

Desarrollo y evaluación de los *scripts* durante la formación profesional

María Dolores Arceo y Eduardo Durante

RESUMEN

Numerosas investigaciones indican que la forma de evaluar el conocimiento ejerce una influencia sustancial en el estilo de aprendizaje de los alumnos. La evaluación de la competencia clínica y, dentro de esta, del razonamiento clínico, que es la guía de toda acción médica, representa un enorme desafío en el marco de la educación médica. Por lo tanto, se debería considerar indispensable la evaluación válida y fiable del razonamiento clínico (RC) durante el proceso de formación médica. Uno de los instrumentos de evaluación que fueron diseñados con este objetivo es el examen de comparación de *scripts* (SCT). En el presente artículo analizaremos el proceso de desarrollo del razonamiento clínico, así como la utilidad del SCT para su medición, y desarrollaremos una breve guía con los pasos por seguir para la construcción de este tipo de examen.

Palabras clave: educación médica, *scripts*, razonamiento clínico, evaluación.

DEVELOPMENT AND EVALUATION OF SCRIPTS DURING MEDICAL TRAINING

ABSTRACT

Research indicates that knowledge drives student learning. The assessment of clinical competence, including clinical reasoning, which is the guide of all medical action, represents a big challenge in medical education. Therefore, valid and reliable assessment of clinical reasoning during clinical development should be considered essential. The script concordance test (SCT) was designed to accomplish this objective. In this article we will analyse the process of development of clinical reasoning and the SCT utility for measuring it. We will also elaborate a guide for the construction of the SCT.

Key words: scripts concordance test, assessment, medical education.

Rev. Hosp. Ital. B.Aires 2013; 33(4): 144-152.

INTRODUCCIÓN

Numerosas investigaciones indican que la forma de evaluar el conocimiento ejerce una influencia sustancial en el estilo de aprendizaje de los alumnos.¹ La evaluación de la competencia clínica, y en particular del razonamiento clínico (RC), representa un enorme desafío en el marco de la educación médica. Por lo tanto, en la formación de médicos competentes, la evaluación debe ser considerada como una parte fundamental en el proceso de aprendizaje de forma tal que permita alcanzar los objetivos planteados y orientar el estilo de aprendizaje de los alumnos. Dentro de esta perspectiva, la evaluación debe ser incluida al inicio de la planificación curricular.

C. Brailovsky define la competencia profesional como “la capacidad de un profesional de utilizar su buen juicio así como también los conocimientos, habilidades, actitudes asociadas a la profesión para solucionar los problemas complejos que se presentan en el campo de su actividad profesional”.² Esta definición enumera

adecuadamente los diversos elementos que se deberían incluir a la hora de evaluar la competencia. A modo de ejemplo: no alcanza solamente con que un profesional conozca las bases conceptuales acerca de cómo se debe realizar la toma de un Papanicolaou y realice la praxia con perfección, sino también es necesario que sepa cuándo y cómo indicarlo, pueda comunicarse con la paciente en forma eficaz, tome en cuenta sus necesidades y el contexto, y además interprete el resultado del estudio y adopte las conductas que sean necesarias, siempre dentro de un marco de incertidumbre. Por lo tanto y según esta definición, ser un profesional de la salud competente incluye contar con diversas subcompetencias tales como habilidades reflexivas, organizativas y técnicas. Sin embargo, la competencia central y más importante del médico consiste en el *razonamiento clínico* (RC) que posibilita la toma de decisiones que conducen a la resolución de situaciones complejas,³ en los diferentes contextos en los que deba actuar.

Kreiter y Bergus⁴ han sugerido que el razonamiento clínico es una especie de “caja negra” y que la evaluación de lo que allí adentro ocurre requiere hacer inferencias a partir de la medición de sus productos (como en el

método tradicional de evaluación, en el cual los exámenes miden la capacidad de resolver problemas) o de sus entradas (un enfoque menos común que incluye la evaluación de la organización del conocimiento y la capacidad de integrar la nueva información). Algunos autores consideran como razonamiento clínico la aplicación del conocimiento, mientras que otros lo ven como una habilidad en sí misma (habilidad cognitiva).⁴ Estas diferentes perspectivas llevaron al desarrollo de múltiples enfoques para la evaluación del razonamiento. A fin de comprender mejor cómo se puede evaluar, desarrollaremos brevemente algunos conceptos actuales sobre el RC a la luz de investigaciones recientes.

RAZONAMIENTO CLÍNICO

Se entiende por RC el conjunto de procesos mentales mediante los cuales el médico plantea un diagnóstico y decide los planes de manejo y el pronóstico en un caso en particular.⁵

El RC se destaca por su función de guía de toda acción médica (el diagnóstico, el estudio de los casos clínicos, la elección de la conducta terapéutica y la estimación del pronóstico de un paciente). Por lo tanto, la evaluación válida y fiable del RC durante el proceso de formación médica es indispensable.³

El RC se lleva a cabo, básicamente, a partir de dos procesos: uno analítico y otro no analítico o basado en reconocimiento de patrones.⁶ Numerosos estudios demostraron que al hacer un diagnóstico, los estudiantes razonan generando sucesivas hipótesis y contrastándolas con sus hallazgos, de manera individual y secuencial, sin poder organizar su conocimiento en estructuras más complejas. Los estudiantes cuentan con escaso conocimiento clínico debido a que estuvieron expuestos a un número reducido de pacientes. En general, su conocimiento es más del tipo biomédico y está organizado en largas listas de datos, sin conexiones entre ellos. Los estudiantes analizan cada una de esas "partes" por separado y las van confirmando o rechazando hasta arribar al diagnóstico. Por ejemplo, si un alumno evalúa un paciente con alopecia pensará: "puede ser alopecia *areata*, tiña *capitis*, alopecia cicatrizal por lupus discoideo, alopecia frontal fibrosante" y para cada uno de esos diagnósticos pensará qué buscar en el examen físico y qué estudios solicitar para confirmar o rechazar cada una de sus hipótesis. A este tipo de RC se lo llama analítico. Por el contrario, los considerados expertos logran organizar su conocimiento en forma de redes conceptuales que les permiten priorizar e interpretar de manera más eficaz los hallazgos clínicos, a través de un proceso no analítico o de reconocimiento de patrones. Cuando un experto se enfrenta a un paciente, rápidamente relaciona el caso con patrones clínicos similares, incluso con casos particulares, a

los que se enfrentó con anterioridad en su práctica clínica. El razonamiento del experto frente al enfermo pasa por una secuencia rápida de eventos dada por la percepción de los elementos clave pertinentes, la activación automática casi inconsciente de las redes de conocimiento que contienen estos elementos clínicos significativos (*script triggering*), el rápido acceso a diferentes hipótesis, la verificación de síntomas y signos que confirman o descartan el diagnóstico y el uso racional de sus conocimientos para guiar de manera apropiada las investigaciones y el tratamiento (*script processing*).¹² Así, al evaluar al paciente descrito previamente, el experto, a diferencia del estudiante, primero examinará visualmente o a través de los sentidos el patrón de distribución de la alopecia e inmediatamente pensará en alopecia androgénica (como si se le apareciera una imagen). Luego, buscará los datos que considere necesarios para confirmar o rechazar su diagnóstico: examinará si el paciente presenta pelos miniaturizados, preguntará sobre los antecedentes familiares, indagará sobre la intensidad de la caída del cabello (que en la alopecia androgénica es lenta y progresiva y, en las otras, más aguda y localizada) y rápidamente llegará a una conclusión diagnóstica. Estas redes conceptuales que los expertos usan en su RC se denominan *illness scripts*.^{6,7,9-11} De esta forma, los *illness scripts* funcionan como organizadores de la estructura cognitiva de cada profesional y permiten una eficiente aplicación de los conocimientos y un rápido arribo a soluciones de manera competente en la mayoría de las situaciones clínicas.^{3,12,13,15} La importancia del contenido y del contexto (ambiente y circunstancias) así como de los factores personales (valores y actitudes) en la creación de estas redes o *scripts* hacen que cada una de ellas sea única para cada individuo. Por lo tanto, los *scripts* son únicos para cada profesional y no son transferibles de docentes a alumnos, ya que no solo incluyen información clínica relevante sino también información percibida inconscientemente de cada ambiente y circunstancias.

Schmidt y cols,⁶ han definido cuatro estadios diferentes del desarrollo e integración de estos procesos cognitivos: los estudiantes pasan desde la creación de un mapa conceptual simple causal de signos y síntomas (estadio 1), a ser capaces de sintetizar el cuadro clínico del paciente en una o dos oraciones (estadio 2). En el estadio 3 ya aparecen los *scripts* de enfermedades que organizan el conocimiento en complejas redes conceptuales de patrones clínicos y tienen una estructura de factores predisponentes, fisiopatología y consecuencias clínicas. En la tabla 1 se presenta un ejemplo acerca de cómo, ante un mismo paciente, organizarían e integrarían el conocimiento dos residentes en diferentes etapas de su formación (Tabla 1).

TABLA 1. Comparación del RC entre un residente de primer año y otro de cuarto en relación con un caso de gota en la rodilla (modificado de J. Bowen. Educational Strategies to Promote Clinical Diagnostic Reasoning. N Engl J Med 2006;355:2217-25).

Historia clínica. Consulta un hombre de 54 años diciendo que la noche anterior lo había despertado un fuerte dolor en la pierna. Luego de realizar el interrogatorio y examen físico, cada residente organiza el conocimiento de la siguiente manera:	
Presentación residente primer año	Presentación residente cuarto año
<p>"Puede ser una infección. Puede ser el comienzo de una artritis reumatoidea o de una artritis gonocócica. Como no recuerda ninguna caída, dudo de que sea una fractura. No sé si la artrosis se puede presentar así, aunque tiene historia de dolor de rodilla" (estadio 1-2).</p>	<p>"El paciente tiene gota aguda. Tuvo varios episodios de dolor agudo intenso de comienzo brusco en una sola articulación, que se ve flogótica al examen físico. Entre los episodios, está asintomático. No veo puerta de entrada para pensar en artritis séptica y el estado general del paciente es bueno. Hubiera esperado que fuera la primera articulación metatarsofalángica, pero se puede presentar en la rodilla. Nada sugiere un problema crónico" (estadio 3).</p>

Mientras que los estadios 1 y 2 se caracterizan por el razonamiento analítico, en el estadio 3 predomina el no analítico, y el estadio 4 representa el almacenamiento de estos *scripts* como *scripts* ejemplares.^{12,7} Los *scripts* comienzan a aparecer cuando los estudiantes se enfrentan con situaciones clínicas reales, y se desarrollan y perfeccionan durante toda la vida profesional.¹² Los alumnos en cualquier estadio usan ambas formas de razonamiento en la práctica clínica: los más avanzados tienen la ventaja de contar con una colección de *scripts* más rica y descansan más sobre este mecanismo que sobre el analítico, excepto cuando se enfrentan a un caso complejo sobre el cual no tienen experiencia.¹³

¿Se puede medir el razonamiento clínico?

Charlin y cols. proponen un método de evaluación llamado Examen de comparación de *scripts* (*Script concordance test [SCT]*), cuyo objetivo es medir la organización del conocimiento clínico en redes conceptuales que se revelan básicamente en la acción y en la toma de decisiones del médico, o sea, la organización de los conocimientos en *scripts*. El examen está compuesto por una serie de viñetas clínicas, que sitúan al examinado en un contexto clínico de incertidumbre dentro del cual debe interpretar datos para tomar decisiones. Para analizar la utilidad de este tipo de examen usaremos la ecuación de utilidad propuesta por Van Der Vleuten,¹⁴ que relaciona las variables más importantes para tener en cuenta al momento de decidir usar un examen:

$$U = (C(p) + V(p) + IE(p) + (p)A) \times 1/C$$

Así, la U es la utilidad de un instrumento y es igual a la suma de la proporción de *confiabilidad* (C), más la pro-

porción de *validez* (V), más la proporción del *impacto educativo* (IE), más la proporción de *aceptabilidad* (A), todo multiplicado por la inversa del costo.²

Cuando nos referimos a la confiabilidad hablamos de la precisión de la medición o la reproducibilidad del puntaje obtenido por el instrumento así como de la medida en que los ítems individuales de un examen (por ejemplo, cada pregunta en un *multiplechoice*) se comportan de manera similar (covarian) dentro de un examen determinado en relación con el desempeño global del sujeto en ese examen. Esta determinación se realiza con métodos estadísticos. Para el caso de preguntas con múltiples respuestas (no dicotómicas) como el SCT se utiliza el conocido método de alfa de Cronbach. El coeficiente varía entre 0 y 1,¹ y 1 es el valor asignado a la máxima confiabilidad del instrumento. Por lo general, un examen se considera confiable cuando el coeficiente alfa de Cronbach es de 0,80.¹⁵ En varias series publicadas sobre el SCT los valores oscilaron desde 0,79 hasta 0,82.¹⁹

En cuanto a la validez, existen diferentes dimensiones que se resumen con este término, por lo que podemos hablar de diferentes tipos de validez: 1) la validez de concepto (constructo), que muestra hasta qué punto el método mide lo que se pretende que mida, 2) la validez de criterio muestra si los resultados de un examen son concordantes con los resultados de otras pruebas ya validadas (es raro encontrar un *gold standard* en educación), o si estos resultados predicen el desempeño futuro de los alumnos y 3) la validez de contenido, que pone de manifiesto el grado de "muestreo" de todos los aspectos que se pretende medir.¹

Hasta el momento, el SCT ha demostrado confiabilidad y validez para diferenciar entre el razonamiento de un experto y el de un novato. Una serie de estudios ha

mostrado la capacidad del SCT para discriminar entre diferentes niveles de experiencia y su aplicabilidad en dominios tan diversos como la cirugía, la medicina familiar, el diagnóstico por imágenes, la oncología y la medicina de urgencias. Para evidenciar la validez de constructo del SCT se realizaron diferentes estudios.^{8,9} En uno de ellos, realizado en un servicio de imágenes, se les pedía a los participantes (radiólogos experimentados, residentes de radiología y estudiantes) que contestaran un examen de comparación de *scripts* basado en la imagen radiológica de un hemitórax radiolúcido unilateral. A partir de un caso clínico y de esa imagen, se les planteaban a los examinados diferentes hipótesis diagnósticas (neumotórax, mastectomía, hiperinsuflación compensatoria y atrapamiento aéreo). Luego se les iba presentando nueva información relevante para el caso, sobre la base de la cual ellos debían marcar el impacto que tenía cada uno de estos nuevos datos sobre la hipótesis planteada (no tenía impacto, la hacía más o menos probable). Los resultados mostraron que los médicos más expertos obtenían mejores puntajes en el examen que aquellos menos expertos e incluso el grado de variabilidad en la respuesta de los expertos dependía de la complejidad del caso planteado y no de cada ítem en particular. Por el contrario, los alumnos presentaban variabilidad con cada ítem, independientemente de la complejidad del caso presentado.⁷ Esto muestra un alto grado de confiabilidad y validez. Este y otros estudios también mostraron que el SCT no presenta el denominado “efecto intermedio”, que consiste en que médicos experimentados obtienen un puntaje igual, o incluso inferior, al de los médicos menos experimentados o de estudiantes en los exámenes que evalúan únicamente conocimiento teórico (como ocurre en exámenes tipo *multiple choice* o en el examen clínico objetivo y estructurado). Este efecto puede explicarse debido a que los alumnos están en una fase de adquisición rápida de conocimientos, por lo cual si se los evalúa con un examen que mide solo conocimientos teóricos pueden obtener mayor puntaje que un experto. Por el contrario, en el SCT, que está dirigido a medir RC (el cual aumenta con la experiencia clínica), ese “efecto intermedio” desaparece, debido a que para resolver tal tipo de examen es necesario interpretar datos clínicos para tomar decisiones en situaciones reales y no la simple aplicación de conocimientos teóricos o de memoria (validez de constructo).^{3,6,16,17} Por último, en un estudio realizado por Brailovsky y cols.,¹⁹ el examen de comparación de *scripts* también demostró que, cuando un examinado ha presentado una buena organización del conocimiento clínico en un momento particular de su formación, es de esperar que muestre buena organización en posteriores mediciones de este tipo de conocimiento. En ese estudio se siguió

a una cohorte de 24 estudiantes de una escuela de medicina hasta el final de su residencia en medicina familiar. Dos años antes de finalizar la residencia se los evaluó con el SCT y luego se los comparó con los resultados que obtuvieron en el examen de certificación. En Canadá, el examen de certificación final consta de tres pruebas de competencia clínica, dos que evalúan el RC (SAMP y SOO) y otra, las habilidades clínicas (ECOE). Los autores asumieron que las pruebas que medían RC tendrían mejor correlación que aquellas que medían habilidades clínicas, con los resultados del SCT tomado previamente. Los resultados fueron consistentes con esta hipótesis. Los coeficientes de correlación de Pearson fueron estadísticamente significativos cuando se compararon las puntuaciones en el SCT con los resultados de SAMP y SOO (que miden RC), respectivamente. Cuando la correlación se hizo con el ECOE (que mide habilidades clínicas) no se encontró una relación estadísticamente significativa.¹⁸

La aceptabilidad de un examen habla acerca de cuán adecuados consideran, tanto los evaluados como los evaluadores, los métodos de evaluación propuestos. En numerosos estudios^{19,8,20} en los cuales se les pregunta a los estudiantes y evaluadores acerca de la aceptabilidad, la gran mayoría acuerda en que consiste en una buena experiencia.

Si bien el SCT es un examen difícil de construir ya que requiere la colaboración de expertos entrenados, su costo es aceptable si tenemos en cuenta que es un examen escrito, que se puede realizar en 90 minutos (tiempo estimado para un examen de 20 casos con 60 ítems en total) y con el cual se pueden evaluar un gran número de examinados de manera simultánea.

Poco se sabe acerca de si este tipo de examen produce algún impacto educativo, por lo que se requieren más estudios al respecto.

Construcción del examen de comparación de *scripts*

El examen consta de varios problemas, presentados en viñetas clínicas cortas, cada una de las cuales es seguida por una serie de ítems. Cada uno de esos ítems aportan nueva información clínica al caso, contienen datos útiles que deben promover la reflexión y permitir la resolución del problema que se presenta, a través del razonamiento, de forma tal que el examinado debe responder acerca del impacto que considera que puede tener esta nueva información sobre la hipótesis o la decisión propuesta. Por ejemplo, si partimos de una viñeta clínica de un paciente que consulta por cansancio, el ítem va a estar compuesto por: 1) La hipótesis diagnóstica que puede ser “hipotiroidismo” y 2) la nueva información presentada que podría ser “anticuerpos para enfermedad celíaca positivos” (Tabla

2). Esta última información tiene un impacto que podría ser positivo, neutro o negativo sobre la hipótesis, según quién la evalúe. Es decir, la enfermedad celíaca es más frecuente en pacientes con hipotiroidismo, pero al mismo tiempo el cansancio podría ser causado por una anemia secundaria a enfermedad celíaca, o el dato de anticuerpos para enfermedad celíaca positivos podría valorarse de tal manera que aleje la probabilidad de hipotiroidismo. Como se puede ver, no hay una sola respuesta correcta a esta viñeta. La respuesta va a depender de quien está respondiendo, del contexto en que esté planteada esta situación y del peso que les dé cada uno de los examinados a los distintos datos. Para la construcción de un SCT es necesario seguir una serie de pasos que pueden resumirse de la siguiente manera:

- 1) Definición de una tabla de especificaciones para definir los contenidos que se van a evaluar: esta tabla se construye considerando todas las áreas del campo que se quiere evaluar, especificando los casos que se desarrollarán y el peso que se le dará a cada uno en la evaluación final, con el fin de garantizar un muestreo y una validez de contenido adecuados para el examen. Por ejemplo, existen tablas de especificaciones para confeccionar exámenes de certificación en el área de medicina familiar desarrolladas por el Colegio Médico de Canadá (http://www.cfpc.ca/red_book_TOC/).²¹
- 2) Una vez construida la tabla de especificaciones, se debe convocar a especialistas en el contenido o disciplina para construir los casos, previo entrenamiento en su elaboración. El examen de comparación de *scripts* debe contar con, al menos, veinte viñetas clínicas, cada una de ellas con tres a cinco ítems para alcanzar una confiabilidad adecuada (mayor de 0.80).¹³⁻¹⁹ La necesidad de este número de viñetas se debe a que, como en todo examen, se debe tener en cuenta la especificidad de caso,²² lo que implica que la resolución adecuada de un caso es específico para ese único caso y no predice el desempeño en otros, aun muy similares. Esto obedece al hecho de que la adquisición de competencias es específica para contenidos o contextos diferentes aunque parezcan similares (la competencia del examen físico en un paciente con insuficiencia cardíaca no predice el desempeño en el examen físico de un paciente con neumonía). Como consecuencia de este fenómeno, es necesario aumentar el número de casos en las evaluaciones para asegurar una adecuada confiabilidad intercasos.

Para la construcción de los casos, en primer lugar se les solicita a los especialistas que identifiquen distintas situaciones clínicas y que para cada una de ellas definan los siguientes aspectos:

- a) La hipótesis relevante, las estrategias de investigación y las opciones de tratamiento que se plantean.
- b) Las preguntas que se realizan ante dichas situaciones, los exámenes físicos y pruebas que piden para resolver el problema.

La información clínica, positiva o negativa, que buscan en sus investigaciones.

Una vez definidos estos puntos, se procede a la construcción de las viñetas como se relata a continuación.

Construcción de las viñetas clínicas

Las viñetas clínicas deben describir un caso clínico o una situación clínica auténtica, representativa y problemática, de forma breve, aportando información básica, poco diferenciada, que permita varias posibilidades de diagnóstico y de toma de decisiones, pero que al mismo tiempo contenga todos los elementos necesarios para arribar a una respuesta congruente (véase ejemplo de viñetas clínicas en Tabla 3).

Es importante que, conservando la ambigüedad de las situaciones clínicas reales, la viñeta sea capaz de ubicar al médico en un contexto clínico similar al de su práctica real y que sea capaz de movilizar (“gatillar”) su razonamiento, es decir, de activar sus redes cognitivas o *scripts*. En aquellos casos en los que el paciente tiene un solo motivo de consulta sin ninguna otra manifestación

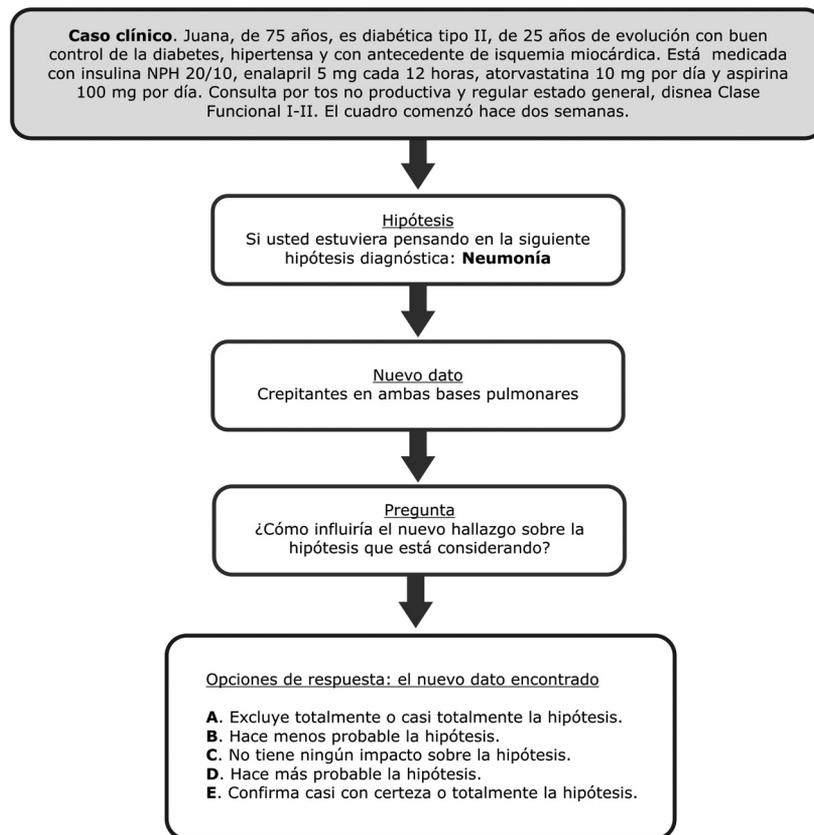
TABLA 2. Ejemplo de un ítem de SCT

Un paciente de 60 años acude a su consultorio de atención primaria por cansancio de varias semanas de duración.

Si usted piensa en la siguiente hipótesis diagnóstica	Y encuentra	Tiene el siguiente efecto sobre su hipótesis				
		A	B	C	D	E
Hipotiroidismo	Anticuerpos para enfermedad celíaca positivos					

- A.** Excluye totalmente o casi totalmente su hipótesis. **B.** Hipótesis menos probable. **C.** No tiene impacto sobre la hipótesis (efecto neutro). **D.** Hipótesis más probable. **E.** Confirma totalmente o casi con certeza la hipótesis.

TABLA 3. Ejemplo de confección de casos e ítems



VIÑETA CLÍNICA → Juana, de 75 años, es diabética tipo II, de 25 años de evolución con buen control de la diabetes, hipertensa y con antecedente de isquemia miocárdica. Está medicada con insulina NPH 20/10, enalapril 5 mg cada 12 horas, atorvastatina 10 mg por día y aspirina 100 mg por día. Consulta por tos no productiva y regular estado general, disnea Clase Funcional I-II. El cuadro comenzó hace dos semanas.

ITEMS →

	Si usted piensa en la siguiente hipótesis diagnóstica	Y encuentra	Tiene el siguiente efecto sobre su hipótesis				
			A	B	C	D	E
	Neumonía	Crepitantes en ambas bases pulmonares					
	Angina de pecho	Ecoestrés sin cambios respecto de estudios previos					
	Insuficiencia cardíaca	Soplo holosistólico 4/6 irradiado a axila y dorso					
	Infarto	ECG c/infradesnivel del CT-T en cara inferior de 2 mm					

- A. Excluye totalmente o casi totalmente su hipótesis
- B. Hipótesis menos probable
- C. No tiene impacto sobre la hipótesis (efecto neutro)
- D. Hipótesis más probable
- E. Confirma totalmente o casi con certeza la hipótesis

↑
Escala de Likert

clínica asociada, la información del enunciado puede completarse con información sobre antecedentes personales o familiares, medicación, etc. En todos los casos es importante dejar en claro, en el enunciado, en qué contexto está siendo atendido al paciente. Como se observa en los ejemplos, las viñetas clínicas pueden ser diseñadas para evaluar el RC sobre diagnóstico, estudios complementarios o tratamiento. Se debe evitar utilizar el mismo caso para evaluar las diferentes instancias, es decir, si se quiere evaluar RC

en diagnóstico y en tratamiento, es necesario realizar dos viñetas diferentes ya que de lo contrario se corre el riesgo de que varios casos se vuelvan interdependientes, lo que viola los principios de construcción del examen.²³

Construcción de los ítems

Una vez definida la viñeta clínica, se procede a la construcción de los ítems. Cada ítem consta básicamente de tres partes:

- a) La hipótesis diagnóstica, la acción de investigación o la opción de tratamiento que es relevante para la situación. El formato de los ítems difiere según el objetivo para valorar (diagnóstico, investigación o tratamiento) (Tabla 4).
- b) La presentación de nueva información (por ejemplo un signo, condición, estudio de imagen o resultado de una prueba de laboratorio) que podría tener un efecto sobre la hipótesis diagnóstica, la acción de investigación o la opción de tratamiento.
- c) Una escala tipo Likert de 5 puntos que está diseñada para que el examinado marque el impacto que tiene la nueva información clínica sobre la hipótesis inicial o la decisión propuesta.

Para la construcción del examen, la elección de cada ítem debe centrarse en los elementos que son más útiles para resolver el problema clínico, y que además

tengan un impacto positivo, neutro o negativo sobre las hipótesis diagnósticas, de investigaciones o terapéuticas de modo que aseguren algún grado de dispersión entre las posibles respuestas.

Es importante que no exista relación de linealidad entre la hipótesis y el nuevo dato presentado, ya que es necesario un proceso de reflexión y no la simple aplicación de conocimientos para la elección de la respuesta. Cada ítem dentro de las viñetas debe ser independiente de los otros. Por ejemplo, si se plantea el siguiente caso: lo consulta Analía de 58 años por pérdida de 5 kg de peso en el último año. Usted sabe que la paciente fue operada de cáncer de mama hace 2 años y toma tamoxifeno desde entonces. Si la hipótesis planteada fuera “metástasis hepáticas” y el nuevo dato para evaluar “elevación de las enzimas hepáticas”, no sería necesario realizar ningún razonamiento para

TABLA 4. Ejemplos de ítems de SCT sobre estudios complementarios y tratamiento

Estudios complementarios

Mujer de 67 años que es atendida en el servicio de urgencias por presentar un episodio sincopal mientras aguardaba para realizar trámites en el banco.

Si usted piensa en la siguiente hipótesis diagnóstica	Y encuentra	Tiene el siguiente efecto sobre su hipótesis				
		A	B	C	D	E
Holter	ECG con taquicardia sinusal, sin otros trastornos del ritmo					
Prueba ergonómica	Ausencia de antecedentes anginosos					
TC de cerebro	Examen físico sin particularidades					

- A.** El nuevo dato encontrado hace que la prueba diagnóstica propuesta no sea útil o esté totalmente contraindicado
- B.** El nuevo dato encontrado hace poco útil, o eventualmente peligrosa, la prueba propuesta
- C.** El nuevo dato encontrado no tiene ningún impacto sobre la prueba propuesta
- D.** El nuevo dato encontrado confirma la utilidad de la prueba propuesta
- E.** El nuevo dato encontrado hace indispensable o casi indispensable la prueba propuesta

Tratamiento

Paciente de 8 años visto en la consulta por otalgia y catarro de vías aéreas superiores. Había recibido tratamiento antibiótico hace 3 semanas por igual motivo.

Si usted piensa en la siguiente hipótesis de tratamiento	Y encuentra	Tiene el siguiente efecto sobre su hipótesis				
		A	B	C	D	E
Amoxicilina	Nivel líquido en oído medio					
Amoxicilina	Hipoacusia					
Beclometasona nasal	IgE normal para la edad					

- A.** Tratamiento totalmente contraindicado o no útil
- B.** Tratamiento poco útil o eventualmente peligroso
- C.** Ningún interés o impacto sobre la opción de tratamiento
- D.** Tratamiento útil
- E.** Tratamiento necesario o absolutamente necesario

TABLA 5. Ejemplo de puntuación SCT.

Respuesta	A	B	C	D	E
Nº de expertos que seleccionan la opción	0	3	6	1	0
Puntuación	0	3/10	6/10	1/10	0
Transformación	0	3/6	6/6	1/6	
Puntuación final	0	0.5	1	0.17	0

Panel de 10 expertos que responden a una pregunta:
Si seis expertos seleccionan la respuesta **C**, tres expertos la respuesta **B** y un experto de respuesta **D**, la opción **C** obtendrá la puntuación máxima de 1 punto ($6/10 = 1$ punto), mientras que a la respuesta B se le asignarán 0.5 puntos ($3/6$) y a la **D** 0.17 puntos ($1/6$). La puntuación total del examen es la suma de las puntuaciones obtenidas en cada ítem multiplicada por 100.

arribar a la respuesta, ni tener en cuenta el caso clínico planteado, ya que existe una linealidad entre la hipótesis y el nuevo dato. En este caso, una hipótesis válida podría ser: “efecto secundario de la medicación”, y la nueva información presentada: hepatograma normal hace 3 meses.

Por otro lado, se debe evitar que dos ítems sucesivos presenten opciones similares o que puedan estar relacionadas entre sí, ya que esto daría lugar a confusión y llevaría a evaluarlas como si estuviesen encadenadas o tuviesen carácter acumulativo. Se debe tener presente que la finalidad de cada ítem es determinar los efectos o el impacto de una información clínica aislada sobre la hipótesis propuesta y no los efectos aditivos de una serie de elementos clínicos. Por ejemplo en el siguiente caso: Marta, de 55 años, diabética en tratamiento con hipoglucemiantes orales, concurre a la consulta con diarrea crónica. Si la primera hipótesis diagnóstica presentada fuera “colon irritable” y la nueva información “anticuerpos antigliadina positivos”, la segunda hipótesis diagnóstica no debería ser “enfermedad celíaca”, ya que si por ejemplo la nueva información fuera anticuerpos antiendomiso negativos, sería muy difícil que el examinado evaluara esta información sin tener en cuenta la información del ítem anterior.

En las viñetas e ítems que contengan descripciones de estudios de imágenes, ECG, etc., es recomendable incluir la imagen correspondiente (en lugar de solo la descripción). Los elementos multimedia, además de hacer más real la situación clínica planteada, son un recurso didáctico con impacto educativo demostrado.

Puntuación y definición de puntaje de aprobación

Luego de su construcción, cada viñeta clínica será sometida a la evaluación de un grupo de expertos (entre 10 y 20, dependiendo del tipo de examen), para descartar términos confusos, validar el examen (verificando la validez aparente

y la relevancia del contenido, la validez de construcción y confiabilidad) y construir la tabla de puntuación.

La tabla de puntuación se construye asumiendo que todas las respuestas dadas por los expertos son aceptadas como válidas. Así, a la respuesta marcada por el mayor número de expertos le corresponde la máxima puntuación y el resto de las respuestas tienen puntajes proporcionales. Por ejemplo, si en un determinado ítem 6 de 10 expertos han seleccionado la opción **C** como correcta, 3 la B y 1 la D, la opción C obtendrá la puntuación máxima de 1 punto ($6/10 = 1$ punto), mientras que a la respuesta B se le asignarán 0.5 puntos ($3/6$) y a la D 0.17 puntos ($1/6$). Las opciones que no han sido elegidas por ningún experto tendrán una puntuación igual a cero (Tabla 5).⁴

Por lo tanto, la puntuación que recibe cada ítem se asigna a partir de las respuestas dadas por los expertos y su valor depende del acuerdo entre ellos. El resultado total del examen está determinado por la suma de las puntuaciones obtenidas en cada ítem. Una puntuación elevada significa mayor coincidencia con los expertos en los diferentes ítems, lo que indica un mayor grado de desarrollo de su razonamiento clínico.²⁰

Por el momento, el SCT no se usa para realizar evaluaciones sumativas, ya que todavía se requieren más estudios para poder establecer un punto de corte adecuado. Se trata de un problema psicométrico complejo que deberá ser resuelto antes de poder utilizar el SCT en instancias de evaluación sumativa.³

CONCLUSIONES

Diseñar un examen de comparación de *script* es una tarea compleja que requiere contar con una serie de expertos y seguir una serie de pasos (véase Puntos clave). Está demostrado que una puntuación elevada en este tipo de examen corresponde a una mayor coincidencia con los expertos en los diferentes ítems e indica un mayor grado de desarrollo del razonamiento clínico.

Sin embargo, es importante recordar que, si bien como instrumento de evaluación el SCT ha demostrado poseer validez de contenido, validez de constructo, validez de predicción, confiabilidad, aceptabilidad y bajos costos,⁸ este tipo de examen no es suficiente si lo que se quiere medir es la competencia clínica global, ya que no examina otras dimensiones importantes como: habilidades comunicacionales, habilidades actitudinales, manejo de situaciones de estrés, etc., que pueden ser evaluadas, por ejemplo, con exámenes orales o con ECOE. Es necesario desarrollar programas de evaluación multidimensionales que incluyan diferentes elementos que se complementen entre sí. La calidad de los instrumentos no se puede determinar de manera aislada sino, más bien, en relación con la de los demás.⁴

Puntos clave para la construcción del SCT

- Definición de tabla de especificaciones para definir los contenidos que se van a evaluar.
 - Construcción de al menos veinte viñetas clínicas con tres a cinco ítems cada una.
 - Viñetas clínicas: deben presentar la ambigüedad de las situaciones clínicas reales, ubicando al médico en un contexto clínico similar al de su práctica habitual, y ser capaces de movilizar el razonamiento (activar redes cognitivas)
 - los ítems cuentan con tres partes:
 - a) la hipótesis diagnóstica, la acción de investigación o la opción de tratamiento
 - b) la presentación de una nueva información que tenga efecto sobre la hipótesis
 - c) una escala tipo Likert de 5 puntos.
 - Evaluación del examen por un grupo de expertos.
 - Confección de la tabla de puntuación.
-

Conflictos de interés: Los autores declaran no tener conflictos de interés.

REFERENCIAS

1. Durante E. La evaluación de los conocimientos: lo que parece ser ¿realmente lo es? *Rev Hosp Ital B. Aires.* 2005;25(1):18-23.
2. Brailovsky C. Educación médica, evaluación de las competencias. En: Universidad de Buenos Aires. Facultad de Medicina. Secretaría de Asuntos Académicos. Aportes para un cambio curricular en Argentina 2001. Buenos Aires: Organización Panamericana de la Salud; 2001. p. 103-20.
3. Gormaz CB, Brailovsky C. Desarrollo del razonamiento clínico en Medicina. *Revista de Docencia Universitaria.* 2012;10(no. especial):177-99.
4. Dory V, Gagnon R, Vanpee D, et al. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ.* 2012;46(6):552-63.
5. Norman GR. The epistemology of clinical reasoning: perspectives from philosophy, psychology, and neuroscience. *Acad Med.* 2000;75(10 Suppl):S127-35. Citado en: Gormaz CB, Brailovsky C. Desarrollo del razonamiento clínico en Medicina. *Revista de Docencia Universitaria.* 2012;10(no. especial):177-99.
6. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med.* 1990;65(10):611-21. Errata en: *Acad Med.* 1992;67(4):287. Citado por: Durante E. La enseñanza en el ambiente clínico: principios y métodos. *Revista de Docencia Universitaria.* 2012;10(no. especial):149-75.
7. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med.* 2000;75(2):182-90.
8. Charlin B. Script questionnaires: their use for assessment of diagnostic knowledge in radiology. *Med Teach.* 1998;20(6):567-71.
9. Boshuizen HP, Schmidt HG. The role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *The Annual Conference of the American Educational Research Association.* Boston, MA; 1990 April 16-20.
10. Van de Wiel MW, et al. Differences in knowledge application by students and medical experts in clinical reasoning. *The Annual Meeting of the American Educational Research Association.* New Orleans, LA; 1994 April 4-8.
11. Schmidt HG, Boshuizen HP. Effects of activation of prior knowledge on the recall of clinical case. *The Annual Conference of the American Educational Research Association.* Boston, MA. 1990 April 16-20.
12. Charlin B, Van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: the script concordance approach. *Eval Health Prof.* 2004;27(3):304-19.
13. Van Der Vleuten CP. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ Theory Pract.* 1996;1(1):41-67.
14. Schmidt HG, Boshuizen HP, Hobus PP. Transitory stages in the development of medical expertise: the "intermediate effect" in clinical case representation studies. En: *Proceedings of the 10th Annual Conference of the Cognitive Science Society.* Montreal, Canada; 1988 August 17-19. Hillsdale, NJ: Cognitive Science Society; 1988. p.139-45.
15. Van der Vleuten CP, Van Luyk SJ, Beckers HJ. A written test as an alternative to performance testing. *Med Educ.* 1989;23(1):97-107.
16. Tornimbeni S, Pérez E, Olaz F, et al. *Introducción a la psicometría.* Buenos Aires: Paidós; 2008.
17. Brailovsky C, Charlin B, Beausoleil S, et al. 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. *Med Educ.* 2001;35(5):430-6.
18. Sibert L, Darmoni SJ, Dahamna B, et al. On line clinical reasoning assessment with Script Concordance test in urology: results of a French pilot study. *BMC Med Educ.* 2006;6:45.
19. Charlin B, Roy L, Brailovsky C, et al. The Script Concordance test: a tool to assess the reflective clinician. *Teach Learn Med.* 2000;12(4):189-95.
20. Durante E. Algunos métodos de evaluación de las competencias: escalando la pirámide de Millar. *Rev Hosp Ital B.Aires.* 2006;26(2):55-61.
21. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak.* 2008;8:18.