

# FIABILIDAD Y VALIDEZ DE LA VERSIÓN ESPAÑOLA DEL TEST DE FRESNO. EVALUANDO LAS COMPETENCIAS DE LOS RESIDENTES DE MEDICINA FAMILIAR Y COMUNITARIA EN PRÁCTICA BASADA EN LA EVIDENCIA.

**Autores:** Gemma Flores Mateo<sup>1</sup>, Josep M Argimon-Pallàs<sup>2</sup>, Josep Jiménez-Villa<sup>2</sup> i Enriqueta Pujol-Ribera<sup>1</sup>

<sup>1</sup> Técnica de investigación del Institut d'Investigació en Atenció Primària Jordi Gol, Barcelona.

<sup>2</sup> Divisió d'Avaluació, Servei Català de la Salut, Barcelona.

**Correspondencia:** [epujol@idiapjgol.org](mailto:epujol@idiapjgol.org)

**Correctores:** Mèncià Benítez y Sebastià Juncosa

*Este trabajo fue galardonado con el primer premio de los VIII Ajuts a la Recerca de la CAMFiC, otorgado en noviembre de 2006.*

**Publicado:** junio 2014

Gemma Flores Mateo, Josep M Argimon-Pallàs, Josep Jiménez-Villa y Enriqueta Pujol-Ribera (2014)

*Fiabilidad y validez de la versión española del test de Fresno. Evaluando las competencias de los residentes de medicina familiar y comunitaria en práctica basada en la evidencia..*

**Available at:** <http://pub.bsalut.net/butlleti/vol32/iss2/4>

*Este es un artículo Open Access distribuido según licencia de Creative Commons (<http://creativecommons.org/licenses/by-nc-nd/2.5/es/>)*

---

<http://pub.bsalut.net/butlleti/vol32/iss2/4>

# RESUMEN

**Objetivos:** Adaptar al español y validar el test de Fresno para evaluar las competencias en práctica basada en la evidencia (PBE). Conocer la efectividad de una intervención formativa a los residentes de medicina familiar y comunitaria (MFIC) en estas competencias.

## Sujetos y métodos

**Diseño:** Adaptación transcultural del test al español siguiendo recomendaciones internacionales. Estudio transversal para evaluar la fiabilidad y la validez del test. Estudio antes-después para analizar la efectividad de la intervención formativa.

**Emplazamiento:** Unitats docents de MFIC de Catalunya.

**Participantes:** Tutores de MFIC (expertos, n=56), residentes de segundo año (principiantes, n=202) y especialistas en MFIC (experiencia intermedia, n=17).

**Medidas:** edad, sexo i cursos de PBE previos a la intervención formativa. El test de Fresno de conocimientos y habilidades en PBE. Medidas de resultados principales: consistencia interna, concordancia intra y entre evaluadores; análisis de los ítems y capacidad de detectar cambios significativos.

**Intervención formativa:** Curso con 4 sesiones interactivas de 4 horas.

**Resultados:** participaron 275 personas. Completaron el pretest el 94,1% (190/202) de los residentes y el 75,2% (152/202) retornó los dos cuestionarios. Alfa de Cronbach 0,88 y 0,77, respectivamente. Fiabilidad entre evaluadores: 0,95

preintervención y 0,85 postintervención e intraevaluador 0,71 y 0,81, respectivamente. La puntuación media: residentes 60,4, expertos 149,8, y grupo intermedio 110,4 ( $P<0,001$ ). La media de diferencia entre antes y después de la intervención fue 47,7 (IC del 95%: 42,8-52,5). Medida del efecto 1,77 (IC del 95%: 1,57-1,95).

**Conclusiones:** La versión española del test de Fresno es una herramienta fiable y válida para evaluar conocimientos y habilidades en PBE.

**Palabras clave:** Evidence-Based Practice; Family practice; professional competence; effectiveness; questionnaires; validation studies

# SUMMARY

**Objectives:** Adapt and validate the test of Fresno into Spanish in order to assess competencies in evidence-based practice (EBP). To assess the effectiveness of an educational intervention for family and community medicine (F&CM) residents in these skills.

## Subjects and methods

**Design:** Transcultural Adaptation of the test into Spanish following international recommendations. Cross-disciplinary study in order to assess the reliability and validity of the test. Before – after study in order to analyze the effectiveness of the training intervention.

**Location:** Educational Units of F&CM from Catalonia.

**Participants:** F&CM Tutors (experts, n = 56), second-year residents (beginners, n =

202) and specialists in F&CM (intermediate experience, n = 17).

**Measures:** age, gender and educational intervention prior to the PBE. The Fresno test knowledge and skills in PBE. Main results measures: internal consistency, concordance intra and between evaluators; analysis of the items and the ability to detect significant changes.

**Educational Intervention:** Course with 4 interactive sessions of 4 hours.

**Results:** participated 269 people. Completed the pre-test 94.1% (188/202) of residents and 75.2% (152/202) of them returned the two questionnaires. Alpha of Cronbach 0.88 and 0.77, respectively. Reliability between evaluators: 0.95 pre-intervention and 0.85 post-intervention and intra-evaluator 0.71 and 0.81, respectively. Average score: 60.4 for residents, 149.8 for experts, and 110.4 for intermediate group (P<0.001). The average difference between before and after the intervention was 47.7 (CI 95%: 42.8-52.5). The effect size 1.77 (CI 95%: 1.57-1.95).

**Conclusions:** The Spanish version of the test of Fresno is a reliable and valid tool for assessing knowledge and skills in EBP.

**Keywords:** Evidence-Based Practice; Family practice; professional competence; effectiveness; questionnaires; validation studies

## INTRODUCCIÓN

La práctica basada en la evidencia (PBE) supone tomar decisiones integrando la mejor evidencia obtenida de la investigación de calidad, las preferencias y

circunstancias de las personas (es decir, sus preocupaciones, expectativas, influencias culturales y características individuales) y la experiencia clínica, con el objetivo de mejorar la efectividad, la seguridad, la equidad y, en definitiva, la calidad de la atención sanitaria<sup>1,2</sup>.

Varios autores<sup>3,4</sup> y organizaciones como el Institute of Medicine<sup>5</sup> proponen que la PBE es una de las competencias clave del currículum de todos los profesionales de la salud. No obstante, hasta ahora, a pesar de los esfuerzos formativos en esta área<sup>6</sup>, no se ha conseguido la aplicación habitual de la PBE a la cabecera del paciente o en la resolución de los problemas clínicos diarios. Entre el 10 y el 40% de pacientes no reciben una atención basada en la mejor evidencia disponible y entre un 11 y un 30% de tratamientos son innecesarios o pueden ser nocivos para los pacientes<sup>7-9</sup>.

Varias revisiones sistemáticas han analizado los resultados de la formación en PBE en las competencias de los participantes<sup>10-14</sup>. En síntesis, estas revisiones, han observado de forma consistente, que la formación en PBE comporta mejoras en los conocimientos en metodología de la investigación y estadística, así como en las actitudes hacia la utilización de la literatura en la toma de decisiones, pero no han encontrado cambios en las conductas. No obstante, estos resultados deben ser interpretados con cautela dadas las limitaciones en la validez interna de la mayoría de estudios<sup>15</sup>. Entre las principales críticas de estos estudios, cabe destacar que se dan pocos detalles sobre cómo se tiene que medir la efectividad de la intervención formativa, sobre el proceso de desarrollo y validación de los cuestionarios utilizados, sobre la forma de administración y sobre el tiempo entre la administración y la intervención<sup>15</sup>.

En cuanto a los instrumentos disponibles para evaluar la formación en PBE hay que destacar la revisión sistemática de Shaneyfelt et al<sup>16</sup>, que incluyó 115 estudios (representativos de 104 instrumentos). Estos autores identificaron varios instrumentos de gran calidad<sup>17-20</sup>, válidos y fiables para evaluar la mayoría de competencias requeridas para la PBE (procedimientos de aprendizaje, conocimientos, habilidades, actitudes y conductas en esta área) de los alumnos individuales y para determinar la efectividad de los programas de formación en PBE. Estos instrumentos disponen de 3 o más pruebas de validez, incluida la capacidad de discriminar entre diferentes niveles de competencias (validez discriminando). Por otra parte, han demostrado su fiabilidad entre evaluadores y comportan una evaluación externa de los resultados (no únicamente autoevaluación). Dadas sus propiedades psicométricas sólidas, se justifica su utilización en evaluaciones formativas y/o sumativas.

Entre los instrumentos de alta calidad, en el presente estudio se ha seleccionado test de Fresno<sup>17</sup>, diseñado para evaluar todas las dimensiones de la PBE. Este test, se inicia con presentación de dos escenarios clínicos que plantean alguna incertidumbre clínica. Incluye siete preguntas abiertas sobre los escenarios de forma que el participante debe: 1) formular una pregunta estructurada siguiendo la estrategia PICO (problema, intervención, comparación y resultados), 2) identificar el diseño de investigación más apropiado para responder a la pregunta y realizar la búsqueda en las bases de datos electrónicas adecuadas; 3) hacer la lectura crítica de los artículos para determinar su pertinencia y validez, y analizar la magnitud y la importancia de los hallazgos de la

investigación. También debe responder cinco preguntas que implican hacer algunos cálculos.

El hecho de plantear preguntas abiertas, en lugar de preguntas con opción múltiple o verdadero-falso, exige a los participantes una respuesta reflexiva y desarrollo de los propios razonamientos.

Aunque el test de Fresno es un instrumento de gran calidad, con pruebas de fiabilidad y validez, este atributos no son fijas, sino que deben evaluarse en relación con la población específica y el contexto<sup>21</sup>. Así pues, un instrumento que ha demostrado tener propiedades psicométricas satisfactorias en una población no necesariamente es adecuado para ser utilizado en otros<sup>22</sup>. La validación del cuestionario de Fresno en diferentes idiomas, diferentes grupos de profesionales, y entornos culturales, permitirá su aplicación generalizada, así como las comparaciones entre países y la evaluación de diferentes métodos de enseñanza.

El objetivo general del presente estudio es realizar la adaptación cultural del cuestionario de Fresno al español y su posterior validación para asegurar la equivalencia de la versión en español comparada con la versión original en inglés.

Se definieron los siguientes objetivos específicos:

1. Hacer la adaptación transcultural del cuestionario en español.
2. Evaluar la fiabilidad del test mediante el análisis de la consistencia interna y la fiabilidad entre e intra evaluadores

3. Evaluar la validez de constructo del test mediante la comparación de las calificaciones obtenidas por los tutores de Medicina Familiar y Comunitaria (MFyC) con instrucción formal en PBE, los residentes de MFyC, y de médicos especialistas en MFyC con experiencia intermedia en PBE.

4. Analizar la capacidad del cuestionario de detectar cambios significativos, comparando las puntuaciones de los residentes de MFyC antes y después de un curso breve sobre PBE.

## PARTICIPANTES Y MÉTODOS

### Diseño

El presente estudio incluyó tres tipos de diseño. La adaptación transcultural del test de Fresno al español siguiendo las recomendaciones internacionales en este tema (traducción directa y retrotraducción, entrevistas cognitivas y método de parafrasear, comparación de las versiones consensuadas con el instrumento original) para comprobar la equivalencia lingüística y semántica con el instrumento original. La fiabilidad y la validez de constructo se evaluó mediante un estudio transversal. Para analizar la capacidad de los test de detectar cambios importantes (responsiveness) se hizo un estudio antes y después (administración del test antes y después de la intervención formativa).

### Emplazamiento

El estudio se realizó en las Unidades docentes (UDD) de Medicina Familiar y Comunitaria (MFyC) de Cataluña, que propusieron el estudio a residentes y tutores. Al inicio del programa docente de MFyC los residentes eligen el centro de atención primaria (CAP) donde realizarán su formación de postgrado. Estos CAPs estando asignados a las UDD de MFyC por criterios geográficos. En el momento de la realización del estudio (año 2007) el número de UDD era 17 y el rango de residentes de cada UD era de 2 a 52. La formación en PBE era parte del contenidos docentes habituales de los residentes de segundo año (n = 202).

### Participantes

El test de Fresno se administró a tres grupos de profesionales: 56 médicos tutores de los residentes de MFyC, con formación reglada en PBE y que participaban regularmente en estudios experimentales (grupo de expertos), 202 residentes de segundo año (grupo de principiantes) y 17 especialistas en MFyC con menor experiencia en PBE (experiencia intermedia).

### Instrumento de medida: El test de Fresno (anexo 1).

Tal como se ha descrito anteriormente, el test de Fresno<sup>17</sup>, se inicia con la presentación de dos escenarios que plantean alguna incertidumbre clínica. Incluye siete preguntas abiertas sobre estos escenarios de forma que el participante debe: 1) formular una pregunta estructurada siguiendo la estrategia PICO (problema, intervención, comparación y resultados), 2) identificar el diseño de investigación más apropiado para responder a la pregunta y realizar la

búsqueda en las bases de datos electrónicas adecuadas; 3) hacer la lectura crítica de los artículos para determinar su pertinencia y validez, y analizar la magnitud y la importancia de los hallazgos de la investigación. También debe responder cinco preguntas que implican hacer algunos cálculos.

La puntuación de las respuestas a las preguntas abiertas se hace mediante rúbricas de evaluación estandarizadas, que especifican criterios explícitos de evaluación para cada pregunta, con el objetivo de minimizar la variabilidad en la misma. Con este procedimientos se establecen cuatro categorías de respuesta: insuficiente, limitada, notable y sobresaliente. Así, cada respuesta tiene una puntuación numérica que puede ir de 0 a 24 puntos en algunas preguntas o 4 puntos en otros. A modo de ejemplo, en el caso de una pregunta con puntuación máxima de 24 puntos, una puntuación < 8 se considera respuesta insuficiente, entre 8 y 15 puntos respuesta limitada, entre 16 y 23 puntos respuesta notable y 24 respuesta excelente. La puntuación máxima posible del test de Fresno es de 212 puntos, que se obtienen de sumar las puntuaciones de todos sus elementos. Las puntuación numéricas de las preguntas permiten establecer puntos de corte para clasificar las respuestas en aprobado o insuficiente. En el presente estudio, dos evaluadores (miembros del equipo investigador) puntuaron las respuestas de forma independiente, después de eliminar los datos de identificación de los participantes. Para homogeneizar la metodología de puntuación hicieron una prueba piloto con 10 cuestionarios. Los procedimientos completos de puntuación de Ramos et al se pueden consultar en la web de BMJ .

(<http://www.bmj.com/content/suppl/2003/02/10/326.7384.319.DC1>).

### **Variables de resultados principal y otras variables**

Como medidas de resultados principales se analizó la consistencia interna del test de Fresno, la concordancia intra y entre evaluadores (*intra-rater inter-rater reliability*); análisis de la capacidad de discriminación y la dificultad de los ítems y la capacidad del test de detectar cambios significativos (*responsiveness*). También se registró edad, sexo, formación previa en PBE y el tiempo necesario para responder a las preguntas del test.

### **Intervención formativa**

La intervención formativa consistía en 4 sesiones interactivas de 4 horas de duración, diseñadas para desarrollar los conocimientos y habilidades de los residentes de segundo año en PBE. El curso incluía lecturas breves, trabajo en grupos pequeños y análisis de casos, siguiendo los pasos propuestos por la formación en PBE (planteamiento de preguntas clínicas estructuradas, búsqueda de la bibliografía, selección y obtención de la evidencia, evaluación crítica de revisiones sistemáticas, ensayos clínicos y estudios de diagnóstico, interpretación de la relevancia y la precisión de los resultados y aplicación de la evidencia a la práctica clínica). Todas las sesiones enfatizaban la aplicabilidad en la práctica clínica diaria y dejaban un tiempo de reflexión y debate sobre este tema. Los docentes eran epidemiólogos y /o especialistas en medicina familiar y comunitaria de las UDD de MFyC. La sesión sobre búsqueda bibliográfica se hizo con la participación de un documentalista.

### **Administración del test de Fresno**

En el grupo de residentes, el test de Fresno se administró antes y después de la intervención formativa. Para evitar un posible sesgo de memoria, que podría sobreestimar el efecto de la intervención, los escenarios y preguntas de cálculo eran diferentes en cada administración. El primer test se administró un mes antes de la intervención formativa, coincidiendo con una sesión sobre la investigación en la atención primaria, dirigida a residentes y profesionales con experiencia intermedia, en la que también se presentaba el presente proyecto. El segundo test se administró el último día de la intervención formativa. En el caso de los tutores, el test de Fresno se administró coincidiendo con una reunión anual de tutores de MFyC.

### **Tamaño de la muestra**

El tamaño de muestra se estimó en base a la capacidad de detectar una mínima diferencia importante entre el test pre y post de 10 puntos (desviación estándar de 20), con una potencia estadística del 90%, una hipótesis bilateral y un error alfa del 5%. Con estas asunciones y considerando unas posibles pérdidas del 20%, el tamaño de la muestra eran 85 participantes.

### **Análisis estadístico**

Se hizo un análisis descriptivo (las variables categóricas se describieron con frecuencias y porcentajes, y las continuas con la media y la desviación estándar). Mediante el análisis de la varianza para datos independientes se compararon las puntuaciones obtenidas por los tres grupos de participantes en base a su experiencia (tutores-expertos, médicos con experiencia intermedia y residentes-principiantes).

El cambio en el conocimiento y las habilidades entre después y antes del curso se calculó como la diferencia entre las puntuaciones, que se compararon con la *t* de Student para datos apareados. Así, una diferencia positiva significa mejora en conocimientos y habilidades.

La capacidad de detectar cambios significativos se analizó con el tamaño del efecto (TE) y la respuesta media estandarizada (RME), el cálculo de las cuales se describirá más adelante.

### **Propiedades psicométricas del test de Fresno**

#### **Consistencia interna**

Se midió con correlación de cada ítem con la puntuación total del test<sup>21</sup> en la que valores superiores a 0,3 muestran evidencia de consistencia interna y con el alfa de Cronbach<sup>23</sup> con un rango de valores entre 0 y 1, en el que valores superiores a 0,7 se consideran evidencia de consistencia interna, en el caso de comparaciones de grupo.

#### **Fiabilidad entre e intra evaluadores (Reliability)**

La fiabilidad entre observadores nos midió en una muestra de 40 participantes (20 respuestas pre-intervención y 20 post-intervención) mediante el coeficiente Kappa para las variables cualitativas y el coeficiente de correlación intraclass en el caso de las cuantitativas<sup>24</sup>. La fiabilidad intra observadores se midió con los mismos coeficientes, analizando 45 respuestas.

### Validez de constructo

El grado de experiencia de los participantes se consideró un indicador clave para evaluar la validez de constructo, bajo el supuesto de que una mayor experiencia debería traducirse en una mejor puntuación en el test de Fresno.

### Capacidad de detectar cambios significativos (*Responsiveness*)

La capacidad de detectar cambios significativos es un tipo de validez de constructo y se analizó con el tamaño del efecto (GE) y la respuesta media estandarizada (RME). La GE se obtuvo a partir de la diferencia de medias entre las puntuaciones post y preintervención, dividida por la desviación estándar de los valores preintervención). Valores positivos reflejan mejora (25) y se define como "pequeña" (GE <0,2), "pequeña a moderada" (GE entre 0,2 y 0,5), "moderado a grande" (GE entre 0,51 y 0,79), "grande" (GE > 0,79), tal como sugiere Cohen (26). La RME se calculó como la media del cambio de puntuaciones dividida por la desviación estándar de este cambio. Como en el caso de la GE, valores positivos indican mejora<sup>27</sup>.

### Dificultad de los ítems

Se evaluó la dificultad de los ítems, con la proporción de respuestas correctas de cada pregunta. Para analizar este aspecto se asignó un punto de corte "puntuación aprobado", para cada pregunta, basado en la propia definición de los autores de test de Fresno<sup>17</sup>, que situaban este punto de corte en el puntaje del medio de la respuesta potente<sup>16-23</sup>. A modo de ejemplo, en el caso de preguntas con puntuación de 0 a 24 el punto de corte era 19,5. Para que un test pueda ser utilizado por participantes con diversos niveles de

experiencia necesario que las preguntas tengan un rango amplio de dificultades, que permitan diferenciar entre ellos.

### Capacidad de discriminación de los ítems

La capacidad de discriminación de los ítems es la diferencia en el porcentaje de respuestas correctas entre dos grupos clasificados según nivel de experiencia. Se puede calcular ordenando los participantes en base a la puntuación total y seleccionando los del cuartil más alto y más bajo. También se calcula para cada ítem y se consideran aceptables valores alrededor del 20%.

El cuestionario se respondió en papel y los datos fueron introducidos electrónicamente al final del curso. El análisis estadístico se realizó con Stata software versión 9.0 (STATA Corp, College Station, TX) y el SPSS software versión 15.0 for Windows (SPSS, Chicago, IL, USA).

### Consideraciones éticas

El estudio se desarrolló siguiendo normas de buena práctica en investigación y los principios de la declaración de Helsinki, y fue aprobado por el Comité Ético de Investigación Clínica del IDIAP. Los participantes recibieron información oral y escrita sobre el estudio y aceptaron voluntariamente su participación. Para mantener la confidencialidad y el anonimato, los datos se guardaron y analizar con una codificación que garantizaba estos aspectos.

# RESULTADOS

## Adaptación transcultural del test de Fresno

La [figura 1](#) enumera los procedimientos realizados. La mayoría de las palabras se tradujeron al español sin dificultad. En base a las respuestas de las entrevistas cognitivas del equipo investigador elaboró la versión final del test, clara, sin ambigüedades y completa. Las pequeñas diferencias se resolvieron en las reuniones de consenso. La única excepción fue la palabra "evidencia", que tiene un significado diferente en español. Los traductores utilizaron "prueba" como una traducción más precisa de evidencia. Sin embargo, dado que la palabra "Evidencia" se ha utilizado ampliamente en español, el equipo investigador encontró más apropiado mantener esta palabra.

## Descripción de la muestra y la participación

Participaron 275 personas: 56 expertos, 17 experiencia intermedia y 202 residentes.

La media de edad de los participantes era 31 años (desviación estándar = 8), el 73% eran mujeres. Entre los residentes 190/202 (94,1%) completaron el test, y 158 (78,2%) devolvieron el post-test. En total, 152/202 residentes (75,2%) volvieron los dos cuestionarios. Los 44 residentes que no devolvieron el segundo cuestionario tenían menor puntuación que los que devolvieron los dos (58,8 vs 63,9;  $P = 0,03$ , respectivamente).

## Consistencia interna

El cuestionario mostró una consistencia interna elevada. La alfa de Cronbach fue de 0,88 y 0,77 antes y después de la intervención, respectivamente. El rango de valores de los coeficientes de correlación ítem-total fue entre 0,36 para la pregunta 1 (formulación de una pregunta) y 0,67 para la pregunta 3 (búsqueda de la evidencia) en el cuestionario pre-test. En el cuestionario post intervención los coeficientes oscilaron entre 0,25 (pregunta 11: el mejor diseño del estudio, diagnóstico) a 0,52 (pregunta 5: relevancia).

## Fiabilidad entre e intra evaluadores

La fiabilidad entre evaluadores fue de 0,95 pre y 0,85 post-intervención e intra-evaluador 0,71 y 0,81, respectivamente ([tablas 1 y 2](#)).

## Dificultad de los ítems y capacidad de discriminación

El rango de dificultad de los ítems osciló entre dificultad moderada y elevada ([tabla 3](#)). La capacidad de discriminación de los ítems se muestra en la [tabla 4](#).

## Validez de constructo

La puntuación media (desviación típica) de los residentes fue 60,4 (DT = 25), de los tutores / expertos 149,8 (DT = 23,2) y la grupo con experiencia intermedia 110,4 (DT = 11,5), con diferencias de medias y comparaciones entre los grupos estadísticamente significativas ( $P < 0,001$  y  $P < 0,01$  respectivamente). Los residentes con formación previa en PBE obtuvieron una puntuación superior y estadísticamente significativa (75,9 vs 57,6;

P <0,001) comparados con el resto de residentes.

Capacidad de detectar cambios significativos (Responsiveness)

La diferencia media en las puntuaciones de los residentes entre el test pre y post-

intervención fue de 47,7 (IC del 95%: 42,8-52,5) puntos. La GE y la RME fueron grandes y se presentan en la [tabla 4](#). La [tabla 5](#) muestra el porcentaje de residentes con puntuación de aprobado antes y después de la intervención formativa.

## DISCUSIÓN

### Resumen de los hallazgos

El presente estudio muestra que la versión española del test de Fresno es una herramienta fiable y válida para evaluar los conocimientos y habilidades en PBE. El cuestionario permite diferenciar los participantes según niveles de experiencia en PBE, la consistencia interna y la fiabilidad son elevadas y la capacidad de detectar cambios significativos (responsiveness) fue mayor (> 1.6) con los análisis realizados.

### Fiabilidad

#### Consistencia interna

En nuestro estudio el alfa de Cronbach fue igual a la del instrumento original (0,88 pre y 0,77 post) y superior al valor 0,7 propuesto por Streiner Y Norman (24) ampliamente aceptado. Por otra parte, las correlaciones de cada ítem con la puntuación total del test fueron superiores a 0,30, excepto para la pregunta 11 post. Estos resultados indican una buena consistencia interna, especialmente si se tiene en cuenta que el test incluye ítems fáciles y difíciles, lo que tiende a reducirla.

#### Fiabilidad entre e intra evaluadores

La fiabilidad entre evaluadores, esencial para un instrumento con preguntas abiertas, sujetas a la interpretación de los evaluadores, fue muy buena, en ambas medidas ([tabla 1](#)). La fiabilidad intra evaluadores fue superior a 0,7, pero inferior a la fiabilidad entre evaluadores, con valores particularmente bajos en las preguntas sobre relevancia y validez interna. Una posible explicación es que la fiabilidad intra evaluador fue evaluada al inicio del proceso de calificación, cuando el evaluador era menos experto y más inconsistente, mientras que la fiabilidad entre evaluadores se evaluó a la mitad del proceso de calificación después de haber calificado más de cien tests y por tanto, con una mayor experiencia. Estos resultados tienen algunas implicaciones: 1) los cambios en conocimientos y habilidades en PBE pueden ser hasta cierto punto erróneos en algunos casos, dada una menor concordancia intra evaluadores en alguna pregunta, lo que podría resultar en un sesgo en favor de la hipótesis nula, en un test para evaluar una intervención educativa, 2) es importante continuar precisando y perfeccionando el proceso de puntuación del test, 3) el entrenamiento previo de los evaluadores es imprescindible y finalmente, 4) la importancia de evaluar la fiabilidad entre e intra evaluadores. Esta

última no fue evaluada en la validación del instrumento original<sup>17</sup>.

### Validez

#### Validez de constructo

En el presente estudio las puntuaciones del test de Fresno han sido superiores en el grupo de participantes de expertos, seguido de los de experiencia intermedia e inferiores en los principiantes. Además, los residentes con instrucción previa en PBE también tuvieron una puntuación más alta que el resto de los residentes. Estos resultados confirman la validez discriminante del test de Fresno.

#### Capacidad de detectar cambios significativos o importantes (Responsiveness)

En el presente estudio la capacidad de detectar cambios significativos del test de Fresno se analizó mediante la distribución de sus puntuaciones. Se obtuvieron dos medidas: el tamaño del efecto (ES) para cuantificar la magnitud del cambio y la media de respuesta estandarizada (MRE) que muestra los cambios estadísticamente significativos<sup>28</sup>. Los valores de estas medidas fueron de 1,77 y 1,65 respectivamente, con resultados similares en el subgrupo de residentes con formación previa en PBE. Por lo tanto, la capacidad de detectar cambios significativos de la versión española del Fresno es muy alta y el test es sensible al cambio en el contexto de un estudio antes-después. Además, dado que se obtuvo una buena respuesta en los dos grupos de diferente experiencia en PBE, podría ser utilizado evaluar profesionales principiantes o con mayor experiencia.

#### Puntos fuertes y limitaciones del estudio

El presente estudio tiene varios puntos fuertes a destacar. En primer lugar, la validación de un cuestionario es un proceso longitudinal, continuo y dinámico que persiste durante su utilización repetida. La validación del test Fresno en español, en una población con diversos grados de experiencia en PBE y diferente a la estudiada por Ramos et al<sup>17</sup>, aporta nuevas evidencias de validez que permiten proponer la generalización de su utilización, la comparaciones entre los países y la evaluación de diferentes métodos de enseñanza. En segundo lugar, el estudio actual midió la capacidad de test para detectar cambios significativos, aspecto no contemplado en la validación original<sup>17</sup>. Esto será de gran utilidad para estimar el tamaño de la muestra necesaria en futuros ensayos clínicos que quieran medir la efectividad de una intervención educativa. En tercer lugar, este estudio se realizó en condiciones que reflejan la práctica habitual, con múltiples profesores a los que se aconsejó no modificar sus sesiones. La participación de varios profesores dificulta la normalización la intervención formativa y la falta de estandarización puede disminuir la capacidad de detectar diferencias, pero resulta coherente con el enfoque pragmático del estudio<sup>29,30</sup>.

Entre las limitaciones potenciales de este estudio hay que considerar las siguientes. Aunque el test fue validado en una muestra con un rango amplio de experiencia en PBE, hubo un predominio de los participantes principiantes y el porcentaje de especialistas en MFyC con experiencia intermedia fue bajo (5%). Por tanto, no podemos descartar que una muestra con una estructura diferente no hubiera dado otros índices psicométricos. Además, las diferencias obtenidas entre el

pre y el post-test podrían ser atribuibles sobre todo a la mejora de los conocimientos y habilidades de los residentes. Es posible que la restricción de la muestra, los sujetos con respuestas completas, también haya influido en las propiedades psicométricas del test. Los 44 residentes que no devolvieron el segundo cuestionario, podrían estar menos comprometidos, menos informados y menos seguros que los que lo devolvieron.

Otros autores han adaptado y validado el test de Fresno en otras poblaciones y han confirmado su fiabilidad y validez<sup>31,32</sup>. En nuestro caso, por razones logísticas, sólo se incluyeron residentes de MFyC. Si bien se podría sugerir que entre los residentes hay más similitudes que diferencias en cuanto a la PBE, la utilización del test de Fresno en residentes de otras especialidades requeriría pruebas de validez y fiabilidad específicas.

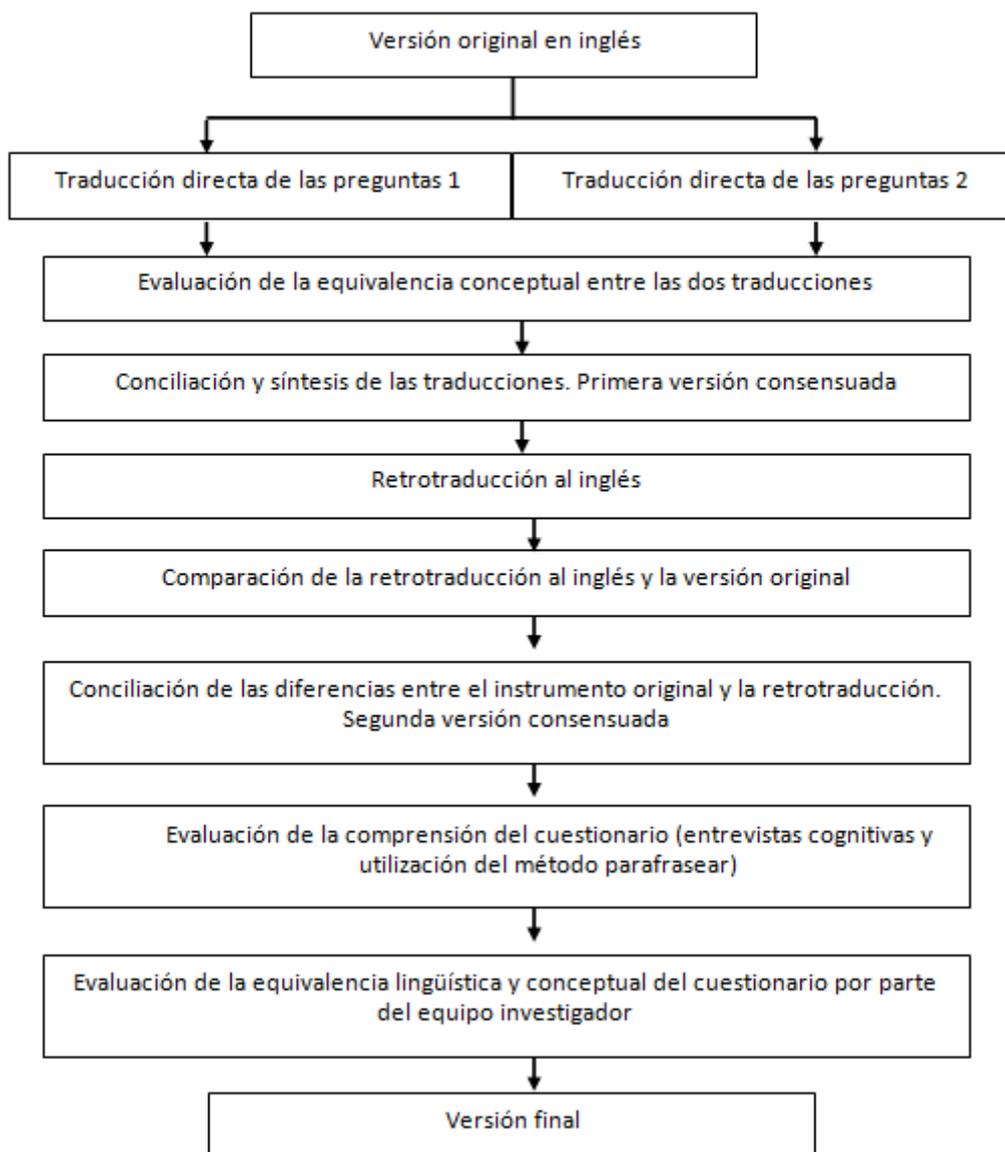
La fiabilidad entre e intra evaluadores, la consistencia interna y la capacidad de discriminación de un test están muy ligadas a la población a la que se aplica y no se puede suponer que estas propiedades se mantienen en poblaciones de características diferentes. La utilización de un instrumento en diferentes poblaciones contribuye a aumentar sus pruebas de validez. El test de Fresno debería aplicarse con nuevos escenarios clínicos, nuevos

ejemplos de preguntas numéricas, y buscando una respuesta y corrección más simple y ágil. La investigación futura podría contemplar el desarrollo y la validación de una versión más corta, utilizando la versión larga como estándar de referencia, o con respuestas de opción múltiple. Aunque la fiabilidad intra evaluadores de la versión española fue buena, habría que mejorar el sistema de puntuación para aumentar esta fiabilidad. También es importante confirmar si la fiabilidad mejora como efecto de la formación continua y la familiarización creciente de los evaluadores con los matices del proceso de calificación del test Fresno. Esto permitiría establecer una medida del tiempo de entrenamiento recomendable para realizar estos procedimientos.

### Conclusiones

Los hallazgos de este estudio indican que la versión española del test de Fresno es un instrumento válido y fiable para evaluar los conocimientos y habilidades en PBE los residentes de MFyC de habla hispana. También un buen instrumento de evaluación, ya que es capaz de detectar los cambios relevantes en los conocimientos y habilidades después de una intervención educativa. Por lo tanto, podría ser utilizado como variable principal de resultados en estudios aleatorizados y controlados.

Figura 1. Proceso de adaptación transcultural del test de Fresno al español



**Tabla 1. Análisis de la fiabilidad de los ítems del test de Fresno. Coeficiente de Correlación Intraclase**

Pregunta	Entre evaluadores		Intra evaluadores	
	Pre intervención	Post intervención	Pre intervención	Post intervención
P1 Plantear la pregunta	0,97	0,86	0,78	0,90
P2 Fuentes de información	0,95	0,87	0,65	0,74
P3 Búsqueda	0,92	0,92	0,65	0,83
P4 Diseño del estudio	0,87	0,73	0,71	0,84
P5 Relevancia	0,88	0,90	0,46	0,34
P6 Validez Interna	0,89	0,76	0,44	0,66
P7 Magnitud del efecto	0,99	0,79	0,72	0,85
P1 a P7	0,95	0,85	0,71	0,81

P: Preguntas.

Tabla 2. Análisis de la dificultad de los ítems del test de Fresno

	Pre-intervención	Post-intervención
P1 Plantear la pregunta	0.29	0.33
P2 Fuentes de información	0.20	0.21
P3 Búsqueda	0.15	0.35
P4 Diseño del estudio	0.38	0.51
P5 Relevancia	0.29	0.38
P6 Validez Interna	0.21	0.58
P7 Magnitud del efecto	0.10	0.20
P8 Sensibilidad	0.54	0.84
Especificidad	0.44	0.77
Valor Predictivo Positivo	0.44	0.79
Valor Predictivo Negativo	0.39	0.78
Cociente de Probabilidad Positivo	0.17	0.46
P9 Reducción Absoluta de Riesgo	0.40	0.84
Reducción Relativa de Riesgo	0.24	0.66
Número Necesario a Tratar	0.25	0.62
P10 Intervalo de Confianza	0.29	0.56
P11 Mejor diseño, estudio diagnóstico	0.21	0.78
P12 Mejor diseño, estudio pronóstico	0.24	0.68

P= Preguntas.

Los valores representan la proporción de participantes que alcanzan la puntuación "aprobado" para cada cada ítem.

**Tabla 3. Análisis de la capacidad de discriminación de los ítems del test de Fresno**

	Pre-intervención	Post-intervención
P1 Plantear la pregunta	0.32	0.29
P2 Fuentes de información	0.15	0.22
P3 Búsqueda	0.16	0.29
P4 Diseño del estudio	0.36	0.46
P5 Relevancia	0.24	0.33
P6 Validez Interna	0.24	0.59
P7 Magnitud del efecto	0.07	0.19
P8 Sensibilidad	0.47	0.82
Especificidad	0.44	0.83
Valor Predictivo Positivo	0.42	0.80
Valor Predictivo Negativo	0.36	0.82
Cociente de Probabilidad Positivo	0.14	0.80
P9 Reducción Absoluta de Riesgo	0.28	0.48
Reducción Relativa de Riesgo	0.21	0.83
Número Necesario a Tratar	0.31	0.67
P10 Intervalo de Confianza	0.24	0.62
P11 Mejor diseño, estudio diagnóstico	0.17	0.54
P12 Mejor diseño, estudio pronóstico	0.20	0.68

P= Preguntas

Los valores representan la diferencia en la proporción de los participantes que responden correctamente entre los del cuartil de puntuación más alta y la proporción de participantes que responden correctamente entre el cuartil de puntuación más baja.

**Tabla 4. Capacidad de detectar cambios significativos (*Responsiveness*) del test de Fresno según grupo**

	Total de residentes (IC 95%)	Residentes con formación en PBE (CI 95%)
Tamaño del efecto*	1.77 (1.57-1.95)	1.78 (1.38-2.17)
Media de respuesta estandarizada**	1.65 (1.47-1.82)	1.60 (1.24-1.96)

**Tabla 5. Porcentaje de residentes con puntuación de aprobado según el test de Fresno antes y después de la intervención formativa e intervalo de confianza de la diferencia.**

	Pre- intervención	Post- intervención	IC 95%
P1 Plantear la pregunta	22.9	28.3	-2.1-13.1
P2 Fuentes de información	12.3	22.3	3.3-16.1
P3 Búsqueda	04.8	22.6	12.0-23.6
P4 Diseño del estudio	22.9	37.7	6.7-22.5
P5 Relevancia	14.4	23.9	2.7-16.3
P6 Validez Interna	28.2	38.2	2.1-18.4
P7 Magnitud del efecto	1.6	14.5	8.4-17.7
P8 Sensibilidad	42.2	82.3	32.2-47.6
Especificidad	34.0	78.5	36.9-52.4
Valor Predictivo Positivo	32.6	81.5	41.5-56.5
Valor Predictivo Negativo	25.9	79.9	46.4-61.1
Cociente de Probabilidad Positivo	8.2	47.8	32.5-46.6
P9 Reducción Absoluta de Riesgo	21.3	81.0	52.7-66.7
Reducción Relativa de Riesgo	06.5	63.9	52.1-64.3
Número necesario a tratar	10.6	60.3	42.7-56.9
P10 Intervalo de Confianza	13.2	48.3	27.7-42.6
P11 Mejor diseño, estudio de diagnóstico	24.0	44.0	11.7-27.8
P12 Mejor diseño, estudio de pronóstico	24.9	34.9	12.0-28.3

Abreviaturas: P: Preguntas. IC=Intervalo de confianza.

### Puntos fuertes y limitaciones del estudio

El presente estudio tiene varios puntos fuertes a destacar. En primer lugar, la validación de un cuestionario es un proceso longitudinal, continuo y dinámico que persiste durante su utilización repetida. La validación del test Fresno en español, en una población con diversos grados de experiencia en PBE y diferente a la estudiada por Ramos et al<sup>17</sup>, aporta nuevas evidencias de validez que permiten proponer la generalización de su utilización, la comparaciones entre los países y la evaluación de diferentes métodos de enseñanza. En segundo lugar, el estudio actual midió la

capacidad de test para detectar cambios significativos, aspecto no contemplado en la validación original<sup>17</sup>. Esto será de gran utilidad para estimar el tamaño de la muestra necesaria en futuros ensayos clínicos que quieran medir la efectividad de una intervención educativa. En tercer lugar, este estudio se realizó en condiciones que reflejan la práctica habitual, con múltiples profesores a los que se aconsejó no modificar sus sesiones. La participación de varios profesores dificulta la normalización la intervención formativa y la falta de estandarización puede disminuir la capacidad de detectar diferencias, pero resulta coherente con el enfoque pragmático del estudio<sup>29,30</sup>.

Entre las limitaciones potenciales de este estudio hay que considerar las siguientes: Aunque el test fue validado en una muestra con un rango amplio de experiencia en PBE, hubo un predominio de los participantes principiantes y el porcentaje de especialistas en MFyC con experiencia intermedia fue bajo (5%). Por tanto, no podemos descartar que una muestra con una estructura diferente no hubiera dado otros índices psicométricos. Además, las diferencias obtenidas entre el pre y el post-test podrían ser atribuibles sobre todo a la mejora de los conocimientos y habilidades de los residentes. Es posible que la restricción de la muestra, los sujetos con respuestas completas, también haya influido en las propiedades psicométricas del test. Los 44 residentes que no devolvieron el segundo cuestionario, podrían estar menos comprometidos, menos informados y menos seguros que los que lo devolvieron.

Otros autores han adaptado y validado el test de Fresno en otras poblaciones y han confirmado su fiabilidad y validez<sup>31,32</sup>. En nuestro caso, por razones logísticas, sólo se incluyeron residentes de MFyC. Si bien se podría sugerir que entre los residentes hay más similitudes que diferencias en cuanto a la PBE, la utilización del test de Fresno en residentes de otras especialidades requeriría pruebas de validez y fiabilidad específicas.

La fiabilidad entre e intra evaluadores, la consistencia interna y la capacidad

de discriminación de un test están muy ligadas a la población a la que se aplica y no se puede suponer que estas propiedades se mantienen en poblaciones de características diferentes. La utilización de un instrumento en diferentes poblaciones contribuye a aumentar sus pruebas de validez. El test de Fresno debería aplicarse con nuevos escenarios clínicos, nuevos ejemplos de preguntas numéricas, y buscando una respuesta y corrección más simple y ágil. La investigación futura podría contemplar el desarrollo y la validación de una versión más corta, utilizando la versión larga como estándar de referencia, o con respuestas de opción múltiple. Aunque la fiabilidad intra evaluadores de la versión española fue buena, habría que mejorar el sistema de puntuación para aumentar esta fiabilidad. También es importante confirmar si la fiabilidad mejora como efecto de la formación continua y la familiarización creciente de los evaluadores con los matices del proceso de calificación del test Fresno. Esto permitiría establecer una medida del tiempo de entrenamiento recomendable para realizar estos procedimientos.

### Conclusiones

Los hallazgos de este estudio indican que la versión española del test de Fresno es un instrumento válido y fiable para evaluar los conocimientos y habilidades en PBE los residentes de MFyC de habla hispana. También un

buen instrumento de evaluación, ya que es capaz de detectar los cambios relevantes en los conocimientos y habilidades después de una

intervención educativa. Por lo tanto, podría ser utilizado como variable principal de resultados en estudios aleatorizados y controlados.

## BIBLIOGRAFÍA

1. Haynes RB, Devereaux PJ, Guyatt GH. Physicians' and patients' choices in evidence based practice. *BMJ* 2002 Jun 8;324(7350):1350.
2. Ubbink DT, Guyatt GH, Vermeulen H. Framework of policy recommendations for implementation of evidence-based practice: a systematic scoping review. *BMJ Open* 2013;3(1).
3. Dawes M, Summerskill W, Glasziou P, Cartabellotta A, Martin J, Hopayian K, et al. Sicily statement on evidence-based practice. *BMC Med Educ* 2005 Jan 5;5(1):1.
4. Guyatt G, Cook D, Haynes B. Evidence based medicine has come a long way. *BMJ* 2004 Oct 30;329(7473):990-1.
5. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academies Press, 2001. 2001.
6. Barzansky B, Etzel SI. Educational programs in US medical schools, 2004-2005. *JAMA* 2005 Sep 7;294(9):1068-74.
7. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003 Oct 11;362(9391):1225-30.
8. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003 Jun 26;348(26):2635-45.
9. Shrank WH, Asch SM, Adams J, Setodji C, Kerr EA, Keesey J, et al. The quality of pharmacologic care for adults in the United States. *Med Care* 2006 Oct;44(10):936-45.
10. Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. *BMJ* 2004 Oct 30;329(7473):1017.
11. Flores-Mateo G, Argimon JM. Evidence based practice in postgraduate healthcare education: a

systematic review. *BMC Health Serv Res* 2007;7:119.

12. Green ML. Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: a critical review of curricula. *Acad Med* 1999 Jun;74(6):686-94.

13. Norman GR, Shannon SI. Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. *CMAJ* 1998 Jan 27;158(2):177-81.

14. Parkes J, Hyde C, Deeks J, Milne R. Teaching critical appraisal skills in health care settings. *Cochrane Database Syst Rev* 2001;(3):CD001270.

15. Hatala R, Guyatt G. Evaluating the teaching of evidence-based medicine. *JAMA* 2002 Sep 4;288(9):1110-2.

16. Shaneyfelt T, Baum KD, Bell D, Feldstein D, Houston TK, Kaatz S, et al. Instruments for evaluating education in evidence-based practice: a systematic review. *JAMA* 2006 Sep 6;296(9):1116-27.

17. Ramos KD, Schafer S, Tracz SM. Validation of the Fresno test of competence in evidence based medicine. *BMJ* 2003 Feb 8;326(7384):319-21.

18. Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer HH, Kunz R. Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and

before and after study of courses in evidence based medicine. *BMJ* 2002 Dec 7;325(7376):1338-41.

19. Taylor R, Reeves B, Mears R, Keast J, Binns S, Ewings P, et al. Development and validation of a questionnaire to evaluate the effectiveness of evidence-based practice teaching. *Med Educ* 2001 Jun;35(6):544-7.

20. Taylor RS, Reeves BC, Ewings PE, Taylor RJ. Critical appraisal skills training for health care professionals: a randomized controlled trial. *BMC Med Educ* 2004 Dec 7;4(1):30.

21. Norman GR, Streiner DL. *Health Measurement Scales: A Practical Guide to their Development and Use*. 4th edition ed. Oxford University Press; 2008.

22. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2(14):i-74.

23. Bland JM, Altman DG. Cronbach's alpha. *BMJ* 1997 Feb 22;314(7080):572.

24. Streiner DL, NGR. *Health measurement scales: A practical guide to their development and use*. (3rd ed.) ed. 2003.

25. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. *Statistics and strategies for evaluation*. *Control Clin Trials* 1991 Aug;12(4 Suppl):142S-58S.

26. Cohen J. Statistical power analysis for the behavioral sciences. ed. 2nd ed. New Jersey: 1988.
27. Bland JM, Altman DG. Statistics Notes: Validating scales and indexes. *BMJ* 2002 Mar 9;324(7337):606-7.
28. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003 Jun;12(4):349-62.
29. Glasgow RE. What does it mean to be pragmatic? Pragmatic methods, measures, and models to facilitate research translation. *Health Educ Behav* 2013 Jun;40(3):257-65.
30. Foster N, Little P. Methodological issues in pragmatic trials of complex interventions in primary care. *Br J Gen Pract* 2012 Jan;62(594):10-1.
31. McCluskey A, Bishop B. The Adapted Fresno Test of competence in evidence-based practice. *J Contin Educ Health Prof* 2009;29(2):119-26.
32. Tilson JK. Validation of the modified Fresno test: assessing physical therapists' evidence based practice knowledge and skills. *BMC Med Educ* 2010;10:38.

# ANEXO 1. VERSIÓN ESPAÑOLA DEL TEST FRESNO SOBRE PRÁCTICA BASADA EN LA EVIDENCIA

Se ruega rellenar todo el cuestionario en una misma sesión. Hay 7 preguntas de respuesta corta, 2 preguntas que requieren una serie de cálculos sencillos, y tres preguntas más de respuesta abierta. Prevea alrededor de 30 minutos para realizar la prueba.

Responda las preguntas 1- 4 basándose en los siguientes escenarios clínicos:

- Acude a su consulta una madre con un lactante de 3 meses que presenta desde hace dos semanas al anochecer, un llanto intenso e inconsolable, gritando y encogiéndose las piernas como si le doliera el vientre. Tras una exploración física completamente normal, le diagnostica de un cólico del lactante. La madre le comenta que otras madres le han aconsejado que le administre un té con extractos de manzanilla, verbena, hinojo y menta balsámica para mejorar los síntomas, pero usted no conoce mucho sobre las propiedades analgésicas de dicho té en el cólico del lactante.
- Acude a su consulta un paciente con acúfenos crónicos. Después un estudio completo se descarta cualquier patología subyacente que explique los acúfenos. El paciente refiere que le dificultan la conciliación del sueño y le solicita una benzodiacepina, pero usted no conoce la evidencia que existe sobre el uso de las benzodiacepinas en el tratamiento de los acúfenos crónicos y su seguridad.

1. Escriba una pregunta clínica para cada escenario de manera que le ayude a realizar una búsqueda bibliográfica y a elegir el mejor artículo de entre los que encuentre.

2. ¿Qué fuentes de información podrían consultar los clínicos para encontrar una respuesta a estas preguntas? Cite todos los posibles tipos o categorías de fuentes de información que conozca. Describa las ventajas y los inconvenientes más importantes de cada una.

3. Elija uno de los escenarios clínicos ¿Qué tipo de estudio (diseño) cree que sería más adecuado para contestar esta pregunta? ¿Por qué?

4. Describa los campos y palabras o términos para realizar una búsqueda bibliográfica en PubMed para localizar estudios que le ayuden a contestar la pregunta clínica. Explique los argumentos en que basa esta estrategia. Describa cómo podría limitar su búsqueda si fuera necesario, y explique sus argumentos.

Las preguntas 5-7 consideran una revisión crítica de la literatura dividida entre la importancia, validez, y la magnitud del tamaño del efecto. Estas pueden ser unas subdivisiones arbitrarias del proceso de la revisión crítica.

5. Cuando encuentra un artículo original ¿qué características del estudio consideraría Vd. relevantes para determinar si le sería útil en su práctica clínica? Incluya ejemplos. (Esta pregunta no se refiere a los aspectos de validez del estudio, ni a la magnitud de los resultados)

6. Cuando encuentra un artículo original, ¿qué características del estudio consideraría usted para determinar la validez de los hallazgos? Haga comentarios si lo considera oportuno. (Ya se ha considerado la relevancia, y la pregunta 7 se referirá a cómo determinar la relevancia de los hallazgos, por tanto en esta pregunta céntrese en la validez del estudio)

7. Cuando encuentra un artículo original, ¿qué aspectos valoraría usted para determinar la magnitud de los resultados? Incluya ejemplos y comentarios si lo cree oportuno. (Ya se ha considerado la relevancia y la validez para esta pregunta, considere cómo determinar el tamaño y el significado del efecto comentado en el estudio.)

8. Un estudio reciente sobre la precisión de los criterios de Centor (adenopatías cervicales, exudado amigdalár, fiebre y ausencia de síntomas catarrales) para el diagnóstico de la faringoamigdalitis estreptocócica, incluyó a 140 pacientes mayores de 14 años con “dolor de garganta” como principal síntoma, 34 de los cuales presentaron frotis positivo (Gold Standard). Entre los diagnosticados 27 presentaban los 4 criterios de Centor positivos. Entre los individuos que no se diagnosticaron de faringoamigdalitis estreptocócica, 8 presentaron los cuatro criterios de Centor positivos. Basándose en este estudio:

Calcule o escriba la fórmula:

En base a estos resultados, la sensibilidad del Centor es \_\_\_\_\_

En base a estos resultados, la especificidad del Centor es \_\_\_\_\_

En base a estos resultados, el valor predictivo positivo del Centor es \_\_\_\_\_

En base a estos resultados, el valor predicativo negativo Centor es \_\_\_\_\_

En base a estos resultados, el cociente de probabilidad positivo \_\_\_\_\_

9. Un estudio de casos y controles reciente mostró que el 10,9% de los pacientes infectados por el HIV vacunados con la vacuna antipneumocócica 23-valente polisacárida, presentaron una neumonía neumocócica al año del seguimiento, mientras que en el grupo de no vacunados la sufrió un 23,8%. Basándose en estos resultados:

Calcule o escriba la fórmula:

La reducción absoluta de riesgo para sufrir una neumonía neumocócica es \_\_\_\_\_

La reducción relativa de riesgo de sufrir una neumonía neumocócica es \_\_\_\_\_

El número que se necesita para tratar (NNT) evitar una neumonía neumocócica es \_\_\_\_\_

10. Una revisión sistemática evaluó el efecto de los antiagregantes orales comparado con el del placebo en la prevención primaria de la enfermedad cardiovascular en pacientes con alto riesgo de enfermedad cardiovascular. Los resultados mostraron que el riesgo relativo de eventos de cardiovasculares era de 0,89 para los individuos que tomaban antiagregantes. Esto indica que los antiagregantes son protectores de sufrir un evento cardiovascular, pero nos preguntamos si esta diferencia es importante en términos estadísticos; para ello consultamos el intervalo de confianza. Indique un ejemplo de un intervalo de confianza que apoyaría la conclusión de que el riesgo relativo de eventos cardiovasculares estaba disminuido entre los individuos que tomaban antiagregantes.

0.89 [95% CI, \_\_\_\_\_]

11. ¿Qué diseño es el mejor para evaluar la eficacia de un nuevo tratamiento?

12. ¿Qué diseño es el mejor para evaluar un el pronóstico de una enfermedad?