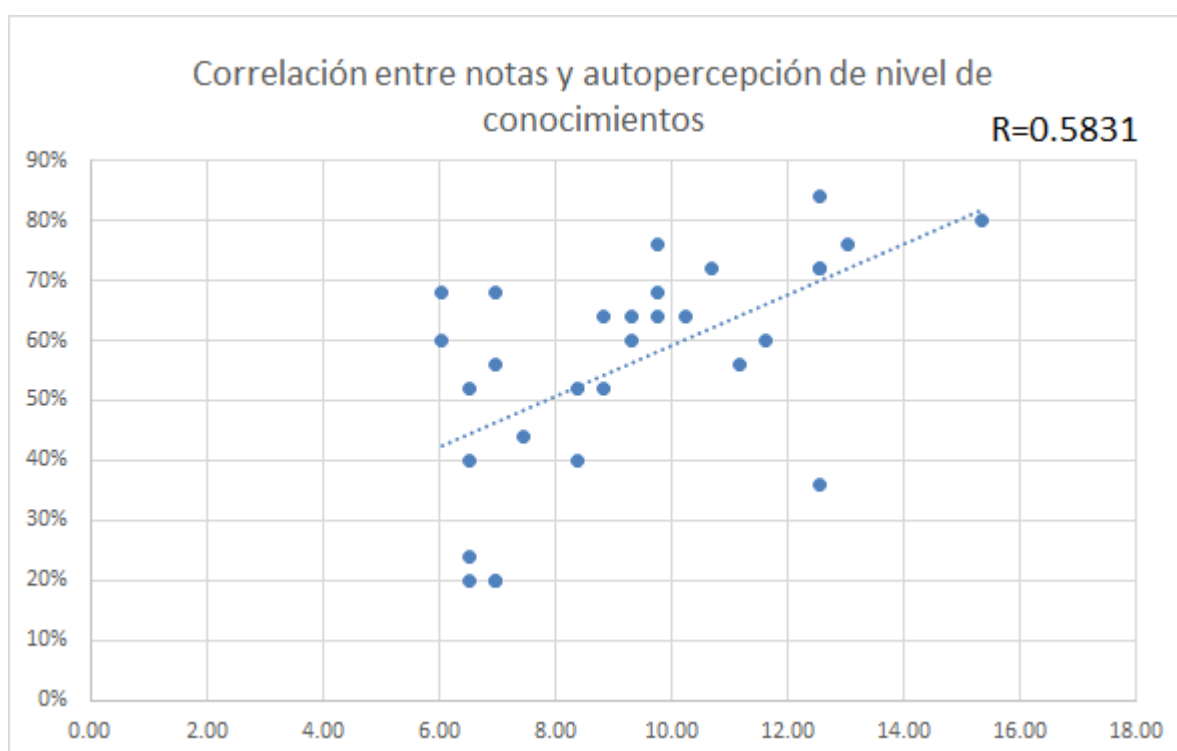


Gráfico n°2. Índice de dificultad de cada ítem evaluado.

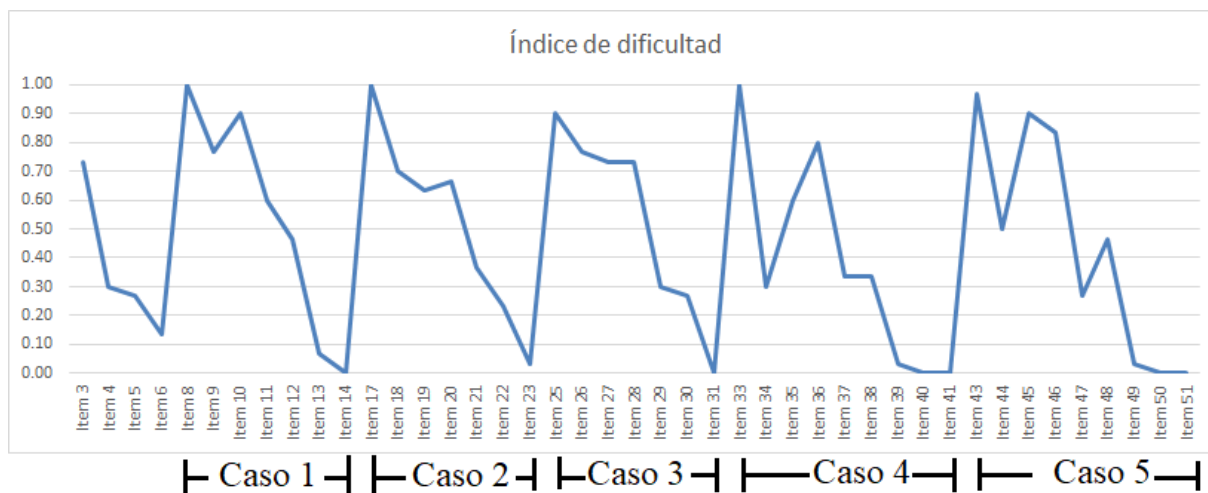


Gráfico n°3. Distribución de ítems según nivel de dificultad

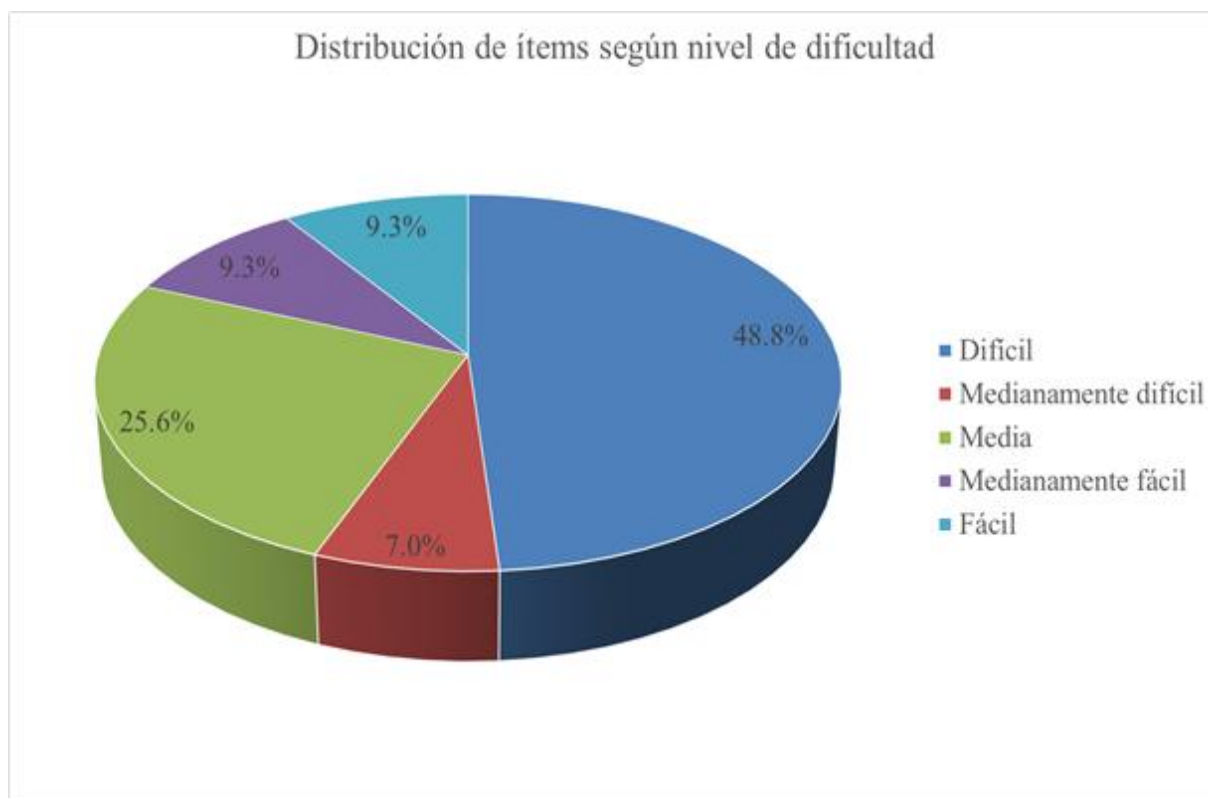




Gráfico n°4. Índice de discriminación según ítem.

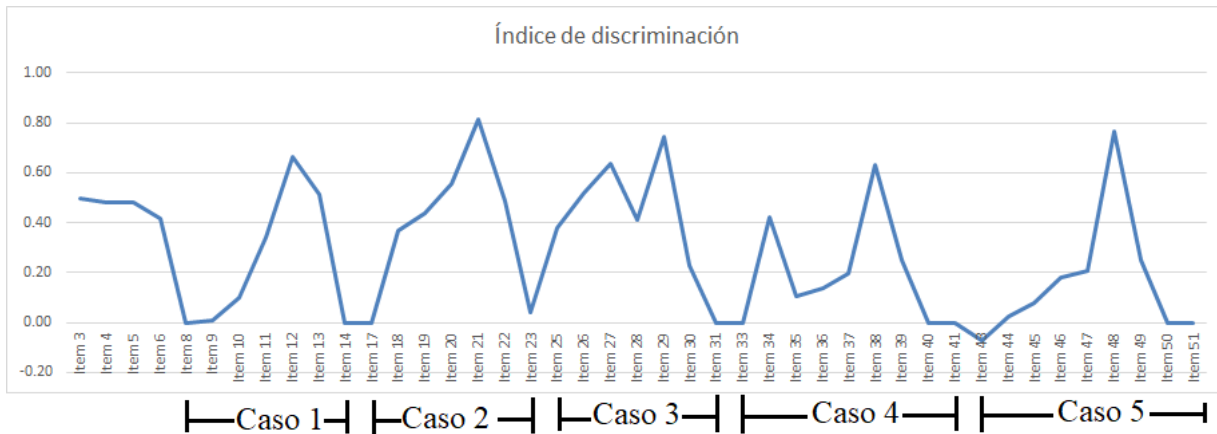
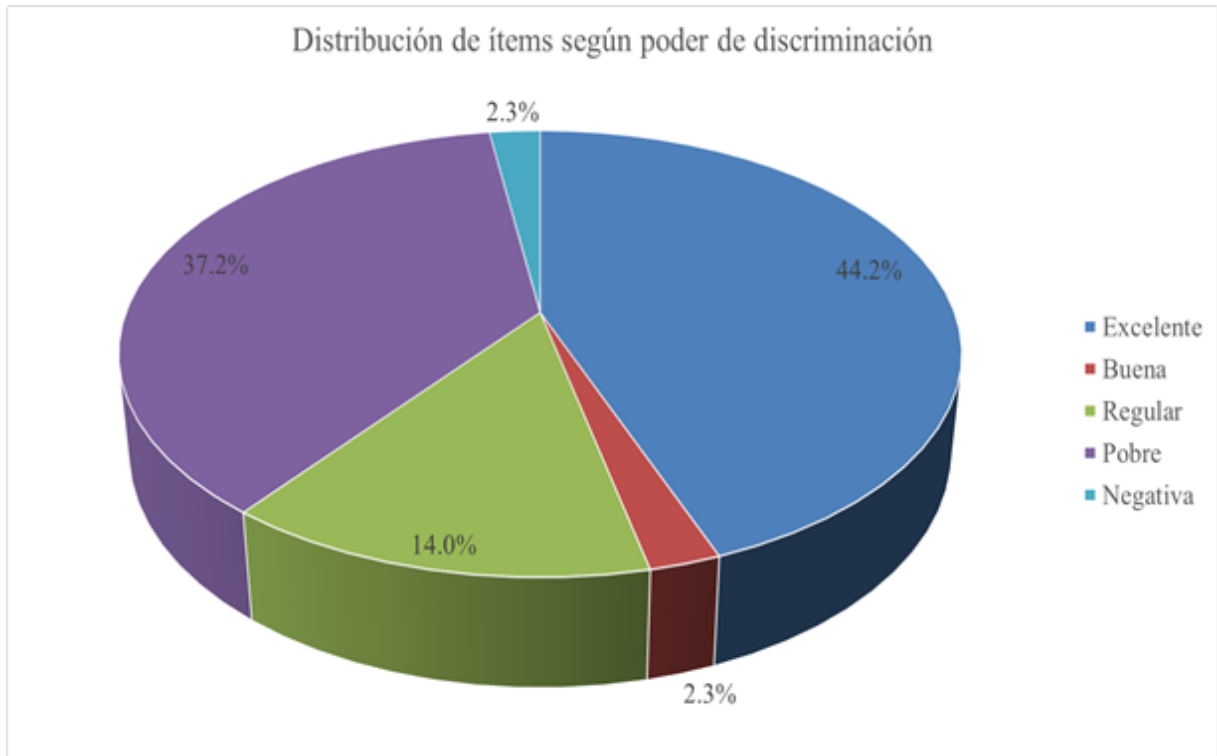


Gráfico n°5. Distribución de ítems según poder de discriminación.



## ANEXO N°1

### Formato de validación por expertos

Estimado validador:

Nos es grato dirigirnos a Usted, a fin de solicitar su colaboración como experto para validar el cuestionario anexo, el cual será aplicado al proyecto: *Validación de una herramienta de evaluación de conocimientos y destrezas en el análisis de gases arteriales*; por cuanto consideramos que sus observaciones y aportes serán de utilidad. Para ello, incluimos los objetivos del estudio, así como un cuadro de operacionalización de las variables que han sido tomadas en cuenta.

Para efectuar la validación de la herramienta, Usted deberá leer cuidadosamente cada enunciado y sus correspondientes alternativas de respuesta. Por otra parte, se le agradece cualquier sugerencia relativa a redacción, contenido, pertinencia y congruencia u otro aspecto que se considere relevante para mejorar el mismo. Gracias por su aporte.

#### Instrucciones:

Los enunciados dentro del cuestionario tienen una asignación de número de ítem, la cual se encuentra al final de cada pregunta. Coloque en cada casilla la letra correspondiente al aspecto cualitativo que le parece que cumple cada ítem dentro del cuestionario según los criterios que a continuación se detallan.

**E**= Excelente / **B**= Bueno / **M**= Mejorar / **X**= Eliminar / **C**= Cambiar

Las categorías por evaluar son redacción, contenido, congruencia y pertinencia. En la casilla de observaciones puede sugerir algún cambio o correspondencia.

Al culminar la valoración, por favor leer y completar los siguientes datos:

*Por medio de la presente hago constar que he revisado, con fines de validación de la herramienta, el cuestionario proporcionado, a los efectos de su aplicación en el proyecto descrito con anterioridad.*

*Luego de hacer las observaciones pertinentes, puedo concluir que el herramienta en cuestión es:*

**DEFICIENTE**

**ACEPTABLE**

**BUENO**

**EXCELENTE**

**Nombres y apellidos:**

**CMP:**

**Firma:**

## Anexo N°1 Objetivos del estudio

### *Objetivo General:*

1. Validar una herramienta para evaluar las actitudes, destrezas y conocimientos de internos de medicina sobre el análisis e interpretación de los gases arteriales.

### *Objetivos específicos:*

1. Evaluar la validez de la herramienta diseñada a través del juicio de expertos.
2. Evaluar la confiabilidad de la herramienta en estudiantes de medicina de una universidad privada.
3. Determinar el Índice de Discriminación y dificultad de las preguntas planteadas en la herramienta de evaluación.
4. Correlacionar el rendimiento del grupo en el test con la autopercepción sobre su nivel de conocimiento en el análisis de gases arteriales.

### Cuadro de operacionalización de las variables

<i>VARIABLE</i>	<i>DEFINICION</i>	<i>DIMENSIONES O TIPO</i>	<i>INDICADORES</i>	<i>ESCALA DE MEDICION</i>
Validez	Capacidad de una herramienta de recolectar información coherente con los objetivos del estudio	Contenido	Opinión de expertos De acuerdo con el juicio de expertos, se revisará: redacción, contenido, congruencia y pertinencia.	Ordinal <b>E</b> = Excelente <b>B</b> = Bueno <b>M</b> = Mejorar <b>X</b> = Eliminar <b>C</b> = Cambiar
		Constructo		Ordinal <b>E</b> = Excelente <b>B</b> = Bueno <b>M</b> = Mejorar <b>X</b> = Eliminar <b>C</b> = Cambiar
Confiabilidad	Exactitud con la que los ítems de la herramienta, recolectan la información requerida por el investigador	Confiabilidad	Coeficiente de confiabilidad	Continua

PREGUNTAS	CRITERIOS A EVALUAR				OBSERVACIONES
Nº ITEM	REDACCIÓN	CONTENIDO	CONGRUENCIA	PERTINENCIA	
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
...					
51					

## ANEXO N°2

Edad:

Sexo:

Colegio de procedencia:

El siguiente test pretende conocer la actitud, habilidad y destreza en cuanto a la toma e interpretación del análisis de gases arteriales; por tanto, se pide responder, con total sinceridad.

---

Considera usted que su entrenamiento en el análisis e Interpretación de Gases Arteriales durante su estancia en pregrado ha sido (i1):

Muy Deficiente ( )

Deficiente ( )

Aceptable ( )

Muy Aceptable ( )

Excelente ( )

Cuál ha sido la metodología de enseñanza que le ha permitido lograr los mejores resultados en su entrenamiento en la lectura de Gases Arteriales (i2):

Clases ( )

Libros ( )

Prácticas Clínicas ( )

ABP ( )

TBL ( )

Rotaciones electivas ( )

Parte 1: Preguntas generales

1. Escriba la ecuación del anión gap (i3):
2. Escriba los valores normales de anión gap (i4):
3. Escriba la Ecuación del Anión GAP urinario (i5):
4. Escriba los valores normales Ecuación del Anión GAP urinario (i6):

## Parte 2: Interpretación

A continuación, se presentan cinco casos distintos con sus resultados de gasometría arterial. Finalizando el cuestionario correspondiente a cada caso, se encuentra una pregunta que evalúa cuan seguro se siente usted sobre sus respuestas, siendo (1) lo mínimo y (5) lo máximo.

*Caso 1:* Un adolescente con antecedente de raquitismo, deformaciones óseas, así como litiasis renal y nefrocalcinosis, presenta un cuadro de infección urinaria y tiene la siguiente gasometría (i7)

pH	6.930	Na	138mmol/L
pCO <sub>2</sub>	7 mmHg	K	1.9 mmol/L
pO <sub>2</sub>	168mmHg	Cl	127mmol/L
HCO <sub>3</sub> <sup>-</sup>	4.8mmol/L		
FiO <sub>2</sub>	0.45mmol/L		

1. ¿Existe un trastorno ácido base? (i8) SI / NO
2. Si existe un trastorno, ¿cuál es la alteración primaria? (i9)  
RESPIRATORIA / METABÓLICA / ACIDEMIA MIXTA / ALCALEMIA
3. Si existe una alteración, ¿cuál es su grado de severidad? (i10)  
LEVE / MODERADA / SEVERA
4. ¿Existe compensación? (i11) SI / NO
5. Si se tienen todos los componentes, escriba el valor de anión gap (i12): .....
6. Escriba el Diagnóstico Ácido Base (i13)  
.....
7. En cada recuadro coloque el diagnóstico en orden de probabilidad donde 1 es menos probable y 5 es el más probable (i14)
  - Cetoacidosis Diabética ( )
  - Acidosis Tubular Renal ( )
  - Injuria Renal Aguda ( )
  - Diarrea Aguda ( )
  - Intoxicación por sustancias desconocidas ( )

¿Qué tan seguro cree estar de sus respuestas? (i15)

- (1) (2) (3) (4) (5)

*Caso 2:*

Paciente de 30 años que fue encontrado inconsciente en la calle es traído a la emergencia, se evidencia respiración anormal y taquipnea. Se toma un AGA que muestra los siguientes valores (i16):

pH	7.280	Na	130mmol/L
pCO <sub>2</sub>	31mmHg	K	3.8mmol/L
pO <sub>2</sub>	83mmHg	Cl	101mmol/L
HCO <sub>3</sub> <sup>-</sup>	15mmol/L	Lactato	1 mmol/L
FiO <sub>2</sub>	0.21mmol/L		

1. ¿Existe un trastorno ácido base? (i17) SI / NO
2. Si existe un trastorno, ¿cuál es la alteración primaria? (i18)  
RESPIRATORIA / METABÓLICA / ACIDEMIA MIXTA / ALCALEMIA
3. Si existe una alteración, ¿cuál es su grado de severidad? (i19)  
LEVE / MODERADA / SEVERA
4. ¿Existe compensación? (i20) SI / NO
5. Si se tienen todos los componentes, escriba el valor de anión gap (i21):
6. Escriba el Diagnóstico Ácido Base (i22)  
.....
8. En cada recuadro coloque el diagnóstico en orden de probabilidad donde 1 es menos probable y 5 es el más probable (i23)
  - Cetoacidosis Diabética ( )
  - Acidosis Tubular Renal ( )
  - Injuria Renal Aguda ( )
  - Diarrea Aguda ( )
  - Intoxicación por sustancias desconocidas ( )¿Qué tan seguro cree estar de sus respuestas?  
(1) (2) (3) (4) (5)

*Caso 3:*

A continuación, se presenta los resultados de AGA de un paciente de 19 años que acude a emergencias por presentar 3 días de diarrea líquida de abundante cantidad al que se le agrega trastorno de sensorio 5 horas antes del ingreso. (i24)

pH	7.320	Na	130mmol/L
pCO <sub>2</sub>	35mmHg	K	4.0mmol/L
pO <sub>2</sub>	90mmHg	Cl	108mmol/L
HCO <sub>3</sub> <sup>-</sup>	18mmol/L	FiO <sub>2</sub>	0.21mmol/L

- ¿Existe un trastorno ácido base? (i25) SI / NO
- Si existe un trastorno, ¿cuál es la alteración primaria? (i26)  
RESPIRATORIA / METABÓLICA / ACIDEMIA MIXTA / ALCALEMIA
- Si existe una alteración, ¿cuál es su grado de severidad? (i27)  
LEVE / MODERADA / SEVERA
- ¿Existe compensación? (i28) SI / NO
- Si se tienen todos los componentes, escriba el valor de anión gap (i29):
- Escriba el Diagnóstico Ácido Base (i30)  
.....
- En cada recuadro coloque el diagnóstico en orden de probabilidad donde 1 es menos probable y 5 es el más probable (i31)
  - Cetoacidosis Diabética ( )
  - Acidosis Tubular Renal ( )
  - Injuria Renal Aguda ( )
  - Diarrea Aguda ( )
  - Intoxicación por sustancias desconocidas ( )

¿Qué tan seguro cree estar de sus respuestas?

- (1)      (2)      (3)      (4)      (5)

*Caso 4:*

Un alumno de cuarto año que se encuentra rotando por su servicio se acerca a preguntarle sobre los resultados de AGA de un paciente de 54 años con insuficiencia renal aguda (i32)

pH	7.15	Na	139mmol/L
pCO <sub>2</sub>	34mmHg	Cl	98 mmol/L
HCO <sub>3</sub> <sup>-</sup>	12 mmol/L		

- ¿Existe un trastorno ácido base? (i33) SI / NO
- Si existe un trastorno, ¿cuál es la alteración primaria? (i34)



RESPIRATORIA / METABÓLICA / ACIDEMIA MIXTA / ALCALEMIA

3. Si existe una alteración, ¿cuál es su grado de severidad? (i35)

LEVE / MODERADA / SEVERA

4. ¿Existe compensación? (i36) SI / NO

5. Existe Tercer Trastorno (i37) SI / NO

6. Si se tienen todos los componentes, escriba el valor de anión gap (i38): .....

7. Describa el Tercer Trastorno (i39) .....

8. Escriba el Diagnóstico Ácido Base (i40)

.....

9. En cada recuadro coloque el diagnóstico en orden de probabilidad donde 1 es menos probable y 5 es el más probable (i41)

- Cetoacidosis Diabética ( )
- Acidosis Tubular Renal ( )
- Injuria Renal Aguda ( )
- Diarrea Aguda ( )
- Intoxicación por sustancias desconocidas ( )

¿Qué tan seguro cree estar de sus respuestas?

(1) (2) (3) (4) (5)

*Caso 5:*

Se tienen los resultados de AGA de un paciente de 74 años con antecedente de EPOC que acude a emergencia por sensación de falta de aire y refiere deposiciones líquidas hace 4 días con fiebre (i42)

pH	7.0	Na	135 mmol/L
pCO <sub>2</sub>	26 mmHg	Cl	112 mmol/L
HCO <sub>3</sub> <sup>-</sup>	8 mmol/L		

1. ¿Existe un trastorno ácido base? (i43) SI / NO

2. Si existe un trastorno, ¿cuál es la alteración primaria? (i44)

RESPIRATORIA / METABÓLICA / ACIDEMIA MIXTA / ALCALEMIA

3. Si existe una alteración, ¿cuál es su grado de severidad? (i45)  
 LEVE / MODERADA / SEVERA
4. ¿Existe compensación? (i46) SI / NO
5. ¿Existe Tercer Trastorno? (i47) SI / NO
6. Si se tienen todos los componentes, escriba el valor de anión gap (i48): .....
7. Describa el Tercer Trastorno (i49) .....
8. Escriba el Diagnóstico Ácido Base (i50)  
 .....
10. En cada recuadro coloque el diagnóstico en orden de probabilidad donde 1 es menos probable y 5 es el más probable (i51)
- Cetoacidosis Diabética ( )
  - Acidosis Tubular Renal ( )
  - Injuria Renal Aguda ( )
  - Diarrea Aguda ( )
  - Intoxicación por sustancias desconocidas ( )

¿Qué tan seguro cree estar de sus respuestas?

- (1)      (2)      (3)      (4)      (5)