



Validez y fiabilidad de un instrumento de medición para un trabajo de investigación

Miguel Ángel Hernández Rivera¹, Manuel Arnoldo Rodríguez Medina², Adán Valles Chávez³ y Salvador Noriega Morales⁴

^{1,2,3}Intituto Tecnológico de Ciudad Juárez,

Av. Tecnológico 1340, Fuentes del Valle,

C.P. 32500, Cd. Juárez, Chihuahua.



Correo correspondal

miguel.hr@cdjuarez.tecnm.mx

⁴Universidad Autónoma de Ciudad Juárez,

Calle Henry Dunant planta alta, Omega, C.P. 32584, Cd. Juárez,

Chihuahua.

Resumen

En el presente artículo se describe el proceso de elaboración de un instrumento de medición que se utilizará en la recolección de los datos para el trabajo de investigación que busca identificar los principales factores que determinan el mejoramiento de la eficacia de la cadena de suministro. El estudio contempla el proceso de validez de contenido mediante el juicio por un experto y la fiabilidad del instrumento. Se utilizó el estadístico de Kendall como índice de concordancia en la validez del contenido y para la fiabilidad se utilizó el coeficiente de Alpha de Cronbach y el de Omega de McDonald para respaldar los resultados del coeficiente de Alpha de Cronbach. El instrumento, que responde a una escala tipo Likert, está conformado por 44 ítems en su primera prueba, posteriormente se adecuó a 44 ítems. Los resultados obtenidos muestran que el instrumento es confiable, permitiendo de esta manera utilizarlo en la investigación.

Abstract

This article describes the process of developing a measurement instrument that will be used in the collection of data for the research work that seeks to identify the main factors that determine the improvement of the efficiency of the supply chain. The study considers the content validity process through the judgment by an expert and the reliability of the instrument. The Kendall statistic is used as an index of agreement in the validity of the content, and for reliability the Cronbach's Alpha coefficient and McDonald's Omega coefficient were used to support the results of the Cronbach's Alpha coefficient. The instrument, which responds to a Likert-type scale, is made up of 44 items in its first test, then it is adapted to 44 items. The results obtained show that the instrument is reliable, thus allowing it to be used in research.

Palabras clave: Instrumento de medición, validez de contenido, juicio por un experto, coeficiente del Alpha de Cronbach, coeficiente de Omega de McDonald.

Keywords: Measurement instrument, content validity, judgment by an expert, Cronbach's Alpha coefficient, McDonald's Omega coefficient.

Introducción

Durante el proceso que se sigue en un trabajo de investigación surge la necesidad de probar la hipótesis que se plantea, mediante la recolección de los datos utilizando un instrumento de medición cuyas cualidades principales son; que su diseño garantice la recolección de la información que se quiere obtener (validez de contenido), y además, debe ser confiable, es decir, que sus mediciones deben tener consistencia a medida que pasa el tiempo (confiabilidad), para que la información que se obtenga sea la que el investigador requiere, reduciendo de esta forma el riesgo de un posible sesgo en los datos recolectados debido al mal diseño del instrumento.

Marco Teórico

De acuerdo a Mousalli K. (2017), la validez de contenido se refiere a la cualidad que tiene el instrumento de medir lo que debe medir. Mencionan Escobar, J. y Cuervo, Á. (2008), que la validez de contenido se establece en dos de las más frecuentes situaciones: el diseño de una prueba y la validación del instrumento construido para una población diferente pero que se adaptó a través de un procedimiento de traducción o equivalencia semántica, siendo necesario validar dichos instrumentos en términos de su contenido y es allí donde la evaluación realizada por expertos cobra especial relevancia debido a que ellos son los que deben modificar o eliminar los ítems que lo requieran, como el caso de las expresiones idiomáticas. Explica Escobar, J. y Cuervo, Á.

(2008), que de acuerdo a Cohen y Swerdik (2001), que la validez de contenido consiste en que tan adecuado es el muestreo que hace una prueba del universo de posibles conductas, de acuerdo con lo que se pretende medir.

Los ítems capturan las dimensiones que la prueba pretende medir. Los ítems seleccionados deben medir las dimensiones del constructo: Un error de validez de contenido puede ser que la dimensión semántica no tuviera ningún ítem que la evaluara, o que los ítems de la dimensión sintáctica sólo evaluaran una parte de ésta, al contrastar con lo que se pretende evaluar en dicha dimensión.

De acuerdo a Escobar y Cuervo (2008), el Coeficiente de concordancia W de Kendall se utiliza cuando se quiere conocer el grado de asociación entre k conjuntos de rangos (Siegel y Castellan, 1995), por lo cual es particularmente útil cuando se les solicita a los expertos asignarles rangos a los ítems, por ejemplo, de 1 a 4. El mínimo valor asumido por el coeficiente es 0 y el máximo 1.

Un acuerdo de 0.55 sería considerado bajo, y se podría inferir que hay dificultad en la clasificación, o que incluso, pueden tener ambigüedad en los indicadores que les permiten decidir en uno u otro sentido. Un acuerdo de 0.7 se considera como el mínimo aceptable. La fiabilidad de un instrumento se puede estimar con el alfa de Cronbach. La fiabilidad mediante el alfa de Cronbach asume que los ítems (medidos en escala tipo Likert) miden un mismo constructo y que están altamente correlacionados (Welch y Comer, 1988).

Entre más se acerca el valor del alfa a 1 mayor es la consistencia interna de los ítems analizados. La fiabilidad de la escala debe obtenerse siempre con los datos de cada muestra para garantizar la medida fiable del constructo en la muestra concreta de investigación.

De acuerdo a Bakieva, González y Sancho, (2014), el co-

eficiente de Alfa de Cronbach es un modelo de consistencia interna que se basa en el promedio de las relaciones entre los ítems. Durante el análisis, tiene la ventaja de evaluar cuanto mejoraría (o empeoraría) la fiabilidad de la prueba si se excluyera un determinado ítem. Como criterio general, George y Mallery (2003), sugieren las recomendaciones siguientes para evaluar los coeficientes de alfa de Cronbach:

- Coeficiente alfa >.9 es excelente.
- Coeficiente alfa >.8 es bueno.
- Coeficiente alfa >.7 es aceptable.
- Coeficiente alfa >.6 es cuestionable.
- Coeficiente alfa >.5 es pobre.
- Coeficiente alfa < .5 es inaceptable.

Mencionan Ventura y Caycho (2017), que en cuanto a las limitaciones al utilizar el coeficiente alfa de Cronbach, se encuentra que está afectado por el número de ítems, el número de alternativas de respuesta y la proporción de la varianza del test de acuerdo a (Domínguez-Lara y Merino-Soto, 2015a). Lo anterior se puede ver en la expresión matemática:

$$\alpha = \frac{K}{K-1} \left[1 - \frac{\sum S_i^2}{S_T^2} \right]$$

- α dónde: α es el símbolo del alfa de Cronbach.
- K el número de ítems.
- $\sum S_i^2$ i suma de las varianzas de cada ítem.
- S_T^2 la varianza total.

También comenta Ventura y Caycho (2017), que Lozano, García-Cueto y Muñiz (2008), mediante un estudio de simulación demostraron que la disminución en las alternativas de respuesta, disminuye la variabilidad de la escala, afectando el coeficiente alfa. Además, el coeficiente alfa presenta también el inconveniente de trabajar con variables continuas, algo que en ciencias sociales no ocurre, lo cual minusvalo-

ra la confiabilidad (Elosua y Zumbo, 2008), además de estar influido por el error muestral (Ledesma, 2004). No obstante, una de sus principales ventajas, es que sólo necesita una aplicación de la prueba para su cálculo según (Schmidt y Ilies, 2003). Menciona Ventura y Caycho (2017), que el coeficiente omega, a diferencia del coeficiente de alfa trabaja con las cargas factoriales según (Gerbing y Anderson, 1988), que son la suma ponderada de las variables estandarizadas, transformación que hace más estable los cálculos (Timmerman, 2005) y refleja el verdadero nivel de fiabilidad. En segundo lugar, no depende del número de ítems tal como se aprecia en su expresión matemática (McDonald, 1999):

$$\omega = \frac{\left(\sum_{i=1}^k \lambda_i \right)^2}{\left(\sum_{i=1}^k \lambda_i \right)^2 + \sum_{i=1}^k (1 - \lambda_i^2)}$$

Dónde:

- ω es el símbolo de coeficiente omega.
- λ_i es la carga factorial estandarizada de i .

También comenta Ventura y Caycho (2017), en tercer lugar, el coeficiente omega es considerado una adecuada medida de la confiabilidad si no se cumple el principio de tal equivalencia, el cual puede incumplirse si los coeficientes de los ítems que conforman una matriz de solución factorial presentan valores muy diferentes según (McDonald, 1999). De acuerdo a Ventura y Caycho (2017), para considerar un valor aceptable de confiabilidad mediante el coeficiente omega, éstos deben encontrarse entre .70 y .90 según (Campo-Arias y Oviedo, 2008), aunque en algunas circunstancias pueden aceptarse valores superiores a .65 (Katz, 2006).

Metodología

El proceso de validación y la fiabilidad del instrumento se llevo a cabo de acuerdo al siguiente diagrama:

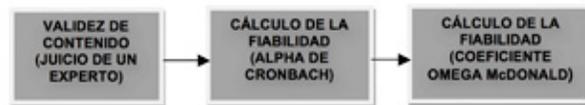


Figura 1. metodología de validación.

La validez de contenido se hizo mediante el juicio de un experto utilizando la tablilla que recomienda Escobar&Cuevo (2008), ver anexo 1.

Una vez que terminaron la evaluación los expertos, se recopilaron los datos obtenidos para calcular el nivel de concordancia entre jueces y para este análisis se utilizó el estadístico de Kendall, con la ayuda del software IBM SPSS Statistics versión 25. Se hizo la prueba de concordancia por cada uno de los cuatro ítems a evaluar en la tablilla. Estos son:

- a) Suficiencia.
- b) Coherencia.
- c) Relevancia.
- d) Claridad.

A continuación se muestra en la Tabla 1, los de resultados obtenidos para cada ítem.

Tabla 1. Resultados obtenidos por cada ítem.

ÍTEM	ÍNDICE DE KENDALL	OBSERVACIONES
Suficiencia	0.775	
Coherencia	0.783	
Relevancia	0.273	
Claridad	0.741	

Fuente: Elaboración propia.

Para determinar la fiabilidad del instrumento se usaron

los datos obtenidos de la aplicación de una muestra de 37 encuestas para calcular el índice del Alpha de Cronbach con la ayuda del lenguaje de programación R versión 4.0.3 (2020-10-10, obteniéndose los resultados que se muestran a continuación:

Raw	alpha	std.alpha	G6(smc)	average_r	S/N	ase	sd	median_r
0.96	0.96	1	0.38	26	0.0091	5.3	0.82	0.37
Lower	alpha	upper	95% confidence boundaries					
0.94	0.96	0.98						

El tamaño de la muestra se calculó considerando un universo de 150 empresas a las que va dirigida la investigación y se utilizó la fórmula de aleatorio simple, cómo se menciona en Bernal (2010). La fórmula es la siguiente:

$$n = \frac{S^2}{\frac{\epsilon^2}{Z^2} + \frac{S^2}{N}}$$

Dónde:

- $S^2 = p(1-p)$ donde p y q serán del 50% (0.50).
- N : Tamaño de la población.
- n : Tamaño necesario de la muestra de empresas.
- Z : Nivel de confianza o margen de confiabilidad (Se sugiere utilizar un margen de confianza alto: 95 % de confianza, para un valor de $Z = 1.96$).
- S : Desviación estándar de la población (estimada mediante una muestra piloto determinar tamaño de la piloto). $S^2 = p(1-p)$ donde p y q serán del 50% (0.50).
- p : Esta probabilidad se estima sobre marcos de muestreo previos o se define. La certeza total siempre es igual a uno, las posibilidades a partir de esto son “ p ” de que si ocurra y “ q ” de que no ocurra ($p + q = 1$). Cuando no tenemos marcos de muestreo previos, usamos un porcentaje estimado de 50%, es decir asumimos que p y q serán del 50% (0.50), y que resulta lo más común, particularmente cuando

seleccionamos por primera vez una muestra en una población (Hernández Sampeiri, Fernandez, & Baptista, 2014).

- ϵ : El error máximo aceptable se refiere a un porcentaje de error potencial que admitimos tolerar de que nuestra muestra no sea representativa de la población (de equivocarnos). Los niveles de error pueden ir de 20 a 1%. Los más comunes son 5% y 1% (uno implica tolerar muy poco error, 1 en 100, por así decirlo; mientras que 5%, es aceptar en 100, 5 posibilidades de equivocarnos) (Hernández Sampeiri, Fernandez y Baptista, 2014).

Los valores que se asignaron son los siguientes:

- $S^2 = p(1-p) = .50(.50) = .25$
- Tamaño del universo = 150
- Error máximo aceptable $\epsilon = 14\%$
- Porcentaje estimado de la muestra = 50%
- Nivel deseado de confianza = 95%

Por lo tanto, el tamaño de la muestra calculado será:

$$N = 37$$

Para respaldar los resultados obtenidos con el índice del Alpha de Cronbach se utilizó el coeficiente de Omega de McDonald y los cálculos se hicieron con el software JASP 0.8.5.1 y se obtuvieron los siguientes resultados:

Reliability Analysis		
Scale Reliability Statistics		
	Cronbach's α	McDonald's ω
scale	0.962	0.964

Note. Of the observations, 37 were used, 0 were excluded listwise, and 37 were provided.

Conclusiones

Los resultados obtenidos al hacer la validez de contenido muestran un valor arriba de 0.7 en la suficiencia, coherencia y claridad, valores aceptables de acuerdo a Escobar y Cuervo (2008), mientras que en la relevancia fue 0.273, por lo tanto el nivel concordancia se considera aceptable. En lo que se refiere a la fiabilidad del instrumento, se obtuvo un coeficiente del Alpha de Cronbach de 0.964 que, de acuerdo a George y Mallery (2003), se considera excelente ya que el resultado es > 0.9 . También se calculó el coeficiente de Omega de McDonald para respaldar el resultado obtenido con el índice del Alpha de Cronbach y fue de 0.964 confirmando lo que menciona Ventura y Caycho (2017) que de acuerdo a Domínguez-Lara y Merino-Soto (2015), en cuanto a las limitaciones del índice del Alpha de Cronbach en cuanto al número de ítems y el número de alternativas de respuesta tal como se muestra en la fórmula matemática para su cálculo, además el coeficiente alfa presenta el inconveniente de trabajar con variables continuas, pero a pesar de que en este instrumento de medición se utilizaron variables ordinarias, el número de ítems (43) sirvió para que el índice del Alpha de Cronbach fuera arriba de 0.9 valor que se confirma al realizar el cálculo del coeficiente del Omega de MacDonald cuyo resultado también sale por arriba de 0.9, confirmando la afirmación de Ventura y Caycho (2017), según (Gerbing y Anderson, 1988), que trabaja con las cargas factoriales y que no depende del número de ítems.

Referencias

- Bakieva, M.; González J. y Sancho, C. (2014). Análisis de Fiabilidad Alfa de Cronbach, InovaMide L4U.
- Bernal, C. (2010). Metodología de la Investigación. México, D.F.: Pearson Educación de México, S.A de C.V.
- Campo-Arias, A. y Oviedo, H. C. (2008). Propiedades psicométricas de una escala: la consistencia interna. *Rev Salud Pública*, 10 (5), pp. 831-839.
- Cohen, R. y Swerdlik, M. (2001). Pruebas y evaluación psicológicas: Introducción a las pruebas y a la medición. México: Mc Graw Hill.
- Domínguez-Lara, S. A. D. y Merino-Soto, C. M. (2015b). Sobre el reporte de confiabilidad del Clarp TDAH, de Salamanca (2010). *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 13 (2), pp. 1316-1317.
- Elosua, P. y Zumbo, B. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, 20 (4), pp. 896-901.
- Escobar, J. y Cuervo, Á. (2008). Validez de Contenido y Juicio de Expertos: Una aproximación a su Utilización. *Avances en la Medición* 6, 27-36.
- George, D. y Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed)*. Boston: Allyn & Bacon.
- Gerbing, D. W. y Anderson J. C. (1988). An update paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, 25 (2), pp. 186-192.
- Hernández Sampeiri, R.; Fernández, C. C y Baptista, L. P. (2014). *Metodología de la Investigación*. México, D.F.: Mc Graw Hill.
- Katz, M. H. (2006). *Multivariable analysis (2a ed.)*. Cambridge: Cambridge University Press
- Lozano, L., García-Cueto, E. & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4 (2), pp. 73- 79.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah: Lawrence Erlbaum Associates, Inc.
- Mousalli, G. (2017). *Los Instrumentos de Evaluación en la Investigación Educativa*. Mérida: Resarchegate management
- Schmidt, F. L.; Le, H. y Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods*, 8, pp. 206-224.