



Universidad del  
**Rosario**



Pontificia Universidad  
**JAVERIANA**  
Bogotá

### **Título**

Construcción de una estrategia para evaluar el razonamiento clínico de residentes de medicina interna mediante el uso de *Script Concordance Test* más *Written Think Aloud*

### **Autores**

Kateir Mariel Contreras Villamizar, Javier Iván Lasso Apraéz, Carlos Andrés Celis Preciado

### **Tutor**

Jesús Armando Sánchez Godoy

Profesor de la Maestría en Educación para profesionales de la salud

### **Evaluadores**

Francisco Olmos y John Alexander Vergel Guerrero

Trabajo de grado presentado como requisito parcial para obtener el título de Magister en  
Educación para profesionales de la salud

Escuela de Medicina y Ciencias de la Salud – Facultad de Medicina  
Maestría en educación para profesionales de la salud  
Universidad del Rosario – Pontificia Universidad Javeriana

Bogotá D.C - Colombia  
2023

## **Título**

Construcción de una estrategia para evaluar el razonamiento clínico de residentes de medicina interna mediante el uso de *Script Concordance Test* más *Written Think Aloud*

## **Autores**

Kateir Contreras <sup>1,3,4,5</sup>

Javier I. Lasso <sup>2,3,4,5</sup>

Carlos Andrés Celis-Preciado <sup>2,3,4,5</sup>

Armando Sánchez <sup>4</sup>

## **Filiaciones institucionales**

<sup>1</sup> Unidad de Nefrología, Servicio de Medicina Interna, Hospital Universitario San Ignacio, Bogotá D.C

<sup>2</sup> Unidad de Neumología, Servicio de Medicina Interna, Hospital Universitario San Ignacio, Bogotá D.C

<sup>3</sup> Departamento de Medicina Interna, Facultad de Medicina, Pontificia Universidad Javeriana

<sup>4</sup> Departamento de Ciencias Fisiológicas, Facultad de Medicina, Pontificia Universidad Javeriana

## **Autor de correspondencia**

Kateir Contreras

Dirección: Kr 7 # 40 - 62 Hospital Universitario San Ignacio, Bogotá D.C, Colombia, 110131

Correo electrónico: [contrerask@javeriana.edu.co](mailto:contrerask@javeriana.edu.co)

**Revista a enviar:** Medical Teacher

## **Resumen**

### ***Propósito***

El razonamiento clínico (RC) es una competencia que debe ser enseñada y evaluada. El *Script Concordance Test* (SCT) evalúa el RC en condiciones de incertidumbre y su validez está demostrada. El *Written Think Aloud* (WTA) es complementario para entender el RC. Este estudio describe el desarrollo y aplicación del SCT-WTA en residentes de medicina interna de una universidad en Bogotá D.C, Colombia.

### ***Métodos***

Estudio de caso instrumental mixto. Durante la elaboración, aplicación y calificación del SCT-WTA, los investigadores llevaron diarios de campo donde consignaron sus percepciones, observaciones y experiencias. El SCT-WTA se aplicó a 11 expertos y 11 residentes. Se hizo codificación y categorización de la información cualitativa y estadística básica para los resultados cuantitativos.

### ***Resultados***

Para la construcción de los casos y las preguntas del SCT fue determinante contar con resultados esperados de aprendizaje, guiones de enfermedad y recurrir a los casos de la práctica diaria, durante la aplicación, la principal dificultad fue la escala de Likert, que fue percibida como confusa en muchos casos. La inversión de tiempo de elaboración, respuesta y calificación fue alta, sin embargo, los expertos y los residentes consideraron que el SCT permitió evaluar contenido, organización y uso de la información. Las respuestas entre residentes y expertos fueron similares.

### ***Conclusiones***

El formato del SCT es atractivo tanto para los expertos como para los residentes, quienes consideran que es una buena herramienta para evaluar el RC, competencia fundamental para el internista y que no cuenta con una cátedra ni evaluación específicas. El SCT-WTA fue percibido como complementario y útil como evaluación formativa.

### ***Palabras clave***

Razonamiento clínico, evaluación en educación, medicina interna, script concordance test, written think aloud.

### **Puntos de práctica**

La evaluación del aprendizaje es fundamental en la educación; la competencia que se puede enseñar se debe evaluar, como es el caso del RC.

El SCT-WTA se percibe como instrumento válido para evaluar el RC en residentes de medicina interna y tiene un valor adicional al fomentar la aplicación del conocimiento, la metacognición y el pensamiento crítico (Charlin & van der Vleuten, 2004).

Su construcción es un proceso dispendioso que amerita la concurrencia de especialistas en el área temática y la revisión por pares, siendo necesario tener en cuenta los resultados de aprendizaje esperados, la activación del conocimiento basado en los guiones de enfermedades y la complejidad de las preguntas, entre otros factores (Fournier et al., 2008).

### **Introducción**

Diferentes iniciativas a nivel mundial pretenden establecer el perfil del médico internista y sus competencias (Pinilla & Cárdenas, 2014; Porcel et al., 2012a). Las competencias en atención, conocimientos clínicos y habilidades comunicativas son necesarias, pero también se deben adquirir habilidades técnicas y de interpretación para integrar los problemas del paciente y organizar los conceptos de otras especialidades (Barrera et al., 2019; Pinilla & Cárdenas, 2014; Porcel et al., 2012b). El RC es una de estas habilidades, sin embargo, los médicos de posgrado durante su formación memorizan gran cantidad de información con poco énfasis en la enseñanza del RC (Richards et al., 2020).

Autores como Norman plantean que el RC es la competencia clínica más importante, porque agrupa los procesos mentales necesarios para plantear un diagnóstico, establecer un pronóstico y decidir planes de manejo (Norman, 2000). Así mismo el RC implica un proceso de pensamiento y síntesis de datos clínicos (Gruppetta & Mallia, 2020) para proponer diagnósticos diferenciales (Thampy et al., 2019) y opciones de tratamiento.

Varios modelos explican el RC. El intuitivo, el deductivo y el probabilístico (Eva, 2005). El modelo intuitivo está soportado en patrones de reconocimiento y redes neuronales, ligado estrechamente a la experiencia. El modelo hipotético-deductivo plantea hipótesis que orientan

la recolección de información adicional para confirmarlas o rechazarlas, el experto plantea mejores hipótesis que el novato. Finalmente, el modelo probabilístico, recurre al uso de guiones mentales para el diagnóstico y toma de decisiones. El modelo que mejor representa el RC es una combinación de los tres. Se ha planteado que el conocimiento, habilidades y actitudes necesarias para aprender el RC efectivo sí son transferibles (Richards et al., 2020); es decir, el RC se enseña (Delany & Golding, 2014; Richards et al., 2020) y, por tanto, se evalúa (Baker et al., 2010; Yassin et al., 2021).

Los aspectos que se pueden enseñar y evaluar del RC son: la adquisición del conocimiento, la organización de la información, los procesos cognitivos y la metacognición. En cada aspecto hay tanto estrategias pedagógicas como evaluativas (Young et al., 2018).

La evaluación impacta tanto al estudiante como al profesor; con los resultados obtenidos, se toman decisiones que afectan positiva o negativamente el aprendizaje. La evaluación del RC es desafiante porque tanto los procesos mentales como los pensamientos y los sentimientos no son directamente observables y usualmente no se miden (Durning et al., 2013). El SCT evalúa la organización de la información y el WTA evalúa la metacognición (Young et al., 2018).

El SCT se ha validado para examinar el RC a través de la capacidad para interpretar información clínica en condiciones de *incertidumbre* (Charlin & van der Vleuten, 2004). Cada ítem incluye un caso viñeta corto con información suficiente y tres preguntas. Cada pregunta propone una hipótesis de diagnóstico, laboratorio o tratamiento. Estas hipótesis pueden ser más o menos probables según la información adicional que se suministre en cada pregunta, de tal forma que no hay una única respuesta para cada pregunta. Las opciones de respuesta están en una escala tipo Likert de 5 puntos que va desde muy probable a muy improbable. El test pretende evaluar qué tan cerca o lejos de la respuesta modal, suministrada por un panel de expertos, está la respuesta del estudiante (Lubarsky et al., 2013). Así, el SCT puede capturar partes del proceso de razonamiento (Arocha & Patel, 2019; Lubarsky et al., 2013; Richards et al., 2020). Para mayor claridad se anexa un ejemplo de pregunta del tipo SCT. Figura 1.

## Figura 1.

### Ejemplo de una Pregunta tipo SCT para Investigar una Hipótesis Diagnóstica

**Caso viñeta:** Una mujer consulta a urgencias por disnea aguda. Tiene antecedentes de diabetes, falla cardíaca y tabaquismo pesado. Recibe sulfato de morfina para el alivio de dolor lumbar crónico intratable. Usted está considerando el diagnóstico de una embolia pulmonar.

Si usted está pensando en	y luego usted encuentra	la conducta (el examen) estaría
p1. hacer un angioTAC	una creatinina en 1.27 mg/dl	-2. -1. 0. +1. +2
p2. pedir una gamagrafía V/Q	una historia de EPOC	-2. -1. 0. +1. +2
p3. ordenar un dímero D	una historia de CA de ovario	-2. -1. 0. +1. +2

↑  
Acción investigativa

↑  
Nueva información

↑  
Escala tipo Likert

-2: fuertemente contraindicado. -1: contraindicado. 0: ni más ni menos indicado. +1: indicado. +2: fuertemente indicado.

Hay estudios que evalúan la utilización del SCT en diversas áreas de la medicina (Carrière et al., 2009; Claessens et al., 2014; Hamui et al., 2018; Kazour et al., 2017; Lubarsky et al., 2013; Talvard et al., 2014), pero no hay trabajos que exploren el SCT en la evaluación del RC, específicamente, en medicina interna.

El SCT se ha validado en varios contextos de evaluación del RC, pero tiene limitaciones como la incapacidad para capturar el curso del pensamiento implicado en la elección de una respuesta dentro de una escala tipo Likert de 5 puntos (Iglesias Gómez et al., 2022). El *think aloud* (TA) durante el SCT corresponde a un componente cualitativo que puede ser escrito o verbal, permite un mejor entendimiento del RC (McBee et al., 2019; Wan et al., 2020) y contribuye a la enseñanza y aprendizaje del RC en el lugar de trabajo (Pinnock et al., 2015; Power et al., 2017).

Desde el primer estudio publicado en 2017 por Power et al., varios trabajos han reportado el empleo del TA asociado al SCT para comprender los procesos cognitivos durante el RC (Ottolini et al., 2021; Patel, 2022; Tedesco-Schneck, 2019, 2021; Wan et al., 2020).

El propósito de este estudio de caso es describir el proceso de desarrollo e aplicación de una herramienta, SCT-WTA, para evaluar y comprender el RC en residentes de medicina interna de una universidad privada de Bogotá con miras a optimizar su enseñanza y evaluación, identificar las dificultades en la potencial implementación de la estrategia de evaluación y aplicar la estrategia a un grupo piloto de residentes.

Para tal fin nos planteamos la siguiente pregunta de investigación. ¿Cómo se desarrolla y se aplica el SCT-WTA como una estrategia de evaluación del RC en residentes de medicina interna de una universidad privada de Bogotá?

## **Métodos**

### *Diseño*

Responder la pregunta de investigación implica la recolección de información de diferentes fuentes con datos cuantitativos, derivados de la aplicación del cuestionario SCT; y datos cualitativos, derivados de los diarios de campo de los investigadores y de las percepciones tanto de los expertos como de los residentes. Por tal razón, basados en el paradigma pragmático, cuyo foco es la utilidad de la información; la adopción de estrategias derivadas de ésta y sus consecuencias; planteamos el trabajo como un estudio de caso instrumental utilizando el método mixto convergente.

El caso fue la elaboración y aplicación de la estrategia de evaluación SCT-WTA en el programa de especialización de medicina interna de la Pontificia Universidad Javeriana, Bogotá D.C., Colombia. Este estudio es el producto de un proyecto de maestría de educación en salud y por eso la elaboración del SCT-WTA estuvo a cargo de cuatro investigadores, tres maestrantes (KC, JL, CC) y un tutor de investigación (JS). La revisión de los casos clínicos del SCT la hicieron 3 pares expertos, especialistas en medicina interna, con más de 3 años de graduados y, al menos, un año de experiencia en el área hospitalaria de una institución universitaria. El panel de referencia para las respuestas estuvo compuesto por once internistas expertos pertenecientes al panel docente del programa de medicina interna y el SCT, también fue respondida por once residentes de medicina interna de diferentes años de formación.

## ***Instrumentos***

### *Diario de campo*

A lo largo de todo el proceso de investigación, los investigadores (KC, JL, CC), registraron sus observaciones, sentimientos y percepciones en un diario de campo, este fue el principal instrumento para el análisis cualitativo.

### *Anotaciones*

En estas se registraron las opiniones expresadas por los expertos durante la revisión por pares y por el panel de expertos y residentes al finalizar la aplicación del SCT-WTA.

## ***Elaboración y aplicación del SCT-WTA***

### *Diseño del SCT*

Siguiendo la recomendación de la guía elaborada por Fournier et al. (2008), los investigadores redactaron 20 casos viñeta cada uno con 3 preguntas. Los temas se escogieron de acuerdo a su frecuencia de presentación en el servicio de urgencias y en la consulta ambulatoria de medicina interna, dando prioridad a las relacionadas con neumología y nefrología por ser las áreas de experticia de los investigadores.

### *Revisión por pares*

Las preguntas se intercambiaron entre los investigadores para afinar la redacción y, posteriormente, el cuestionario con 20 casos y 60 preguntas fue enviado a revisión por correo electrónico a tres profesores expertos de medicina interna escogidos selectivamente por su afinidad con los investigadores y disposición. Para la revisión usaron una rúbrica con 11 ítems (Fournier et al., 2008) sobre complejidad e independencia entre las preguntas, claridad de la redacción, escenario clínico propuesto, relevancia y ambigüedad entre otras. Adicionalmente, los revisores dejaron comentarios escritos acerca de sus impresiones del examen y del proceso de revisión.

### *Aplicación del SCT*

Una vez se hizo la revisión por pares, vino la fase de optimización de la prueba. Se eligieron los 15 casos con mejor puntuación en la rúbrica, la cual debía ser superior al 80%. El

cuestionario definitivo conformado por estos 15 casos fue enviado a los profesores internistas del panel de expertos, cuyas respuestas serían la referencia para la calificación del mismo test aplicado a los residentes. En ambos casos los cuestionarios se enviaron por correo electrónico a internistas y residentes que aceptaron participar en el estudio. El consentimiento informado se envió en DocuSign® y el enlace para responder el examen se compartió en la plataforma Google Forms®. El examen fue respondido en línea, de manera anónima y en su tiempo libre. Los datos solicitados fueron edad, género, año de graduación como médico, año de formación para el caso de los residentes y año de graduación como internistas para los expertos. A ambos grupos se les pidió hacer WTA en dos casos viñeta (6 preguntas). Finalmente, en ambos grupos se hizo una pregunta abierta de respuesta obligatoria acerca de la experiencia vivida con el cuestionario.

### *Calificación del SCT*

La calificación de cada pregunta del examen respondido por los residentes se hizo con el sistema de puntuación agregada ponderada, usando una tabla de referencia (Wan et al., 2020) para calcular el valor de las respuestas con relación a la respuesta modal de los expertos.

Para asegurar la credibilidad de la investigación, utilizamos métodos mixtos y diferentes fuentes de información, profesores, residentes y diarios de campo de los investigadores (Shenton, 2004). Se describe cada una de las etapas del estudio para permitir la reproducibilidad y la transferibilidad de los hallazgos. El uso de diarios de campo permitió la trazabilidad de la información y contrastar las percepciones de los investigadores, lo que contribuyó a la confiabilidad. Los resultados del SCT fueron revisados por dos de los investigadores por separado para disminuir la probabilidad de error.

### *Análisis de los datos*

Teniendo en cuenta la complejidad del RC y la necesidad de evaluarlo, narramos nuestra experiencia como investigadores durante las diferentes fases de construcción y aplicación de SCT-WTA. Quisimos conocer más acerca del proceso para aproximarnos a la viabilidad de su implementación como forma de evaluación en el programa de medicina interna.

Para el análisis cualitativo, la información obtenida de los diarios de campo de los investigadores, las anotaciones de las opiniones de los expertos y los residentes, fue incluida en el programa Atlas Ti 22®, después del análisis de la literatura y de familiarizarnos con los datos, se llevó a cabo un proceso de codificación deductiva y se identificaron los códigos y las categorías de interés (Braun & Clarke, 2006).

Las respuestas escritas de residentes y expertos en el WTA fueron exportadas a Microsoft Word y analizadas, identificando aquellas que tuvieran respuesta correcta con razonamiento incorrecto, respuesta incorrecta con razonamiento correcto y respuesta incorrecta con razonamiento incorrecto. Igualmente se analizó si la explicación de las respuestas estaba basada en la fisiopatología, guías de práctica clínica o en reconocimiento de patrones y experiencia.

Para el análisis cuantitativo se utilizó Microsoft Excel®, se realizó estadística básica para las variables sociodemográficas y se tabularon los resultados de expertos para el cálculo de la respuesta modal y la puntuación agregada y con esta se calificó el SCT de los residentes (Charlin et al., 2000; Fournier et al., 2008; Lubarsky et al., 2013).

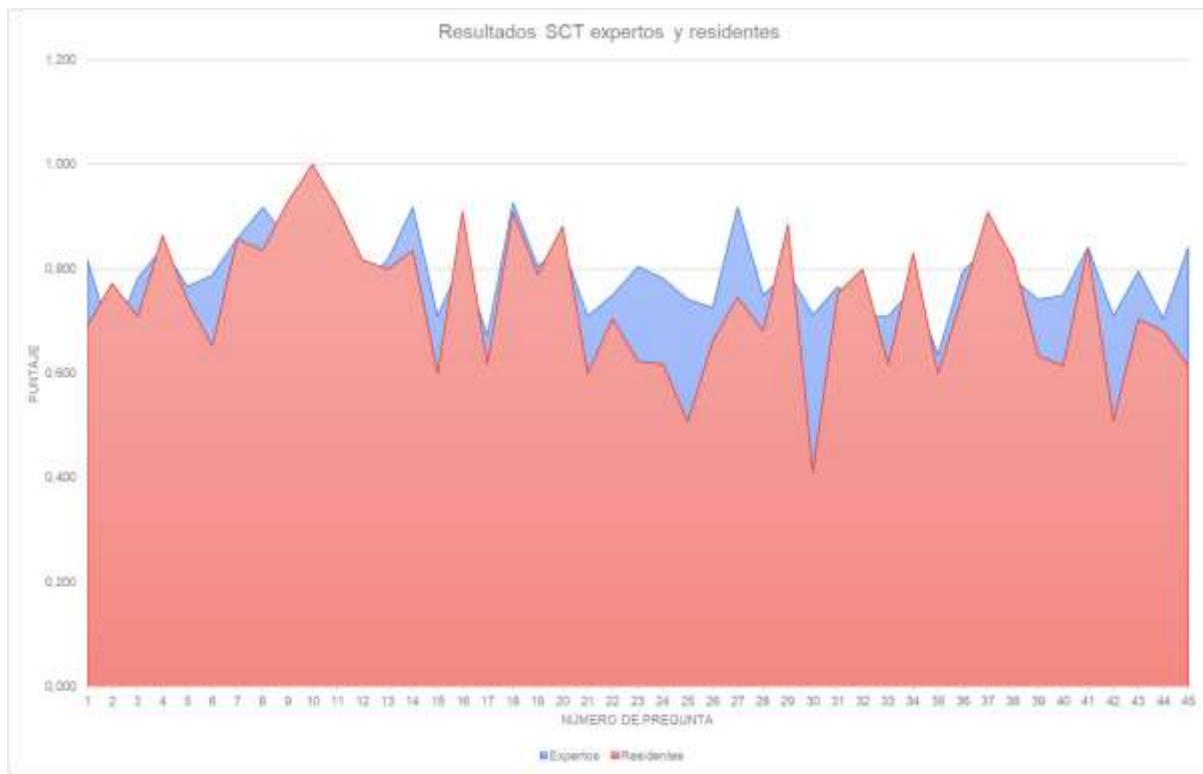
## **Resultados**

El SCT se aplicó a 11 expertos (edad promedio 35 años, 55% mujeres, tiempo promedio de ejercicio de medicina interna 6 años) y a 11 residentes de medicina interna (edad promedio 27 años, 64% mujeres, tiempo promedio de ejercicio de medicina 2,6 años) de todos los niveles de formación (2 de primer año, 5 de segundo año y 4 de tercer año).

El puntaje promedio de los expertos fue de 78,8 (SD 4,85) y el de los residentes de 73,9 (SD 4,71), con variaciones en varias preguntas como lo muestra la figura 2.

## Figura 2.

Puntajes promedio para cada pregunta de expertos y residentes. A pesar de los puntajes finales similares, se observan diferencias en varias preguntas.



## Categorías

En el análisis cualitativo de la información obtenida de los diarios de campo de los investigadores, las anotaciones y las percepciones de expertos y residentes, se definieron 11 códigos que agrupamos en 3 categorías que se explican a continuación.

### Elaboración del SCT

Esta categoría hace referencia a los aspectos necesarios para construir una pregunta usando la aproximación de características clave (Bordage & Zacks, 1984), en la que la elección de la pregunta se enfoca en los elementos más útiles para resolver un problema clínico y en el método para la elaboración del SCT (Charlin et al., 2000; Farmer & Page, 2005).

Esta categoría se formó a partir de cinco códigos: contexto clínico, complejidad, guion de enfermedad, resultados de aprendizaje esperados y expertos.

El contexto clínico expresado como caso viñeta, de acuerdo con la teoría del *Script* (Charlin et al., 2007), moviliza el conocimiento organizado en redes neuronales. Las preguntas deben

estar relacionadas con el contexto clínico el cual debe ser suficiente para suscitar el razonamiento, pero en un ambiente de incertidumbre clínica, es decir que produzcan variabilidad dentro de un panel de expertos.

Durante la escritura de los casos la dificultad mayor fue construir los escenarios clínicos suficientes para la finalidad evaluativa (“Me parece difícil encontrar los escenarios de incertidumbre y escoger cuál será la frase que mejor refleje el escenario que deseo evaluar”, JL). La evaluación de los escenarios por parte de los examinados nos mostró lo siguiente: información insuficiente (“Creo que el caso podría tener una descripción un poco más detallada“, par evaluador (PE) 1); información innecesaria para responder (“...si bien el caso clínico puede correlacionarse con las preguntas, estas se podrían responder incluso sin leer el caso...” PE1); ambigüedad versus incertidumbre clínica (“...podrían ser más precisos en algunos escenarios ...enriquecer la información de claves para favorecer la comprensión y evitar ambigüedad...” PE2).

El código guion de enfermedad hace referencia a la organización de la información o del conocimiento en patrones de reconocimiento, para nuestro caso, enfermedades. Es claro que el SCT evalúa el RC a partir del uso de la información que activa el conocimiento previo; por lo cual, no se puede desligar del conocimiento o de la carga cognitiva. Desde este punto de vista construir las preguntas también representa una dificultad porque no se puede evaluar el proceso sin tener la información, pero esta no debe ser factual como datos numéricos. El SCT no está diseñado para medir la capacidad de recordar datos descontextualizados (Lubarsky et al., 2013). La evaluación de las preguntas reveló que la pregunta debe evaluar el uso de la información (“...las preguntas no deben evaluar si sabe o no sabe determinada información sino cómo se utiliza...” CC) y es necesario tener un bagaje de conocimientos para hacer y responder las preguntas (“...para construir las preguntas es claro que sí se debe partir de unos conocimientos básicos ...” CC).

El código resultados de aprendizaje esperados hace referencia al *blueprint* sobre el cual se van a construir las preguntas. Está relacionado directamente con los objetivos de aprendizaje de cada asignatura, como se ha recomendado para la construcción de los ítems del SCT (Charlin et al., 2000; Hornos et al., 2013; Lubarsky et al., 2013). En la construcción de los casos se tuvieron en cuenta los resultados de aprendizaje esperados para los residentes de medicina interna (“... usé un listado de resultados esperados de aprendizaje de acuerdo con situaciones que he visto recientemente en clínica...” KC).

Para delimitar un poco el espectro de lo evaluable nos enfocamos en temas de neumología y nefrología. La opinión de los expertos revisores fue confirmatoria de esta recomendación (“... el uso de los resultados esperados del aprendizaje es fundamental para poder desarrollar la pregunta...PE 3”, “...el SCT evalúa un aspecto específico del razonamiento clínico que es la habilidad para interpretar la información médica bajo condiciones de incertidumbre”, “...las preguntas deben tener un propósito de evaluación...” PE 3).

El código complejidad hace referencia al grado de dificultad de las preguntas y del formato empleado, las preguntas con mucha dificultad o con casos infrecuentes y las preguntas con respuesta obvia o que no generen variabilidad de conductas no son útiles (Charlin et al., 2000, 2007). Así, algunas preguntas fueron consideradas no evaluativas por ser de baja complejidad (“...todas las opciones son de una complejidad similar...PE 1”, “...las opciones son de baja complejidad...PE 1”, “casos de baja complejidad en los que la situación de incertidumbre no está ... (PE 1)”, “...la pregunta no es diferente de una pregunta del tipo de escogencia múltiple...” PE 1).

El código experto hace referencia al primer grupo de docentes que revisaron las preguntas aplicando una rúbrica propuesta para tal fin (Fournier et al., 2008). En general, los escenarios clínicos se consideraron pertinentes, desafiantes e informativos teniendo en cuenta el contexto propuesto y se anotó claridad en la redacción de estos, condiciones que se han descrito como necesarias en la elaboración del SCT (Charlin et al., 2000).

### *Aplicación del SCT*

El formato del SCT es satisfactorio en cuanto a las propiedades psicométricas de las puntuaciones obtenidas y congruente con la base teórica del SCT (Dory et al., 2012), pero suele ser confuso, especialmente, cuando es la primera vez que se contesta la prueba (Lubarsky et al., 2013). La distribución en columnas que busca evaluar el impacto de la nueva información en la probabilidad de un diagnóstico o tratamiento concreto, puede causar confusión entre los examinados, ya que esto no refleja el RC de la vida real (van den Broek et al., 2012), y las respuestas de los examinados están influidas por su comprensión de términos específicos, su deseo de obtener información adicional, su desacuerdo con la acción prevista, sus faltas de conocimiento subyacentes y su deseo de demostrar confianza o humildad (Gawad et al., 2021).

La presencia de una escala tipo Likert en la tercera columna es uno de los aspectos más distintivos de la prueba (Charlin et al., 2000) y varios estudios han demostrado correlaciones

relativamente débiles entre las puntuaciones del SCT y las obtenidas en los exámenes de preguntas múltiples, lo que respalda la idea de que el SCT mide un constructo diferente al de las pruebas que exploran el recuerdo puro del conocimiento proposicional (Lubarsky et al., 2011). La escala Likert de 5 puntos permite que cualquier respuesta refleje una opinión aceptable y los ítems se construyen para dispersar las respuestas entre todos los valores de la escala.

Se encontró que expertos y residentes señalaron falta de comprensión de la escala (“...no me quedan claras las opciones de la escala de Likert...” PE 1) y una relación no clara con el tipo de pregunta (“...la escala de Likert no es clara cuando está hablando de paraclínicos o conductas...” PE2, “...la escala de Likert no siempre coincide con el tipo de pregunta...” PE2), derivando esto en dificultad a la hora de responder (“...agotador el estar leyendo la escala de Likert...”PE 3, KC), señalando además posibles formas de ajustar estas situaciones (“...se propone ajustar la escala de Likert a cada pregunta...”、“...poder realizar la escala de Likert con anclas que sean claras, no ambiguas...”JL).

#### *Ventajas y Desventajas*

El SCT ha demostrado ser una herramienta de evaluación válida y fiable del RC (Charlin et al., 2007). A diferencia de las preguntas de selección múltiple, evalúa la organización del conocimiento y la toma de decisiones en situaciones que simulan la vida real (Fournier et al., 2008).

Bajo el código ventajas se incluyeron las apreciaciones respecto a los beneficios del SCT. Se consideró que la prueba SCT es una buena estrategia para evaluar el RC, dado que expone al evaluado a situaciones de incertidumbre, muy similares a las que se viven en la práctica diaria (“...me parece más evaluativo que los tipos de preguntas usuales que implican memoria...” R3, “...algunas opciones de conducta no están del todo bien o mal lo cual hace que sea más difícil escoger la respuesta, que es lo que usualmente se ve en la práctica clínica...” R4).

En las respuestas se apreció que los casos con respuestas concordantes fueron aquellos que mostraban situaciones con manejos ampliamente aceptados (“...sitúa a los examinados ante situaciones clínicas auténticas en las que tienen que interpretar datos para tomar decisiones...” KC) y el SCT fue calificado como “...una herramienta útil para la evaluación formativa...” R5, dado que permite discutir con razones las respuestas.

Dentro de las posibles desventajas que se han descrito están que el formato clásico del SCT sólo capta la respuesta del estudiante en la escala de 5 puntos, pero el RC real y el procesamiento del pensamiento en la elección de una respuesta no se registran ni se examinan (Iglesias Gómez et al., 2022). El formato de las preguntas y las opciones de respuesta generó confusión entre expertos y residentes (“...se debe dejar más claro específicamente el diagnóstico, laboratorio o conducta a los que se refieren las preguntas. Uno lo puede inferir, pero a veces es confuso...” Experto (E) E1) y las anclas de la escala de Likert fueron percibidas como una dificultad, ya que el mantener la misma escala en todas las preguntas hace que, en algunos casos, no coincida el escenario con la escala (“...en muchas ocasiones puede no ser fácil de interpretar la misma escala en contextos de evaluación, diagnóstico o tratamiento...” E2, Residente (R)1) y generó confusión entre expertos y residentes, se ajustó durante la fase de revisión; sin embargo, siguió siendo confusa en varios casos (“...la escala de Likert no es clara cuando no está hablando de paraclínicos o conductas...” E3). Se utilizó la palabra “consideración” para referirse a un diagnóstico, paraclínico o tratamiento, esto también fue confuso (“...en varias preguntas es difícil entender a cuál consideración se refieren...” R2).

La construcción del SCT consume tiempo y se percibe como mayor que el que se gasta en la construcción de preguntas de selección múltiple (“...tuve que realizar la revisión en 3 momentos diferentes, dado que me sentía exhausto por la complejidad del examen, sentía que debía estar mentalmente tranquilo y descansado para hacerlo...” PE 3, “...implica un trabajo largo y multipersonal...” CC).

El tiempo que se necesita para contestar la prueba fue percibido como una dificultad para los expertos. El tiempo promedio por pregunta fue de 1,5 minutos. El tiempo y la atención necesarios para calificar el SCT son mayores y esto puede suscitar errores (“...al no haber respuestas correctas, todas las opciones de respuesta tienen un puntaje y deben contabilizarse...” JL). Igualmente se consideró que hacer una sola prueba para evaluar el RC en general, es complejo y quizá no sea posible (“...solo se puede evaluar el RC temático...” JL, “...el SCT no evalúa destrezas psicomotoras...” KC).

Para los investigadores, ajustar la complejidad de las preguntas fue un reto (“...me parece difícil encontrar los escenarios de incertidumbre y escoger cuál será la frase que mejor refleje el escenario que deseó evaluar...JL”, “...surgían dudas acerca de cómo estaban planteadas las preguntas, lo que para uno era un planteamiento de incertidumbre para otro podía ser de ambigüedad...JL”, “...no se puede separar la evaluación del RC del conocimiento...” JL)

Para hacer las preguntas, es fundamental mantener la independencia entre ellas para que no se interpreten como preguntas agregativas; al revisar el WTA, encontramos respuestas que sugerían que la información suministrada se había interpretado como agregativa e interdependientes.

En la literatura no se encontraron bancos de preguntas tipo SCT para medicina interna. En el código oportunidades se anotó que adquirir la experiencia para generar este tipo de bancos sería un trabajo útil para desarrollar (“...he tratado de buscar modelos de preguntas tipo SCT. Existen algunas preguntas, pero no para evaluar el RC en medicina interna...” CC, “...gran oportunidad para iniciar la creación de un banco de preguntas...” CC).

Las preguntas más frecuentes, son las de selección múltiple con única respuesta y las preguntas con texto enriquecido, estas usualmente evalúan contenido y algunas tomas de decisiones, las preguntas del SCT amplían el panorama a situaciones clínicas de incertidumbre, convirtiéndose en un instrumento para complementar otros dominios en la evaluación del residente (“...los *scripts* pueden tratar cuestiones relativas al diagnóstico, a la realización de pruebas complementarias o al tratamiento...” JL).

### ***Aporte de WTA al SCT***

El análisis de las anotaciones permitió identificar cómo los expertos y los residentes explicaban sus respuestas. Los expertos utilizaron más la experiencia clínica y el reconocimiento de patrones o guiones de enfermedad para explicar su razonamiento, “...no influyó la respuesta, dado que el paciente tiene franca sobrecarga hídrica...” E10.

Por su parte, los residentes solían explicar su respuesta desde el punto de vista fisiopatológico y de guías de manejo clínico, “...en el contexto del paciente la lesión renal aguda se ve explicada más probablemente por síndrome cardiorrenal tipo 1, en el que el mecanismo fisiopatológico de la hipertensión venosa renal se impacta positivamente con el uso de diurético...” R8, “...el cuadro clínico es sugestivo de asma; por lo tanto, mi consideración hubiese sido hacer la espirometría con prueba de broncodilatador. El hecho de que luego la espirometría salga normal, no cambiaría la conducta de haberla tomado en primera instancia...R9”

El WTA permitió identificar en los residentes preguntas con respuesta correcta pero con RC incorrecto, “...esperaría que si el verdadero diagnóstico es asma y el paciente se encuentra sintomático y sin tratamiento, la espirometría mostrara un patrón obstructivo con respuesta a broncodilatador...R6”, “...no creo que sea inapropiado descartar la patología previa, si bien no es la principal sospecha diagnóstica...R7”, así como preguntas con RC incorrecto a pesar de respuesta SCT correcta, “...la sensibilidad y especificidad de la prueba de metacolina para asma disminuiría la posibilidad del diagnóstico sin descartarlo...R5”, “...se puede hacer pruebas terapéuticas con broncodilatadores, no hubiera iniciado el esteroide inhalado debido a que no he encontrado objetivamente broncoobstrucción...R10” y preguntas con RC adecuado con respuesta SCT subóptima, “...no siempre se puede documentar la bronco reactividad en asma, a veces se contempla la prueba terapéutica...R11”.

Así mismo, el WTA fue útil para identificar la interpretación errónea de la pregunta tanto por los expertos como por los residentes, debido a falta de claridad en la viñeta clínica o en el enunciado, “...no puedo tomar una consideración sobre si está bien hacer o no la prueba de metacolina si la condición es que ya se la hice y el resultado es negativo. La pregunta está mal formulada...” E9), “...no es claro a qué consideración se refiere la pregunta. Si mi consideración era hacer una espirometría el resultado normal no cambiaría la consideración porque igual ya se hizo. Si la consideración fuera una sospecha de asma ahí sí la consideración sería menos probable...” R10).

## **Discusión**

El SCT es una herramienta diseñada para la evaluación del RC en situaciones de incertidumbre, su validez y utilidad están documentadas en varios contextos de pregrado y posgrado, pero existe poca evidencia de su implementación en residentes de medicina interna (Brailovsky et al., 2001; Charlin et al., 1998).

Mediante la recopilación de experiencias, percepciones y anotaciones de los investigadores, profesores y residentes de medicina interna, identificamos aspectos importantes para la elaboración y aplicación del SCT, así como las ventajas y desventajas que podrían afectar su implementación en un programa de medicina interna.

Durante la elaboración del SCT por parte de los investigadores se observó que usar un *blueprint*, establecer previamente los resultados de aprendizaje esperados y plantear los casos viñeta basados en la práctica clínica diaria, fueron estrategias útiles y concuerdan con las

recomendaciones ofrecidas por varios artículos (Charlin et al., 2000, 2007; Lubarsky et al., 2011, 2013). Otro aspecto importante fue la elaboración de las preguntas, pues su resolución implica un proceso cognitivo que consta de cuatro pasos secuenciales (comprensión, recuerdo, toma de decisiones y coincidencia de respuestas), resaltando la importancia de revisar cuidadosamente la formulación de cada pregunta, dado que el motivo de la respuesta errada puede ser la falta de comprensión del enunciado, no necesariamente un razonamiento incorrecto (Gawad et al., 2021; Willis, 2015).

En nuestro estudio, el formato de las preguntas y la escala de Likert se percibieron como los aspectos que causaron mayor confusión, por lo que recomendamos poner un ejemplo de caso y pregunta al inicio de la prueba para facilitar la comprensión del nuevo formato. Dado que los casos tienen tres preguntas, idealmente distribuidas en diagnóstico, exámenes y tratamiento, la mayor dificultad encontrada fue la de adecuar una escala Likert cuyas expresiones respondieran por igual a todas las categorías de preguntas. Creemos que varias de las preguntas que ofrecieron dificultades para expertos y residentes tuvieron que ver con este aspecto y esto puede afectar el proceso de toma de decisiones e influir la respuesta. En la literatura se recomienda, y nosotros lo hemos corroborado, que se deben utilizar escalas que hagan referencia a probabilidad de un diagnóstico, utilidad de un examen o indicación de un tratamiento por separado (Fournier et al., 2008).

Dentro de los comentarios acerca del SCT, se encontró que el recuerdo de casos similares atendidos en sus prácticas clínicas ayudó a responder la prueba, hallazgo que se ha encontrado en otros estudios (Gawad et al., 2021) y apoya la importancia del aprendizaje en el lugar de trabajo para la enseñanza del RC.

Con respecto a los resultados cuantitativos del SCT en expertos y residentes, se observó que las respuestas de los dos grupos fueron similares, no obstante, el tamaño de la muestra no nos permitió hacer una comparación entre los grupos; sin embargo, pudimos apreciar que a diferencia de lo que se describe en algunos trabajos acerca de mejores puntajes en expertos o entre niveles de entrenamiento (Charlin et al., 2007; Fournier et al., 2008), en nuestra investigación no lo encontramos, tal vez podríamos explicarlo por la baja complejidad de las preguntas según opinión de algunos participantes o porque los temas que empleamos eran muy frecuentes y están presentes temprano en la formación de los residentes.

Contar con equipos de pares con tiempo protegido para que participen en la elaboración de los casos y las preguntas puede ser de utilidad, dado que el tiempo consumido en la construcción puede convertirse en una limitante (Fournier et al., 2008). Este es un punto de partida para el desarrollo de bancos de preguntas tipo SCT como ya lo han hecho otros autores en disciplinas como pediatría, o subespecialidades como cardiología o gastroenterología (Hornos et al., 2013).

Por otra parte, estudios recientes han señalado que con la incorporación del WTA al SCT se obtienen beneficios adicionales al complementar el abordaje cuantitativo del SCT (Power et al., 2017) y proporcionar información sobre la comprensión de los estudiantes de un escenario clínico determinado (Wan et al., 2020).

Nuestro análisis de las anotaciones del WTA permitió identificar cómo los expertos y los residentes explicaban sus respuestas, ya fuera mediante experiencia clínica y el reconocimiento de patrones o guiones de enfermedad o desde el punto de vista fisiopatológico y de guías de manejo clínico, respectivamente. El WTA nos permitió identificar en los residentes preguntas con respuesta correcta, pero con RC incorrecto, así como la interpretación errónea de la pregunta tanto por los expertos como por los residentes, debido a falta de claridad en la viñeta clínica o en el enunciado.

La incorporación del WTA puede optimizar el SCT demostrando cuándo las respuestas correctas de los examinandos se basan en conjeturas o incertidumbres en lugar de un razonamiento clínico sólido. El WTA también puede mejorar el SCT permitiendo a los examinandos justificar respuestas que de otro modo se habrían considerado incorrectas e identificando preguntas que se interpretan mal con frecuencia para evitar incluirlas en futuros exámenes. El WTA también tiene un valor significativo a la hora de diferenciar entre las variaciones aceptables en las respuestas de los clínicos expertos y la desviación asociada a una justificación defectuosa o a una mala interpretación de las preguntas; esto podría mejorar la fiabilidad del SCT.

De esta forma, el SCT-WTA podría tener un papel potencial como una actividad de aprendizaje distintiva y estructurada puesto que permite discutir las respuestas a través del proceso cognitivo empleado para la solución, brinda oportunidad para la toma de conciencia y la autoevaluación (Patel, 2022) y, desde el punto de vista de la evaluación programática, si se documenta de forma sistemática, la rica información procedente de la reflexión escrita del

WTA en el SCT también podría servir de base para la toma de decisiones importantes para la progresión de los estudiantes en los programas de formación (Wan et al., 2020).

Dentro de las limitaciones del estudio cabe anotar que es un estudio piloto, y no realizamos WTA a la totalidad de las preguntas, por lo que es probable que hayamos perdido información acerca de la fortalezas y debilidades de esta prueba.

## **Conclusiones**

Para implementar el SCT como una estrategia de evaluación de RC para residentes de medicina interna es necesario tener presentes los resultados de aprendizaje esperados, la activación del conocimiento basado en los guiones de enfermedades y la complejidad de las preguntas. El formato del SCT es atractivo tanto para los expertos como para los residentes, quienes consideran que es una nueva forma de evaluar una competencia que se da por hecho durante la formación médica. Sin embargo, el formato es desconocido y su elaboración ofrece dificultades. La inversión de tiempo de elaboración, respuesta y calificación fue alta. Su utilidad en general es aceptada porque se considera que evalúa contenido, organización y uso de la información. Las respuestas entre residentes y expertos fueron similares. El WTA mostró ser complementario al SCT y se propone como una estrategia de evaluación con intenciones formativas y de retroalimentación.

## **Agradecimientos**

Al panel de expertos y residentes del Departamento de Medicina Interna de la Pontificia Universidad Javeriana de Bogotá, D.C.

## **Aprobación ética**

Se presentó y recibió aprobación por el comité de ética e investigaciones de la Pontificia Universidad Javeriana y del Hospital Universitario San Ignacio, Bogotá D.C, Colombia (N. 1045 de 2022)

## **Conflicto de interés**

Ninguno

## **Contribuciones de los autores**

KC, CC, JL participaron en la concepción de la idea de investigación, diseño del protocolo, recolección de los datos, análisis de resultados y elaboración del manuscrito.

AS participó en la revisión del protocolo, panel de expertos y en la revisión del manuscrito.

## Referencias

Arocha, J. F., & Patel, V. L. (2019). Methods in the Study of Clinical Reasoning. En *Clinical Reasoning in the Health Professions*. Edited by Joy Higgs, Gail M Jensen, Stephen Loftus and Nicole Christensen (Fourth Edition, pp. 345-366). Elsevier Limited.

Baker, E., Ledford, C., & Liston, B. (2010). Teaching, Evaluating, and Remediating Clinical Reasoning. *Academic Internal Medicine Insight*, 8(1), 12-17.

Barrera, E., Ruiz-Bedoya, C., & Villar-Centeno, J. (2019). El deber de actualizar el currículo en medicina interna. *Acta Médica Colombiana*, 44, 31-33.

<https://doi.org/10.36104/amc.2019.1061>

Bordage, G., & Zacks, R. (1984). The structure of medical knowledge in the memories of medical students and general practitioners: Categories and prototypes. *Medical Education*, 18(6), 406-416. <https://doi.org/10.1111/j.1365-2923.1984.tb01295.x>

Brailovsky, C., Charlin, B., Beausoleil, S., Coté, S., & Van der Vleuten, C. (2001).

Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: An experimental study on the script concordance test. *Medical Education*, 35(5), 430-436. <https://doi.org/10.1046/j.1365-2923.2001.00911.x>

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>

Carrière, B., Gagnon, R., Charlin, B., Downing, S., & Bordage, G. (2009). Assessing clinical reasoning in pediatric emergency medicine: Validity evidence for a Script Concordance Test. *Annals of Emergency Medicine*, 53(5), 647-652.

<https://doi.org/10.1016/j.annemergmed.2008.07.024>

CHARLIN, B. (1998). Script questionnaires: Their use for assessment of diagnostic knowledge in radiology. *Medical Teacher*, 20(6), 567-571.

<https://doi.org/10.1080/01421599880300>

Charlin, B., Boshuizen, H. P. A., Custers, E. J., & Feltovich, P. J. (2007). Scripts and clinical reasoning. *Medical Education*, 41(12), 1178-1184. <https://doi.org/10.1111/j.1365->

2923.2007.02924.x

- Charlin, B., Brailovsky, C., Leduc, C., & Blouin, D. (1998). The Diagnosis Script Questionnaire: A New Tool to Assess a Specific Dimension of Clinical Competence. *Advances in Health Sciences Education: Theory and Practice*, 3(1), 51-58. <https://doi.org/10.1023/A:1009741430850>
- Charlin, B., Roy, L., Brailovsky, C., Goulet, F., & van der Vleuten, C. (2000). The Script Concordance test: A tool to assess the reflective clinician. *Teaching and Learning in Medicine*, 12(4), 189-195. [https://doi.org/10.1207/S15328015TLM1204\\_5](https://doi.org/10.1207/S15328015TLM1204_5)
- Charlin, B., & van der Vleuten, C. (2004). Standardized assessment of reasoning in contexts of uncertainty: The script concordance approach. *Evaluation & the Health Professions*, 27(3), 304-319. <https://doi.org/10.1177/0163278704267043>
- Claessens, Y.-E., Wannepain, S., Gestin, S., Magdelein, X., Ferretti, E., Guilly, M., Charlin, B., & Pelaccia, T. (2014). How emergency physicians use biomarkers: Insights from a qualitative assessment of script concordance tests. *Emergency Medicine Journal: EMJ*, 31(3), 238-241. <https://doi.org/10.1136/emered-2012-202303>
- Delany, C., & Golding, C. (2014). Teaching clinical reasoning by making thinking visible: An action research project with allied health clinical educators. *BMC Medical Education*, 14(1), 20. <https://doi.org/10.1186/1472-6920-14-20>
- Dory, V., Gagnon, R., Vanpee, D., & Charlin, B. (2012). How to construct and implement script concordance tests: Insights from a systematic review. *Medical Education*, 46(6), 552-563. <https://doi.org/10.1111/j.1365-2923.2011.04211.x>
- Durning, S. J., Ratcliffe, T., Artino, A. R., van der Vleuten, C., Beckman, T. J., Holmboe, E., Lipner, R. S., & Schuwirth, L. (2013). How is clinical reasoning developed, maintained, and objectively assessed? Views from expert internists and internal medicine interns. *The Journal of Continuing Education in the Health Professions*, 33(4), 215-223. <https://doi.org/10.1002/chp.21188>
- Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Medical Education*, 39(1), 98-106. <https://doi.org/10.1111/j.1365-2929.2004.01972.x>
- Farmer, E. A., & Page, G. (2005). A practical guide to assessing clinical decision-making skills using the key features approach. *Medical Education*, 39(12), 1188-1194. <https://doi.org/10.1111/j.1365-2929.2005.02339.x>
- Fournier, J. P., Demeester, A., & Charlin, B. (2008). Script concordance tests: Guidelines for construction. *BMC Medical Informatics and Decision Making*, 8, 18. <https://doi.org/10.1186/1472-6947-8-18>

- Gawad, N., Wood, T. J., Cowley, L., & Raiche, I. (2021). How do cognitive processes influence script concordance test responses? *Medical Education*, *55*(3), 354-364. <https://doi.org/10.1111/medu.14416>
- Gruppetta, M., & Mallia, M. (2020). Clinical reasoning: Exploring its characteristics and enhancing its learning. *British Journal of Hospital Medicine (London, England: 2005)*, *81*(10), 1-9. <https://doi.org/10.12968/hmed.2020.0227>
- Hamui, M., Ferreira, J. P., Torrents, M., Torres, F., Ibarra, M., Ossorio, M. F., Urrutia, L., & Ferrero, F. (2018). [Script Concordance Test: First nationwide experience in pediatrics]. *Archivos Argentinos De Pediatría*, *116*(1), e151-e155. <https://doi.org/10.5546/aap.2018.e151>
- Hornos, E. H., Pleguezuelos, E. M., Brailovsky, C. A., Harillo, L. D., Dory, V., & Charlin, B. (2013). The practicum script concordance test: An online continuing professional development format to foster reflection on clinical practice. *The Journal of Continuing Education in the Health Professions*, *33*(1), 59-66. <https://doi.org/10.1002/chp.21166>
- Iglesias Gómez, C., González Sequeros, O., & Salmerón Martínez, D. (2022). Clinical reasoning evaluation using script concordance test in primary care residents. *Anales De Pediatría*, *97*(2), 87-94. <https://doi.org/10.1016/j.anpede.2022.06.005>
- Kazour, F., Richa, S., Zoghbi, M., El-Hage, W., & Haddad, F. G. (2017). Using the Script Concordance Test to Evaluate Clinical Reasoning Skills in Psychiatry. *Academic Psychiatry: The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, *41*(1), 86-90. <https://doi.org/10.1007/s40596-016-0539-6>
- Lubarsky, S., Charlin, B., Cook, D. A., Chalk, C., & van der Vleuten, C. P. M. (2011). Script concordance testing: A review of published validity evidence. *Medical Education*, *45*(4), 329-338. <https://doi.org/10.1111/j.1365-2923.2010.03863.x>
- Lubarsky, S., Dory, V., Duggan, P., Gagnon, R., & Charlin, B. (2013). Script concordance testing: From theory to practice: AMEE guide no. 75. *Medical Teacher*, *35*(3), 184-193. <https://doi.org/10.3109/0142159X.2013.760036>
- McBee, E., Blum, C., Ratcliffe, T., Schuwirth, L., Polston, E., Artino, A. R., & Durning, S. J. (2019). Use of clinical reasoning tasks by medical students. *Diagnosis (Berlin, Germany)*, *6*(2), 127-135. <https://doi.org/10.1515/dx-2018-0077>
- Norman, G. R. (2000). The epistemology of clinical reasoning: Perspectives from philosophy, psychology, and neuroscience. *Academic Medicine: Journal of the Association of*

*American Medical Colleges*, 75(10 Suppl), S127-135.

<https://doi.org/10.1097/00001888-200010001-00041>

- Ottolini, M. C., Chua, I., Campbell, J., Ottolini, M., & Goldman, E. (2021). Pediatric Hospitalists' Performance and Perceptions of Script Concordance Testing for Self-Assessment. *Academic Pediatrics*, 21(2), 252-258.  
<https://doi.org/10.1016/j.acap.2020.10.003>
- Patel, R. (2022). General practice trainees' learning experiences of formative think-aloud script concordance testing. *Education for Primary Care: An Official Publication of the Association of Course Organisers, National Association of GP Tutors, World Organisation of Family Doctors*, 33(4), 229-236.  
<https://doi.org/10.1080/14739879.2022.2057240>
- Pinilla, A., & Cárdenas, F. (2014). Evaluación y construcción de un perfil de competencias profesionales en medicina interna. *Acta Médica Colombiana*, 39(2), 165-173.
- Pinnock, R., Young, L., Spence, F., Henning, M., & Hazell, W. (2015). Can Think Aloud Be Used to Teach and Assess Clinical Reasoning in Graduate Medical Education? *Journal of Graduate Medical Education*, 7(3), 334-337.  
<https://doi.org/10.4300/JGME-D-14-00601.1>
- Porcel, J. M., Casademont, J., Conthe, P., Pinilla, B., Pujol, R., & García-Alegría, J. (2012a). Core competencies in internal medicine. *European Journal of Internal Medicine*, 23(4), 338-341. <https://doi.org/10.1016/j.ejim.2012.03.003>
- Porcel, J. M., Casademont, J., Conthe, P., Pinilla, B., Pujol, R., & García-Alegría, J. (2012b). Core competencies in internal medicine. *European Journal of Internal Medicine*, 23(4), 338-341. <https://doi.org/10.1016/j.ejim.2012.03.003>
- Power, A., Lemay, J.-F., & Cooke, S. (2017). Justify Your Answer: The Role of Written Think Aloud in Script Concordance Testing. *Teaching and Learning in Medicine*, 29(1), 59-67. <https://doi.org/10.1080/10401334.2016.1217778>
- Richards, J. B., Hayes, M. M., & Schwartzstein, R. M. (2020). Teaching Clinical Reasoning and Critical Thinking: From Cognitive Theory to Practical Application. *Chest*, 158(4), 1617-1628. <https://doi.org/10.1016/j.chest.2020.05.525>
- Shenton, A. (2004). Strategies for Ensuring Trustworthiness in Qualitative Research Projects. *Education for Information*, 22, 63-75. <https://doi.org/10.3233/EFI-2004-22201>
- Talvard, M., Olives, J.-P., & Mas, E. (2014). [Assessment of medical students using a script concordance test at the end of their internship in pediatric gastroenterology]. *Archives De Pédiatrie: Organe Officiel De La Société Française De Pédiatrie*, 21(4), 372-376.

<https://doi.org/10.1016/j.arcped.2014.01.012>

- Tedesco-Schneck, M. (2019). Use of Script Concordance Activity With the Think-Aloud Approach to Foster Clinical Reasoning in Nursing Students. *Nurse Educator*, 44(5), 275-277. <https://doi.org/10.1097/NNE.0000000000000626>
- Tedesco-Schneck, M. (2021). Script Concordance and the Think-Aloud Approach Through the Lens of Cognitive Apprenticeship Theory. *Nurse Educator*, 46(6), E193-E197. <https://doi.org/10.1097/NNE.0000000000001008>
- Thampy, H., Willert, E., & Ramani, S. (2019). Assessing Clinical Reasoning: Targeting the Higher Levels of the Pyramid. *Journal of General Internal Medicine*, 34(8), 1631-1636. <https://doi.org/10.1007/s11606-019-04953-4>
- van den Broek, W. E. S., van Asperen, M. V., Custers, E., Valk, G. D., & Ten Cate, O. T. J. (2012). Effects of two different instructional formats on scores and reliability of a script concordance test. *Perspectives on Medical Education*, 1(3), 119-128. <https://doi.org/10.1007/s40037-012-0017-0>
- Wan, M. S. H., Tor, E., & Hudson, J. N. (2020). Examining response process validity of script concordance testing: A think-aloud approach. *International Journal of Medical Education*, 11, 127-135. <https://doi.org/10.5116/ijme.5eb6.7be2>
- Willis, G. B. (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford University Press.
- Yassin, B. A. G., Almothaffar, A., Aladhadh, M., Hussein, M. F., Ghafour, Z. A., & Gorial, F. I. (2021). Studying Clinical Reasoning of Internal Medicine Residents in Medical City Campus Using Script Concordance Test. *International Medical Journal*, 28(3), 280-283.
- Young, M. E., Dory, V., Lubarsky, S., & Thomas, A. (2018). How Different Theories of Clinical Reasoning Influence Teaching and Assessment. *Academic Medicine: Journal of the Association of American Medical Colleges*, 93(9), 1415. <https://doi.org/10.1097/ACM.0000000000002303>

## **Anexos**

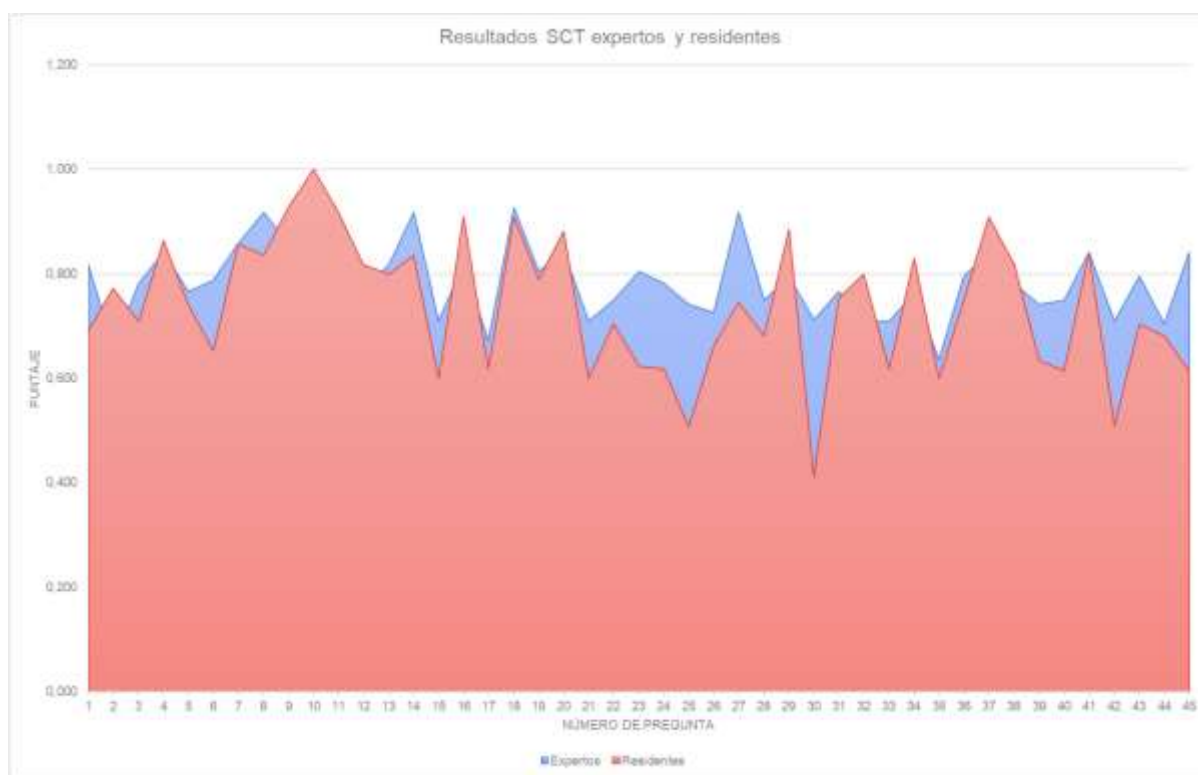
**Caso viñeta:** Una mujer consulta a urgencias por disnea aguda. Tiene antecedentes de diabetes, falla cardíaca y tabaquismo pesado. Recibe sulfato de morfina para el alivio de dolor lumbar crónico intratable. Usted está considerando el diagnóstico de una embolia pulmonar.

Si usted está pensando en	y luego usted encuentra	la conducta (el examen) estaría
p1. hacer un angioTAC	una creatinina en 1.27 mg/dl	-2. -1. 0. +1. +2
p2. pedir una gamagrafia V/Q	una historia de EPOC	-2. -1. 0. +1. +2
p3. ordenar un dímero D	una historia de CA de ovario	-2. -1. 0. +1. +2



-2: fuertemente contraindicado. -1: contraindicado. 0: ni más ni menos indicado. +1: indicado. +2: fuertemente indicado.

**Figura 1.** Ejemplo de pregunta del SCT



**Figura 2.** Puntajes promedio para cada pregunta de expertos y residentes. A pesar de los puntajes finales similares, se observan diferencias en varias preguntas.