

ARTÍCULO ORIGINAL

Ruta metodológica para la validación de instrumentos en investigación de ingeniería

Methodological pathway for instrument validation in engineering research

Abanto, R.¹ , Cabrera, F.² , González, J.³ , Ruiz, J.⁴ , y Rodríguez, J.⁵ 

RESUMEN

El proceso de validación de instrumentos de medición es fundamental en la investigación, particularmente en el campo de la ingeniería, para garantizar que los datos obtenidos sean confiables, este proceso consta de dos fases principales: una cualitativa y una cuantitativa, la fase cualitativa se basa en el juicio de expertos, quienes evalúan la pertinencia del instrumento, para cuantificar este consenso y reducir la subjetividad, se emplea el coeficiente V de Aiken, posteriormente, la fase cuantitativa se enfoca en medir la fiabilidad interna del instrumento utilizando el Alfa de Cronbach, con un valor aceptable entre 0.7 y 0.9 tras la validación, se evalúa la normalidad de los datos utilizando pruebas como la de Shapiro-Wilk para muestras pequeñas (menos de 50) o la de Kolmogorov-Smirnov para muestras grandes (50 o más), esta evaluación es crucial para la selección de las pruebas estadísticas adecuadas: si los datos siguen una distribución normal, se aplican pruebas paramétricas como la t de Student; de lo contrario, se opta por pruebas no paramétricas como la U de Mann-Whitney.

Palabras clave: Coeficiente V de Aiken; Alfa de Cronbach; Prueba de normalidad.

ABSTRACT

The process of validating measurement instruments is fundamental to research, especially in engineering, to ensure that the data obtained is reliable. This process consists of two main phases: a qualitative and a quantitative one. The qualitative phase is based on the judgment of experts, who evaluate the instrument's relevance. To quantify this consensus and reduce subjectivity, Aiken's V coefficient is used. Subsequently, the quantitative phase focuses on measuring the instrument's internal reliability using Cronbach's Alpha, with an acceptable value between 0.7 and 0.9. After validation, the normality of the data is evaluated using tests such as Shapiro-Wilk for small samples (less than 50) or Kolmogorov-Smirnov for large samples (50 or more). This evaluation is crucial for selecting the appropriate statistical tests: if the data follows a normal distribution, parametric tests like the t-Student are applied; otherwise, non-parametric tests like the Mann-Whitney U are chosen.

Keywords: Coefficient V of Aiken's; Cronbach's Alpha; Normality Test.

* Autor para correspondencia

¹ Universidad Nacional de Trujillo, Perú. Email: rabantod@unitru.edu.pe ; jgonzalez@unitru.edu.pe ; jrodriguezmon@unitru.edu.pe; jsirlupu@unitru.edu.pe

² Universidad Señor de Sipán, Perú. Email: csernaqueflorro@uss.edu.pe

INTRODUCCIÓN

La validación de instrumentos es fundamental en la investigación, especialmente en ingeniería, para asegurar la credibilidad y fiabilidad de las mediciones (Santos et al., 2025), por ello se presenta una ruta para validar y analizar la consistencia de los instrumentos de investigación en ingeniería, este proceso garantiza que el instrumento evalúe lo que se diseñó y asegura la calidad y fiabilidad de los datos (Lages y Martínez, 2023), lo que impacta la validez de los resultados; la construcción de un instrumento incluye fases cualitativas y cuantitativas (Díaz y Terrazas, 2023) para evaluar la claridad, pertinencia y suficiencia de los ítems. La consistencia interna y la validación son clave para obtener resultados confiables y relevantes (Cascaes et al., 2015).

Esta ruta, centrada en técnicas como, el juicio de expertos asegura la adecuación y fiabilidad de los instrumentos (Galicia et al., 2017), pues especialistas evalúan la claridad, coherencia, relevancia y suficiencia de los ítems (Rodríguez, 2021), apoyado en las herramientas virtuales, que han optimizado este proceso al facilitar la organización y registro de opiniones, agilizando tiempos y reduciendo ambigüedades (Molina et al., 2016; López et al., 2019). Para cuantificar el consenso entre expertos se usa el coeficiente V de Aiken, esta medida robusta incorpora un enfoque probabilístico que reduce la subjetividad y permite calcular intervalos de confianza, generalmente con el método score de Wilson, mejorando la precisión (Rodríguez, 2015), evaluando la validez de contenido, cuantificando el acuerdo sobre la relevancia de los ítems en una escala de 0 a 1, destacando por su simplicidad y solidez metodológica (Robles, 2018).

Una vez validado el contenido, es esencial evaluar la fiabilidad del instrumento, específicamente su consistencia interna, usando el Alfa de Cronbach (Roco et al., 2024), un valor entre 0.7 y 0.9 es generalmente aceptable, aunque desde 0.6 puede ser tolerable en ciertos contextos, por ello es importante determinar los intervalos de confianza del Alfa (Caycho, 2017), ya que ofrecen una visión más precisa de su valor real. El Alfa puede variar por factores como el número de opciones en escalas Likert, la presencia de valores atípicos y el tamaño muestral (Toro et al., 2022).

Para una validación integral, se consideran otras pruebas estadísticas:

Pruebas de normalidad (Kolmogorov-Smirnov y Shapiro-Wilk); esenciales para evaluar si los datos siguen una distribución normal, basándose en la hipótesis nula de normalidad (Roco et al., 2023). Kolmogorov-Smirnov con corrección de Lilliefors evalúa el ajuste a una distribución

normal mediante la prueba de significancia de la hipótesis nula (Domínguez, 2018). *Pruebas de correlación (Pearson y Spearman)*, que evalúa relaciones lineales, mientras que Spearman analiza relaciones monotónicas sin requerir linealidad (Apaza et al., 2022).

Test de Chi-cuadrado; útil para analizar la relación entre variables cualitativas, comparando datos observados con los esperados bajo la hipótesis de independencia; el valor del estadístico y su p-valor indican la existencia de asociación (Cerda y Villaroel, 2007).

En consecuencia, el objetivo de esta investigación es proponer y describir una ruta metodológica para la validación de instrumentos en ingeniería, integrando el juicio de expertos y el análisis de consistencia interna, este estudio se centra en la propuesta procedural, validando su operatividad técnica mediante el cálculo de intervalos de confianza y la verificación de supuestos de normalidad para asegurar la fiabilidad de los datos en ciencias aplicadas

Es importante precisar que la ruta metodológica propuesta en este trabajo está diseñada primordialmente para investigaciones cuantitativas en el ámbito de la ingeniería que requieren el uso de instrumentos estructurados de medición, si bien los procedimientos descritos proporcionan un marco robusto para garantizar la validez y fiabilidad técnica, su aplicación en enfoques de carácter cualitativo o mixto debe considerarse de forma parcial y complementaria, quedando siempre condicionada al paradigma epistemológico que sustente la investigación; asimismo, aunque esta ruta es transversal a diversos campos de la ingeniería donde se empleen escalas de medición de actitudes, conocimientos o percepciones, su validez se circscribe a contextos donde la cuantificación de variables sea el eje central del análisis estadístico, previniendo así extrapolaciones indebidas en estudios de naturaleza puramente fenomenológica o interpretativa

MATERIALES Y MÉTODOS

Para garantizar la calidad y fiabilidad de los resultados, se valida el instrumento mediante juicio de expertos utilizando el coeficiente V de Aiken, que mide el consenso sobre la relevancia de los ítems. Participaron entre 5 expertos, quienes evaluaron relevancia, claridad y coherencia en escala ordinal (1 a 5). Se aplicó la fórmula original de Aiken y la modificada por Penfield y Giacobbi (2004), además del intervalo de confianza del valor V según el método de Wilson. Los criterios de interpretación serán: liberal ($V_0=0.50$), moderado ($V_0=0.70$) y estricto ($V_0=0.80$). La

fiabilidad interna se mide con el Alfa de Cronbach, que evalúa la consistencia entre ítems. Se considerará aceptable un valor entre 0.70 y 0.90, y admisible desde 0.60 en ciertos casos, según Nunnally y Bernstein (1994), se calcula también intervalos de confianza del Alfa con el método de Feldt (1965), factores como el tipo de escala Likert, valores atípicos y tamaño muestral pueden influir en su valor. A través de una prueba piloto, que incluya al menos 30 encuestados; previo al análisis inferencial, se verificará la normalidad de los datos. Para muestras < 50 , se usará la prueba de Shapiro-Wilk; para muestras ≥ 50 , la Kolmogorov-Smirnov con corrección de Lilliefors. Si $p > 0.05$, se asume distribución normal; si $p \leq 0.05$, se rechaza la normalidad. Según los resultados, se aplicarán pruebas paramétricas (como t de Student y correlación de Pearson) o no paramétricas (como Mann-Whitney U, Kruskal-Wallis y correlación de Spearman).

Validación por juicio de expertos (V de Aiken)

Fórmula original de Aiken (1980, 1985)

$$V = \frac{\sum_{i=1}^r (r_i - l)}{n(c - 1)}$$

r_i Puntuación otorgada por el juez i

c Número de categorías de la escala

l Valor mínimo de la escala

n Número total de jueces

El coeficiente V de Aiken para determinar la validez de contenido del instrumento, un panel de entre 3 y 10 expertos en la temática evaluó la relevancia, claridad y coherencia de cada ítem, en una escala ordinal (de 1 a 5).

V de Aiken modificado (Penfield y Giacobbi, 2004)

$$V = \frac{\bar{X} - l}{k}$$

Donde:

V Índice de validez de contenido (modificado)

\bar{X} Media de las puntuaciones asignadas por los jueces al ítem

l Valor mínimo posible en la escala de calificación

k Rango efectivo de la escala, definido como $c - 1$, es decir, el número total de opciones menos 1

Intervalo de confianza de V (score de Wilson)

El intervalo de confianza (IC) indica con qué grado de certeza podemos afirmar que el valor de V se encuentra dentro de un rango determinado.

Límite inferior

$$L = \frac{2nkV + z^2 - z * \sqrt{4nkV(1 - V) + z^2}}{2 * (nk + z^2)}$$

Límite superior

$$U = \frac{2nkV + z^2 + z * \sqrt{4nkV(1 - V) + z^2}}{2 * (nk + z^2)}$$

Donde:

- L, U Límite inferior y superior del intervalo de confianza, respectivamente
n Número de jueces
k Rango de la escala (k=c-1)
V Valor del coeficiente V de Aiken modificado
z Valor crítico de z según el nivel de confianza (por ejemplo: 1.96 para 95%)

No existe una "única tabla oficial" universal de intervalos de confianza (IC) para el V de Aiken, pero sí hay tablas críticas desarrolladas por Aiken (1985) para valores mínimos aceptables de V según el número de jueces y categorías de la escala, basadas en el valor crítico de V = 0.50; sin embargo, con la propuesta de Penfield y Giacobbi (2004) y Soto y Segovia (2009), se reemplazan esas tablas por intervalos de confianza construidos analíticamente, usando el método de Wilson.

Dicho eso, algunos autores sí presentan tablas prediseñadas de límites inferiores críticos de V para diferentes combinaciones de jueces y categorías, como orientación.

Tabla de referencia: Valores críticos del límite inferior del IC95% de V de Aiken, (*Rango de escala: 1-5; k = 4*)

Tabla 1

Valores mínimos aceptables del límite inferior del Intervalo de Confianza (IC 95%) para el coeficiente V de Aiken según el número de jueces.

N.º de jueces (n)	Límite inferior Mínimo aceptable (IC95%)
3	0.75
4	0.78
5	0.79
6	0.80
7	0.81
8	0.82
9	0.83
10	0.84

Nota: El coeficiente V de Aiken, que solía depender de tablas críticas, ahora se analiza con intervalos de confianza (IC) para mayor precisión. A pesar de esto, las tablas siguen siendo útiles como guía, ya que establecen el límite inferior mínimo aceptable del IC según el número de jueces, facilitando la validación de un instrumento.

Tabla 2

Criterios para el nivel de V

Nivel de V_0	Nombre del criterio	Significado
$V_0=0.50$	Criterio liberal Vincent, (1981)	Acepta ítems con baja concordancia entre jueces. Útil en estudios exploratorios o con pocas restricciones. Mayor tolerancia a ítems con validez moderada.
$V_0=0.70$	Criterio moderado o recomendado Charter, (2003)	Considera válidos solo ítems con alta concordancia entre jueces. Es el criterio más comúnmente usado en validación de instrumentos.
$V_0=0.80$	Criterio estricto	Exige muy alta concordancia, se usa en estudios donde se requiere máxima precisión o rigurosidad (evaluación clínica, psicometría avanzada, educación de alta calidad).

Nota: El coeficiente V de Aiken tiene tres criterios de validación: el liberal ($V_0=0.50$) para estudios exploratorios con baja concordancia; el moderado ($V_0=0.70$), el más común, que exige alta concordancia; y el estricto ($V_0=0.80$), que requiere la máxima concordancia para estudios de alta precisión.

EVALUACIÓN DE FIABILIDAD INTERNA (ALFA DE CRONBACH)

El análisis de consistencia interna es un proceso que evalúa la fiabilidad de una prueba, examinando cómo cada ítem contribuye al conjunto, y puede complementarse con herramientas gráficas que permiten visualizar patrones de respuesta y el impacto de incluir o eliminar ítems en la escala (Ledesma et al., 2022).

De acuerdo con Soler y Soler (2012), el análisis de consistencia interna permite evaluar la fiabilidad de un instrumento identificando ítems problemáticos y mostrando cómo su eliminación

o la estandarización de escalas mejora la coherencia de las respuestas, optimizando así la precisión del instrumento.

Para Oviedo y Campo (2005), la consistencia interna como la medida de la correlación entre ítems de una escala, evaluada principalmente mediante el alfa de Cronbach. Este coeficiente debe aplicarse en escalas unidimensionales y su valor puede variar según la población, por lo que debe interpretarse con cautela y reportarse en cada estudio.

Posterior a la aplicación piloto del instrumento (mínimo 30 encuestados), se calcula el coeficiente alfa de Cronbach para estimar la consistencia interna del instrumento.

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_T^2} \right)$$

α Coeficiente alfa de Cronbach

k Número de ítems del instrumento

σ_i^2 Varianza de cada ítem

σ_T^2 Varianza total del test (suma total de ítems por persona)

Tabla 3

El valor del alfa fue interpretado como sigue:

Valor de α	Interpretación (Nunnally & Bernstein, 1994)
$\alpha < 0.6$	Mala consistencia (debe revisarse el instrumento)
$0.6 \leq \alpha < 0.70$	Aceptable en estudios exploratorios
$0.7 \leq \alpha < 0.80$	Buena consistencia
$0.8 \leq \alpha < 0.90$	Muy buena consistencia
$\alpha \geq 0.9$	Excelente, aunque podría implicar redundancia

Nota: Se considera aceptable $\alpha \geq 0.70$, deficiente $\alpha < 0.60$ y posible redundancia $\alpha \geq 0.90$, según Nunnally & Bernstein (1994).

Método basado en la distribución F (Feldt, 1965)

Este es el método clásico y el más utilizado. El intervalo se calcula usando fórmulas relacionadas con la distribución F de Fisher (Feldt, 1965).

Fórmulas (IC del 95%):

Si α es el valor calculado, con k ítems y n sujetos, el intervalo de confianza al 95% se estima así:

$$\text{Límite inferior} = 1 - \left(\frac{(1-\alpha)F_{1-\frac{\alpha}{2}, v_1, v_2}}{\alpha} \right); \text{Límite superior} = 1 - \left(\frac{(1-\alpha)F_{\alpha, \frac{v_1, v_2}{2}}}{\alpha} \right)$$

Donde:
 $v_1 = n - 1$
 $v_2 = (n - 1)(k - 1)$
 F_{p, v_1, v_2} = es el valor crítico de la distribución F para el nivel de confianza deseado

PRUEBA DE NORMALIDAD

Antes de realizar el análisis estadístico inferencial, se verificó la distribución de los datos mediante pruebas de normalidad:

Tabla 4

Pruebas de normalidad

Prueba	Tamaño muestral	ventajas
Shapiro-Wilk	$n < 50$	Más potente para muestras pequeñas
Kolmogorov-Smirnov	$n \geq 50$	Bueno para distribuciones generales
Anderson-Darling	Cualquier n	Sensible en las colas
Lilliefors	$n \geq 50$	Adaptación de KS para μ y σ
		desconocidos

Nota: Normalidad evaluada con prueba de Shapiro-Wilk ($n < 50$). Se asumió distribución normal si $p > 0.05$ y no normal si $p \leq 0.05$.

Shapiro-Wilk ($n < 50$)

El **p-valor** fue el criterio de decisión:

- Si $p > 0.05$: los datos se asumen con distribución normal
- Si $p \leq 0.05$: se rechaza la hipótesis de normalidad

La prueba de Shapiro-Wilk fue desarrollada por Samuel Shapiro y Martin Wilk (1965).

Evalúa si una muestra proviene de una población con distribución normal, comparando el orden de los datos observados con los valores esperados de una distribución normal.

Es una prueba muy potente y sensible para tamaños de muestra pequeños o moderados (usualmente $n < 50$, aunque puede usarse hasta $n \approx 2000$ en software como SPSS o R).

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$x_{(i)}$ Valores ordenados de menor a mayor (orden estadístico)

\bar{x} Media de la muestra

a_i Constantes derivadas de los coeficientes esperados de los cuantiles de la normal

n Tamaño de la muestra

W Estadístico de Shapiro-Wilk (varía entre 0 y 1)

Kolmogorov-Smirnov ($n \geq 50$)

La prueba de Kolmogorov-Smirnov (K-S) compara la distribución empírica acumulada de los datos observados con la función de distribución acumulativa de una distribución teórica

(normal, uniforme, etc.). Fue desarrollada por Kolmogorov (1933) y extendida por Smirnov (1939).

Para que sea adecuada como prueba de normalidad, se aplica normalmente con la corrección de Lilliefors, ya que los parámetros (media y desviación estándar) son estimados de la muestra, no fijados.

$$D = \sup_x |F_n(x) - F(x)|$$

Donde

D Estadístico de la prueba

$F_n(x)$ Función de distribución acumulada empírica (de los datos)

$F(x)$ Función de distribución acumulada teórica (normal estándar)

\sup Supremum: el valor máximo de la diferencia entre ambas funciones

SELECCIÓN DE PRUEBAS ESTADÍSTICAS

La elección de pruebas estadísticas depende de la distribución de los datos:

(Flores Ruiz, Miranda Novales, & Villasís Keever, 2017), la prueba estadística adecuada garantiza resultados válidos y confiables, depende del diseño del estudio, tipo de variables y distribución de datos, un mal uso puede invalidar el análisis y afectar la credibilidad científica.

(Sánchez Solis, Raqui Ramírez, Huaroc Ponce, & Huaroc Ponce, 2024), la prueba de normalidad asegura un análisis estadístico válido. Shapiro-Wilk se usa en muestras <50 y Kolmogorov-Smirnov en >50 . Determina si aplicar pruebas paramétricas o no paramétricas, y combinar métodos numéricos y gráficos mejora su precisión.

A) Si los datos son normales (paramétricas):

Comparación de medias: t de Student,

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Donde:

$\bar{X}_1; \bar{X}_2$ Medias de los grupos 1 y 2.

$s_1^2; s_2^2$ Varianzas muestrales de los grupos.

$n_1; n_2$ Tamaños de las muestras.

Grados de libertad (df):

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

(Fórmula de Welch-Satterthwaite para varianzas desiguales).

Casos específicos

a) Muestras con varianzas iguales $s_1^2 = s_2^2$, se usa una varianza combinada s_p^2

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \text{ donde } s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

Grados de libertad (df): $df = n_1 + n_2 - 2$

b) Muestras pareadas (datos dependientes)

Para comparar mediciones antes/después en los mismos sujetos:

Si las varianzas se asumen iguales

$$t = \frac{\bar{D}}{s_D \sqrt{n}}$$

Donde:

\bar{D} Media de las diferencias entre pares.

s_D Desviación estándar de las diferencias

n Número de pares.

Grados de libertad $df = n - 1$

Interpretación

Valor crítico: Comparar el valor t calculado con el valor crítico de la tabla t (según df y nivel de significancia α , ej. 0.05).

Decisión:

Si $|t| > t_{\text{crítico}}$, se rechaza la hipótesis nula H_0 : no hay diferencia entre medias).

Si $p\text{-valor} < \alpha$, la diferencia es significativa.

Correlación: Pearson

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

Interpretación de r :

+1: Correlación positiva perfecta.

-1: Correlación negativa perfecta.

0: No hay correlación lineal.

Valores intermedios (ej.: 0.7, -0.3): Intensidad de la relación.

B) Si los datos no son normales (no paramétricas):

Comparación de medianas: **Mann-Whitney U, Kruskal-Wallis**

Mann-Whitney U

Es una prueba no paramétrica para comparar dos grupos independientes, alternativa a la prueba t de Student para muestras independientes.

Supuestos:

Muestras independientes

Variables ordinales, de intervalo o razón

No requiere normalidad

Se tienen dos grupos con tamaños n_1 y n_2 . Se combinan todos los datos y se les asignan ranks (rangos).

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_2$$

R_1 suma de rangos del grupo 1

R_2 suma de rangos del grupo 2

Se escoge el menor entre U_1 y U_2 como estadístico de prueba

$$U = \min (U_1, U_2)$$

Si las muestras son grandes ($n > 20$), se puede aproximar a una normal:

$$Z = \frac{U - \mu_U}{\sigma_U}$$

Donde

$$\mu_U = \frac{n_1 n_2}{2}, \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Kruskal-Wallis H

Es una extensión del Mann-Whitney para más de dos grupos independientes.

Alternativa no paramétrica al ANOVA.

Supuestos:

Grupos independientes

Medición ordinal o superior

No requiere homogeneidad de varianzas

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

N número total de observaciones

k número de grupos

R_i suma de rangos del grupo i

n_i número de observaciones en el grupo i

Corrección por empates

Si hay rangos empataos, se ajusta H :

$$H_{Corregida} = \frac{H}{1 - \frac{\sum T}{N^3 - N}}$$

Donde $T = t^3 - t$ para cada conjunto de empates (t : número de observaciones empatadas)

Distribución:

Para $k \geq 3$ y $n_i \geq 5$, H sigue una distribución χ^2 con $k-1$ grados de libertad.

Correlación Spearman: La correlación de Spearman es una medida no paramétrica del grado de asociación monótona entre dos variables. Evalúa si a medida que una variable aumenta, la otra tiende a aumentar (o disminuir), sin requerir que la relación sea lineal

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Donde:

n : Número de observaciones.

$d_i = R(X_i) - R(Y_i)$ diferencia entre los rangos de cada par de valores

Casos especiales:

Si hay empates: Se usa la fórmula de Pearson aplicada a los rangos:

$$r_s = \frac{cov(R_x, R_y)}{\sigma R_x \sigma R_y}$$

Donde:

R_x, R_y rangos de las variables X y Y

$cov(R_x, R_y)$ covarianza entre los rangos

$\sigma R_x \sigma R_y$ desviaciones estándar de los rangos

Es fundamental precisar que el presente estudio se define como una evaluación empírica de la ruta metodológica propuesta, por tanto, los análisis estadísticos presentados en las secciones posteriores no pretenden establecer hallazgos sustantivos sobre variables finales de ingeniería, sino que funcionan como evidencia demostrativa del desempeño y rigor de la ruta aplicada a un instrumento piloto, de este modo, el objeto real de evaluación es el procedimiento de validación en sí mismo, asegurando su aplicabilidad y reproducibilidad científica

RESULTADOS

La propuesta de esta ruta metodológica establece de forma secuencial las etapas necesarias para garantizar la validez y fiabilidad de un instrumento de medición en el campo de la ingeniería, se plantea un proceso estructurado que inicia con el juicio de expertos empleando el coeficiente V de Aiken y sus intervalos de confianza mediante el método score de Wilson, considerando criterios liberal, moderado y estricto para la toma de decisiones sobre cada ítem.

Para el Ítem 1 en los criterios de Suficiencia, Relevancia y Claridad.

$$V_{ij} = V_{1S} = \frac{\sum_{i=1}^r (r_i - l)}{n(c-1)} = \frac{(4-1) + (4-1) + (5-1) + (5-1) + (4-1)}{5 * (5-1)} = 0,850$$

$$V_{ij} = V_{1R} = \frac{\sum_{i=1}^r (r_i - l)}{n(c-1)} = \frac{(5-1) + (4-1) + (5-1) + (5-1) + (5-1)}{5 * (5-1)} = 0,950$$

$$V_{ij} = V_{1C} = \frac{\sum_{i=1}^r (r_i - l)}{n(c-1)} = \frac{(4-1) + (4-1) + (5-1) + (4-1) + (4-1)}{5 * (5-1)} = 0,800$$

$V_{ij} = V$ para ítem i y criterio j ; para $i = 1 - 20$ y $J = \text{Suficiencia}(S), \text{Relevancia}(R)$ y $\text{Claridad}(C)$

Luego, para todos los ítems:

$$V = \frac{\sum_{i=1}^r (r_i - l)}{n(c-1)} = 0,896$$

Los valores obtenidos mediante el coeficiente V de Aiken confirman una alta validez de contenido del instrumento, al situarse todos por encima de 0,80. El promedio general de 0,896 evidencia un sólido consenso entre los expertos sobre la claridad, relevancia y suficiencia de los ítems, garantizando que el cuestionario es técnicamente apto para medir las variables de investigación de manera precisa.

Tabla 5*Validación del instrumento con V-Aiken*

Var iable	Dime nsión Subdimensión	Í tem	Sufici encia	Relev ancia	Cla ridad	Pro medio V	
VI	(SD1)	1	0.850	0.950	0.80	0.86	
		2	0.900	1.000	0.80	0.90	
		3	0.900	0.950	0.85	0.90	
		4	0.850	0.850	0.85	0.85	
	(SD2)	5	0.950	0.900	0.95	0.93	
		6	0.850	0.850	0.90	0.86	
		7	0.850	0.950	1.00	0.93	
	(SD3)	8	0.900	0.800	1.00	0.90	
		9	1.000	0.850	0.90	0.91	
		0	0.900	0.900	0.90	0.90	
VD	(SD1)	1	0.850	0.900	0.90	0.88	
		2	0.850	0.850	0.90	0.86	
		3	0.900	0.850	0.95	0.90	
		4	0.850	0.850	0.90	0.86	
	(SD2)	5	0.950	0.950	0.95	0.95	
		6	1.000	0.850	0.90	0.91	
		7	0.900	0.850	1.00	0.91	
	(SD3)	8	0.800	0.850	0.90	0.85	
		9	0.900	0.900	0.90	0.90	
		0	0.900	0.900	-	0.90	
TOTAL						0.89	
6						6	

Nota: Instrumento diseñado y aplicado a 5 jueces, donde la variable independiente (VI) tiene 2 dimensiones (D), 3 subdimensiones (SD) y 9 ítems; mientras que la variable dependiente (VD) posee 2 dimensiones, 3 subdimensiones y 11 ítems.

Tabla 6*Fiabilidad del instrumento con Alfa de Cronbach*

Encuesta do	Ítems																			
	I	I	I	I	I5	I	I	I	I	I1	I2									
1	2	3	4	1	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	
E1	5	2	2	3	5	5	3	3	3	4	4	3	3	3	4	4	4	4	3	3
E2	5	3	3	3	5	5	3	4	4	3	4	3	3	4	4	4	3	3	3	4
E3	5	1	2	4	5	4	3	3	3	4	4	3	3	3	3	4	4	3	3	3
E4	5	1	1	4	5	5	3	3	4	3	3	3	4	3	4	4	3	3	4	3
E5	5	3	3	4	5	4	4	4	3	4	3	3	3	3	4	4	4	4	3	3
E6	5	2	3	2	5	5	4	3	3	3	3	4	4	3	4	3	3	3	3	3
E7	5	2	3	2	5	5	4	4	3	3	3	3	4	3	4	4	3	3	4	3
E8	5	1	1	3	5	4	3	3	3	3	4	3	3	3	4	4	4	3	3	4
E9	5	1	3	3	5	5	3	4	4	3	3	3	4	3	3	4	4	4	3	3
E10	5	3	3	3	5	4	4	3	3	4	4	4	4	3	4	4	4	4	3	3
E11	5	2	3	3	5	5	3	4	3	4	3	3	4	3	4	4	4	4	3	3
E12	5	1	1	3	5	5	4	3	4	3	3	4	4	3	3	3	4	4	3	4
E13	5	1	3	3	5	5	3	4	3	4	4	3	4	3	4	3	3	3	3	4
E14	5	3	3	3	5	5	4	3	4	3	4	4	4	3	3	3	3	3	3	4
E15	5	2	3	3	5	4	4	4	4	3	3	3	3	3	4	4	4	4	3	3
E16	5	1	1	4	4	4	4	4	3	3	4	4	3	4	3	4	3	3	4	4
E17	5	3	3	4	5	4	4	3	4	4	4	4	3	3	4	3	3	3	4	3
E18	5	2	2	1	3	3	4	4	4	4	3	4	3	4	4	3	3	3	4	4
E19	5	3	3	1	5	5	4	3	4	3	3	4	4	3	3	4	3	3	4	4
E20	5	3	3	1	5	5	4	3	3	4	4	3	4	4	3	4	3	3	3	3
E21	5	2	3	1	4	5	3	4	3	4	3	3	4	3	3	3	3	3	3	4
E22	5	2	3	1	5	4	4	4	4	4	3	3	3	3	3	4	3	3	3	4
E23	5	3	3	2	4	5	3	3	3	3	3	3	3	4	3	3	4	4	3	3
E24	5	3	3	2	4	4	4	3	4	4	4	4	3	3	4	4	4	4	3	3
E25	5	2	3	2	4	4	4	4	3	3	3	4	3	3	4	3	3	3	3	4
E26	5	3	3	2	5	4	4	4	4	3	3	3	3	3	3	4	4	4	3	3
E27	5	3	3	2	5	5	4	4	3	3	3	4	4	4	3	3	3	3	3	4
E28	5	3	3	2	5	4	4	3	3	3	3	3	3	3	3	4	3	3	3	4
E29	5	2	3	2	5	4	4	3	4	3	4	4	4	3	4	4	3	3	3	3
E30	5	3	3	2	5	4	4	4	3	3	3	3	3	3	4	4	4	4	3	3

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_T^2} \right) = \frac{30}{30-1} \left(1 - \frac{21,02}{379,2} \right) =$$

$$\alpha = 1,03448275862069 * 0,944567510548523 = 0,977138804015714$$

El instrumento posee una validez de contenido óptima (V de Aiken = 0,896) y una consistencia interna excelente (Alfa de Cronbach = 0,977), garantizando que es altamente preciso y apto para su aplicación.

Se incorpora la evaluación de la consistencia interna a través del Alfa de Cronbach con sus intervalos de confianza calculados mediante el método de Feldt, lo que permite estimar la estabilidad del instrumento antes de su aplicación definitiva, el diseño contempla la verificación de supuestos estadísticos previos, particularmente la normalidad de los datos (Shapiro-Wilk o Kolmogorov-Smirnov), a fin de seleccionar pruebas paramétricas o no paramétricas de manera justificada; con ello, la ruta metodológica ofrece un marco replicable, ordenado y adaptable a distintos contextos de investigación.

Límites de fiabilidad (F para α IC del 95%) y un alfa de Cronbach 0,977138804015714:

Cálculo de grados de libertad:

$$F_{p, V_1, V_2} = \text{es el valor crítico de la distribución F con } V_1 \text{ y } V_2 \text{ grados de libertad y probabilidad } p$$

$$V_1 = n - 1 = 30 - 1 = 29; V_2 = (n - 1)(k - 1) = (30 - 1)(20 - 1) = 551$$

$$\alpha/2 = 0,025$$

$$F_{0,975;29;551} = \text{Valor Crítico inferior (2,5\%)} \\ F_{0,025;29;551} = \text{Valor Crítico superior (97,5\%)}$$

Tabla 7

Valores críticos la fiabilidad

Función Excel	Valor Crítico
=+DISTR.F.INV(0,025;29;551)	1,603418431744420
=+DISTR.F.INV(0,975;29;551)	0,547956532238049

Para una prueba de hipótesis con $\alpha=0.05$ (95% de confianza), los valores críticos de F son:

Rechazamos H_0 si $F < 0,547956532238049$ o $F > 1,603418431744420$

$$\text{Límite inferior} = 1 - \left(\frac{(1-0,977138804015714)(1,603418431744420)}{0,977138804015714} \right) = 1 - 0,0366560630129257$$

$$\text{Límite superior} = 1 - \left(\frac{(1-0,977138804015714)(0,547956532238049)}{0,977138804015714} \right) = 1 - 0,0668959321704030$$

$$IC_{95\%}[0,9331040678295970; 0,9633439369870740]$$

Esto significa que, con un 95% de confianza, el verdadero valor del coeficiente Alfa de Cronbach se encuentra entre 0,9331040678295970 y 0,9633439369870740, lo que confirma la robusta fiabilidad del instrumento.

DISCUSIÓN

Los resultados obtenidos en la validación del instrumento mediante el coeficiente V de Aiken reflejan un alto nivel de consenso entre los expertos, con un valor global de $V = 0,896$, lo que supera ampliamente el criterio moderado ($V_0 = 0,70$) recomendado por Charter (2003) y se acerca al criterio estricto ($V_0 = 0,80$). Esto confirma la validez de contenido del instrumento en cuanto a suficiencia, relevancia y claridad de los ítems, tal como señalan Galicia et al. (2017) y Rodríguez Medina et al. (2021). La consistencia en las evaluaciones de los cinco jueces respalda la adecuación del instrumento para medir las variables de interés en el contexto de la ingeniería, reduciendo la subjetividad mediante un enfoque cuantitativo robusto (Penfield & Giacobbi, 2004; Merino Soto & Livia Segovia, 2009).

La novedad de este estudio reside en la implementación de una ruta metodológica secuencial que optimiza el juicio de expertos mediante herramientas virtuales (Molina et al., 2016) y trasciende el uso convencional del coeficiente V de Aiken al incorporar un enfoque probabilístico con intervalos de confianza de Wilson, permitiendo una toma de decisiones más robusta y precisa (Rodríguez, 2015; Robles, 2018). Esta propuesta integral se complementa con el análisis de la consistencia interna mediante el Alfa de Cronbach y sus intervalos de confianza (Roco et al., 2024; Caycho, 2017), junto con una rigurosa evaluación de la normalidad (Kolmogorov-Smirnov y Shapiro-Wilk) y asociación de variables (Pearson, Spearman y Chi-cuadrado), asegurando que la elección de las pruebas estadísticas finales sea metodológicamente coherente con la naturaleza de los datos.

La fiabilidad interna medida through el alfa de Cronbach resultó en un valor de $\alpha = 0,977$, con un intervalo de confianza del 95% entre 0,933 y 0,963, calculado mediante el método de Feldt (1965); este valor no solo supera el umbral de 0,70 establecido por Nunnally y Bernstein (1994), sino que se sitúa en el rango de excelencia, lo que indica una consistencia interna muy alta, no obstante, como advierten Oviedo y Campo (2005) y Cascaes et al. (2015), valores excesivamente altos pueden sugerir redundancia en los ítems, por lo que en futuras aplicaciones podría

considerarse una revisión para optimizar la longitud del instrumento sin comprometer su confiabilidad.

La metodología propuesta integra validación cualitativa, cuantificación del consenso, evaluación de la fiabilidad y verificación de supuestos, que ofrece un marco replicable y riguroso para la validación de instrumentos en ingeniería, tal como destacan Lages Ruiz y Martínez Trujillo (2023) y Díaz Vega y Terrazas Rodríguez (2023); este enfoque sistemático no solo mejora la credibilidad de los datos, sino que también facilita la adaptación del instrumento a distintos contextos de investigación.

CONCLUSIONES

La implementación de la ruta metodológica propuesta, que integra de forma secuencial la validación por juicio de expertos mediante el coeficiente V de Aiken y la evaluación de la fiabilidad interna a través del Alfa de Cronbach, demuestra ser un procedimiento robusto y eficaz para garantizar la validez y confiabilidad de instrumentos de medición en el ámbito de la investigación en ingeniería. Los resultados obtenidos, con un V de Aiken global de 0,896 y un Alfa de Cronbach de 0,977 (IC95%: 0,933 - 0,963), confirman que el instrumento desarrollado posee una excelente validez de contenido y una consistencia interna muy alta, superando los criterios establecidos en la literatura especializada.

Este estudio evidencia la importancia de emplear métodos mixtos (cualitativos y cuantitativos) en la validación de instrumentos, tal como recomiendan autores como Díaz y Terrazas (2023) y Cascaes et al. (2015). El uso del coeficiente V de Aiken, complementado con intervalos de confianza mediante el método de Wilson, permitió cuantificar el consenso entre expertos de manera precisa y reducir la subjetividad inherente al juicio de expertos. Asimismo, el cálculo del Alfa de Cronbach junto con sus intervalos de confianza mediante el método de Feldt (1965) proporcionó una estimación más exacta y confiable de la consistencia interna del instrumento.

La aplicabilidad de esta ruta metodológica se extiende más allá del contexto inmediato de este estudio, ofreciendo un marco replicable, ordenado y adaptable para la validación de instrumentos en diversas áreas de la ingeniería y otras disciplinas. La incorporación de pruebas de normalidad y la subsequent selección justificada de pruebas estadísticas (paramétricas o no

paramétricas) aseguran que los análisis inferenciales se basen en supuestos válidos, mejorando así la credibilidad y solidez científica de los hallazgos futuros.

REFERENCIAS BIBLIOGRÁFICAS

- Apaza Zúñiga, E., Cazorla Chambi, S., Condori Carbajal, C., Arpasi Meléndez, F. R., Tumi Figueroa, I., Yana Viveros, W., & Quispe Coaquira, J. E. (2022). La Correlación de Pearson o de Spearman en caracteres físicos y textiles de la fibra de alpacas. *Revista de Investigación Veterinaria*, 33(3), e22908. doi:10.15381/rivep.v33i3.22908
- Oviedo, H. C., & Campo Arias, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(2), 572-580.
- Robles Pastor, B. F. (2018). Índice de validez de contenido: Coeficiente V de Aiken. *Pueblo continente*, 29(1), 193-197.
- Aiken, L. R. (1980). Content Validity and Reliability of Single Items or Questionnaires. *Educational and Psychological Measurement*, 44(4), 955–959. doi:10.1177/001316448004000419
- Cascaes da Silva, F., Gonçalves, E., Valdivia Arancibia, B. A., Graziele Bento, G., da Silva Castro, T. L., Soleman Hernandez, S. S., & da Silva, R. (2015). Estimadores de consistencia interna en las investigaciones en salud: el uso del coeficiente alfa. *Revista Peruana de Medicina Experimental y Salud Pública*, 32(1), 129-138. Obtenido de http://www.scielo.org.pe/scielo.php?pid=S1726-46342015000100019&script=sci_abstract
- Caycho Rodríguez, T. (2017). Confidence Intervals for Cronbach's alpha coefficient: contributions to pediatric research. *Acta Pedátrica de México*, 4, 291-292. doi:10.18233/APM38No4pp291-2941440. Obtenido de <https://www.redalyc.org/journal/4236/423659205011/html/>
- Cerda L., J., & Villaroel del P., L. (2007). Interpretación del test de Chi-cuadrado (X2) en investigación pediátrica. *Revista chilena de pediatría*, 414-417. doi:10.4067/S0370-41062007000400010.
- Charter, R. A. (2003). A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *J Gen Psychol*, 130(3), 290-304. doi:10.1080/00221300309601160

- Díaz Vega, G., & Terrazas Rodríguez, L. D. (2023). Construcción y validación de un instrumento para evaluación de crisis familiares. *Acta Médica Peruana*, 40(4), 314-319. doi:10.35663/amp.2023.404.2696
- Dominguez Lara, S. (2018). Magnitud del efecto para pruebas de normalidad en investigación en salud. *Revista en Investigación Médica*, 7(27), 92-93. doi:10.22201/facmed.20075057e.2018.27.1776
- Flores Ruiz, E., Miranda Novales, M. G., & Villasís Keever, M. Á. (2017). El protocolo de investigación VI: cómo elegir la prueba Mestadística adecuada. *Estadística inferencial. Revista alergia México*, 64(3), 364-370. Obtenido de https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2448-91902017000300364
- Galicia Alarcón, L. A., Balderrama Trápaga, J. A., & Edel Navarro, R. (2017). Validez de contenido por juicio de expertos: propuesta de una herramienta virtual. *Apertura* (Guadalajara, Jal.), 9(2), 42-53. doi:10.32870/Ap.v9n2.993
- Lages Ruiz, J., & Martínez Trujillo, N. (2023). Validación de instrumentos para estudio de referenciación en enfermería oftalmológica. *Revista Cubana de Enfermería*. 2023, 39, e5652. Obtenido de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-03192023000100003
- Ledesma, R., Molina Ibañez, G., & Valero Mora, P. (2022). Análisis de consistencia interna mediante Alfa de Cronbach: un programa basado en gráficos dinámicos. *Revista Psico-USF*, 7(2), 143-152. Obtenido de <https://www.scielo.br/j/pusf/a/psJ44DfZRngHRYRp9C9PTKg/?format=pdf&lang=es>
- López Fernández, R., Avello Martínez, R., Palmero Urquiza, D. E., Sánchez Gálvez, S., & Quintana Álvarez, M. (2019). Validación de instrumentos como garantía de la credibilidad en las investigaciones científicas. *Revista Cubana de Medicina Militar*, 48(2), 441-450. Obtenido de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0138-65572019000500011
- Merino Soto, C., & Livia Segovia, J. (2009). Intervalos de confianza asimétricos para el índice la validez de contenido: Un programa Visual Basic para la V de Aiken. *Anales de psicología*, 25(1), 169-171. Obtenido de <https://revistas.um.es/analesps/article/view/71631/69111>

- Molina Achury, N., Forero Nieto, S., Ramos Caballero, D. M., Benavides Piracón, J., & Quintana Cortés, M. A. (2016). Diseño y validación de un instrumento de evaluación de condiciones de salud y trabajo de los fisioterapeutas en Colombia. *Revista de la Facultad de Medicina Humana*, 64, S59-67. doi: 10.15446/revfacmed.v64n3Supl.51655
- Nunnally, J. C., & Bernstein, I. B. (1994). *Psychometric Theory*. McGraw-Hill Companies, Incorporated. doi:10.1177/014662169501900308
- Penfield, R. D., & Giacobbi, P. R. (2004). Applying a Score Confidence Interval to Aiken's Item Content-Relevance Index. *Measurement in Physical Education and Exercise Science*, 8(4), 213-225. doi:10.1207/s15327841mpee0804_3
- Roco Videla, Á., Vladimir Flores, S., Olguin, M., & Maureira Carsalade, N. (2024). Alpha de Cronbach y su intervalo de confianza. *Nutrición Hospitalaria*, 41(1), 270-271. doi:10.20960/nh.04961
- Roco Videla, Á., Landabur Ayala, R., Maureira Carsalade, N., & Olguin Barraza, M. (2023). ¿Cómo determinar efectivamente si una serie de datos sigue una distribución normal cuando el tamaño muestral es pequeño? *Nutrición hospitalaria*, 40(1), 234-235. doi:10.20960/nh.04519
- Rodríguez Cordón, J. (2015). Analizando la V de Aiken Usando el Método Score con Hojas de Cálculo. 1-7. Obtenido de <https://www.researchgate.net/publication/277556053>
- Rodríguez Medina, M. A., Poblano Ojinaga, E. R., Alvarado Tarango, L., González Torres, A., & Rodríguez Borbón, M. I. (2021). Validation by Expert Judgment of an Evaluation Instrument for Evidence of Conceptual Learning. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 12(22), e240. doi:10.23913/ride.v11i22.960
- S. Feldt, L. (1965). The Approximate Sampling Distribution of Kuder-Richardson Reliability Coefficient Twenty. *Psychometrika*, 30(3), 589-600. doi:10.1007/BF02289499
- Sánchez Solis, Y., Raqui Ramirez, C. E., Huaroc Ponce, E. J., & Huaroc Ponce, N. M. (2024). Importancia de Conocer la Normalidad de los Datos Utilizados en los Trabajos de Investigación por Tesistas. *Revista Tecnológica-Educativa Docentes*(17), 404-413. doi:10.37843/rted.v17i2.554
- Santos, P. A., Chuquisengo, E., & Vasquez, A. E. (2025). Diseño y validación de un instrumento para medir el uso de la herramienta Julius AI en estudiantes universitarios peruanos. *Revista Espacios*, 46(2), 204-212. doi:10.48082/espacios-a25v46n02p16

- Soler Cárdenas, S. F., & Soler Pons, L. (2012). Usos del coeficiente alfa de Cronbach en el análisis de instrumentos escritos. *Revista Médica Electrónica*, 34(1), 1-6. Obtenido de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1684-18242012000100001
- Toro, R., Peña Sarmiento, M., Avendaño, B. L., & Mejía Vélez, S. (2022). Análisis Empírico del Coeficiente Alfa de Cronbach según Opciones de Respuesta, Muestra y Observaciones Atípicas. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 2(63), 17-30. doi:10.21865/RIDEP63.2.02
- Vincent Cicchetti, D., & Sparrow, S. A. (1981). Developing Criteria for Establishing Interrater Reliability of Specific Items: Applications to Assessment of Adaptive Behavior. *American Journal of Mental Deficiency*, 86(2), 127-37.