¿Necesitamos nuevos instrumentos para la evaluación de los residentes de Medicina de Familia?

Francisco Escobar Rabadán[a], Jesús López-Torres Hidalgo[a]

[a] Especialistas en Medicina de Familia y Comunitaria. Profesores Asociados de la Facultad de Medicina de Albacete

Correspondencia: Francisco Escobar Rabadán, Centro de Salud Universitario Zona IV, C/ Seminario, 4, 02006 Albacete, fjescobarr@sescam.jccm.es

Recibido el 2 de septiembre de 2008

Aceptado para su publicación el 21 de septiembre de 2008

RESUMEN

La tutorización de residentes debe incluir una evaluación de la adquisición de competencias a lo largo de su formación. El actual programa de la especialidad establece un sistema de evaluación formativa con el fin de monitorizar la consecución de los objetivos docentes, para la que se proponen instrumentos como la autovaloración de objetivos y actividades realizadas, análisis de registros clínicos, análisis de casos clínicos, demostraciones de técnicas diagnósticas y terapéuticas, análisis de los registros anotados en el Libro de especialista en formación, análisis de otros tipos de registros referentes a actividades teórico-prácticas, vídeo-grabaciones, evaluación clínica objetiva y estructurada, no descartando otros sistemas de valoración de objetivos docentes. Por otra parte, para la evaluación sumativa, además de los anteriores instrumentos, se proponen el registro de cumplimiento de actividades, el registro de asistencia a actividades programadas, evaluaciones escritas multitest y otros que puedan ser propuestos por la Comisión de Evaluación.
Proponemos el uso de un nuevo instrumento: el Test de Concordancia de Scripts, basado en una teoría cognitiva del desarrollo de la pericia clínica ("teoría de los scripts"). Este test sitúa a los examinados ante situaciones clínicas escritas, pero auténticas, en la que tienen que interpretar datos para tomar decisiones. Pertenece al tipo de simulaciones escritas, pudiendo ser realizada con papel u ordenador, y puede ser usado en no graduados, graduados o en educación médica continuada. El test está diseñado para probar si la organización del conocimiento clínico permite decisiones clínicas adecuadas, ya sean diagnósticas, de investigación o terapéuticas.
Palabras clave. Evaluación. Medicina familiar y Comunitaria

ABSTRACT

Do we need new tools to evaluate Primary Care house physicians?
The education of house physicians should include an evaluation of the competence they have acquired during their training. The current programme for this speciality establishes a training evaluation system whose aim is to monitor the achievement of teaching objectives. For this purpose tools such as self-assessment of objectives and activities performed, analysis of clinical records, analysis of case studies, demonstrations of diagnostic and treatment techniques, analysis of records in the Specialist in Training Book, analysis of other records concerning theoretical-practical activities, video recordings, objective, structured clinical evaluation, as well as other systems for the evaluation of teaching objectives. Furthermore for the total evaluation, in addition to the above mentioned tools, the following are proposed: a record of the compliance to the activities, a record of attendance to the programmed activities, written multi-test assessments and others that the Evaluation Committee may propose.
We propose the use of a new tool: the Script Concordance Test, based on a cognitive theory of the development of clinical expertise ("the scripts theory"). This test places examinees in written, but authentic, clinical situations in which they must interpret data to make decisions. It is a written simulation test and can be either paper- or computer-based and can be used in undergraduate, postgraduate, or continuing medical education. It is designed to probe whether the clinical knowledge of examinees is efficiently organized for making appropriate clinical decisions, whether these are related to diagnosis, investigation or treatment.
Key words. Evaluation. Family Practice.

INTRODUCCIÓN

Es difícilmente asumible que pueda existir tutorización sin un seguimiento pormenorizado de la formación del residente, que inexcusablemente debe incluir una evalua-

ción de sus competencias. Afortunadamente el actual programa de la especialidad[1] establece la necesidad de un contacto estructurado y continuado entre tutor y residente a lo largo de los cuatro años que durará la especialización, así como estancias en el centro de salud en cada uno de ellos. El programa establece un sistema de evaluación formativa, encomendada al tutor, a realizar por medio de reuniones periódicas entre éste y el residente con el fin de monitorizar la consecución de los objetivos docentes, con un mínimo de cuatro por año. Se proponen distintos instrumentos: autovaloración de objetivos y actividades realizadas, análisis de registros clínicos, análisis de casos clínicos, demostraciones de técnicas diagnósticas y terapéuticas, análisis de los registros anotados en el Libro de especialista en formación, análisis de otros tipos de registros referentes a actividades teórico-prácticas, vídeo-grabaciones, evaluación clínica objetiva y estructurada (ECOE), no descartando otros sistemas de valoración de objetivos docentes. En cuanto a la evaluación sumativa, además de los anteriores instrumentos, se proponen: registro de cumplimiento de actividades, registro de asistencia a actividades programadas, evaluaciones escritas multitest, y otros que puedan ser propuestos por la Comisión de Evaluación. Todos ellos constituyen métodos de evaluación que gozan de un amplio consenso entre expertos en este campo[2].

Pero, ¿podemos incorporar nuevos instrumentos de evaluación que nos permitan llevar a cabo esta tarea y que pueda ser entendido por parte de los residentes como una prolongación de su actividad clínica?

TEORÍA DE LOS SCRIPTS

Una parte significativa de la competencia de los médicos tiene como base la capacidad de enfrentarse con la incertidumbre. En un encuentro clínico no están disponibles de entrada todos los datos para resolver un problema. Éstos deben ser recopilados para formular el problema y resolverlo. Además, los datos pueden ser confusos, contradictorios y mal definidos, y a menudo se caracterizan por información imperfecta, inconsistente o incluso insegura. La capacidad para razonar en contextos de incertidumbre y resolver problemas pobremente definidos es el sello de la competencia profesional. Una valoración extensa del razonamiento clínico debería incluir por tanto instrumentos que midan la capacidad para resolver racionalmente problemas mal definidos[3].

Por otra parte, no podemos dejar de considerar que el médico clínico trata problemas que no siempre tienen una solución fácil. En el núcleo de la competencia profesional se encuentran el juicio y la perspicacia, que descansan sobre el conocimiento tácito. Este tipo de conocimiento no es visible ni tangible, por lo que no es fácil de evaluar, y constituye la piedra de toque de la competencia práctica profesional. Solo se pone de manifiesto en la acción, en situaciones auténticas en las que los médicos tienen que reflexionar sobre preocupaciones reales. En la medicina clínica, los médicos diestros y experimentados difieren de los que lo son menos en que tienen redes elaboradas de conocimiento adecuadas a sus tareas habituales. Estas redes, conocidas como scripts, están organizadas para alcanzar metas relativas a temas de diagnóstico, estrategias de investigación u opciones de tratamiento. Comienzan a aparecer cuando los estudiantes se enfrentan con sus primeros casos clínicos y posteriormente son desarrolladas y refinadas durante toda la vida clínica.

En situaciones específicas los clínicos movilizan conjuntos de conocimientos almacenados previamente (sus scripts) que son usados para comprender la situación y actuar de acuerdo a las metas específicas, como diagnóstico, investigación o tratamiento. Los scripts de clínicos con experiencia varían en sus detalles, porque cada uno tiene su experiencia clínica, pero son similares en sus elementos esenciales. Si no fuera este el caso, los clínicos serían incapaces de comunicar eficientemente acerca de enfermedades o pacientes, y no alcanzarían el mismo diagnóstico en situaciones similares. Los scripts contienen información acerca de las conexiones que unen los ítems de conocimiento (características clínicas) relacionadas con una enfermedad. Son estas conexiones, en situaciones diagnósticas, las que permiten a una persona tomar decisiones con relación a la fuerza o debilidad de una hipótesis o decidir si una característica clínica nunca está asociada con tal hipótesis, en cuyo caso la hipótesis tiene que ser rechazada. Similares conexiones son usadas para manejar decisiones de investigación o tratamiento[4].

Según la teoría de los scripts, los médicos utilizan estas estructuras de conocimiento especialmente adaptadas a las tareas que realizan comúnmente para procesar activamente información, a fin de confirmar o rechazar hipótesis u opciones de manejo. Los clínicos están por tanto haciendo continuamente juicios cualitativos sobre la significación de los datos que recogen. Cada uno de estos juicios puede ser medido, proporcionando un método de valoración del razonamiento sobre problemas mal definidos y en contextos de incertidumbre.

TEST DE CONCORDANCIA DE SCRIPTS

El Test de Concordancia de Scripts (TCS), propuesto por Charlin et al[5], es un instrumento de valoración basado en esta teoría cognitiva del desarrollo de la pericia clínica. Sitúa a los examinados ante situaciones clínicas escritas, pero auténticas, en la que tienen que interpretar datos para tomar decisiones. Pertenece al tipo de simulaciones escritas que puede ser realizada con papel u ordenador. Puede ser usado en no graduados, graduados o en educación médica continuada. Con el TCS se presenta al examinado una serie de problemas del paciente, proporcionándole elementos de información específicos, y se le pide que tome decisiones diagnósticas, de investigación o terapéuticas. El test está diseñado para probar si la organización del conocimiento clínico permite decisiones clínicas adecuadas, e intenta valorar la significación de las conexiones entre diferentes ítems, más que valorarlos aisladamente. El sistema de puntuación está diseñado para medir la distancia que existe entre los scripts del estudiante y de un panel de expertos.

La construcción del TCS requiere la colaboración de un pequeño número de expertos, siendo habitualmente suficiente con dos en la etapa de producción de los ítems. En una entrevista informal se les pide que describan algunas situaciones clínicas que son representativas de su campo y son problemáticas. Para cada situación se debe especificar:
a) La hipótesis relevante, estrategias de investigación y opciones de tratamiento.
b) Las preguntas que ellos realizan, los exámenes físicos que llevan a cabo y las pruebas que piden para resolver el problema.
c) Por último, qué información clínica, positiva o negativa, buscarían para sus investigaciones.

En el TCS no hay necesidad de buscar datos clínicos inusuales, siendo posible discriminar entre los examinados con datos comunes que requieren interpretación. Consta de varios problemas del paciente, presentados en viñetas cortas, cada una de las cuales va seguida por una serie de ítems relacionados. Cada parte del test está basado en un caso clínico descrito en unas pocas frases. La descripción puede ser sencilla o muy detallada, según la pregunta planteada, pero debe contener sistemáticamente toda la información necesaria para que un experto haga una elección informada de la situación. A veces puede ser necesario describir una evolución de dicha situación. En la tabla 1 se muestra un ejemplo de TCS.

El formato de los ítem difiere según el objetivo a valorar (diagnóstico, investigación o tratamiento), y para una viñeta dada, los ítem se reagrupan por formato. Cada ítem consta de tres partes:
1) Hipótesis diagnóstica, acción de investigación u opción de tratamiento, que es relevante a la situación.
2) Presenta nueva información (por ejemplo un signo, condición, estudio de imagen o resultado de test de laboratorio) que podría tener un efecto sobre la hipótesis diagnóstica, la acción de investigación o la opción de tratamiento.
3) Escala tipo Likert de 5 puntos (se han usado también escalas de 7 puntos).

Se pueden establecer otros formatos para otras situaciones, como dar un pronóstico o proporcionar un consejo.

La elección de las preguntas se centra en los elementos que son más útiles para resolver un problema clínico. Cada ítem está construido para que sea necesario reflexionar para responderlo, y cada uno es independiente de los otros. Para evitar que los examinados obtengan información acumulativa acerca del paciente de las preguntas siguientes, la hipótesis u opciones cambian para cada pregunta. También está claramente especificado en las instrucciones para los participantes que, dentro de las viñetas, cada ítem es independiente de los otros. La meta de cada ítem no es determinar el efecto aditivo de una serie de elementos de información clínica, sino determinar el efecto de cada ítem aislado sobre una hipótesis, acción u opción de tratamiento. Para prevenir un efecto indicador sobre los examinados, los ítems son construidos para ofrecer respuestas entre todos los valores de la escala tipo Likert.

Una vez construido, el test se remite a un grupo de expertos, que servirá también para elaborar un sistema de puntuación. Se les pide que identifiquen los ítems que encuentran confusos o irrelevantes, los cuales serán descartados o reescritos. El proceso de puntuación del test está basado en el principio de que la respuesta de cualquier experto refleja solo su opinión, y aquellas respuestas para las que no hay acuerdo entre todos los expertos no deberían ser descartadas. Es decir, cualquier respuesta dada por un experto tiene un valor intrínseco, incluso si otros expertos no están de acuerdo con él. Por lo tanto, las puntuaciones para cada ítem son computadas a partir de las frecuencias dadas para cada punto de la escala tipo Likert por los expertos, y su valor depende del acuerdo entre ellos.

Así, el test mide la concordancia entre los scripts de los examinados y los del panel de expertos.
El resultado del test viene determinado por la suma de las puntuaciones obtenidas en cada ítem. Por conveniencia de interpretación se recomienda convertir todas las puntuaciones para alcanzar una puntuación máxima de 100. Ésta refleja que el examinado da a cada ítem la respuesta que dan la mayoría de los expertos. Las puntuaciones aumentan con la experiencia clínica (médicos con experiencia puntúan mejor que residentes, y ambos mejor que estudiantes). Una explicación podría ser que el TCS explora la capacidad de interpretación de datos en la toma de decisiones clínicas, una destreza que tiene más que ver con la competencia clínica que el simple recordatorio de datos.

Esta circunstancia contrasta con los tradicionales exámenes de valoración de conocimientos, en los que se presenta el llamado "efecto intermedio", consistente en que los clínicos experimentados puntúan a duras penas mejor, o incluso lo hacen peor, que clínicos menos experimentados o estudiantes. Y es que, en medicina clínica, los médicos expertos y diestros difieren de los que lo son menos en que tienen redes elaboradas de conocimiento (lo que hemos denominado scripts) que se adecuan a las tareas que tienen que realizar regularmente. Esta clase de conocimiento se revela solo en la acción, cuando los médicos tienen que reflexionar sobre situaciones reales. En este sentido, el TCS sitúa a los examinados ante situaciones clínicas problemáticas en las que deben responder a preguntas planteadas por expertos en esas situaciones y tienen que interpretar datos para tomar decisiones. Permite demostrar la organización del conocimiento clínico, tratando de verificar si el conocimiento de los examinados es más elaborado que disperso. En este sentido, el test valora cómo los ítem de conocimiento están estructurados y conectados, incluyendo la naturaleza de la asociaciones entre ítem de conocimiento, más que la acumulación de los mismos. Cuando los clínicos son medidos con este instrumento, las puntuaciones reflejan su nivel de competencia y experiencia, sin verse afectado por el "efecto intermedio"[6]. Y hay que subrayar que no todos los clínicos son capaces de organizar adecuadamente sus conocimientos. El TCS permitiría identificar a aquellos residentes que tienen organizado su conocimiento para un uso eficiente en su trabajo clínico, aquéllos cuyo conocimiento alcanza una estructura elaborada. Además, cuando un examinado ha mostrado 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[7].

VENTAJAS Y LIMITACIONES DEL TCS

El método de puntuación agregado utilizado en el TCS, en comparación con métodos que buscan el consenso entre expertos, está más próximo a la realidad de la práctica clínica, es más rico y tiene más capacidad para discriminar a los examinados en su nivel de destreza[8]. Se han señalado, sin embargo, posibles problemas inherentes a este método de puntuación. En este sentido, Bland et al[9] llaman la atención acerca del hecho de no recompensar con puntos sobre la base de la corrección de la respuesta de los examinados. Así, para un ítem en el que todos los expertos respondieran +1 en una escala tipo Likert de 5 puntos, los examinados que respondieran –2 o +2 obtendrían la misma puntuación: 0. De manera que el examinado que está de acuerdo con el experto en la dirección del impacto pero no en el grado del mismo recibiría la misma puntuación que quien no sabe la dirección ni el grado de impacto. Estos autores hacen hincapié en que se premie el conocimiento que demuestre el examinado sobre relaciones definidas objetivamente que son importantes para la solución del problema clínico, por lo que aconsejan seleccionar una única "mejor respuesta" y medir las diferencias respecto de ésta de las respuestas de los examinados. Basan su recomendación además en la constatación de la similitud de las puntuaciones obtenidas con sistemas agregados y de "mejor respuesta", ya sea seleccionando la moda o la media de las respuestas de los expertos.

El panel de expertos debe estar constituido por médicos con buena experiencia clínica en el campo que queremos valorar, más que expertos en pequeñas partes de él. Ahora bien, la selección de los mismos dependerá de la situación a valorar: por ejemplo, para valorar el conocimiento que hayan alcanzado los residentes de medicina de familia en ginecología, los expertos pueden ser seleccionados entre especialistas en ginecología, pero también entre médicos de familia con práctica en este campo. Debido a las variaciones entre las respuestas de los miembros del panel, incluso entre grupos homogéneos de médicos, éste tiene que ser suficientemente grande como para proporcionar puntuaciones fiables del propio panel y por tanto fiables para los examinados. Si está formado por menos de diez expertos, puede que no se consiga una valoración fiable del razonamiento clínico. Con paneles de más de diez miembros se ha puesto de

manifiesto una fiabilidad aceptable, recomendándose para exámenes de alto nivel que esté conformado por veinte miembros. Reunir más expertos sólo produce un beneficio marginal en términos de propiedades psicométricas, de forma que se admite la necesidad de reunir un mínimo de diez miembros, pero reunir más de veinte no ofrece ventajas[3].

La valoración con el TCS tiene tres principios básicos, relacionados con los componentes inherentes del test: la tarea que se exige a los examinados, la forma en que son registradas sus respuestas y la forma en que éstas son transformadas en puntuaciones.
La tarea representa una situación clínica auténtica, y es un desafío, incluso para un experto, ya sea porque la viñeta no contiene todos los datos necesarios para proporcionar una solución o porque pueden ser defendidas varias actitudes. Varias opciones, ya sean diagnósticas, de manejo o de actitud, son relevantes. Los ítems se refieren a preguntas que los expertos se plantean a sí mismos para buscar una solución.

El formato de respuesta está de acuerdo a lo que se conoce sobre el proceso de razonamiento clínico, toda vez que el método de puntuación tiene en cuenta la variación de las respuestas entre los miembros del panel de expertos. En el rico contexto de la vida clínica real, tales como los descritos en las viñetas e ítem del TCS, las respuestas de éstos varían sustancialmente, lo que está en concordancia con los hallazgos de la investigación sobre razonamiento clínico: en situaciones similares, los profesionales no recogen exactamente los mismos datos y no siguen exactamente las mismas líneas de pensamiento, mostrando una variación sustancial en la resolución de cualquier caso real o simulado.

Una ventaja del TCS con relación a otros test es que a los miembros del panel se les pide que contesten preguntas que son muy similares a las que se plantean ellos mismos en su propio trabajo clínico. Además, en contraste con otras pruebas que requieren una revisión de conocimientos para una realización óptima, un clínico puede realizarlo en cualquier momento sin ninguna preparación previa. Estas dos razones explican por qué, en la práctica, no es difícil reunir miembros para conformar paneles de referencia.

Hipótesis	Si el paciente cuenta o encontramos en la exploración	La hipótesis será
Gota	Antecedente de artritis de primera articulación metatarso-falángica hace 3 meses	-2 -1 0 +1 +2
Artritis por depósito de pirofosfato cálcico	Antecedente de caída hace dos días	-2 -1 0 +1 +2
Artritis reumatoide	En una analítica de hace 1 año tenía ácido úrico de 9,2 mg/dl	-2 -1 0 +1 +2
Artritis gonocócica	Antecedente de uretritis hace 5 años	-2 -1 0 +1 +2
Artritis reactiva	Lumbalgia y rigidez matutina en las últimas semanas	-2 -1 0 +1 +2
Artritis psoriásica	Ausencia de lesiones cutáneas	-2 -1 0 +1 +2
Artritis traumática	“Cajón anterior” en rodilla derecha	-2 -1 0 +1 +2
Gonartrosis	“Peloteo rotuliano” en rodilla derecha	-2 -1 0 +1 +2

Tabla 1. Viñeta clínica: PGS, hombre de 45 años, acude a consulta por dolor e hinchazón en su rodilla derecha, de unas 24 horas de evolución.

BIBLIOGRAFÍA

1. Comisión Nacional de la Especialidad de Medicina Familiar y Comunitaria, Consejo Nacional de Especialidades Médicas, Ministerio de Sanidad y Consumo, Ministerio de Educación Cultura y Deporte. Programa de la especialidad de Medicina Familiar y Comunitaria. Disponible en: http://www.fisterra.com/material/premios/programa_mfyc_2002.pdf. Con acceso el 24 de mayo de 2008.
2. Frohna JG, Kalet A, Kachur E, Zabar S, Cox M, Halpern R, Hewson MG, Yedidia MJ, Williams BC. Assesing residents'competency in care management: report of a consensus conference. Teach Learn Med 2004; 16: 77-84.
3. Gagnon R, Charlin B, Coletti M, Sauvé E, van der Vleuten C. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? Med Educ 2005; 39: 284-291.
4. Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: theory and application for clinical reasoning instruction and research. Acad Med 2000; 75: 182-190.
5. Charlin B, Roy L, Brailowsky C, Goulet F, van der Vleuten C. The Script Concordance test: a tool to asses the reflective clinician. Teach Learn Med 2000; 12: 189-195.
6. Charlin B, Brailowsky CA, Brazeau-Lamontagne L, Samson L, Leduc C, van der Vleuten C. Script questionnaires : their use for assessment of diagnostic knowledge in radiology. Med Teach 1998; 20: 567-571.
7. Brailowsky C, Charlin B, Beausoleil S, Coté S, van der Vleuten C. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study of the script concordance test. Med Educ 2001; 35: 430-436.
8. Charlin B, Desaulniers M, Gagnon R, Blouin D, van der Vleuten C. Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity. Teach Learn Med 2002; 14: 150-156.
9. Bland AC, Kreiter CD, Gordon JA. The psychometric properties of five scoring methods applied to the script concordance test. Acad Med 2005; 80: 395-399.