



## ORIGINAL

# Validación psicométrica de la Escala de gravedad de fatiga en médicos residentes mexicanos<sup>☆</sup>



Rocío Elizabeth Duarte<sup>a,\*</sup>, Eduardo Velasco<sup>b</sup>, Juan José Sánchez-Sosa<sup>b</sup>  
y Lucina Isabel Reyes-Lagunes<sup>b</sup>

<sup>a</sup> Escuela de Ciencias de la Salud, Universidad del Valle de México, Campus Lomas Verdes, Ciudad de México, México

<sup>b</sup> Unidad de Investigaciones Psicosociales, División de Investigación y Posgrado, Facultad de Psicología, Universidad Nacional Autónoma de México, Ciudad de México, México

Recibido el 12 de septiembre de 2017; aceptado el 8 de noviembre de 2017

Disponible en Internet el 13 de abril de 2018

### PALABRAS CLAVE

Fatiga;  
Residencia médica;  
Medición;  
Validación  
psicométrica;  
Invarianza

**Resumen** Los médicos residentes están continuamente expuestos a la fatiga, que puede ocasionarles problemas laborales, académicos y personales, por consiguiente, es importante contar con instrumentos confiables y válidos para medirla. Una de las escalas más utilizadas es la Escala de gravedad de fatiga, sin embargo, hay versiones con una y distinto número de reactivos en distintas poblaciones. Por lo que el propósito de este trabajo fue traducir, adaptar de manera culturalmente relevante y comparar las versiones con distinto número de reactivos de la escala, para identificar la más adecuada en médicos residentes mexicanos. Se trabajó con tres grupos diferentes de participantes; el primero para la traducción y adaptación conformado por siete jueces expertos en el idioma, el segundo para la validación psicométrica con 100 médicos residentes y, el tercero, para el análisis factorial confirmatorio y prueba de invarianza por tipo de programa (especialidad y alta especialidad) con 182 médicos residentes. Se probó la distribución, discriminación, confiabilidad y estructura de la escala, en un análisis exploratorio y confirmatorio, quedando una escala final de 6 reactivos con buen ajuste ( $\chi^2[9] = 9,39$ ,  $p = 0,40$ ; CFI = 0,99; RMSEA = 0,02), con una confiabilidad = 0,91 e invarianza factorial estricta por tipo de programa de residencia.

© 2018 Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

<sup>☆</sup> Una versión preliminar de la escala se presentó en el XVI Congreso Mexicano de Psicología Social pero fue con una muestra de población general y no hay conflicto de intereses.

\* Autor para correspondencia.

Correo electrónico: [drarocioduarte@hotmail.com](mailto:drarocioduarte@hotmail.com) (R.E. Duarte).

**KEYWORDS**

Fatigue;  
 Medical residency;  
 Assessment;  
 Psychometric  
 validation;  
 Measurement  
 invariance

**Psychometric validation of the Fatigue Severity Scale in Mexican medical residents**

**Abstract** Resident physicians are continually exposed to fatigue, which can lead to work, academic and personal problems, so it is important to have reliable and valid instruments to measure it. One of the most used is Fatigue Severity Scale, but, there are versions with different number of items in different populations. Therefore, the purpose of this work was to translate, adapt culturally relevant and compare versions with different numbers of items, to identify the most appropriate in Mexican resident physicians. With three different groups of participants; first one for translation by seven expert judges in the language, second for psychometric validation with 100 resident physicians, and third for confirmatory factor analysis and invariance test by type of program (specialty and high specialty) with 182 resident physicians. Distribution, discrimination, reliability and factor structure of the scale were tested in an exploratory and confirmatory analysis, leaving a final scale of 6 items with good fit ( $\chi^2[9] = 9,39$ ,  $p = 0,40$ ; CFI = 0,99; RMSEA = 0,02), with a reliability = 0,91 and strict factorial invariance by type of residence program.

© 2018 Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Introducción**

La residencia médica ha sido identificada como un periodo en el que se viven altas demandas de tiempo, cargas de trabajo y responsabilidades en la formación de médicos y cirujanos<sup>1</sup>, por lo que se han investigado las consecuencias de estas demandas en la salud física, salud psicológica, bienestar general, rendimiento académico y laboral y relaciones interpersonales de sus estudiantes<sup>1-3</sup>.

Una de las consecuencias que han sido identificadas previamente en la literatura es la experiencia de fatiga<sup>4</sup>, que es la sensación de cansancio físico y falta de energía, diferente de la debilidad física, que ocurre como consecuencia de la realización de esfuerzos físicos y mentales<sup>5,6</sup> y su incidencia en esta población es alta y supera la de la población general<sup>7,8</sup>.

La fatiga que sufren los médicos residentes puede generarles problemas de tres tipos: los laborales por errores médicos<sup>2</sup>; personales como abuso de sustancias, daño en las relaciones sociales, trastornos de la salud mental y accidentes automovilísticos<sup>9-12</sup> y, problemas en el rendimiento académico con deterioro de memoria, habilidades de resolución de problemas y motricidad fina, así como la reducción de la calidad de la enseñanza a los alumnos de menor grado<sup>13,14</sup>.

Por lo que resulta necesario hacer investigaciones a profundidad sobre la fatiga y sus consecuencias en los médicos residentes, con esta finalidad es necesario contar con formas confiables y válidas para evaluar la fatiga. Tradicionalmente, la fatiga es evaluada de dos maneras distintas, la primera con métodos objetivos en los que se utilizan aparatos para realizar distintas pruebas al organismo<sup>15</sup>, la segunda con métodos subjetivos para realizar la medición de fatiga, en la que se considera a los instrumentos psicométricos.

Dado que los métodos objetivos son más complejos y costosos para su aplicación, es más común utilizar los métodos subjetivos<sup>5,16</sup>, sin embargo, su desarrollo es difícil debido a la variabilidad en la experiencia de las molestias

reportadas<sup>5,16,17</sup>. Existe un amplio número de instrumentos psicométricos diseñados para medir la fatiga, pero, solo se han estudiado las propiedades psicométricas de algunos de ellos<sup>18</sup>.

Los instrumentos diseñados para medir fatiga pueden clasificarse de acuerdo con sus dimensiones en dos tipos, los unidimensionales y los multidimensionales. Las escalas unidimensionales se centran en la gravedad de la fatiga. Las escalas multidimensionales de fatiga buscan explorar de manera amplia las afecciones causadas por la fatiga en el funcionamiento cognitivo, conductual y social de la persona, lo que se conoce como tipos de fatiga<sup>18</sup>.

Algunos de los instrumentos que existen para medir fatiga para los que se han estudiado sus características psicométricas se presentan a continuación en la [tabla 1](#).

Además de los instrumentos presentados en la tabla una de las escalas más utilizadas es la Escala de severidad de la fatiga (FSS)<sup>25</sup> la cual mide gravedad de fatiga de manera unidimensional con nueve reactivos tipo Likert con siete opciones de respuesta en términos de acuerdo y desacuerdo.

Esta escala se ha utilizado y validado psicométricamente en pacientes con distintos tipos de enfermedades crónicas entre las que se encuentran esclerosis múltiple, eventos vasculares cerebrales, lupus, hepatitis C y VIH, además de población general<sup>26</sup> también se ha adaptado a diversos idiomas en varios países del mundo entre los que se encuentran Noruega<sup>27</sup>, Turquía<sup>28</sup>, Italia<sup>29</sup> y Colombia<sup>30</sup> y en todos los casos demuestra tener adecuadas propiedades psicométricas en términos de discriminación, confiabilidad y validez<sup>18</sup>.

La principal ventaja de esta escala es su extensión breve, que en el contexto de los médicos residentes es especialmente valiosa, porque no constituye una interrupción larga a su trabajo; la versión original cuenta con nueve reactivos en los que se evalúa si la fatiga afecta la motivación, el ejercicio y funcionamiento físico, las actividades de la vida cotidiana y la vida social; hay versiones con siete reactivos desarrolladas en pacientes con VIH y esclerosis múltiple angloparlantes en las que no incluyen los problemas en la

**Tabla 1** Instrumentos psicométricos de fatiga con dimensiones, número de reactivos y consistencia interna

Nombre	Medición: dimensiones, reactivos y confiabilidad
Escala de fatiga analógica-visual VAS-F <sup>19</sup>	Dos dimensiones energía y fatiga con 18 reactivos en escala analógica visual y consistencia interna de entre 0,91-0,96
Escala de fatiga FS <sup>20</sup>	Dos dimensiones fatiga física y fatiga mental con 11 reactivos tipo Likert con cuatro opciones de respuesta y una consistencia interna entre 0,88 y 0,90
Escala de impacto de la fatiga FIS <sup>21</sup>	Tres dimensiones fatiga física, fatiga mental y fatiga psicosocial con 40 reactivos tipo Likert con cinco opciones de respuesta de 0 no hay problema a 4 problema extremo, con consistencia interna entre 0,88 y 0,98 y evidencias de validez convergente con diferentes subescalas de un inventario de salud física SF-36
Inventario de síntomas de fatiga FSI <sup>22</sup>	Tres dimensiones intensidad, duración e impacto en la calidad de vida, 13 reactivos tipo Likert de 11 puntos, con consistencia interna entre 0,93 y 0,95, convergente con la escala de síntomas POMS
Inventario de fatiga multidimensional MFI-20 <sup>23</sup>	Cinco dimensiones fatiga general, fatiga física, la actividad reducida, reducción de la motivación y la fatiga mental, con 20 reactivos tipo Likert con cinco opciones de respuesta, consistencia interna 0,53-0,93, evidencia de validez convergente con la escala VA de fatiga
Escala de fatiga en cáncer de Schwartz SCFS <sup>24</sup>	Cuatro dimensiones física, emocional, cognitiva, y temporal, con 28 reactivos tipo Likert con cinco opciones de respuesta, con consistencia interna de 0,97

Fuente: elaboración propia.

motivación ni el ejercicio físico<sup>26,31</sup>, también existe una versión con ocho reactivos realizada en pacientes de esclerosis múltiple italoportales que no incluye los problemas a la motivación<sup>29</sup>.

Por lo que el primer objetivo del presente estudio es traducir y adaptar de manera culturalmente relevante la FSS y validarla psicométricamente en médicos residentes mexicanos, el segundo objetivo es comprobar cuál es la estructura factorial más adecuada para los médicos residentes con análisis factorial confirmatorio y el tercero es realizar una prueba de invarianza por niveles de residencia para garantizar que la prueba no tiene funcionamiento diferencial por grado y si es así comparar las medias de la variable latente fatiga, debido a que en la literatura se reporta que los estudiantes en los primeros niveles de residencia sufren más fatiga que en otros<sup>32</sup>.

## Material y método

Para cumplir con los objetivos se realizaron tareas en tres fases consecutivas: traducción y adaptación culturalmente relevante, validación psicométrica y análisis factorial confirmatorio con prueba de invarianza por tipo de programa de residencia, las cuales se reportan a continuación.

Se trabajó con tres grupos diferentes de participantes, uno para cada fase del estudio. El primero para la traducción y adaptación culturalmente relevante estuvo conformado por siete jueces, con edades entre los 28 y 60 años de edad, todos psicólogos con experiencia en medición y manejo de idiomas inglés y español.

El segundo grupo para la validación psicométrica conformado por 100 médicos residentes, todos de hospitales públicos, 74 mujeres y 26 hombres, con edades entre los 25 y 37 años y una edad promedio de 29,48 años y una desviación estándar de 2,41, 76 de programas de especialidad médica (primeros tres niveles) y 24 de programas de alta

o subespecialidad (niveles entre 4 y 6). Para determinar el tamaño de la muestra se decidió exceder en 10 veces con participantes el número de reactivos de acuerdo con la propuesta de Nunnally y Bernstein<sup>33</sup>.

Los médicos que participaron en esta fase del estudio provenían de hospitales en tres regiones del país en las que se seleccionó aleatoriamente un estado, para el norte se tuvo el estado de Sinaloa, para el centro la Ciudad de México y para el sur Michoacán en proporciones iguales.

Se trabajó con médicos de cuatro programas de especialidad para tener representación de cada uno de los bloques en los que se clasifican las especialidades en México, del bloque quirúrgico se trabajó con 19 estudiantes de Ginecología y Obstetricia, del bloque clínico 19 estudiantes de Medicina de Urgencias, del bloque Salud Pública y Atención Primaria 20 estudiantes de Medicina Familiar y del bloque Especialidades Auxiliares y de Diagnóstico 18 estudiantes de Medicina Física y Rehabilitación.

En cuanto a los médicos de programas de subespecialidad se trabajó con 10 estudiantes de Cardiología, 5 estudiantes de Oncología en la Ginecología, 4 estudiantes de Neonatología y 5 estudiantes de Biología de la Reproducción Humana.

Los programas en los que se realizó el estudio se seleccionaron de forma no probabilística intencional de acuerdo con el permiso otorgado por las autoridades de los hospitales en los que se realizó el estudio y la participación de los estudiantes fue voluntaria, como criterio de exclusión se consideró que los estudiantes tuvieran algún diagnóstico de enfermedades agudas o crónicas al momento de la aplicación del instrumento.

El tercero para el análisis factorial confirmatorio y la prueba de invarianza conformado por 182 médicos residentes, 100 mujeres y 82 hombres, con edades entre los 25 y 42 años y una edad promedio de 29,46 años y una desviación estándar de 2,42, 101 de programas de especialidad médica (primeros tres niveles) y 81 de programas de alta o subespecialidad (niveles entre 4 y 6). El tamaño de la muestra para esta fase del estudio se consideró con base en

un análisis de potencia estadística calculado para GFI<sup>34</sup> en el que se consideró tener nueve reactivos, 45 grados de libertad, un  $\alpha$  de 0,05 y una potencia deseada de 0,80 que daba como resultado 157 participantes, también se cuidó exceder en 20 veces con participantes el número de reactivos de acuerdo con la propuesta Jackson<sup>35</sup>.

Los participantes también provenían de las mismas tres regiones del país en las que se seleccionó aleatoriamente un estado, para el norte se tuvo el estado de Sinaloa con 40 médicos, para el centro la Ciudad de México con 100 médicos y para el sur Michoacán con 42 médicos.

En esta fase se contó también con médicos de cuatro programas de especialidad para volver a tener representación de cada uno de los bloques, del bloque quirúrgico se trabajó con 24 estudiantes de Cirugía General, del bloque clínico 25 estudiantes de Medicina Interna, del bloque Salud Pública y Atención Primaria 36 estudiantes de Medicina Familiar y del bloque Especialidades Auxiliares y de Diagnóstico 16 estudiantes de Medicina Física y Rehabilitación.

En cuanto a los médicos de programas de subespecialidad se trabajó con estudiantes de: Algología 7, Cardiología 9, Cardionemología 3, Cirugía Bariátrica 4, Cirugía laparoscópica 4, Diabetes y Metabolismo 6, Gerontología Médica 4, Insuficiencia Cardíaca 6, Medicina del Dolor 5, Medicina Genómica 4, Medicina Materno-Fetal 7, Neonatología 4, Neurofisiología 3, Neuropsiquiatría 4, Oncología Pediátrica 4, Psiquiatría 2 y Trastornos del Sueño 5.

De nuevo en esta fase los programas en los que se realizó el estudio se seleccionaron de forma no probabilística intencional, de acuerdo con el permiso otorgado por las autoridades de los hospitales en los que se realizó el estudio y la participación de los estudiantes fue voluntaria, como criterio de exclusión se consideró que los estudiantes tuvieran algún diagnóstico de enfermedades agudas o crónicas al momento de la aplicación del instrumento.

Para la primera fase se usó la versión traducida de la escala y se colocó en un formato para la evaluación de expertos que incluía la opción de aceptación o modificaciones a la misma.

Para la segunda fase se usó la versión traducida y adaptada producto de la fase anterior, la cual tiene nueve afirmaciones con siete opciones de respuesta de tipo Likert pictórico<sup>36</sup> rotuladas con de acuerdo y en desacuerdo en los extremos.

Para la tercera fase se usó la versión validada producto de la fase anterior, la cual tiene nueve afirmaciones con siete opciones de respuesta de tipo Likert pictórico<sup>36</sup> rotuladas con de acuerdo y en desacuerdo en los extremos y una consistencia interna con el coeficiente Alpha de Cronbach 0,94.

También se aplicó la Escala de impacto de fatiga validada en población mexicana<sup>37</sup>. Con una consistencia interna con el coeficiente alpha de Cronbach 0,93.

En la primera fase se revisó la traducción y adaptación del instrumento, con esta finalidad dos traductores independientes tradujeron los reactivos de la escala, posteriormente se integró la primera versión que se revisó con la opinión experta de siete jueces entrenados en psicometría e idioma inglés, que compararon esta versión con la original y con la definición de fatiga de la misma y determinaron si lo que se había traducido literalmente del inglés al español en cada reactivo era lo que se deseaba medir. Se buscó que para

cada reactivo hubiera al menos un 80% de acuerdo acerca de la redacción junto con una confiabilidad interjueces alpha de Krippendorff de 0,88<sup>38</sup>. Posteriormente se retradujo al idioma inglés y se evaluó su equivalencia con la original.

En la segunda fase de validación psicométrica se contactó a los participantes en los hospitales con previa autorización de los responsables de su programa de especialidad, se les dio una explicación detallada del estudio, tiempos y procedimiento, posteriormente si accedían a participar se les pidió que firmaran el formulario de consentimiento informado y se les dio el instrumento para que lo contestaran. Una vez obtenidos los datos se siguió el procedimiento para validación psicométrica culturalmente relevante propuesto por Reyes-Lagunes y García y Barragán<sup>39</sup>, que incluye análisis de distribución, discriminación, direccionalidad, confiabilidad y estructura factorial de los reactivos de la escala.

Para la tercera fase del análisis factorial confirmatorio y prueba de invarianza, de nuevo, se contactó a los participantes en los hospitales con previa autorización de los responsables de su programa de especialidad, se les dio una explicación detallada del estudio, se les pidió que firmaran el formulario de consentimiento informado y se les dio el instrumento para que lo contestaran, es importante mencionar que en esta fase la batería incluía los dos instrumentos usados en el estudio para que pudieran responderse de forma simultánea. Una vez obtenidos los datos se analizó cuál era la estructura factorial más adecuada y se hizo una prueba de invarianza por tipo de programa de residencia siguiendo el procedimiento propuesto por Milsap y Olivera-Aguilar<sup>40</sup>, que consiste en evaluar cuatro modelos anidados cada uno más restrictivo que el anterior. En un primer paso se prueba el ajuste de la configuración cuidando que la estructura factorial sea la misma para los grupos, posteriormente se tiene la invarianza métrica o débil en la que las cargas factoriales se restringen a igualdad, se sigue con invarianza fuerte en el que las cargas factoriales y los interceptos se restringen a igualdad y finalmente en invarianza estricta las cargas factoriales, los interceptos y las varianzas únicas de los reactivos se restringen a igualdad y se compara el ajuste de los modelos. Si no hay diferencias significativas en la chi cuadrada ni cambios mayores a 0,01 en el CFI, se considera que existe invarianza<sup>41</sup>. Se busca tener invarianza estricta debido a que de esta manera se puede considerar que los cambios solo se deben a la variable latente y no a sesgo en la medición<sup>42</sup>.

## Resultados

### Adaptación culturalmente relevante

Los reactivos con un 80% de acuerdo acerca de la redacción y una confiabilidad interjueces alpha de Krippendorff de 0,88 se presentan en la [tabla 2](#).

### Validación psicométrica

Se hizo un análisis de frecuencias de cada uno de los reactivos, se encontró que todas las opciones de respuesta tuvieron frecuencia, posteriormente se calcularon el sesgo, la curtosis y la prueba Z de Kolmogorov-Smirnov y los resultados muestran que todos tenían una distribución sesgada. Lo que se consideró para los siguientes análisis ([tabla 3](#)).

**Tabla 2** Reactivos aprobados por los jueces*Reactivo*

1. Cuando estoy fatigado disminuye mi motivación
2. El ejercicio aumenta mi fatiga
3. Me fatigo fácilmente
4. Mi funcionamiento físico se entorpece con la fatiga
5. La fatiga me causa problemas frecuentemente
6. La fatiga no me permite tener un funcionamiento físico consistente
7. La fatiga me impide llevar a cabo ciertas tareas y responsabilidades
8. La fatiga es uno de los tres síntomas que menos me permiten hacer cosas
9. La fatiga me estorba en mi trabajo, con mi familia o con mi vida social

Posteriormente se probó el poder discriminativo de los reactivos, es decir, su capacidad para detectar diferencias en las cantidades del atributo que se está midiendo, lo que de acuerdo con Magnusson<sup>43</sup> puede hacerse al probar

que existe una relación sistemática entre las puntuaciones del reactivo y el total de la prueba. Para esto se hicieron dos análisis, en primer lugar, una comparación de grupos extremos con respecto al total de la prueba (puntuaciones inferiores al primer cuartil y puntuaciones superiores al tercer cuartil) con la prueba U de Mann Whitney, la misma que se seleccionó debido a que es adecuada a la distribución de los reactivos por ser no paramétrica, y se conservaron todos los reactivos por discriminar significativamente.

El segundo análisis fue una correlación entre el total de la prueba y cada uno de los reactivos, también se hizo una versión corregida de la misma en la cual el total de la prueba no incluía al reactivo, solo se conservaron aquellos con correlaciones superiores a 0,30 en ambos análisis, que en este caso fueron todos (tabla 4).

Posteriormente para conocer la estructura de la escala se hizo un análisis factorial exploratorio de ejes principales con rotación oblicua, adecuado para identificar las variables latentes que subyacen a los reactivos<sup>44,45</sup>, se eligió este método de extracción debido a que no se basa en el supuesto de normalidad, y el tipo de rotación debido a que este es más preciso y provee más información que la

**Tabla 3** Análisis de distribución de los reactivos de la Escala de severidad de fatiga FSS adaptada

Reactivo	Análisis de distribución frecuencia de los reactivos								Prueba de normalidad		
	Frecuencia en porcentaje por opción de respuesta								Z	p con corrección de Lilliefors	
	En desacuerdo		De acuerdo		Asimetría	Curtosis					
FSS1	2	2	4	5			74	7	6	-1,28	4,80
FSS2	5	3	3	78	4	3	4	-0,32	3,92	0,40	p < 0,001
FSS3	5	1	7	73	8	4	3	-0,14	4,08	0,37	p < 0,001
FSS4	4	3	6	75	6	3	3	-0,19	4,14	0,38	p < 0,001
FSS5	5	3	73	7	6	2	4	1,28	3,34	0,41	p < 0,001
FSS6	4	4	5	75	7	1	4	-0,74	3,96	0,40	p < 0,001
FSS7	4	3	63	19	6	2	3	1,07	3,06	0,34	p < 0,001
FSS8	3	4	58	23	8	1	3	0,81	2,24	0,31	p < 0,001
FSS9	3	3	7	74	6	4	3	0,20	4,00	0,38	p < 0,001

**Tabla 4** Análisis de discriminación de los reactivos de FSS versión adaptada

Reactivo	Análisis de discriminación					
	Comparación de grupos extremos			Correlación reactivo total	Correlación reactivo total sin el reactivo	
	U Mann Whitney	gl	p			
FSS1	35,00	43	p < 0,001	0,76**	0,69**	
FSS2	80,00	43	p < 0,001	0,70**	0,61**	
FSS3	12,00	43	p < 0,001	0,94**	0,92**	
FSS4	2,50	43	p < 0,001	0,93**	0,91**	
FSS5	72,00	43	p < 0,001	0,85**	0,80**	
FSS6	22,50	43	p < 0,001	0,84**	0,79**	
FSS7	28,50	43	p < 0,001	0,83**	0,77**	
FSS8	51,00	43	p < 0,001	0,79**	0,73**	
FSS9	12,00	43	p < 0,001	0,79**	0,73**	

\*\* p ≤ 0,01.



**Tabla 5** Resultados del análisis factorial exploratorio de ejes principales con rotación oblicua del FSS adaptado

Reactivo	Factor 1, fatiga
FSS3	0,95
FSS4	0,94
FSS5	0,84
FSS6	0,83
FSS7	0,79
FSS8	0,77
FSS9	0,74
FSS1	0,71
FSS2	0,64
% Varianza explicada	0,65
Autovalor o valor Eigen	5,88

rotación ortogonal. Se encontró una solución unidimensional con los nueve reactivos que explica el 65% de la varianza (tabla 5).

Para verificar la pertinencia del análisis se calculó la medida de adecuación muestral Kaiser, Meyer, Olkin (KMO= 0,81) y la prueba de esfericidad de Bartlett ( $X^2_{(36)} = 948,78$ ,  $p < 0,001$ ) y, se llegó a la conclusión de que la matriz de correlaciones de la que parte fue adecuada. Posteriormente, se calculó la confiabilidad total de la escala alpha de Cronbach y se obtuvo un valor adecuado de consistencia interna ( $\alpha = 0,94$ ).

### Análisis factorial confirmatorio y prueba de invarianza

Posteriormente se hizo un análisis factorial confirmatorio de máxima verosimilitud robusta, método seleccionado por la falta de normalidad en los datos<sup>12</sup>, probando tres modelos comparativos, todos con una dimensión, pero con diferente número de reactivos, uno con los 9, otro con 8 reactivos y uno con 7. La identificación de los modelos se hizo fijando la carga factorial del primer reactivo a uno para definir la métrica de la variable latente<sup>46</sup> con el software R y el paquete lavaan<sup>47</sup>.

Para evaluar el ajuste de los modelos se utilizaron los siguientes índices de bondad de ajuste: chi cuadrada normada, dividida por los grados de libertad (debería ser menor a 3,00 con una probabilidad no significativa; entre menor cociente mejor ajuste). Índice de bondad de ajuste Lisrel GFI, índice de ajuste comparativo de Bentler CFI (para que exista un buen ajuste, los valores de GFI y CFI deben ser  $\geq 0,95$ ). Y raíz cuadrada media del error de aproximación RMSEA y SRMR (los valores de RMSEA y SRMR deben ser  $\leq 0,08$ , para indicar un buen ajuste).

Los resultados de los tres modelos mostraron falta de ajuste global para explicar los datos, por lo que se analizaron las faltas de ajuste local mediante el uso de índices de modificación y se encontraron asociados con los reactivos 1,3 y 4, entonces se tomó la decisión de modificar el modelo eliminando estos reactivos, ya que alcanzar ajuste mediante la inclusión de los mismos puede reflejar más la idiosincrasia de la muestra que la relación entre variables<sup>48</sup>. El modelo modificado con una dimensión y seis reactivos cuenta con adecuado ajuste, en todos los índices,

a continuación, se presentan los índices de ajuste para los cuatro modelos evaluados (tabla 6).

A continuación, se presentan las cargas factoriales estandarizadas de la escala en las cuatro soluciones en la tabla 7.

Para determinar si hay funcionamiento diferencial de los reactivos por tipo de programa de residencia se hizo una prueba de invarianza factorial de medición y se alcanza invarianza factorial estricta ya que no hay diferencias significativas en el ajuste de los modelos anidados (tabla 8).

Después de probar la invarianza factorial se compararon las medias de la variables latentes y se encontró que no hay diferencias estadísticamente significativas ( $Z = -0,32$ ,  $p = 0,09$ ).

Como evidencia de validez concurrente se calculó la correlación con la puntuación total de la Escala de impacto de fatiga y se encontró significativa media ( $r = 0,50$ ,  $p < 0,001$ ).

### Discusión

Los primeros tres objetivos del presente trabajo eran traducir, adaptar y validar psicométricamente la FSS en una población de médicos residentes mexicanos. Para ello, se buscó tener una traducción adecuada al lenguaje sin sacrificar la evaluación de los constructos que la escala original propone. La versión final cuenta con 6 reactivos, una confiabilidad alpha de Cronbach = 0,91 y todos los reactivos se encuentran relacionados significativamente. Lo que muestra que se trata de la medición de un mismo constructo.

Mediante la validación psicométrica, a través del análisis factorial exploratorio de ejes principales, se encontró que la estructura que emerge libremente de la población tiene una dimensión con nueve reactivos como en la estructura de la escala original (Krupp y LaRocca<sup>25</sup>, 1989) con una adecuada consistencia interna. Sin embargo, en el análisis factorial confirmatorio se encontró que los reactivos 1, 3 y 4 no fueron adecuados.

La falta de ajuste con los reactivos 3 y 4 que indican que problemas en el funcionamiento físico pueden explicarse por dos posibles razones, la primera es que son reactivos que surgieron en poblaciones afectadas por enfermedades crónicas, lo que no corresponde con el estado de salud de los médicos residentes, la segunda es que contestar afirmativamente a estos reactivos puede implicar una disminución de la capacidad de trabajo (por ejemplo, la destreza motriz de un cirujano) y en México eso no es aceptado porque el trabajo se ve como algo positivo al ser un deber que permite contribuir con la resolución de necesidades de la familia y los amigos<sup>49</sup>.

De manera congruente se puede explicar que el reactivo 2 que habla de ejercicio físico que no había funcionado en las versiones validadas en poblaciones afectadas por enfermedades crónicas de Johansson et al.<sup>31</sup> (2014) y Lerdal et al.<sup>26</sup> funciona en los médicos que cuentan con un estado de salud y capacidad de actividad física distinta.

La falta de ajuste en el reactivo 1 que habla sobre problemas en la motivación corresponde con lo encontrado por Johansson et al.<sup>31</sup> y Lerdal et al.<sup>26</sup>, lo que puede deberse a que la motivación puede ser un constructo mucho más amplio que el hecho de tener energía para realizar una

**Tabla 6** Índices de ajuste para el análisis factorial confirmatorio de los modelos

Modelo	$\chi^2$	p	$\chi^2/\text{gl}$	CFI	GFI	RMSEA	IC 90%	SRMR
9 reactivos	$\chi^2(27) = 100,05$	$p < 0,001^{**}$	3,70	0,90	0,87	0,14	(0,11-0,18)	0,06
8 reactivos	$\chi^2(20) = 68,82$	$p < 0,001^{**}$	3,44	0,92	0,89	0,11	(0,09-0,14)	0,05
7 reactivos	$\chi^2(14) = 44,72$	$p < 0,001^{**}$	3,19	0,94	0,92	0,14	(0,09-0,18)	0,03
6 reactivos	$\chi^2(9) = 9,39$	$p = 0,40$	1,04	0,99	0,98	0,02	(0,00-0,10)	0,01

\*\*  $p < 0,01$ .

**Tabla 7** Cargas factoriales estandarizadas para los cuatro modelos

Reactivo	Modelo 9 reactivos		Modelo 8 reactivos		Modelo 7 reactivos		Modelo 6 reactivos	
	Factor único	Error estándar	Factor único	Error estándar	Factor único	Error estándar	Factor único	Error estándar
FSS1	0,60	0,05	-	-	-	-	-	-
FSS2	0,59	0,06	0,58	0,07	-	-	0,53	0,07
FSS3	0,79	0,04	0,78	0,04	0,77	0,04	-	-
FSS4	0,76	0,05	0,75	0,06	0,74	0,06	-	-
FSS5	0,84	0,02	0,85	0,03	0,86	0,03	0,85	0,03
FSS6	0,89	0,02	0,90	0,03	0,90	0,03	0,88	0,03
FSS7	0,88	0,02	0,89	0,02	0,89	0,02	0,91	0,02
FSS8	0,80	0,03	0,81	0,03	0,81	0,03	0,82	0,03
FSS9	0,84	0,02	0,85	0,03	0,85	0,03	0,87	0,03
A		0,93		0,93		0,93		0,91

**Tabla 8** Prueba de diferencia de la  $\chi^2$  en los modelos de la dimensión fatiga

Modelo	$\chi^2$	gl	$\Delta\chi^2$	$\Delta\text{gl}$	p	CFI	$\Delta\text{CFI}$
Configuración	34,07	18	-	-	-	0,99	-
Métrica	36,70	23	2,49	5	0,77	0,99	0
Fuerte	38,43	28	1,62	5	0,89	0,99	0
Estricta	55,02	34	10,35	6	0,11	0,98	0,01

actividad, ya que, de acuerdo con Santrock<sup>50</sup> la motivación es el conjunto de razones por las que las personas se comportan de las forma en que lo hacen.

La extensión de la escala la convierte en una versión corta que incluye solo a los reactivos con las mejores propiedades psicométricas para la población meta, el haber obtenido una versión corta es una ventaja ya que es más práctica para aplicar a los médicos residentes que se encuentran inmersos en grandes cargas de trabajo y no disponen de mucho tiempo para contestar instrumentos<sup>1</sup>.

Es importante destacar que en el presente trabajo se obtuvo una nueva versión de la FSS, por lo que los resultados obtenidos en ella no son comparables con los resultados obtenidos de otras versiones de la misma escala, ya que al cambiar el número de reactivos, la definición del constructo cambia y se basa en los resultados obtenidos por población de médicos residentes mexicanos, también se recomienda realizar estudios que aporten otras evidencias de validez para la escala, como aquellos realizados con algún criterio externo.

Una evidencia de validez concurrente es la correlación media con la Escala de impacto de fatiga.

En cuanto al cuarto objetivo se pudo observar que existe invarianza factorial estricta para comparar a los alumnos de programas de especialidad con los de alta o subespecialidad, lo que significa que no existe sesgo que favorezca a ningún grupo y las diferencias se deberían a fatiga, sin embargo, se encontró que no hay diferencias, lo que va en contra de los resultados de Papp et al.<sup>32</sup>, lo que sugiere que las diferencias pueden estar más en las causas de la fatiga y no en su efecto en el organismo.

## Limitaciones

Es importante resaltar que en este estudio no se utilizaron pruebas físicas para medir la fatiga y se consideraron solo medidas de autorreporte por lo que es conveniente hacer estudios que las incluyan, ya que a pesar de que el tratamiento estadístico riguroso de los datos ayuda a controlar el error en las mediciones, se ha observado que en el caso de autorreporte de salud las personas pueden exagerar o subestimar los síntomas por tener percepciones alteradas debido a su propio estado.

En el presente estudio se presentó evidencia sobre la confiabilidad de la escala por medio de su consistencia interna con el coeficiente alpha de Cronbach, sin embargo, en futuras investigaciones sería conveniente probar la estabilidad temporal de la prueba en el mismo individuo mediante su reproducibilidad test-retest.

También se sugiere considerar en futuros estudios si los médicos tienen consumo de sustancias tóxicas recientes o continuadas, realizar un cribado psiquiátrico para conocer efectos de la comorbilidad, acompañar el estudio con datos sobre la cantidad y calidad del sueño reciente, y registrar el momento en el que se realiza el estudio especialmente en

relación con las horas trabajadas, ya que todo esto puede influir en los resultados sobre la fatiga y en el presente estudio, solo se tomó como criterio de exclusión que los participantes no tuvieran algún diagnóstico de enfermedades crónicas o agudas.

Otro aspecto que considerar es que los médicos residentes mexicanos que participaron en este estudio lo hicieron de manera voluntaria, y provienen de gran diversidad de especialidades, pero no de todos los programas que se ofrecen en el país, por lo que resulta conveniente realizar más investigaciones a futuro en las que se trate de hacer procesos de selección aleatoria para evitar sesgos de autoselección y por políticas institucionales antes de generalizar los resultados obtenidos.

Si la escala quisiera usarse en otros contextos resulta conveniente hacer estudios sobre sus propiedades psicométricas, aun en la misma cultura si se va a trabajar con médicos con algún diagnóstico clínico, ya que para poblaciones con enfermedades parece que la estructura de reactivos puede ser distinta considerando las diferentes versiones validadas que existen en ellas y los reactivos eliminados de la que aquí se presenta.

Se debe destacar que la prueba de invarianza demuestra que la escala es una medida insesgada por nivel de especialidad, y por consiguiente es adecuado usarla para comparar la fatiga entre médicos de especialidad y médicos de alta o subespecialidad, sin embargo, si va a usarse para otras variables atributivas deben hacerse estudios que demuestren que funciona de la misma manera para todos los grupos.

## Conclusión

Los resultados encontrados muestran que la FSS tiene propiedades psicométricas que la hacen una medida confiable y válida para el propósito de evaluar fatiga en población de médicos residentes y puede ser usada en investigaciones o evaluación que contribuyan a mejorar los efectos de la residencia sobre su salud física, salud psicológica, bienestar general, rendimiento académico y laboral y relaciones interpersonales.

## Financiación

Financiado por el programa de Becas CONACYT.

## Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

## Bibliografía

- Slavin S, Shoss M, Broom MA. A program to prevent burnout, depression, and anxiety in first-year pediatric residents. *Acad-Pediatr*. 2017;17:456-8.
- Malta KS, Bittencourt AT, Kapp AC, Leite N. Sleep deprivation and drowsiness of medical residents and medical students. *Rev Col Bras Cir*. 2016;43:438-44.
- Pereira-Lima K, Loureiro SR. Burnout, anxiety, depression, and social skills in medical residents. *Psychol Health Med*. 2015;20:353-62.
- Avidan AY. Sleep and Fatigue Countermeasures for the Neurology Resident and Physician. *Continuum (Minneapolis)*. 2013;19:204-22.
- Gottschalk M, Kümpfel T, Flachenecker P, Uhr M, Trenkwalder C, Holsboer F. Fatigue and regulation of the hypothalamo-pituitary-adrenal axis in multiple sclerosis. *Arch Neurol*. 2005;62:277-80.
- Viner R, Christie D. Fatigue and somatic symptoms. *BMJ (Clinical Research Ed.)*. 2005;330:1012-5.
- Leiter MP, Frank E, Matheson TJ. Demands, values, and burnout: relevance for physicians. *Can Fam Physician*. 2009;55:1224-5.
- West CP, Tan AD, Shanafelt TD. Association of resident fatigue and distress with occupational blood and body fluid exposures and motor vehicle incidents. *Mayo Clin Proc*. 2012;87:1138-44.
- Buysse V, Sparkman KL, Wesley PW. Communities of practice: Connecting what we know with what we do. *Exceptional Children*. 2003;69:263-77.
- Canadian Association of Internes & Residents. Summary of Key Findings of the National Resident Survey. Paper presented at: National Resident Survey 2012 April 29-May 3; Alberta Canada: Canadian Association of Internes & Residents Annual Member Survey; 2012.
- Wallace JE, Lemaire JB, Ghali GA. Physician wellness: a missing quality indicator. *Lancet*. 2009;374:1714-21.
- West SG, Taylor AB, Wu W. Model fit and model selection in structural equation modeling. En: Hoyle RH, editor. *Handbook of structural equation modeling*. New York: Guilford Press; 2012.
- Dimitris KD, Taylor CD, Fankhauser RD. Resident work-week regulations: Historical review and modern perspectives. *J Surg Educ*. 2008;08:1931-7204.
- Feddock CA, Hoellein AR, Wilson JF, Caudill TS, Griffith CH. Do pressure and fatigue influence resident job performance? *Med Teach*. 2007;29:495-7.
- Ali MR, Khaleque A, Khanam M, al-Shatti A, Ahmed RU. Critical flicker frequency of mentally retarded and normal persons. *Percept Mot Skills*. 1994;79:1235-8.
- Kos D, Kerckhofs E, Carrea I, Verza R, Ramos M, Jansa J. Evaluation of the Modified Fatigue Impact Scale in four different European countries. *MultScler*. 2005;11:76-80.
- Armutlu K, Keser I, Korkmaz N, Akbiyik DI, Sümbüloğlu V, Güney Z. Psychometric study of Turkish version of Fatigue Impact Scale in multiple sclerosis patients. *J Neurol Sci*. 2007;15:64-8.
- Whitehead L. The measurement of fatigue in chronic illness: A systematic review of unidimensional and multidimensional fatigue measures. *J Pain Symptom Manage*. 2009;37:107-28.
- Lee KA, Hicks G, Nino-Murcia G. Validity and reliability of a scale to assess fatigue. *Psychiatry Res*. 1991;36:291-8.
- Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, et al. Development of a fatigue scale. *J Psychosom Res*. 1993;37:147-53.
- Fisk JD, Ritvo PG, Ross L, Haase DA, Marrie TJ, Schlech WF. Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. *Clin Infect Dis*. 1994;18:79-83.
- Fisk JD, Pontefract A, Ritvo PG, Archibald CJ, Murray TJ. The impact of fatigue on patients with multiple sclerosis. *Can J Neurol Sci*. 1994;21:9-14.
- Smets EM, Garssen B, Bonke B, de Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosom Res*. 1995;39:315-25.
- Schwartz AL. The Schwartz Cancer Fatigue Scale: testing reliability and validity. *Oncol Nurs Forum*. 1998;25:711-7.
- Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale Application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol*. 1989;46:1121-3.



26. Lerdal A, Kottorp A, Gay C, Aouizerat BE, Portillo CJ, Lee KA. A 7-item version of the fatigue severity scale has better psychometric properties among HIV-infected adults: an application of a Rasch model. *Qual Life Res.* 2011;20:1447–56.
27. Lerdal A, Kottorp A. Psychometric properties of the Fatigue Severity Scale—Rasch analyses of individual responses in a Norwegian stroke cohort. *Int J Nurs Stud.* 2011;48:1258–65.
28. Gök K, Cengiz G, Erol K, Özgöçmen S. The Turkish version of Multidimensional Assessment of Fatigue and Fatigue Severity Scale is reproducible and correlated with other outcome measures in patients with systemic sclerosis. *Archives of Rheumatology.* 2016;31:329–32.
29. Ottonello M, Pellicciari L, Giordano A, Foti C. Rasch analysis of the fatigue severity scale in Italian subjects with multiple sclerosis. *J Rehabil Med.* 2016;48:597–603.
30. Bernal L, Riveros F, Vinaccia S, Quiceno JM. Estructura factorial y consistencia interna de la Escala de Severidad de Fatiga en población colombiana con enfermedades crónicas. *Enfermería Global.* 2017;46:44–9.
31. Johansson S, Kottorp A, Lee KA, Gay CL, Lerdal A. Can the Fatigue Severity Scale 7-item version be used across different patient populations as a generic fatigue measure - a comparative study using a Rasch model approach. *Health Qual Life Outcomes.* 2014;12:1–9.
32. Papp KK, Stoller EP, Sage P, Aikens JE, Owens J, Avidan A, et al. The effects of sleep loss and fatigue on resident-physicians: A multi-institutional mixed-method study. *Acad Med.* 2004;79:394–406.
33. Nunnally JC, Bernstein IH. *Teoría psicométrica.* New York: McGraw-Hill; 1994. p. 256.
34. MacCallum RC, Hong S. Power analysis in covariance structure modeling using GFI and AGFI. *Multivariate Behav Res.* 1997;32:193–210.
35. Jackson DL. Revisiting sample size and number of parameter estimates: some support for the N: q hypothesis. *Struct Equ Model.* 2003;10:128–41.
36. Reyes-Lagunes LI. Las redes semánticas naturales, su conceptualización y su utilización en la construcción de instrumentos. *Rev Psic Soc y Pers.* 1993;9:81–97.
37. Duarte RE, Velasco AE, Sánchez-Sosa JJ, Reyes-Lagunes LI. Adaptación y validación de la escala de impacto de fatiga. *Act Inv Psic.* 2017;7:2585–92.
38. Krippendorff K. Computing Krippendorff's Alpha-Reliability. 2011 [consultado 24 Ene 2017]. Disponible en [http://repository.upenn.edu/asc\\_papers/](http://repository.upenn.edu/asc_papers/) 43.
39. Reyes-Lagunes LI, García-y-Barragán LF. Procedimiento de validación psicométrica culturalmente relevante: un ejemplo (capítulo 27). En: García T, editor. *La Psicología Social en México.* Vol. XII. México: Asociación Mexicana de Psicología Social. 2008. p. 625-36.
40. Millsap RE, Olivera-Aguilar M. Investigating measurement invariance using confirmatory factor analysis (capítulo 12). En: Hoyle RH, editor. *Principles and practice of structural equation modeling (4th ed.).* New York, NY: The Guilford Press; 2012. p. 380-92.
41. Kline R. *Principles and practice of structural equation modeling.* 3rd Edition NY: Guilford Publications; 2016. p. 366.
42. DeShon RP. Measures are not invariant across groups without error variance homogeneity. *Psychology Science.* 2004;46:137–49.
43. Magnusson D. *Teoría de los tests.* México: Trillas; 1998.
44. Fabrigar LR, MacCallum RC, Wegener DT, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods.* 1999;4:272–99.
45. Schmitt TA. Current methodological considerations in exploratory and confirmatory factor analysis. *J Psychoeduc Assess.* 2011;29:304–21.
46. Kenny DA, Milan S. *Handbook of structural equation modeling.* En: identification: a nontechnical discussion of a technical issue. New York: Guilford Press; 2012.
47. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *J StatSoftw.* 2012;48:1–36.
48. MacCallum RC, Roznowski M, Necowitz LB. Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychol Bull.* 1992;111:490–504.
49. Díaz-Guerrero R. *Psicología del mexicano (6. areimp.).* México: Trillas. 2003.
50. Santrock J. *Psicología de la educación.* México: Mc Graw-Hill; 2002.