



## Artículo de revisión

# Interpretación y evaluación de las propiedades de medición de los cuestionarios contestados por los pacientes

Interpretation and evaluation of the measurement properties of patient-reported outcomes

Agles Cruz-Peralta,\* Valentín Herrera-Alarcón†

\* Cirujano Dentista, Maestro en Educación.

† Cirujano Cardioráquico.

Hospital Regional «Lic. Adolfo López Mateos», ISSSTE. Universidad Nacional Autónoma de México. México.

**Citar como:** Cruz-Peralta A, Herrera-Alarcón V. Interpretación y evaluación de las propiedades de medición de los cuestionarios contestados por los pacientes. Alergia Asma Inmunol Pediatr. 2022; 31 (1): 27-31. <https://dx.doi.org/10.35366/104883>

## RESUMEN

Para medir la condición de salud de las personas tenemos instrumentos físicos como las básculas, el termómetro, esfigmomanómetro, etcétera; sin embargo, cuando se requiere medir la calidad de vida, la satisfacción, el dolor, la depresión y otras características con diferentes grados de subjetividad utilizamos cuestionarios, índices, escalas o inventarios. Los instrumentos de resultados informados por los pacientes (*Patient-Reported Outcomes [PRO]*) son cuestionarios o escalas que recogen la percepción de la condición de salud de un paciente, directamente de él mismo, sin la interpretación de un clínico o cualquier otra persona. Estos instrumentos deben poseer ciertas propiedades de medición que nos garanticen una adecuada medición. La confiabilidad informa la precisión de las mediciones, mientras que la validez es sobre la exactitud. La sensibilidad aplica sólo a instrumentos que miden cambio. Otro aspecto importante es su pertinencia para la población y contexto de la medición, además sus puntuaciones deben tener características que permitan interpretar claramente el resultado de su aplicación. Para comparar los resultados de pacientes con diferencias culturales o idiomáticas es necesario que se realice una adaptación del instrumento.

**Palabras clave:** Viabilidad, confiabilidad, validez, adaptados culturalmente, sensibles al cambio.

## ABSTRACT

*To measure people's health condition we have physical instruments such as scales, thermometer, sphygmomanometer, etc., however, when it is required to measure quality of life, satisfaction, pain, depression and other characteristics with different degrees of subjectivity, we use questionnaires, indices, scales or inventories. The instruments patient-reported outcomes (PRO) are questionnaires or scales that collect the perception of the health condition of a patient, directly from himself; without interpretation by a clinician or anyone else. These instruments must have certain measurement properties that guarantee an adequate measurement. Reliability informs about the precision of measurements, validity about accuracy. Responsiveness applies only to instruments that measure change. Another important aspect is their relevance to the population and context of the measurement, in addition, their scores must have characteristics that allow the result of their application to be clearly interpreted. To compare the results of patients with cultural or idiomatic differences, it is necessary to carry out an adaptation of the instrument.*

**Keywords:** Feasibility, reliability, validity, cross-cultural adaptation, responsiveness.

Recibido: 13/01/2022. Aceptado: 26/01/2022.

Correspondencia:

**Agles Cruz-Peralta**

E-mail: [agles.cruz@issste.gob.mx](mailto:agles.cruz@issste.gob.mx)



## INTRODUCCIÓN

Los instrumentos de medición como las escalas, índices, cuestionarios, algoritmos, etcétera, se utilizan para calificar o cuantificar atributos, cualidades, propiedades o constructos (conceptos completamente teóricos) que son imposibles de medir o cuantificar de manera directa.<sup>1,2</sup>

Los resultados informados por los pacientes (PRO) se definen como el reporte de la condición de salud de un paciente que proviene directamente de él mismo, sin la interpretación de un clínico o cualquier otra persona.<sup>3,4</sup> Los PRO requieren que los pacientes asignen respuestas a preguntas o afirmaciones relacionadas con sus percepciones o actividades. Estas respuestas son combinadas de forma que se crean puntuaciones que generalmente se suman y son usadas para medir conceptos como función, desempeño de roles, peso de síntomas, bienestar físico, psicológico, social, gravedad, etcétera.<sup>1,5</sup>

En la actualidad, el proceso de medición forma parte de la práctica clínica diaria de los profesionales de la salud. El resultado de aplicar un cuestionario o escala puede determinar una decisión pronóstica, diagnóstica o terapéutica. No obstante, en la mayoría de los casos se desconoce el origen o la capacidad real del instrumento para evaluar la condición bajo estudio.<sup>6,7</sup>

La Sociedad Internacional para la Fármaco-Economía y la Investigación de Resultados Reportados por los Pacientes (ISPOR) y el consenso internacional de taxonomía, terminología y definición de propiedades de medición de resultados reportados por los pacientes (COSMIN) han realizado los esfuerzos más importantes para sistematizar la metodología para la elaboración y validación de estas herramientas.<sup>4,8-10</sup>

Desde el año 2009 la *United States Food and Drug Administration* (FDA) se ha interesado en el desarrollo y la validación de instrumentos autoaplicables, con la finalidad de usar la información obtenida con ellos en la industria farmacéutica en las especificaciones de la ficha técnica de medicamentos aprobados. Inclusive ha publicado guías para evaluar los instrumentos PRO existentes, modificados o de reciente creación.<sup>11</sup>

Los indicadores métricos para determinar su calidad son la viabilidad, confiabilidad, validez, sensibilidad al cambio, adaptación cultural e interpretabilidad.

### ¿A qué se refiere la viabilidad (feasibility) de un instrumento?

Es la propiedad de un instrumento que informa sobre la probabilidad de utilización futura del mismo, calificándolo como útil, aceptado y contextualizado. El instrumento deberá ser: breve, claro, sencillo (fácil de contestar y

calificar), amenidad del formato y adecuación del mismo a la población que se pretende medir; interpretación de resultados lógica y no complicada, registro y codificación fácil, clara y precisa. Los instrumentos que requieren cálculos complejos para obtener el resultado, que implican altos costos o un largo tiempo para ser contestados carecen de esta calidad.<sup>12,13</sup> Los desarrolladores deben dar a conocer la experiencia del paciente respecto a la facilidad de contestar el cuestionario y la percepción del profesional en cuanto a su utilidad en la investigación o en la práctica clínica,<sup>14</sup> el tiempo promedio requerido para contestarlo, la necesidad de condiciones particulares en las cuales haya que poner al sujeto antes de iniciar y durante el procedimiento, si su aplicación es individual o grupal y la forma, método y tiempo requerido para calificar el puntaje de la escala.<sup>15</sup> Los detalles operacionales y los criterios de un índice determinan que las evaluaciones que se realicen con el mismo sean reproducibles. Asimismo, es de suma importancia saber la función clínica del instrumento: describir un estado, denotar un cambio, estimar un pronóstico, expresar una guía.<sup>2</sup>

En conclusión, la viabilidad nos indica el uso correcto del instrumento, el objetivo y las características de los datos que se pueden obtener.

### ¿Cómo saber si un instrumento es confiable y qué significa?

En cualquier medición, los datos obtenidos dependen de tres fuentes de variación cuantificables: el que realiza la medición, aquello que está siendo medido y el instrumento (cuestionario) con el que se está realizando la medición.

Se debe considerar la población que los desarrolladores originales usaron para crear el instrumento: edad promedio, espectro de enfermedad, enfermedades asociadas, tiempo de evolución y características del atributo a medir en cuanto a su objetividad, ya que estos factores pueden afectar los coeficientes de confiabilidad y hacer que el instrumento funcione diferente en otras poblaciones. En cuanto al instrumento, se debe considerar la variación que resulta de la estructura de éste.

La confiabilidad es la propiedad de mostrar resultados similares, libres de error, en mediciones repetidas. Asumiendo que las condiciones en las cuales se realiza la aplicación de la escala y el estado del atributo se mantienen estables en el tiempo o en aplicaciones simultáneas por diferentes evaluadores previa estandarización.<sup>16</sup>

De lo anterior se desprenden los diferentes tipos de confiabilidad y la forma de medir e interpretarlos:

**Estabilidad del instrumento:** quiere decir que el instrumento por sí mismo no agrega variación a los datos. Las puntuaciones obtenidas por una primera aplicación

del instrumento se pueden replicar en una segunda aplicación; lo que se comprueba con un método de prueba postprueba o *test-retest* al obtener un coeficiente de correlación entre la primera y segunda medición (se utiliza coeficientes de correlación de Pearson, Spearman o de preferencia coeficiente de correlación intraclass). Un resultado de 0.8 o más se considera bueno. Para medir correctamente esta «propiedad del instrumento» debemos mantener sin cambio el atributo en cuestión y usar un intervalo de tiempo que asegure que no queden efectos de la primera medición (usualmente son de siete a 15 días). También es importante que en ambas veces el contexto de aplicación sea el mismo. Como sinónimo de esta propiedad se ha utilizado repetibilidad y fiabilidad intraevaluador.<sup>17</sup>

**Replicabilidad o reproducibilidad del instrumento:** mide la capacidad del instrumento para dar evaluaciones similares cuando se aplica por diferentes evaluadores previa estandarización y en un contexto similar. Se ha utilizado como sinónimo fiabilidad entre observadores.<sup>18</sup> Cuando utilizamos PRO esta evaluación no se requiere, ya que son de autorreporte. En términos generales a esta propiedad se le conoce como confiabilidad.

**Consistencia interna:** se utiliza como sinónimo fiabilidad. Mide la relación que tienen las preguntas o afirmaciones de un cuestionario o escala entre sí (interrelación), significa que todos ellos miden el mismo atributo, es decir, son «consistentes entre sí».

Cuando un instrumento está compuesto por diferentes subescalas, cada una de las cuales pretende medir una dimensión diferente del fenómeno, debe evaluarse la consistencia interna de cada una de ellas. La evaluación de la consistencia interna requiere la aplicación del instrumento en una sola ocasión.<sup>12</sup> Si la interrelación es alta, se dice que tiene alto grado de homogeneidad y por lo tanto todos los ítems miden el mismo constructo. Sin embargo, a mayor número de ítems se sobreestima la confiabilidad (instrumentos con 20 o más ítems frecuentemente dan coeficiente  $\alpha$  de Cronbach de 0.90). Por otro lado, cuando hay ítems redundantes se disminuye el coeficiente, pero antes de decidir eliminar algún ítem debe evaluarse si ello afectaría a la validez del cuestionario, ya que podría ser preferible mantener ésta aun a costa de una consistencia interna ligeramente menor.<sup>12</sup> Cuando se tienen varias escalas que miden el mismo constructo que muestran similares valores de consistencia interna se puede suponer que las interrelaciones entre los ítems son mayores en aquélla que está compuesta por el menor número de ítems. Una consistencia interna adecuada va de 0.70 a 0.90.<sup>19</sup> No aplica a instrumentos de evaluación global que sólo tienen un ítem, a instrumentos pictográficos, escalas visuales análogas e instrumentos que miden rendimiento.

Se incrementará el coeficiente de confiabilidad si:

1. Eliminan los ítems que muestran correlaciones corregidas bajas con la puntuación total.
2. Eliminan los ítems que muestran redundancia (correlaciones entre ítems altas).
3. Existe mayor variación en las puntuaciones en los ítems individuales y en la escala global (aplicación a población heterogénea en relación con el atributo o característica que se pretende medir).
4. Existe mayor número de opciones de respuesta (polítómica vs dicotómica).

**Consistencia interna y validez:** la consistencia interna es un indicador indirecto de la validez del instrumento en la población de interés. La consistencia interna reportada en la población en la que se elaboró el instrumento o en otro estudio no necesariamente será la misma en la población de interés de un nuevo estudio, por lo cual siempre se debe hacer la medición y el informe. Los resultados observados pueden indicar la necesidad de adaptaciones de la escala para poblaciones específicas.<sup>6,20</sup>

### ¿Qué significa la validez de un instrumento?

La confiabilidad es una condición indispensable, pero no suficiente para garantizar la validez de un instrumento.<sup>20</sup> Un instrumento para ser válido requiere ser confiable, sin embargo, la confiabilidad no asegura la validez.

La validez es el grado en que un resultado obtenido refleja el fenómeno bajo estudio o la capacidad del instrumento para medir la cualidad para la que fue construido.<sup>7</sup> La confiabilidad indica la calidad de las mediciones, mientras que la validez señala la calidad de las inferencias que se pueden hacer a partir de las puntuaciones obtenidas con un instrumento.

¿Qué significan los diferentes tipos de validez?

¿Cómo saber que tan válido es un instrumento?

**Validez lógica o aparente.** Grado en que parece coherente que un cuestionario, una parte de él o un ítem evalúa adecuadamente lo que pretende medir. Algunos instrumentos han demostrado su validez en apariencia, con el uso a través del tiempo, pero es necesario contar con evidencias más sólidas.

**La validez de contenido.** Se define como el grado en que un instrumento mide todas las áreas que representan el concepto de interés. La evidencia debe demostrar que el instrumento representa el universo o la totalidad de los contenidos del fenómeno que pretende medir: «qué tanto la muestra de preguntas del instrumento representa el concepto bajo estudio». El objetivo es que el instrumento

contenga una muestra representativa de todos los dominios o áreas del concepto que se pretende medir. Para probar la validez de contenido es necesario aportar evidencia empírica de que los ítems y los dominios son pertinentes, relevantes y exhaustivos en relación al concepto que se pretende medir, al uso deseado y en la población determinada.<sup>21</sup> Así pues, la validez de contenido se gesta desde el inicio de la elaboración de un instrumento, estableciendo los límites y discriminando lo que pertenece o no al evento de estudio. La definición del atributo que se desea medir expresa: «¿Qué sí es? ¿Qué no es? ¿Qué partes lo forman?», de tal manera que se evita que se traslape con otros constructos o que se omitan partes importantes en su medición. El contexto de uso del instrumento determina «¿para qué queremos medir el constructo?» y la siguiente pregunta «¿en quién lo queremos medir?». La respuesta a estas interrogantes indicará la población de interés, inclusive las características de formato y la manera de administración.<sup>22</sup>

**Validez de criterio.** Este tipo de validez corresponde a un estudio de proceso denominado validación de pruebas diagnósticas. Requiere de un criterio externo (estándar de oro) que mida de forma adecuada el mismo atributo que se desea medir con el nuevo instrumento.<sup>23</sup> Esta evidencia se utiliza con mayor frecuencia para instrumentos para reporte por el clínico (*Clinician-Reported Outcome Assessments [ClinRO]*).<sup>9</sup> La validez de criterio se puede evaluar de dos formas: validez concurrente o validez predictiva. La primera es el grado en que se relaciona el resultado del nuevo instrumento con el resultado del estándar de oro o prueba de referencia, siendo ambas administradas simultáneamente en forma ciega e independiente. La validez predictiva valora hasta qué punto la nueva medida es capaz de predecir correctamente un resultado futuro, como puede ser la muerte, el alta hospitalaria, etcétera.<sup>23</sup>

**Validez de constructo.** Este tipo de validez es la alternativa cuando se requiere validar un instrumento que mide un atributo o característica que no puede evaluarse directamente con un estándar de oro válido o que usarlo no es posible por consideraciones éticas, por ejemplo, si implica una biopsia de cerebro no necesaria para el manejo del paciente. Se realiza obteniendo evidencia indirecta de que el nuevo instrumento realiza adecuadamente la medición para la que fue elaborado. El método es elaborar hipótesis *a priori* sobre el comportamiento teórico del atributo que se quiere medir con el nuevo instrumento, seguido de la contrastación empírica de los supuestos.<sup>24,25</sup>

La validez de constructo es similar a la validez de criterio, en cuanto a que utiliza criterios externos como referencia para evaluar el nuevo instrumento, pero la diferencia consiste en que estos criterios externos (que pueden ser otros cuestionarios, pruebas de laboratorio o gabinete, etcétera) miden atributos diferentes, aunque similares, o con alguna relación al que es el objetivo del nuevo instru-

mento, por lo cual no cabe esperar altas correlaciones y no aplica un estudio de pruebas diagnósticas. El objetivo consiste en demostrar una correlación directa o inversamente proporcional entre las puntuaciones obtenidas con el nuevo instrumento y las obtenidas con los criterios externos con base en conjeturas teóricas, de tal forma que esa relación sostenga una hipótesis creada *a priori*.<sup>26</sup> Por ejemplo, si el instrumento PRO pretende medir bienestar del paciente, sus puntuaciones serán divergentes con las puntuaciones de otro instrumento validado que mida gravedad de la enfermedad (mayor gravedad menor bienestar) y convergente con un instrumento validado que mida calidad de vida (a mayor calidad de vida mayor bienestar); por lo cual se denomina validez convergente o divergente.<sup>25,27</sup>

**Validez estructural.** Clásicamente la validez estructural se ha considerado un tipo de validez de constructo, aunque en la actualidad se concibe por separado. Este tipo de validez analiza la estructura del instrumento determinando su dimensionalidad, para esto se utiliza con mayor frecuencia el análisis factorial, que se obtiene con un estudio transversal. Con esta técnica estadística se pueden seleccionar los ítems de un cuestionario y determinar qué conjunto de ellos evalúan mejor el atributo que se pretende medir mediante las interrelaciones entre ellos y el constructo o atributo de interés.<sup>25,27</sup>

### ¿Qué es la sensibilidad al cambio (responsiveness) y cómo se evalúa?

Esta propiedad expresa la capacidad del nuevo instrumento para detectar cambios clínicos importantes del atributo que se pretende medir a través del tiempo; por lo tanto, aplica a cuestionarios, índices o escalas que pretenden evaluar intervenciones, útiles para estudios de eficacia de tratamiento, o bien para la práctica clínica. Es importante que para que un instrumento sea capaz de identificar estas diferencias de «antes y después» es indispensable que anteriormente haya demostrado que es estable mediante una prueba postprueba. Para obtener la evidencia de que el instrumento recién desarrollado tiene esta capacidad, se requiere de un estudio longitudinal que demuestre, mediante un criterio externo, que se presentó «cambio real» en el estado de salud de los pacientes en estudio. A esto se le denomina método de anclaje.<sup>28</sup>

Cuando un instrumento validado se necesita aplicar en un idioma o una región diferente de la que le dio origen es necesario realizar una adaptación al nuevo contexto (*Translation and Cross-Cultural Adaptation*), con la finalidad de que el instrumento en el nuevo idioma o con las correcciones que se le hayan realizado conserve una equivalencia conceptual y semántica con el original.<sup>29</sup>

Por último, los puntos obtenidos al aplicar el instrumento deben ser fácilmente interpretables; por ejemplo,

el CV-6 es un instrumento para medir la evolución de los pacientes con vitílico, a medida que se incrementa el número o tamaño de las lesiones se incrementa la puntuación; entonces, si hay incremento en el número o tamaño de las lesiones a través del tiempo, se puede entender que la enfermedad está empeorando y viceversa, si no hay cambio se considera estable.<sup>26</sup>

## REFERENCIAS

1. Lara-Muñoz MC, Ortega-Soto H. ¿La clinimetría o la psicometría? Medición en la práctica psiquiátrica. *Salud Mental*. 1995; 18 (4): 33-40. Disponible en: [http://www.revistasaludmental.mx/index.php/salud\\_mental/rt/metadata/574/0](http://www.revistasaludmental.mx/index.php/salud_mental/rt/metadata/574/0)
2. Lara-Muñoz MC. Psiquiatría-4. México: Intersistemas SA de CV; 2003. Disponible en: <https://es.slideshare.net/iniberto69/evaluacion-clinica-en-psiquiatria-pac>
3. Patrick DL, Burke LB, Gwaltney CJ et al. Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1-elicitating concepts for a new PRO instrument. *Value Heal*. 2011; 14 (8): 967-977. doi: 10.1016/j.jval.2011.06.014.
4. Walton MK, Powers JH, Hobart J et al. Clinical outcome assessments: conceptual foundation-report of the ISPOR clinical outcomes assessment-emerging good practices for outcomes research task force. *Value Heal*. 2015; 18 (6): 741-752. doi: 10.1016/j.jval.2015.08.006.
5. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: The ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO task force report. *Value Heal*. 2009; 12 (8): 1075-1083. doi: 10.1111/j.1524-4733.2009.00603.x.
6. Viola K, Nijsten T, Krishnamurthy K. "Validation" of outcome measures in dermatology. *J Invest Dermatol*. 2013; 133 (10): 1-4. doi: 10.1038/jid.2013.332.
7. Alarcón MAM, Muñoz N S. Medición en salud: Algunas consideraciones metodológicas. *Rev Med Chil*. 2008; 136 (1): 125-130. doi: 10.4067/S0034-98872008000100016.
8. Terwee CB, Bot SDM, de Boer MR et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007; 60 (1): 34-42. doi: 10.1016/j.jclinepi.2006.03.012.
9. Powers JH, Patrick DL, Walton MK et al. Clinician-reported outcome assessments of treatment benefit: report of the ISPOR clinical outcome assessment emerging good practices task force. *Value Heal*. 2017; 20 (1): 2-14. doi: 10.1016/j.jval.2016.11.005.
10. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, De Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012; 21 (4): 651-657. doi: 10.1007/s11136-011-9960-1.
11. Herdman M, Fernández N. Los resultados comunicados por los pacientes en los ensayos clínicos. Lo importante para el paciente es lo que él percibe. En: Dal-Ré R, Carné X, Gracia D. Luces y sombras en la investigación clínica. Madrid: TRIACASTELA; 2013. pp. 325-346. Disponible en: <https://www.fundaciogrifols.org/documents/4662337/4688901/cap12.pdf/873460ff-b4a2-407b-b0bd-fc5db5d2d102>
12. Argimon Pallás JM, Jiménez Villa J. Diseño de cuestionarios. En: Argimon JM, Jiménez Villa J. Métodos de investigación clínica y epidemiología. 2a ed. Barcelona, España: Hacourt; 2000. pp. 155-166.
13. Hernández-Sampieri R, Fernández-Collado C BP. Recolección de los datos cuantitativos. En: Metodología de la investigación. 4a ed. México: McGraw Hill; 2008. pp. 292-319.
14. Carvajal A, Centeno C, Watson R, Martínez M, Sanz Rubiales Á. ¿Cómo validar un instrumento de medida de la salud? *An Sist Sanit Navar*. 2011; 34 (1): 63-72. doi: 10.4321/S1137-66272011000100007.
15. Sánchez R, Echeverry J. Validación de escalas de medición en salud. *Rev Salud Pública*. 2004; 6 (302): 302-318. doi: 10.1590/S0124-00642004000300006.
16. Rajeswaran J, Blackstone EH. Patient-reported outcomes and importance of their appropriate statistical analyses. *J Thorac Cardiovasc Surg*. 2015; 150 (3): 461-462. doi: 10.1016/j.jtcvs.2015.07.043.
17. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005; 19 (1): 231-240.
18. Prieto G, Delgado AR. Fiabilidad y validez. *Papeles del Psicólogo*. 2010; 31 (1): 67-74. doi: 10.4067/S0718-09342002005100014.
19. Streiner DL. Being inconsistent about consistency: when coefficient alpha does and doesn't matter. *J Pers Assess*. 2003; 80 (3): 217-222. doi: 10.1207/S15327752JPA8003\_01.
20. Campo-Arias A, Oviedo HC. Propiedades psicométricas de una escala: la consistencia interna. *Rev Salud Pública*. 2008; 10 (5): 831-839. doi: 10.1590/S0124-00642008000500015.
21. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018; 27 (5): 1159-1170. doi: 10.1007/s11136-018-1829-0.
22. Cruz-Avelar A, Cruz-Peralta ES. Metodología para la construcción de instrumentos de medición en salud. *Alergia Asma Inmunol Pediatr*. 2017; 26 (3): 100-105.
23. Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Estudios de proceso (prueba diagnóstica). *Rev Med Inst Mex Seguro Soc*. 2011; 49 (2): 163-170. Disponible en: <http://www.mediographic.com/pdfs/imss/im-2011/im12k.pdf>
24. Magnusson D. Validez. En: Magnusson D, editor. Teoría de tests: psicometría diferencial, psicología aplicada, orientación vocacional. 2a ed. México: Trillas; 1990. pp. 237-267.
25. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. Cambridge: Cambridge University Press; 2011.
26. Peralta-Pedrero ML, Herrera-Bringas D, Torres-González KS, Morales-Sánchez MA, Jurado Santa-Cruz F, Cruz-Avelar A. Development and validation of a new scoring tool to evaluate the clinical evolution of adult patients with nonsegmental vitiligo. *Dermatology*. 2021; 237 (6): 952-960. doi: 10.1159/000511890.
27. Streiner DL, Norman GR. Validity. In: Streiner DL, editor. Health measurement scales. New York: Oxford University Press; 1989. pp. 11-18.
28. Polit DF. Assessing measurement in health: Beyond reliability and validity. *Int J Nurs Stud*. 2015; 52: 1746-1753. doi: 10.1016/j.ijnurstu.2015.07.002.
29. Ortiz-Gutiérrez S, Cruz-Avelar A. Translation and cross-cultural adaptation of health assessment tools. *Actas Dermosifiliogr*. 2018; 109: 202-206. Available in: <https://doi.org/10.1016/j.adengl.2018.02.003>