

Metodología para la construcción de instrumentos de medición en salud

Agles Cruz-Avelar,* Eduardo Sinaí Cruz-Peralta**

RESUMEN

Introducción: Los instrumentos de medición como las escalas, índices y cuestionarios, se utilizan para calificar o cuantificar atributos, cualidades, propiedades o constructos (conceptos completamente teóricos), que es imposible medir o cuantificar de otra forma. **Preparación:** Antes de proceder a medir algo, debemos definir de manera clara y precisa lo que se quiere medir, esto es definir el constructo. Debe explicitarse el objetivo de la medición del constructo, contestando la pregunta ¿para qué lo queremos medir? y la siguiente pregunta ¿en quién lo queremos medir? **Elaboración del instrumento:** Respecto a la función a desempeñar por el instrumento, se requerirán consideraciones metodológicas específicas. Para la elaboración del primer borrador, se identifican los indicadores que la evidencia científica avala como representantes del constructo. **Procedimientos para determinar la confiabilidad y validez:** Consistencia interna, *test-retest*, variabilidad entreevaluadores. Validez de apariencia, de contenido, constructo y criterio.

Palabras clave: *Test-retest*, Constructo, Instrumento de Medición en Salud.

ABSTRACT

Introduction: Measuring instruments such as scales, indices and questionnaires are used to qualify or quantify attributes, qualities, properties or constructs (completely theoretical concepts), which is impossible to measure or quantify otherwise. **Preparation:** Before proceeding to measure something, we must define clearly and precisely what we want to measure, this is to define the construct. The objective of the measurement of the construct must be made explicit, answering the question: what do we want to measure it for? And the next question, in whom do we want to measure it? **Elaboration of the instrument:** With respect to the role to be played by the instrument, specific methodological considerations will be required. For the elaboration of the first draft, the indicators that the scientific evidence endorses as representatives of the construct are identified. **Procedures for determining reliability and validity:** Internal consistency, *test-retest*, inter-rater variability. Face validity, content, construct and criteria.

Key words: *Test-retest*, construct, health measurement instruments.

* División de Estudios de Postgrado, Facultad de Medicina.

** Hospital General de México.

INTRODUCCIÓN

Aunque en diferentes niveles, cuando se entrevista a un paciente y se decide que tiene o no una enfermedad, se está realizando un proceso de medición. Medir es asignar un valor a objetos o sucesos de acuerdo con ciertas reglas.¹ Si necesitamos determinar la tensión arterial de un paciente o su nivel de albúmina en sangre, no tenemos ninguna duda ya que contamos con instrumentos específicos para ello; cuando las medidas o cuantificaciones tienen alta probabilidad de ser engañosas, existe dificultad para separar los sujetos o manifestaciones normales de los considerados anormales como sucede con algunos síntomas clínicos aislados, el bienestar, la satisfacción del paciente, grado de funcionalidad, etc. es necesario contar con otro tipo de instrumentos.²

Los instrumentos de medición como las escalas, índices y cuestionarios, se utilizan para calificar o cuantificar atributos, cualidades, propiedades o constructos (conceptos completamente teóricos), que es imposible medir o cuantificar de otra forma.³

Las características deseables en un instrumento son: fácil aplicación, amabilidad del formato, breve, claro, fácil interpretación de resultados, de fácil puntuación, registro y codificación. Las no deseables: que requieran cálculos complejos para obtener el resultado, que implique alto costo, largo tiempo para ser aplicado.⁴

PREPARACIÓN

Antes de proceder a medir algo debemos definir de manera clara y precisa lo que se quiere medir, esto es definir el constructo. Es importante tener presente, que un mismo problema puede definirse desde diferentes perspectivas teóricas y por lo tanto puede haber diferentes definiciones de un mismo constructo. En ocasiones es necesario construir la definición, seleccionar una de varias disponibles o simplemente explicitarla. Esto se puede hacer mediante consulta a expertos, revisión bibliográfica o ambos.⁵

Es fundamental establecer los límites y discriminar lo que pertenece o no al evento de estudio. Sabremos que nuestra definición es útil cuando podamos contestar ¿qué sí es? ¿qué no es? ¿qué partes lo forman? De tal manera que evitemos que se traslape con otros constructos o que se omitan partes importantes en su medición. Además, de la revisión bibliográfica se desprenderá la certeza de que no existe ya tal instrumento, lo cual entre otras cosas justifica su elaboración.

Debe explicitarse el objetivo de la medición del constructo, contestando la pregunta ¿para qué lo queremos medir? y la siguiente pregunta ¿en quién lo queremos medir? La respuesta a estas preguntas nos dará el camino para definir la población de interés y con ello las características de formato y manera de administración.³

Llegado este momento, ya tenemos los elementos básicos para hacer un modelo teórico con el cual podremos explicar la complejidad del constructo, ¿qué partes lo definen?, ¿cómo se expresa en términos operativos? ¿cuáles son los indicadores que nos permitirán medirlo? y anticipar ¿cómo se comportará la medición del atributo en la población de interés en circunstancias particulares?⁶

ELABORACIÓN DEL INSTRUMENTO

Del significado del constructo, propósito de la medición y población de estudio, dependerá en gran medida el contenido del instrumento.

Respecto a la función a desempeñar por el instrumento, se requerirán consideraciones metodológicas específicas, por ejemplo, en el área de diagnóstico se puede tratar de un instrumento para detección y por lo tanto será relevante su sensibilidad, para identificar personas en estadios preclínicos o clínicos tempranos o para categorización de los pacientes, siendo de especial importancia su capacidad de discriminación. En el área de pronóstico, un instrumento tendrá que demostrar su capacidad para predecir eventos específicos a futuro. En el área de tratamiento, será trascendente la capacidad de detectar cambio en el estado de salud a través del tiempo.⁷ Es esencial tomar en consideración estos aspectos, ya que van a definir si será un instrumento para ser llenado por el médico, por el paciente, a distancia con dispositivos electrónicos o inclusive mediante software.

Para la elaboración del primer borrador, se identifican los indicadores que la evidencia científica avala como representantes del constructo. Por ejemplo, si se piensa en un instrumento para identificar personas con mayor susceptibilidad de desarrollar urticaria crónica, será factible si se cuenta con estudios originales, donde se determinen los factores de riesgo y más aún si contamos con sus riesgos relativos podría pensarse inclusive en un instrumento para cuantificar riesgo, los ítems serían los factores de riesgo. En cambio, si lo que se necesita es medir el trato digno durante la atención médica y en la bibliografía no se ha determinado en forma operativa y válida cuáles son los componentes de trato digno, se tendrá que iniciar con la operacionalización del concepto. En este caso se tendría que empezar por revisiones bibliográficas, para determinar los dominios que componen el concepto de trato digno en la atención médica y luego proceder a validarlos mediante consensos formales con expertos en bioética. En un segundo paso, se tendrían que elaborar indicadores que derivarán de esos dominios para con ellos elaborar los ítems. También se puede dar el caso de que se necesite utilizar metodología cualitativa como grupos focales o entrevistas a profundidad, si lo que se necesita medir son percepciones, capacidades, desempeño de roles o responsabilidades

de los pacientes. Se pueden elaborar instrumentos para categorizar a los pacientes mediante criterios clínicos, que pueden ser elaborados por sociedades profesionales reconocidas internacionalmente.^{8,9}

Una vez que se han identificado los indicadores del constructo deberán ser transformados en ítems, se aconseja que al menos sean el doble de los que se piensa debe contener el instrumento, pero esto depende de qué tan subjetivo sea lo que mide cada ítem, si son signos o síntomas de enfermedad tan sólo variará

la redacción, pero si se mide algo más subjetivo como calidad de la atención en etapas posteriores, seguro se eliminarán gran parte de ellos.¹⁰

En la *figura 1* se puede observar la secuencia para realizar el proceso antes descrito.

Se debe decidir si se manejarán afirmaciones o preguntas. En el caso de preguntas decidir el número de alternativas de respuesta. Las decisiones para el formato derivarán del tipo de población que lo va a contestar: si el usuario será personal profesional o técnico se usa ter-

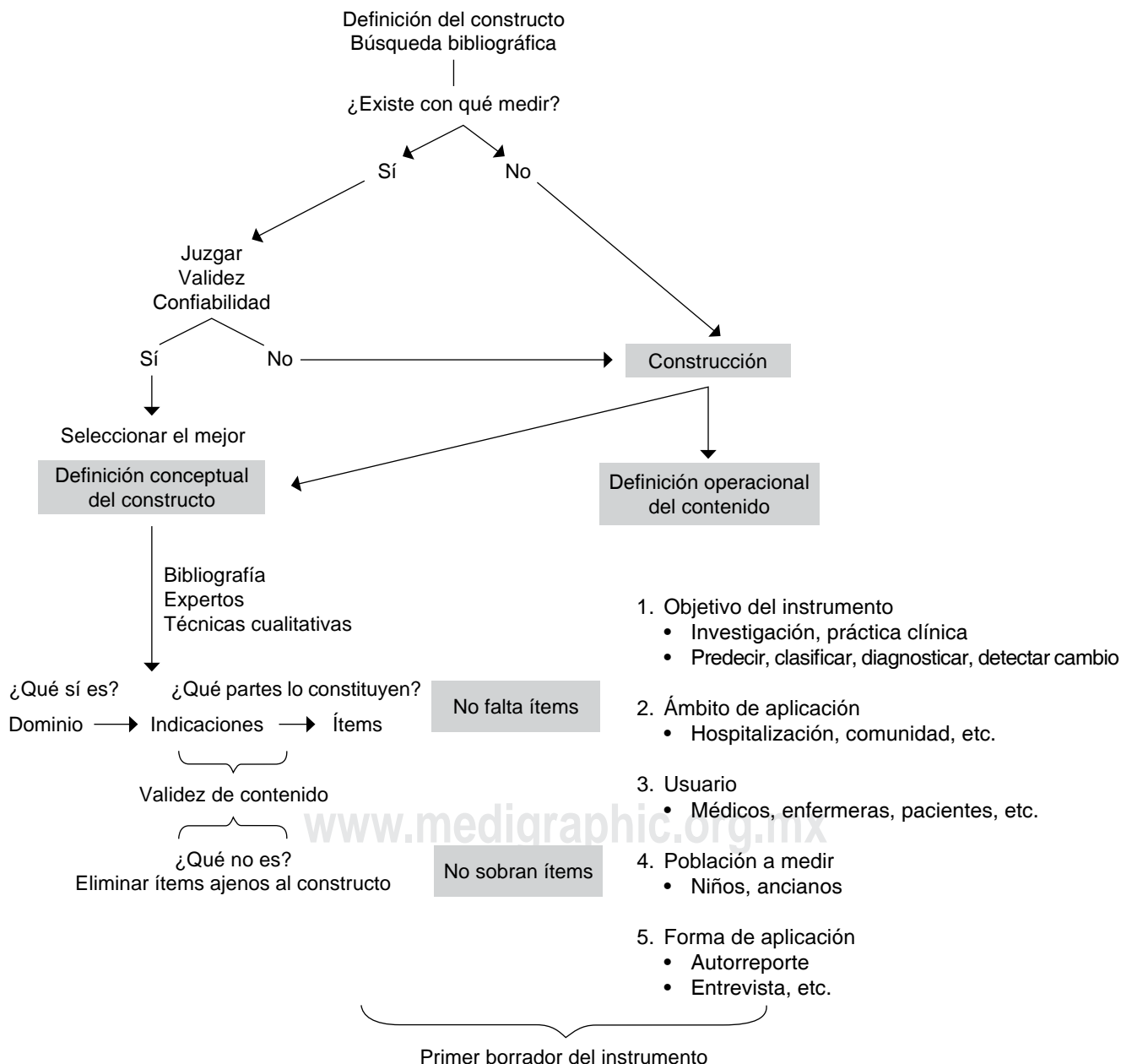


Figura 1. Proceso para elaborar un instrumento de medición en salud.

minología habitual en el campo, si es para pacientes se redactará en función a la edad y escolaridad, piloteando para determinar las palabras de uso común.

Revisar que los ítems se hayan redactado de forma correcta, evitando que puedan propiciar respuestas socialmente aceptadas, que los ítems o las alternativas de respuesta no sean mutuamente excluyentes, que no se formulen preguntas negativas, etc. De la misma forma se deberá revisar la redacción de las instrucciones, que sean claras y apropiadas a la población en quien se piensa aplicar.

Elaborado el primer borrador se debe someter a prueba de usuario. No existe una regla para determinar el tamaño de muestra, dependiendo de la objetividad de los ítems, se podrán aplicar a grupos de 10 en 10, hasta que estemos seguros de que todos los ítems son entendidos por el usuario, con el significado que se les atribuyó cuando fueron elaborados, que no sean ambiguos, identificar y eliminar la carga afectiva. Para ver la potencial utilidad de un ítem, sus respuestas deben variar en los diferentes pacientes y tener un rango de respuesta sin patrones, lo cual también refleja que la muestra de personas participantes, cubre el espectro del concepto objetivo del instrumento.⁸ Es útil la entrevista cognitiva mediante la cual documentaremos los ajustes que se van realizando.¹¹

Debe elaborarse a la par un manual donde se detalle contexto de aplicación, capacitación necesaria, tiempo de llenado y características clinimétricas del instrumento.

Es indispensable asegurarse de que la última versión del instrumento que va a ser sometido a validación, contenga los ítems o preguntas que representen totalmente el evento que se pretende medir, lo que le dará la validez de contenido.¹²

Para que el instrumento de medición sea útil, se necesita someterlo a pruebas empíricas que aportarán evidencias de confiabilidad y validez.

PROCEDIMIENTO PARA DETERMINAR LA CONFIABILIDAD

La confiabilidad es la propiedad de mostrar resultados similares, libres de error, en mediciones repetidas.¹³ Asumiendo que las condiciones en las cuales se realiza la aplicación de la escala y el estado del atributo, se mantienen estables en el tiempo¹⁴ o en aplicaciones simultáneas por diferentes evaluadores, previa estandarización.

El paso siguiente previa estandarización será determinar:

- a) Si el instrumento que generamos es capaz de ofrecer la misma información, cuando se aplica por un mismo observador dos veces consecutivas (se recomienda 15 días entre mediciones), en ausencia de cambios en el atributo que se está midiendo.¹⁵

- b) Si cuando se aplica por más de una persona al mismo tiempo pero en forma independiente a un mismo sujeto, se obtienen datos muy similares.

La primera prueba se denomina test-retest y la finalidad es obtener evidencias de que el instrumento por sí mismo no ofrece variaciones ajenas a la cantidad del atributo que se pretende medir; diríamos que el instrumento es «estable». En este caso se elimina la variación que podría provenir de los aplicadores. Como se obtiene con dos mediciones a través del tiempo, se deben ofrecer pruebas de que el atributo no ha sufrido cambios entre una y otra medición.

La segunda prueba mide la reproducibilidad del instrumento.¹⁶ Pretende verificar si la prueba puede ser aplicada por personas diferentes, por ejemplo dos enfermeras, o inclusive de diferente formación académica, una enfermera y una licenciada en nutrición. El objetivo de esto, es determinar qué tanto afecta a los resultados la variabilidad introducida por un segundo o tercer observador.

Por otro lado, es lógico pensar que los ítems contenidos en un instrumento se interrelacionan, pues todos buscan medir lo mismo, esto es la fiabilidad o consistencia interna. Esta propiedad nos informa del grado de homogeneidad que tiene un instrumento; a mayor homogeneidad mayor consistencia. Cuando el instrumento pretende medir atributos con alto grado de subjetividad y que no existe estándar de oro para una validez de criterio, esta propiedad da una medida indirecta de validez (aunque estrictamente no es una medida de validez), cuando un instrumento no demuestra su homogeneidad es muy probable, que los ítems midan constructos diferentes y se pone en duda su validez.¹⁷ Esta prueba se realiza aplicando el instrumento al menos en 100 personas.¹⁵

PROCEDIMIENTO PARA DETERMINAR 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.⁸

La validez de un instrumento se genera ofreciendo evidencias que pueden ser de diferentes tipos:

La validez de contenido es el grado en que un instrumento representa la totalidad de los contenidos del fenómeno que pretende medir. Demostrar que no omite ningún ítem ni incluye alguno extraño al atributo de interés. Esto se puede documentar haciendo explícito cómo se definió el constructo, los dominios, indicadores e ítems, ejemplo en *cuadro I*.

La validez de constructo se establece mediante la elaboración de hipótesis, en relación a cómo se comportarán las puntuaciones obtenidas con el instrumento en diferentes circunstancias:

Cuadro I. Cuadro sugerido para transitar desde un constructo hasta la elaboración de los ítems.¹⁸

Constructo	Dimensión	Indicador	Ítems
Riesgo de cáncer de piel Definición conceptual: Conjunto de características de una persona que hacen incrementar la probabilidad de desarrollar cáncer de piel	Características constitucionales	Fototipo de Fitzpatrick	¿Cuál es el color de tu piel?
	Antecedentes personales y familiares	Número de nevos melanocíticos	¿Aproximadamente cuántos lunares tiene en total en el cuerpo?
	Exposiciones ambientales	Antecedentes personales y/o familiares y personales de cáncer de piel	¿Su padre, madre, o hermanos tienen o han tenido cáncer de piel? ¿Usted tiene o ha tenido cáncer de piel?
		Exposiciones al sol Ingestión de agua contaminada con arsénico	¿Ha sufrido quemaduras de sol? ¿Cuántas veces? ¿Ha vivido en lugares con playa, en la montaña? ¿Ha consumido agua de pozo por 10 años o más?

- a) Aplicando instrumentos que miden constructos diferentes, por ejemplo, si se aplica el cuestionario para medir trato digno por enfermería y el cuestionario de Maslach Burnout; si el cuestionario que construimos realmente mide trato digno, con los resultados se demostrará que a mayor Burnout (desgaste por el trabajo) menor trato digno (validez divergente).
- b) Comparando grupos extremos, por ejemplo, se construyó un cuestionario para medir riesgo de cáncer de piel, para generar evidencia de que mide lo que dice medir. Se aplicó en dos grupos similares en características sociodemográficas; uno de pacientes con cáncer de piel y el otro sin cáncer de piel y se demostró diferencia clínica y estadísticamente significativa en la puntuación.¹⁸
- c) Si se crea un instrumento para predecir muerte en niños que se ingresan a cuidados intensivos. Se aplica a una muestra de pacientes y se siguen a través del tiempo y se verifica si la puntuación obtenida fue capaz de predecir la mortalidad.

La validez de constructo sólo se limita por la creatividad y conocimiento del investigador, ya que consiste en demostrar que el instrumento funciona para lo que fue creado. La metodología de la investigación y la estadística, son los instrumentos para generar y someter a prueba empírica estas hipótesis.

La validez de criterio es la más robusta, pero requiere de que el atributo que pretende medir el instrumento ya tenga un estándar de oro y consiste en demostrar que el nuevo instrumento mide el atributo de manera comparable, como lo hace el estándar de oro y puede ser de dos tipos:

- a) Validez concurrente: grado en que se relaciona la nueva medida con la de referencia, siendo ambas administradas simultáneamente.
- b) Validez predictiva: cuando el criterio de referencia no esté disponible hasta un tiempo después, por ejemplo, desarrollo de una enfermedad y valora hasta qué punto la nueva medida es capaz de predecirlo correctamente.

La validez de criterio es un estudio de proceso que se realiza con metodología para pruebas diagnósticas.¹⁹

La validez de un instrumento es cuestión de grado, instrumentos como el Apgar nunca se sometieron a una validación formal, sin embargo, hasta ahora tienen validez de apariencia pues se ha usado durante mucho tiempo por considerarse útil. La validez es cuestión de grado, y las evidencias pueden ser desde las más débiles como la validez de apariencia y contenido hasta la más robusta que es la validez de criterio.²⁰

Por último cuando un instrumento es aceptado por los usuarios y lo consideran útil, aceptado y contextualizado hablamos de otra característica que es la viabilidad.¹⁵

BIBLIOGRAFÍA

- Lara-Muñoz MC, Ortega-Soto H. ¿La clinimetría o la psicometría? Medición en la práctica psiquiátrica. *Salud Ment.* 1995; 18(4): 33-40. http://www.revistasaludmental.mx/index.php/salud_mental/rt/metadatos/574/0.
- Jenicek M. *Identificación de los casos de enfermedad. Clinimetría y diagnóstico.* In: Epidemiología. La Lógica de La Medicina Moderna; 1996, pp. 79-120.
- Lara-Muñoz MC. *Psiquiatría-4.* Editorial Intersistemas SA de CV, México: Asociación Psiquiátrica Mexicana AC; 2003. Disponible en: <https://es.slideshare.net/iniberto69/evaluacion-clinica-en-psiquiatria-pac>.

4. Argimon Pallás JM, Jiménez Villa J. *Diseño de cuestionarios*. In: Hacourt, ed. *Métodos de investigación clínica y epidemiológica*. 2a ed. Es: Hacourt; 2000. pp. 155-166.
5. Arribas M. Diseño y validación de cuestionarios. *Matronas Profesión*. 2004; 5(17): 23-29. http://www.enferpro.com/documentos/validacion_cuestionarios.pdf.
6. McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. *BMC Med*. 2011; 9: 86. <http://www.biomedcentral.com/1741-7015/9/86>.
7. Powers JH 3rd., Patrick DL, Walton MK, Marquis P, Cano S, Hobart J et al. Clinician-reported outcome assessments of treatment benefit: report of the ISPOR clinical outcome assessment emerging good practices task force. *Value Health*. 2017; 20(1): 2-14. doi:10.1016/j.jval.2016.11.005.
8. Alarcón AM, Muñoz S. Medición en salud: algunas consideraciones metodológicas Some methodological issues about measurements in health. *Rev Med Chile*. 2008; 136: 125-130. doi: 10.4067/S0034-98872008000100016.
9. Streiner DL. A checklist for evaluating the usefulness of rating scales. *Can J Psychiatry*. 1993; 38 (2): 140-148.
10. Hernández-Sampieri R, Fernández-Collado C BP. *Recolección de los datos cuantitativos*. In: McGrawHill, ed. *Metodología de La Investigación*. 4ta ed. México; 2008. pp. 92-319.
11. 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 2 - Assessing respondent understanding. *Value Health*. 2011; 14 (8): 978-988. doi: 10.1016/j.jval.2011.06.013.
12. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E 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-Eliciting concepts for a new PRO instrument. *Value Health*. 2011; 14 (8): 967-977. doi: 10.1016/j.jval.2011.06.014.
13. Kaplan RM. *Pruebas psicológicas : principios, aplicaciones y temas*. 6a ed. (Thomson, ed.). México, D.F.; 2006.
14. Rudner ML, Schafer WD. *What teachers need to know about assessment*. Washington, D.C.; 2002. https://www.researchgate.net/publication/237666611_What_Teachers_Need_to_Know_about_Assessment.
15. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, De Vet HC. 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.
16. Prieto G, Delgado AR. Fiabilidad y validez [Reliability and Validity]. *Papeles del Psicólogo*. 2010; 31 (1): 67-74. doi: 10.4067/S0718-09342002005100014.
17. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, 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.
18. Morales-Sánchez MA, Peralta-Pedrero ML, Domínguez-Gómez MA. Validation of a questionnaire to quantify the risk for skin cancer. *Gac Med Mex*. 2014; 150 (5) 409-419.
19. Lifshitz A, Talavera JO, Rivas-Ruiz R. Estudios de proceso (prueba diagnóstica). In: *Investigación clínica: en busca de la certidumbre*. 2013; 51: 16-23. http://revistamedica.imss.gob.mx/es/system/files/recurso_diverso/rm2013-suplemento1-completo_0.pdf.
20. Polit DF. Assessing measurement in health: beyond reliability and validity. *Int J Nurs Stud*. 2015; 52 (11): 1746-1753. doi: 10.1016/j.ijnurstu.2015.07.002.

Dirección para correspondencia:
Agles Cruz Avelar
Circuito de Postgrados, Unidad de Postgrado,
Edificio G 2º piso, oficina 221 Zona Cultural,
Ciudad Universitaria, Deleg. Coyoacán,
04510, Ciudad de México.
Cel: 5529223006
E-mail: aglesavelar@gmail.com