

Evaluación de la confiabilidad, validez y objetividad en instrumentos de medición: un estudio comparativo en áreas sociales y de salud

Evaluation of reliability, validity and objectivity in measurement instruments: a comparative study in social and health areas

Ma. de Lourdes E. García-Vargas ^a, Héctor D. Molina-Ruíz ^b

Abstract:

This study analyzes 10 measurement instruments and emphasizes the importance of using a robust questionnaire with appropriate metrics to ensure reliability, validity, and objectivity, which enhance the accuracy of the results. Reliability refers to the degree to which repeated applications of an instrument to the same individual or object yield consistent results. Objectivity is achieved when the instrument remains free from the biases and tendencies of the researcher. Validity refers to the extent to which the instrument accurately measures what it is intended to measure. The research has a descriptive and exploratory scope with a non-experimental longitudinal design, the instruments were applied in social and health areas in different Spanish-speaking countries, it was shown that all use metrics necessary for the instruments to offer results with greater precision, the use of Cronbach's alpha and Omega was appreciated to achieve reliability, it is important to consider that the different validity tests such as expert opinion, property analysis, exploratory factor analysis, confirmatory factor analysis and correlation analysis gave greater certainty when applying the proposed instruments. Objectivity was achieved in the instruments since there was no improvisation when applying the questionnaires, the same conditions and the same instructions were maintained to avoid bias. The proper use of existing metric tools ensures the instrument's validity, concluding that the instruments are robust and suitable for their intended purpose. It is recognized that there is no perfect measurement, but the measurement error must be reduced to tolerable limits, so the necessary metrics must be analyzed so that the error is reduced to tolerable limits.

Keywords:

Measuring instrument, objectivity, validity, reliability

Resumen:

Este estudio analiza 10 instrumentos de medición y subraya la importancia de utilizar un cuestionario robusto con métricas apropiadas para asegurar la confiabilidad, validez y objetividad, lo cual incrementa la exactitud en los resultados. La confiabilidad se refiere al grado en que las aplicaciones repetidas de un instrumento a un mismo individuo u objeto arrojan resultados consistentes. La objetividad se alcanza cuando el instrumento está libre de los sesgos y tendencias del investigador. La validez indica el grado en que el instrumento mide con precisión lo que pretende medir. La investigación tiene un alcance descriptivo y exploratorio con un diseño no experimental de corte longitudinal, los instrumentos fueron aplicados en áreas sociales y de salud en diferentes países de habla hispana, se demostró que todos utilizan métricas necesarias para que los instrumentos ofrezcan resultados con mayor precisión, se apreció el uso del alfa Cronbach y Omega para lograr la confiabilidad, es importante considerar que las diferentes pruebas de validez como la opinión de expertos, análisis de propiedades, análisis factorial exploratorio, análisis factorial confirmatorio y análisis de correlación dieron mayor certeza al aplicar los instrumentos propuestos. Se logró la objetividad en los instrumentos ya que no hubo improvisación al aplicar los cuestionarios, se mantuvieron las mismas condiciones y las mismas instrucciones para evitar sesgos. Se demuestra que el uso adecuado de herramientas métricas existentes otorga validez al instrumento y se concluye que son robustos y adecuados para lo que fueron diseñados. Se reconoce que no existe una medición perfecta, pero el error debe reducirse a límites aceptables mediante el análisis de métricas adecuadas.

Palabras Clave:

Instrumento de Medición, objetividad, validez, confiabilidad

^a Autor de Correspondencia, Universidad Autónoma del Estado de Hidalgo | Escuela Superior Tepeji | Tepeji del Río-Hidalgo | México, <https://orcid.org/0000-0002-1724-4744>, Email: maría_garcia10476@uaeh.edu.mx

^b Universidad Autónoma del Estado de Hidalgo | Escuela Superior Tepeji | Tepeji del Río-Hidalgo | México, <https://orcid.org/0000-0003-4657-3237>, Email: hmolina@uaeh.edu.mx

Introducción

El conocimiento se puede explorar desde una perspectiva empírica, experiencia en la materia puede ser examinado desde una perspectiva de primera, segunda o tercera personas. Desde la perspectiva de primera persona, los datos se centran en la práctica subjetiva y en la evaluación auto reportada. Por lo general, esta información se obtiene a través de cuestionarios, escalas, entrevistas y grupos focales.

Un cuestionario es un instrumento de recopilación de datos formado por un conjunto de preguntas o ítems relacionados con una o varias áreas específicas, aborda múltiples constructos y se estructura con diferentes tipos de preguntas, como cuestiones de opción múltiple, preguntas abiertas, cerradas, de verdadero o falso, etc., obtiene información sobre características, actitudes, creencias o comportamientos de los individuos; los puntajes obtenidos se examinan para realizar análisis estadísticos sencillos o complejos, de acuerdo a los objetivos planteados.

En una investigación, existen diferentes herramientas para recolectar datos, estos instrumentos deben ser confiables, objetivos y validados. Un instrumento se considera confiable cuando al aplicarse de forma repetida al mismo individuo u objeto produce los mismos resultados, la objetividad se fortalece con la estandarización para aplicar el instrumento y la validez se refiere a la capacidad del instrumento para medir lo que pretende medir

Marco teórico

Existen diferentes herramientas para recolectar datos; estos instrumentos deben ser confiables. Algunas formas para determinar la confiabilidad son, por coeficiente, por ejemplo, la medida de estabilidad (confiabilidad por test-retest), el método de formas alternativas o paralelas, el método de mitades partidas y medida de consistencia interna denominada coeficiente alfa Cronbach. La confiabilidad es variable de acuerdo al número de indicadores que incluye el instrumento de medición, entre más indicadores, más variación [1],[2].

La objetividad se fortalece con la estandarización; para aplicar el instrumento se deben presentar las mismas condiciones, generar las mismas instrucciones para todos los participantes y evaluar de manera objetiva. Para lograrlo, es importante capacitar al personal que está a cargo para que conozca el proceso, el objetivo de aplicar el instrumento y el instrumento mismo para que se desempeñe de una manera adecuada.

La evidencia de validez es la capacidad del instrumento para medir lo que pretende medir. Un instrumento válido debe medir de manera precisa y exacta la variable psicológica de interés, por lo tanto, un

instrumento de medición se valida por contenido, por criterio y por constructo [3],[4]. Estas formas de validez pueden combinarse entre sí ya que, si existe mayor evidencia de concordancia en los resultados, más cercano estará el instrumento de representar las variables que pretende medir. La validez de contenido se refiere al grado en que un instrumento refleja un dominio específico de información de lo que se mide; un instrumento requiere tener representados a prácticamente todos o la mayoría de los componentes del dominio de contenido de las variables a medir. El dominio de contenido de una variable, por lo regular, está justificado en la revisión de la literatura y antecedentes [5]. La evidencia sobre la validez de contenido se obtiene al contrastar el universo de ítems frente a los ítems presentes en el instrumento de medición, para determinar la validez de contenido se realiza un análisis minucioso de los elementos o ítems de la prueba y se evalúa si realmente representan y cubren de manera adecuada el concepto o el constructo que se desea medir, se establece mediante la opinión de expertos en el campo, quienes evalúan y juzgan la adecuación de los ítems [6].

En 1975, Lawshe propuso un modelo para determinar la validez de contenido que consiste en organizar panelistas para la evaluación de contenido integrado por especialistas para evaluar competencias, conocimientos, habilidades, funciones u otro tipo de elementos distintivos de la capacidad de un sujeto que va a ser evaluado. De acuerdo a este modelo, se entrega un ejemplar de la prueba o del conjunto de ítems a analizar y donde emiten su opinión en tres categorías: esencial, útil y no necesario. El autor propone que estas categorías se asocien con la habilidad, conocimiento o competencia medidos por el ítem para el desempeño de una tarea, una vez que los panelistas anotan su opinión respecto a cada ítem en las tres categorías citadas, se debe determinar el número de coincidencias en la categoría "esencial", se requieren acuerdos entre los jueces, de hecho, más del 50% de acuerdos debe ocurrir entre jueces en esta categoría para considerar que el ítem tiene un cierto grado de validez de contenido. Para establecer el consenso de los panelistas en la categoría "esencial" adicionalmente, debe pensarse que no se trata solamente que el CVR (el grado de acuerdo entre los expertos sobre la relevancia de un ítem para medir un constructo) sea positiva, sino que sea estadísticamente significativa, por lo cual Lawshe presenta una tabla con datos atribuidos a Schipper en un comunicado personal con el autor, pero no explica el modelo utilizado para obtener los valores de la tabla. Dentro de la tabla se tiene que para cinco panelistas el valor mínimo de CVR es 1 y para 40 panelistas el valor mínimo de CVR es 0.29, se nota que no hay propuesta para menos de 5 panelistas, de hecho, hasta 7 jueces o panelistas se requieren para tener un

valor de CVR igual a 1, es decir, un consenso perfecto entre los jueces. Una vez calculada la CVR de todos los ítems y aceptados los que tienen valores superiores a los mínimos propuestos por Lawshe, se calcula la media de CVR y con ello se obtiene el Índice de Validez de Contenido de toda la prueba (Content Validity Index, CVI), y se debe interpretar como la concordancia entre la capacidad (habilidad, competencia, conocimiento, etc.) solicitada en un dominio específico y el desempeño solicitado en la prueba que trata de medir dicho dominio, por lo tanto el CVI es el promedio de los CVR de todos los ítems aceptables.

De acuerdo a Tristán [7], el modelo de Lawshe ha demostrado ser de utilidad para dictaminar la calidad de un instrumento en función de la validez de contenido, pero tiene el problema de no aplicarse cuando se tiene un reducido panel de expertos, por lo que Tristán propuso un modelo alternativo que corrige los problemas del modelo de Lawshe y de fácil interpretación, con base en una normalización y la definición de un nuevo indicador del acuerdo entre los jueces. El modelo alternativo, de uso muy simple, ha sido empleado para dictaminar la validez de contenido de pruebas objetivas; recientemente se ha utilizado para dictaminar la validez de contenido de pruebas para pacientes de enfermería.

Para realizar un análisis cualitativo de validez de contenido con el modelo alternativo al propuesto por Lawshe, se sugieren estos pasos: Convocar a un panel para realizar la validez de contenido integrado por especialistas o expertos que conozcan del dominio que se va a evaluar, desde el punto de vista profesional, académico, de competencias, etc. Someter al escrutinio de los especialistas los ítems de un cuestionario o de un banco de ítems, los especialistas deberán clasificar a los ítems en tres categorías como “esencial”, “importante, aunque no indispensable”, “no necesario”, o nomenclaturas equivalentes. Para cada ítem contabilizar el total de casos clasificados por los panelistas como “esencial”. Determinar el CVR como proporción de acuerdos en la categoría esencial, respecto del número de panelistas involucrados en el escrutinio. Considerar como aceptables los ítems cuyo CVR sea superior o igual a 0.58. Estos valores son los que pueden integrar el instrumento o el banco. Los ítems no aceptables deberán ser revisados, corregidos y sometidos a un nuevo dictamen de parte de los panelistas. Determinar el CVI como promedio de los CVR de todos los ítems de cuestionario o del banco de datos (incluyendo aceptables y no aceptables). Se considera aceptable el conjunto de ítems si su CVI es superior a 0.58. Cuando el CVI se calcule solamente con los ítems aceptables, de entrada,

se sabe que su valor será superior a 0.58 y sólo se tiene que reportar el valor promedio obtenido.

Es importante considerar que la validez del contenido de un instrumento de medición por juicio de expertos es poco conocida. La técnica consiste básicamente, en solicitar la demanda de un juicio hacia un objeto, un instrumento, material de enseñanza u opinión respecto a un aspecto concreto a una serie de personas expertas en la materia [8]. La realización adecuada desde un punto de vista metodológico constituye a veces el único indicador de validez de contenido del instrumento de recogida de datos o de información cualitativa [9]; de ahí que resulte de gran utilidad en la valoración.

Con el objetivo de conocer más opciones en el momento de validar un instrumento de medición por expertos con orientación cualitativa, se describe el método y se citan autores especialistas en el tema para aprobarlo. Es importante destacar, como lo refieren Díaz et al, [10], la selección del número de especialistas depende de aspectos como la facilidad para acceder a ellos o la posibilidad de conocer expertos suficientes sobre la temática objeto de estudio. Por otra parte, autores como Escobar y Cuervo [11], señalan que el número de jueces que se deben emplear en un juicio depende del nivel de experticia y de la diversidad del conocimiento.

En relación con los métodos para recoger la información que tienen los jueces, se incluyen desde los más simples hasta los que implican un alto nivel de estructuración, pueden realizarse de manera individual, grupal o a través del método Delphi*, técnica que ofrece un alto nivel de interacción entre los expertos. El método individual consiste en obtener información de cada uno de los expertos sin que los mismos estén en contacto. Entre las técnicas grupales se encuentra la nominal y la de consenso, en las que se requieren la presencia de expertos y un nivel de acuerdo, que es mayor en el caso del consenso.

Después de recopilar la información brindada por los profesionales expertos se identifican los ítems como “entendibles” ya que, según los comentarios de los expertos, los ítems se presentan de forma clara en donde se puede verificar la terminología para su correcta redacción, así mismo la mayoría de entrevistados coincide en que los ítems y su clasificación son adecuados para identificar la validación de un juicio, instrumento u objeto solicitado. Por lo anterior, se concluye que el instrumento se encuentra bien elaborado, es posible que presente desviaciones mínimas que se identifican como ajustes y que finalmente se modifican.

* En el método Delphi cada juez realiza la evaluación individualmente y se recogen las opiniones de manera individual y anónima; luego de analizar las respuestas, se le

envía a cada uno de ellos la mediana obtenida y se le pide que reconsidere su juicio hasta llegar a un consenso.

En la fase final del proceso de consulta a los expertos se elaboran las conclusiones del juicio que serán utilizadas para la descripción en términos de validez y fiabilidad del instrumento de medición, sin desestimar la presencia de variables individuales como la personalidad o las habilidades sociales de los jueces que pueden generar sesgos a favor de uno o varios aspectos del mismo [12]. Una vez finalizada la evaluación de los expertos, se consideran sus aportaciones para realizar las modificaciones oportunas, ya que sus sugerencias avalan una concordancia entre el diseño del instrumento metodológico que se valida, su eficacia con respecto al objetivo para el que ha sido creado el constructo [13]. La duración del proceso de una validación por juicio de expertos es variable, depende, en primer lugar, de la disponibilidad de varios especialistas de la materia y, en segundo lugar, de las apreciaciones y aportaciones de los mismos que inciden directamente en el porcentaje de cambios en el objeto de validación. La calidad de los resultados depende, del cuidado que se ponga en la elaboración del instrumento de medición y en la elección de los expertos consultados, estos investigadores se caracterizan por haber realizado estudios sobre el tema seleccionado [14]. La selección de expertos con una distribución heterogénea garantiza mayor significancia al estudio, una vez diseñado el primer cuestionario, se somete a revisión de 5 académicos con el objetivo de localizar faltas y corregirlas antes de su aplicación.

La validez de criterio se refiere a la medida en que los resultados de una prueba o instrumento se correlacionan o están relacionados con un criterio externo o estándar conocido. Este criterio externo puede ser una medida ya establecida y aceptada en el campo o un resultado objetivo que se desea predecir, por lo tanto, la validez de criterio evalúa si un instrumento o prueba puede predecir o estimar con precisión un determinado comportamiento o resultado [15] y se establece mediante la comparación de los puntajes obtenidos en el instrumento con los puntajes del criterio externo, utilizando análisis estadísticos como la correlación. La magnitud y la dirección de la correlación indican la fuerza de la relación entre el instrumento y el criterio externo [16]. Se establece "la validez de un instrumento de medición al comparar sus resultados con los de algún criterio externo que pretende medir lo mismo" [17], generalmente la pregunta que se responde con la validez de criterio es: ¿En qué grado el instrumento comparado con otros criterios externos mide lo mismo? [18]. La evidencia sobre la validez de criterio se obtiene al comparar los resultados de la aplicación del instrumento de medición frente a los resultados de un criterio externo, por lo tanto, la validez de criterio es la

medida en que los resultados de una prueba o instrumento se correlacionan o están relacionados con un criterio externo o estándar conocido.

La validez de constructo se refiere a la medida en que un instrumento o prueba mide el constructo teórico o abstracto que se pretende evaluar. Los constructos son conceptos psicológicos, como la inteligencia, la personalidad, la motivación, la compasión, etc., que no pueden ser medidos directamente, la validez de constructo implica acumular evidencias de que las medidas utilizadas están relacionadas con el constructo teórico subyacente. Debe explicar el modelo teórico empírico que está detrás de la variable de interés. A esta validez le concierne el significado del instrumento, explicar que está midiendo y como opera [19]. Cabe resaltar que algunas preguntas que se deben responder con la validez de constructo son: ¿El concepto teórico está realmente reflejado en el instrumento? ¿Qué significan las puntuaciones del instrumento? ¿El instrumento mide el constructor y sus dimensiones? ¿Por qué? ¿cómo opera el instrumento? La evidencia sobre la validez de constructo se puede determinar mediante el análisis factorial, las correlaciones con otros constructos relacionados, la consistencia interna de los ítems y la comparación con teorías existentes [20].

Método

El estudio se llevó a cabo con un enfoque cuantitativo, ya que se midieron y estimaron las magnitudes de los problemas de investigación. Se orientó en el estudio de 10 instrumentos de medición y evaluarlos con el fin de explicar su condición. Es un estudio con alcance descriptivo porque ya se describen las características del fenómeno y lo que se busca, es exponer su presencia en un determinado grupo, el artículo tiene como finalidad establecer y describir el comportamiento de las variables en cuestión. Las variables estimadas fueron confiabilidad, contenido, validez, criterio, dentro de diez instrumentos de medición seleccionados aleatoriamente. En cuanto al diseño, es una investigación no experimental porque no existe manipulación de las variables, de corte longitudinal ya que se mide en un solo momento y en un tiempo único.

Se analizan 10 instrumentos de medición de origen latinoamericano aplicado en diferentes áreas social y de salud, en varios países de Latinoamérica en los años 2021-2024, la forma en que validan sus instrumentos se concentra en la tabla no. 1

Tabla 1

Revisión crítica de los instrumentos de medición en diferentes áreas.

Autor, año, país	Muestra	Edad	Muestreo	Alpha Cromb	Contenido	Validación	Criterio
Tenorio, et al. (2020) Chile	48 niños de Santiago, La Calera y Rancagua	14-24	Muestreo no-Prob.	Cognición = 0,919, Motor = 0,926, Lenguaje = 0,912 y Socio emocional = 0,863	Si	Si	Si
Tejada, et al.(2003) Oviedo, Esp.	791 inmigrantes exogrupal 418 endogrupal	Sub escala exogrupal 30 a 31 años, endogrupal 41-82 años	Aleatorio simple	Sub escala exogrupal de 0.78, endogrupal 0.73.	Si	Si	Si
Lara, et al.(2021)Chile	70 estudiantes	No se describe	Conveniencia	0.791	Si	Si	Si
Martínez, et al. (2020)Méx.	505 directivos	Promedio de 44.6	Conveniencia	0.868	Si	Si	Si
Chemisquy, et al.(2022) Argentina	307	18-54	Conveniencia	0.87	Si	Si	Si
Palacios (2021) México	309 jóvenes	14-24	No prob.	0.81	Si	Si	Si
Benavides y Ruiz(2024) Perú	584	No se describe	Muestreo Prob.	0.72 reeval. cognitiva. 0.74 supresión	Si	Si	Si
Gamarra, et al.(2024) Perú	522	60 a 93 años	No Prob.	0.8	Si	Si	Si
Alvarado, et al.(2024) Méx.	991	12-32 años	No prob.	0.81-0.84	Si	Si	Si
Páez, et al.(2024) Colombia	364	18-28	No prob.	0.7	Si	Si	Si

Resultados

Analizar los datos generales obtenidos permite caracterizar de manera general las diferentes métricas seleccionadas en cada instrumento para resumir que los instrumentos aplicados son robustos considerando el grado de confiabilidad, validez y objetividad, asimismo, se realiza el contraste de medias de cada factor determinado como autor, año, país, muestra, edad y tipo de muestreo. Los coeficientes de los 10 cuestionarios expresaron inter correlación, el grado de confiabilidad fue alto de acuerdo al alfa de Cronbach superior y Omega. En el instrumento de Tenorio la confiabilidad de consistencia interna se realizó para tres dimensiones y la puntuación total resultó adecuada para fases del desarrollo de herramientas (sustentado en su estructura interna y en su contenido) [21]. La estabilidad de los instrumentos fue buena dado que los indicadores se relacionan apropiadamente con sus dimensiones y la consistencia interna fue óptima.

El estudio de Martínez et al, validó el instrumento por juicio de especialistas para evaluar el grado de

relevancia, claridad y representatividad de las dimensiones, subdimensiones e indicadores [22]. Los especialistas se seleccionaron utilizando como criterios la publicación de investigaciones y disposición para participar en la investigación, la validez de constructo se realizó con la técnica del análisis factorial exploratorio (AFE), verificando la pertinencia de los datos. Los resultados revelan que los ítems están representados en el modelo factorial, manifestándose un solo factor, lo cual corresponde con el modelo teórico propuesto. Los índices de bondad de ajuste del modelo tetra factorial del PASS-Q fueron aceptables. Las cargas factoriales estandarizadas resultaron estadísticamente significativas, superando el valor mínimo recomendado ($\lambda = 0.40$). En la validez discriminante y concurrente se encontró respaldo empírico a las correlaciones del PASS-Q con el ABQ y SPSS. El modelo tetra factorial demostró invarianza configural, métrica y escalar en relación al género. El PASS-Q muestra adecuadas evidencias de validez y fiabilidad, recomendando su uso. Los resultados de Palacios muestran la estructura integrada por tres

factores mediante un análisis factorial confirmatorio [25], referentes a la amabilidad, complacencia y adaptación conductual. Las puntuaciones alcanzan un índice de consistencia interna de 85. La correlación ítems-factor revaluación cognitiva/supresión presentan buen índice de homogeneidad, y valores desde 0.534 hasta 0.730, y 0.718 hasta 0.834, respectivamente. En el análisis factorial confirmatorio se ratificó el modelo bifactorial independiente.

Por su parte, Gamarra [23] en Perú analizó evidencias de validez, confiabilidad e invarianza factorial de la Escala de Bienestar Psicológico, se encontraron adecuados índices de ajuste para un modelo de tres factores correlacionados. Se encontró relación entre puntuaciones de la BIESP-A y el WHO-5 ($r=0.504$), lo que es evidencia de validez convergente. Se acertó evidencia de confiabilidad con coeficientes alfa y omega (>0.80).

Conclusiones

No hay duda de que estas metodologías de validación proporcionan innumerables ventajas para evaluar y ajustar un instrumento de medición. Ahora bien, por el carácter cualitativo de las experiencias de investigación que se presentan en algunos casos, el factor “*subjetividad*” incide en un alto grado en las respuestas de los jueces, puesto que sus diferentes perspectivas pueden hacer que estos últimos se desvíen del objetivo específico del constructo. Por este motivo, resulta imprescindible partir de una formulación clara de los objetivos a fin de que no se generen imprecisiones.

Las bondades y practicidad de la técnica Delphi [24] también se le atribuye subjetividad por considerarla un simple procedimiento que no provee confianza de la opinión como fuente de información. Sin embargo, la prospectiva tiene la valiosa condición de no obviar el entorno económico, social y político en el que se desenvuelven. El método Delphi ofrece un adecuado canal para comprender áreas de conocimiento complicadas, dinámicas, imprecisas y con falta de información.

Se puede apreciar el uso del alfa Cronbach en todos los instrumentos de medición para lograr la confiabilidad de los instrumentos además que los autores manejaron algunos criterios diferentes para evaluar sus cuestionarios, sin embargo, es importante considerar que las diferentes pruebas realizadas demostraron mayor certeza para utilizar el instrumento. Por lo tanto, la evidencia demuestra que el uso adecuado de herramientas métricas existentes aprueba que los instrumentos son robustos y adecuados para lo que requirieron medir y se concluye que los cuestionarios y elementos que los integran representan y miden con precisión el constructo que pretendieron evaluar.

Referencias

- [1] Hernández, R., Fernández, C. y Baptista, P. (2010). Metodología de la investigación (6ta.edición). México: McGraw-Hill-Interamericana Editores, S.A. de C.V.
- [2] March, T. (2015). Caracterización de la validez y confiabilidad en el constructo metodológico de la investigación social. *REDHECS*, 20(10), 107-127.
- [3] Martínez, C., José, I., Palacios, A., Gloria E. & Juárez, L., Gibran. (2020). Análisis de validez de constructo del instrumento: “Enfoque Directivo en la Gestión para Resultados en la Sociedad del Conocimiento”. *RETOS. Revista de Ciencias de la Administración y Economía*, 10(19), 153-165. <https://doi.org/10.17163/ret.n19.2020.09>
- [4] Hernández, R., Fernández, C. y Baptista, P. (2010). Metodología de la investigación (6ta.edición). México: McGraw-Hill-Interamericana Editores, S.A. de C.V
- [5] Hernández, R., Fernández, C. y Baptista, P. (2010). Metodología de la investigación (6ta.edición). México: McGraw-Hill-Interamericana Editores, S.A. de C.V
- [6] Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual review of clinical psychology*, 5(1), 1-25.
- [7] Tristán, A. (2008). Modificación al modelo de Lawshe para el dictamen cuantitativo de la validez de contenido de un instrumento objetivo. *Avances en medición*.
- [8] Cabero A. J. y Llorente C., M. C. (2013), La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información (TIC). En *Eduweb. Revista de Tecnología de Información y Comunicación en Educación*, 7 (2) pp.11-22. Disponible en <http://tecnologiaedu.us.es/tecnoedu/images/stories/jca107.pdf>
- [9] Escobar, J. y Cuervo, M. A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. En *Avances en Medición*, 6, pp. 27-36. Disponible en http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3_Juicio_de_expertos_27-36.pdf
- [10] Díaz F. Y, Cruz R. M, Pérez Pravia M.C, Ortiz Cárdenas T. El método criterio de expertos en las investigaciones educacionales: visión desde una muestra de tesis doctorales. *Revista Cubana de Educación Superior* [Internet]. 2020 [Citado 03/01/2021];39(1): [Aprox. 2p.]. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0257-43142020000100018&nrm=iso
- [11] Escobar, J. y Cuervo, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. En *Avances en Medición*, 6, pp. 27-36. Disponible en http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3_Juicio_de_expertos_27-36.pdf
- [12] Escobar, J. y Cuervo, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. En *Avances en Medición*, 6, pp. 27-36. Disponible en http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3_Juicio_de_expertos_27-36.pdf
- [13] Prieto, G. y Delgado, A. (2010). Fiabilidad y validez. En *Papeles del Psicólogo*, 31 (1), pp. 67-74. Disponible en www.papelesdelpsicologo.es/vernumero.asp?id=1797

- [14] López, E. (2018). El método Delphi en la investigación actual en educación: una revisión teórica y metodológica. *Educación XX1* [Citado 17/12/2020];21(1):17-40. Disponible en: <https://www.redalyc.org/pdf/706/70653466002.pdf>
- [15] Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual review of clinical psychology*, 5(1), 1-25
- [16] Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual review of clinical psychology*, 5(1), 1-25
- [17] Hernández, R., Fernández, C. y Baptista, P. (2010). Metodología de la investigación (6ta.edición). México: McGraw-Hill-Interamericana Editores, S.A. de C.V.
- [18] Hernández, R. y Mendoza Ch.P. (2018). Metodología de la investigación (1ª. edición) México: Mc. Graw -Hill Education Editores, S.A. de. C.V.
- [19] Hernández, R., Fernández, C. y Baptista, P. (2010). Metodología de la investigación (6ta.edición). México: McGraw-Hill-Interamericana Editores, S.A. de C.V.
- [20] Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual review of clinical psychology*, 5(1), 1-25
- [21] Tenorio, B., Arango, B., Aparicio, A., Rosas, R., & Strasser, K. (2020). Evidencia de Confiabilidad y Validez del Test de Aprendizaje y Desarrollo Infantil (TADI) para Evaluación de Niñas y Niños Chilenos con Síndrome de Down. *Psykhé* (Santiago), 29(1), 1-16. <https://dx.doi.org/10.7764/psykhe.29.1.1378>
- [22] Martínez, C., Isaías, I., Palacios, A., & Juárez, L. (2020). Análisis de validez de constructo del instrumento: “Enfoque Directivo en la Gestión para Resultados en la Sociedad del Conocimiento”. *RETOS. Revista de Ciencias de la Administración y Economía*, 10(19), 153-165. <https://doi.org/10.17163/ret.n19.2020.09>
- [23] Gamarra, P., Quiroz, F. & Alomía, E. (2024). Evidencias psicométricas de la Escala de Bienestar Psicológico (BIEPS-A) en una muestra de adultos mayores peruanos. *Acta Colombiana de Psicología*, 27(1), 97-115
- [24] López, E. (2018). El método Delphi en la investigación actual en educación: una revisión teórica y metodológica. *Educación XX1* [Citado 17/12/2020];21(1):17-40. Disponible en: <https://www.redalyc.org/pdf/706/70653466002.pdf>
- [25] Palacios, J. (2021). Evidencias de validez y confiabilidad de la escala de flexibilidad en jóvenes mexicanos. *Persona*, (024 (1)), 27-45.