



Desarrollo y validación preliminar de un cuestionario de razonamiento estadístico en estudiantes de ciencias del deporte

Development and preliminary validation of a statistical reasoning question among sports science students

Autores

María Luisa Zambrano-Rojas¹
 Eliana Patricia Cuellar-Carvajal²
 Jorge Enrique Correa-Bautista³

^{1,2,3} Universidad de Cundinamarca
 (Colombia)

Autor de correspondencia:
 Jorge Enrique Correa-Bautista
jorgeecorrea@ucundinamarca.edu.co

Recibido: 15-04-26
 Aceptado: 26-05-26

Cómo citar en APA

Zambrano Rojas, M. L., Cuellar-Carvajal, E. P., & Correa-Bautista, J. E. (2026). Desarrollo y validación preliminar de un cuestionario de razonamiento estadístico en estudiantes de ciencias del deporte. *Retos*, 81, 524-537. <https://doi.org/10.47197/retos.v81.119251>

Resumen

Introducción: El razonamiento estadístico es crítico para la toma de decisiones basada en la evidencia en las ciencias del deporte. No obstante, se carece de instrumentos psicométricamente válidos para esta población.

Objetivo: Desarrollar y aportar evidencia de validez de contenido, estructura interna y confiabilidad de un cuestionario para evaluar el razonamiento estadístico dirigido a estudiantes de las ciencias del deporte.

Métodos: Se realizó un estudio de validación psicométrica, que integró un consenso Delphi tri-tápico (n=13, expertos) para la generación de ítems. La validez de contenido se cuantificó través del índice de validez de contenido por ítem (IVC-I), índice de validez de contenido general (IVC-G) y Kappa modificado (K*). Una muestra de (n=99) estudiantes (21,9± 2,04 años) completó el instrumento. Se ejecutó un análisis factorial exploratorio con estimación robusta de mínimos cuadrados ponderados diagonalmente. La confiabilidad se determinó través del alfa de Cronbach y coeficiente de correlación intraclase (CCI). **Resultados:** La validez de contenido fue óptima (IVC-G: 0,93; K*>0,84). El análisis factorial exploratorio indicó una estructura bifactorial parsimonioso de 10 reactivos, que explicó el 62,4% de la varianza total. La consistencia interna fue aceptable ($\alpha = 0,746$), y la estabilidad temporal adecuada (ICC= 0,820; IC95%:0,672-0,813). **Conclusiones:** El cuestionario REI-Sports evidencia propiedades psicométricas de validez y confiabilidad sólidas para evaluar el razonamiento estadístico en estudiantes de las ciencias del deporte. Se requieren estudios adicionales para confirmar su estructura factorial y su aplicabilidad en otros contextos.

Palabras clave

Encuestas y cuestionarios; método Delphi; psicometría; razonamiento estadístico; validación.

Abstract

Introduction: Statistical reasoning is critical for evidence-based decision-making in the sports sciences. However, there is a lack of psychometrically valid instruments for this population.

Objective: To develop and provide evidence of content validity, internal structure, and reliability for a questionnaire designed to assess statistical reasoning among sports science students.

Methods: A psychometric validation study was conducted, incorporating a three stage Delphi consensus (n=13, experts) for item generation. Content validity was quantified using the item-level content validity index (IVC-I), the general content validity index (IVC-G), and modified Kappa (K*). A sample of (n=99) students (21.9 ± 2.04 years) completed the instrument. An exploratory factor analysis was performed using a robust estimate of diagonally weighted least squares. Reliability was determined using Cronbach's alpha and the intraclass correlation coefficient (ICC).

Results: Content validity was excellent (CVI-G: 0.93; K* > 0.84). Exploratory factor analysis indicated a parsimonious two-factor structure comprising 10 items, which explained 62.4% of the total variance. Internal consistency was acceptable ($\alpha = 0.746$), and test-retest reliability was adequate (ICC = 0.820; 95% CI: 0.672–0.813). **Conclusions:** The REI-Sports questionnaire demonstrates solid psychometric properties of validity and reliability for assessing statistical reasoning in sports science students. Further studies are needed to confirm its factor structure and applicability in other contexts.

Keywords

Delphi method; statistical reasoning; surveys and questionnaires; psychometrics; validation.

Introducción

El razonamiento estadístico (RE) se constituye en una competencia fundamental para interpretar datos, evaluar evidencia y tomar decisiones informadas en las disciplinas científicas (Friedrich et al., 2022). El RE se entiende como la capacidad de coordinar la interpretación descriptiva, procedimientos e inferencias para otorgar sentido a la información estadística (Garfield, 2002; Garfield & Gal, 1999). Este implica reconocer patrones de variabilidad, modelar incertidumbres, establecer relaciones entre variables y evaluar conclusiones derivados de un conjunto de datos. Actualmente, existe una tendencia por incluir la formación en estadística en los currículos de pregrado, en donde, el RE, se utiliza como un objetivo de aprendizaje (Sabbag et al., 2025). Sin embargo, la presencia de contenidos estadísticos dentro de las rutas de aprendizaje de los programas de pregrado no avala la comprensión conceptual, ni la aplicación de conocimientos estadísticos en contextos reales (Garfield & Ben-Zvi, 2008). Diversos estudios han confirmado estas limitaciones en el aprendizaje y la interpretación de resultados estadísticos en estudiantes universitarios (Hussain et al., 2019; McGrath, 2014; Zieffler et al., 2008; Garfield & Ahlgren, 1988). En las ciencias del deporte, esta competencia adquiere una relevancia particular, dado que la práctica profesional se sustenta en la interpretación de evidencia cuantitativa derivada de diseños experimentales, observacionales y longitudinales reportados en la literatura científica (Corbin, 2021).

El modelo de comprensión del RE propuesto por Garfield (2003) describe cinco niveles: (1) Razonamiento idiosincrásico, caracterizado por el uso fragmentado de términos estadísticos, sin comprenderlos completamente, a menudo aplicados de forma incorrecta. (2) Razonamiento verbal. Comprensión superficial sin aplicación funcional. (3) Razonamiento transicional, que integra parcialmente una o dos dimensiones de un proceso estadístico. (4) Razonamiento procedimental, centrado en la ejecución correctamente de interpretación descriptiva y procedimientos estadísticos, sin la comprensión profunda del proceso. (5). Razonamiento de procesos integrados, en que el estudiante coordina reglas, interpretación descriptiva y evidencia para generar inferencias adecuadas. Esta taxonomía, ha sido un referente en la investigación en educación estadística y ha evidenciado que la mayoría de los estudiantes de pregrado no alcanzan el nivel integrativo dentro del RE. En un reciente estudio a través de autorreportes en 171 profesores universitarios de estadística en Estados Unidos, identificaron las temáticas que generan mayor dificultad de aprendizaje en estadística: (a) aleatoriedad, (b) teoría de la probabilidad, (c) la ley de los grandes números, (d) tendencia central, (e) valores p , (f) distribuciones de muestreo y (g) variación y desviación estándar. Al igual, que la comprensión conceptual (en comparación con la procedimental); la comprensión de relaciones entre interpretación descriptiva, el "conectar y usar" para la resolución de problemas; la resolución de problemas aplicados; y la identificación de la estructura lógica de los problemas (Silva & Sarnecka, 2025). Igualmente, Delmas et al. (2007) reportaron que al menos el 60 % de los estudiantes universitarios que completaron un curso de formación en estadística, aún no comprendían la causalidad, los valores p y los intervalos de confianza. Estas dificultades en el aprendizaje estadístico se asocian con una baja de capacidad de abstracción y una poca integración de múltiples representaciones de datos (Begum et al., 2025; Conway et al., 2019).

Para evaluar el RE se han desarrollado instrumentos como ARTIST (Assessment Resource Tool for Improving Statistical Thinking), la cual es una herramienta diseñada para medir el pensamiento estadístico básico en estudiantes de educación superior y secundaria (Garfield & DelMas, 2010). La escala CAOS (Comprehensive Assessment of Outcomes in a first Statistics course) diseñada para evaluar el cambio en la comprensión de estudiantes luego de cursos básicos en estadística (Chance et al., 2004; Delmas et al., 2007). La prueba de evaluación de razonamiento estadístico (SRA) que evalúa la capacidad de razonar en estadística y probabilidad (Garfield, 2003). Así mismo, se ha medido el RE en poblaciones específicas, como estudiantes de secundaria (Chan & Ismail, 2014), y profesionales de las ciencias de la salud (Gaviria-Bedoya et al., 2025). Otros instrumentos, han sido utilizados para evaluar el RE como ejercicios prácticos en textos y por computadora (Schwob et al., 2025), pruebas de razonamiento basado en casos (Perner, 2008), y pruebas de lápiz y papel (Marco, 2023; He & Lao, 2018). A pesar del creciente interés por la evaluación del RE en distintos niveles educativos y contextos, persisten limitaciones relacionadas con la disponibilidad de instrumentos que integren simultáneamente robustez psicométrica, especificidad disciplinar y pertinencia sociocultural. Las ciencias del deporte, exhibe una notable cantidad de instrumentos de auto reporte en áreas como la psicología del deporte (Cantón-Chirivella et al., 2025; Ceballos-Bernal & Correa-Bautista, 2025), el bienestar percibido (Sepdanius et al., 2025) y la gestión educa-

tiva (Carranza-Bautista & Giakoni-Ramírez, 2025). En el contexto de la didáctica del deporte se evidencia una escasez de herramientas validadas que examinen la capacidad del estudiantado para analizar e interpretar indicadores cuantitativos fisiológicos, biomecánicos y de rendimiento. En Colombia, no se dispone actualmente de instrumentos validados que evalúen el RE desde una perspectiva integrativa en estudiantes universitarios de las ciencias del deporte configurando un vacío metodológico que restringe la investigación educativa en la enseñanza de la estadística y el desarrollo estrategias pedagógicas orientadas al fortalecimiento de las competencias analítico-inferenciales, en educación superior.

Frente a ello, el presente estudio tuvo como objetivo desarrollar y aportar evidencia de validez de contenido, estructura interna y confiabilidad de un cuestionario para evaluar el razonamiento estadístico integral dirigido a estudiantes de las ciencias del deporte.

Método

Este estudio evaluó las propiedades psicométricas del cuestionario REI-Sports en Colombia con un diseño transversal de validación psicométrica, utilizando previamente la metodología Delphi de panel único (Galanis, 2018). Se siguió los estándares del consenso internacional en el desarrollo de instrumentos de medición (COSMIN) (Gagnier et al., 2021) de manera complementaria se tuvo en cuenta los criterios para realizar reportes de estudios Delphi en ciencias sociales y salud, DELPHYSTAR (Nasa et al., 2021; Niederberger et al., 2024). Por un lado, los criterios COSMIN orientaron el diseño global del estudio y DELPHISTAR se aplicó específicamente en la fase de validación de contenido mediante el método Delphi. El Comité de Ética, Bioética e Integridad en Investigación - CEBII y la Dirección de Protección de Datos Personales de la Universidad de Cundinamarca aprobaron la realización del estudio (PI-No 661/2025). Todos los estudiantes firmaron el consentimiento informado y declararon su voluntad de participación. El proceso de diseño y validación del instrumento REI-Sports se muestra en la figura No 1.

Figura 1. Esquema del Diseño y Validación del Instrumento REI-Sports.



Fuente: Elaboración propia realizada en NotebookLM

Generación de dimensiones y elementos

El diseño del cuestionario REI-Sports se inicia a partir de la selección de los dominios que se pretenden medir. Un dominio se refiere al concepto, atributo o comportamiento no observado que constituye el objeto del estudio (Hinkin, 1995). En este caso se propusieron dos dimensiones: a) Alfabetización y comprensión estadística; b) Interpretación y razonamiento estadístico integrado, resultado de la revisión de literatura disponible entre el año 1999 y el 2026. Con base en esta fundamentación, se diseñaron 15 preguntas (reactivos) tipo opción múltiple, con cuatro opciones de respuesta y única respuesta correcta (Heemskerk et al., 2008). Se definieron seis temas (TESTLET) comunes relacionados con las ciencias del deporte: 1) evaluación fisiológica del rendimiento aeróbico, 2) control de cargas en el entrenamiento deportivo, 3) la relación entre variables deportivas, 4) probabilidad y prevención de lesiones deportivas, 5) análisis comparativo e interpretación estadística y 6) razonamiento probabilístico. Una

vez terminado el borrador final del instrumento REI-Sports, se procedió a realizar una validación cognitiva de este, a través de entrevistas semiestructuradas (Beatty & Willis, 2007; Kamp et al., 2018; Shoozan & Mohamad, 2024) con una muestra de seis estudiantes universitarios (4 hombres; 2 mujeres) con el fin de auditar la validez del contenido y el nivel de interpretabilidad de las preguntas. Para minimizar el sesgo de respuesta y asegurar la validez aparente, se implementó la técnica de parafraseo inverso (Evans et al., 2010), donde cada participante expresó cada ítem en su propio léxico, manifestando posibles ambigüedades. Así mismo, se aplicó el protocolo de indagación por sondeo (Probing) (DeJonckheere & Vaughn, 2019) a través de una escala ordinal de entendimiento de cuatro niveles (donde: 1) No se entiende y 4) buen entendimiento). Aquellos ítems que obtuvieron calificaciones ≤ 3 fueron ajustados. Con el fin de minimizar el error relacionado con la legibilidad gramatical del instrumento, se calculó el índice de Legibilidad de Szigriszt-Pazos, como adaptación del algoritmo Flesch-Kincaid para el idioma español (Szigriszt, 2001) para identificar el nivel de dificultad de lectura de los 15 reactivos del instrumento. Se obtuvo una puntuación global de 64,2/100, lo que permite clasificar el nivel de dificultad de lectura como estándar, de acuerdo con la escala propuesta por Flesch-Kincaid, donde puntajes de 90-100 es fácil, 60-70 es estándar (accesible para la mayoría), y 0-30 es confuso (Kincaid et al., 1975). Este resultado confirma la congruencia entre la complejidad sintáctica y el perfil académico de la población objeto de la aplicación.

Estudio Delphi

Se conformó un panel multidisciplinario de expertos por muestreo no probabilístico por conveniencia, entre el 2 de febrero y el 15 de marzo del 2026. La identificación de los mejores perfiles de expertos se efectuó a través de registros institucionales y redes de contacto. Se seleccionó un total de (n=13) expertos (35.5%, fueron mujeres), con una media de edad años de $40,9 \pm 9,41$, con $15,6 (\pm 8,6)$ años de experiencia profesional y docente, el 46,2% con formación en ciencias del deporte, un 38,5% con formación doctoral, y el 23% con formación en matemáticas. El tamaño de este panel de expertos se ajusta a las recomendaciones para los estudios de validación de contenido (n = 10-15) (Kellerhuis et al., 2025).

Ronda 1

Se conformó un comité organizador de tres docentes para gestionar la primera ronda. Los expertos recibieron por correo electrónico el instrumento preliminar y una encuesta de calificación de cuatro niveles para cada reactivo: (1) no es relevante, (2) requiere una revisión mayor, (3) requiere una revisión mínima; (4) muy relevante y no requiere revisión. Al igual, se permitió presentar observaciones de mejora para cada reactivo de acuerdo con las respuestas reportadas. Las respuestas se ordenaron y se tabularon para determinar el consenso general entre los jueces expertos. El valor del consenso general de la primera ronda para los 15 reactivos fue de 86,6% en lo que se refiere a que son relevantes y no requieren ajustes superando el criterio de aceptación sugerido de $\geq 60\%$ (Niederberger et al., 2024).

Ronda 2

En la segunda ronda, los panelistas fueron convocados a una sesión virtual sincrónica, donde se socializaron los resultados consolidados de la sesión previa y se presentó la versión ajustada del instrumento. La sesión se centró en evaluar los criterios de claridad, relevancia y representatividad para cada ítem (Traynor & Christopherson, 2024). Para la cuantificación de las respuestas se utilizó una escala de tipo Likert de 1 a 4: donde 1) "no es representativo"; 2) "poco representativo"; 3) "representativo", 4) "muy representativo". Este procedimiento permitió obtener las métricas necesarias para el cálculo del índice de validez de contenido y la depuración del instrumento.

Ronda 3

Una tercera ronda fue convocada para la socialización del consolidado final de los resultados y la discusión técnica de factores críticos identificados durante el proceso de validación. Los 13 panelistas expertos evaluaron de manera independiente el instrumento definitivo, ratificando su calidad psicométrica en los criterios de claridad, relevancia, representatividad. La independencia en el juicio, sumado al análisis del consenso de estos, permitió afinar los contenidos del instrumento. Luego, de la evaluación de los panelistas expertos, el instrumento REI-Sports fue digitalizado mediante la plataforma Microsoft Forms, y distribuido por correo electrónico y WhatsApp para facilitar la accesibilidad del registro. La recolección de los datos se llevó a cabo en entornos controlados (aulas de clases), con la supervisión de

un auxiliar de investigación entrenado. Una vez finalizada la recolección, los datos crudos fueron exportados y consolidados en una matriz en Excel en formato CSV (codificación UTF-8), lo que aseguró la interoperabilidad de la información para el procesamiento estadístico de los datos.

Validación del instrumento

Una vez recibidas y tabuladas las respuestas de los jueces expertos se procedió a calcular el acuerdo inter-jueces. Para cuantificar el consenso, se calculó el índice de validez de contenido por ítem (IVC-I) y el índice de validez de contenido general de la escala (IVC-G). Adicionalmente, para fortalecer el rigor estadístico y corregir el posible acuerdo por azar, se estimó la probabilidad de acuerdo por el azar (p_c) y el coeficiente de kappa modificado (k^*) (Polit & Beck, 2006). Se calcularon medidas de tendencia central y dispersión en las variables continuas; y se desarrollaron tablas de frecuencias relativas y absolutas para variables categóricas. Se utilizó el módulo de fiabilidad del software Amazing Statistics Program (JASP) de Jeffreys versión 0.19.0.3 (JASP Team., 2024) para estimar el coeficiente de α -Cronbach, el coeficiente de correlación intraclase (CCI); al igual que el módulo de factor para el análisis factorial exploratorio (AFE). Para calcular los valores de acuerdo Inter jueces se utilizó el programa Excel® Versión 18 (Microsoft Corporation. (2021)

Selección de participantes

A pesar de la ausencia de un consenso definitivo frente al cálculo del tamaño de muestral óptimo para estudios de fiabilidad (Borg et al., 2022), el presente estudio utilizó una muestra por conveniencia de ($n = 99$) de profesionales en formación pertenecientes a los últimos semestres (tercer y cuarto año) del programa de ciencias del deporte. El muestreo se realizó en una universidad pública con sede en Soacha-Colombia.

La muestra compuesta por adultos jóvenes, entre los 19 a 30 años con una media de edad del $21,94 \pm 2,04$ años, con una representación de género femenino del 23,6%. Los criterios de inclusión contemplaron ser estudiantes activos de las asignaturas de fundamentos de estadística, bioestadística aplicada a las ciencias del deporte y la actividad física y a la asignatura de tecnología aplicada al deporte. El proceso de captación se llevó a cabo mediante invitación directa y personalizada. Se empleó una estrategia de comunicación multicanal que incluyó el envío de correos institucionales, difusión por redes sociales y mensajería instantánea por WhatsApp durante el primer semestre del 2026. Los participantes recibieron información detallada sobre los objetivos del estudio y el carácter voluntario de su colaboración. Posteriormente, se obtuvo el consentimiento informado por escrito, seguido de un cuestionario socio-demográfico. Tras la aceptación, el instrumento REI-Sports fue digitalizado mediante la plataforma Microsoft Forms en una encuesta única autoadministrada, la cual fue compartida en un enlace directo. Los participantes completaron el instrumento REI-Sports bajo supervisión directa para evitar distracciones y asegurar condiciones de uniformidad en la recolección de los datos. A los 15 días se replicó el mismo protocolo para determinar el test-retest.

Análisis de la estructura interna

La estructura interna del instrumento se evaluó mediante un análisis factorial exploratorio (AFE). Inicialmente, la viabilidad de la factorización se verificó a través del índice de KMO, y prueba de esfericidad de Bartlett. Para justificar la selección de los análisis estadísticos posteriores, se calculó el coeficiente de Mardia. Este análisis confirmó la ausencia de normalidad multivariante en los datos.

Consecuentemente, se implementó el método de mínimos cuadrados ponderados diagonalmente robustos (RDWLS) con rotación oblicua Oblimin (Li, 2016a). Durante este proceso, se aplicó un criterio de reducción de reactivos. Este procedimiento aseguró la retención exclusiva de ítems parsimoniosos, funcionales y con una sólida consistencia interna (Fan, 1998).

Análisis de confiabilidad y estabilidad

La confiabilidad se estimó mediante la consistencia interna, utilizando el coeficiente de alfa de Cronbach (Dunn et al., 2014), además de la estabilidad temporal mediante un procedimiento de prueba test-retest (intervalo de 15 días) analizando el coeficiente de correlación intraclase (CCI).

Resultados

El diseño de la escala REI-Sports, se rigió por un enfoque nomológico sustentado en una revisión exhaustiva de la literatura científica. El constructo se operacionalizó en dos dominios del RE: a) Alfabetización y comprensión estadística; b) Interpretación y razonamiento estadístico integrado. Estos dominios se distribuyeron en 15 reactivos, los cuales abordaron seis temas de mayor dentro de las ciencias del deporte.

Por otra parte, el juicio de expertos mediante la metodología Delphi se llevó a cabo entre febrero a abril del 2026. El panel evaluador quedó constituido por 13 expertos (35,5% mujeres) con formación en distintas disciplinas científicas. Los jueces fueron escogidos bajo los criterios estrictos idoneidad académica y trayectoria profesional. El (38,5%) de la muestra contaba con formación doctoral, con una experiencia profesional y docente media de 15,6 ($\pm 8,33$) años. Las características sociodemográficas y profesionales del panel se presentan en la Tabla No. 1.

Tabla 1. Caracterización sociodemográfica y profesional de los panelistas (n= 13)

VARIABLES SOCIODEMOGRÁFICAS	n	%
Sexo		
Masculino	8	61,5
Femenino	5	35,5
Edad		
≤ 30	2	15,3
30-45	8	61,5
> 45	3	23,2
Promedio \pm SD	40,9 ($\pm 9,41$)	
Experiencia profesional/docente	15,6 ($\pm 8,33$)	
Área de desempeño		
Ciencias del deporte	6	46,2
Ciencias de la salud	3	23,1
Ciencias Naturales y Exactas	2	15,3
Ciencias económicas y administrativas	1	7,7
Ciencias Humanas y sociales	1	7,7
Nivel de Formación		
Especialización	3	23
Maestría	5	38,5
Doctorado	5	38,5

Fuente: Elaboración propia

Todos los panelistas participaron en la retroalimentación y el consenso del instrumento en la ronda 1. Basado en la retroalimentación de los panelistas, 5 de los ítems (33,3%) fueron ajustados. Por ejemplo, en la pregunta No 3, los panelistas sugirieron reemplazar la "media" y "desviación estándar" por "mediana" y "rango intercuartílico" en el enunciado. El consenso general de la primera ronda fue de 86,6%. Ver Tabla No. 2.

Tabla 2. Acuerdos de la primera ronda de expertos (n = 13)

Ítems	Consenso (%) *	Observaciones	Ajustes
1	Relevante y no requiere ajustes (92%) Requiere una mínima revisión (8%)	Es una temática relevante de las ciencias del deporte	No se realiza ajuste
2	Relevante y no requiere ajustes (83,3%) Requiere una mínima revisión (16,7%)	La diferencia indica asimetría positiva	No se realiza ajuste
3	Relevante y no requiere ajustes (75%) Requiere una mínima revisión (12,5%) Requiere una revisión mayor (12,5%)	Se sugiere reemplazar la "media" y "desviación estándar" por "mediana" y "rango intercuartílico" en el enunciado	Se realiza el ajuste
4	Relevante y no requiere ajustes (66%) Requiere una mínima revisión (23%) No es relevante (11%)	Ajustar el enunciado	Se realiza el ajuste
5	Relevante y no requiere ajustes (83,3%) Requiere una mínima revisión (16,7%)	El ítem mide la sensibilidad de diferentes estadísticos (media vs. mediana y RIC)	No se realiza ajuste
6	Relevante y no requiere ajustes (83,3%) Requiere una mínima revisión (16,7%)	Error en el enunciado frente a la correlación lineal ($r = 0,68$) en la imagen ($r = -0,63$)	Se realiza el ajuste
7	Relevante y no requiere ajustes (92%) Requiere una mínima revisión (8%)	La opción seleccionada resume adecuadamente que r^2 expresa la fracción de variación de la potencia vinculada estadísticamente al porcentaje de grasa	No se realiza ajuste
8	Relevante y no requiere ajustes (92%) Requiere una mínima revisión (8%)	Se puede realizar asociaciones, pero no causalidad directa ni predicción exacta	No se realiza ajuste



9	Relevante y no requiere ajustes (75%) Requiere una mínima revisión (16,6%) Requiere una revisión mayor (8,3%)	Evalúa correctamente probabilidad condicional	No se realiza ajuste
10	Relevante y no requiere ajustes (92%) Requiere una mínima revisión (8%)	No se puede concluir sin tener la tasa de exposición de los no lesionados	No se realiza ajuste
11	Relevante y no requiere ajustes (92%) Requiere una mínima revisión (8%)	Se sugiere ampliar el enunciado	Se realiza ajuste
12	Relevante y no requiere ajustes (92%) Requiere una mínima revisión (8%)	Un tamaño del efecto alrededor de 0,45 se considera pequeño-moderado,	No se realiza ajuste
13	Relevante y no requiere ajustes (92%) Requiere una mínima revisión (8%)	Se solicita incluir un enunciado a la pregunta	Se realiza el ajuste
14	Relevante y no requiere ajustes (92%) Requiere una mínima revisión (8%)	El ítem evalúa correctamente la comparación global de distribuciones	No se realiza ajuste
15	Relevante y no requiere ajustes (92%) Requiere una mínima revisión (8%)	El ítem focaliza adecuadamente la noción de probabilidad	No se realiza ajuste

Nota. * Solo se presentan las respuestas a los enunciados que tuvieron consenso.

Fuente: Elaboración propia

En la segunda ronda, los panelistas evaluaron los criterios de claridad, relevancia y representatividad en los 15 ítems. En la tercera ronda se realizó una discusión técnica sobre los factores críticos relacionados con el proceso, y se socializa la versión final del instrumento. Los 13 panelistas expertos, ratificaron la calidad en la validez aparente del instrumento.

Las evaluaciones de los panelistas de la segunda ronda, permitió calcular el índice de validez de contenido global (IVC-G) con un valor obtenido de 0,93, valor que supera el umbral de excelencia de 0,90 establecido en la literatura (Polit & Beck, 2006). A nivel individual, los índices por ítems oscilaron entre 0,77 y 0,92, la claridad, pertinencia y representatividad, obtuvieron un (IVC-I = 0,92), pertinencia (IVC-I = 0,94), representatividad (IVC-I = 0,92), lo que indica que las respuestas de los panelistas fueron homogéneas. Para corregir la probabilidad de acuerdo por azar, se calculó el coeficiente Kappa modificado (k^*). Como se presenta en la tabla 3, todos los ítems alcanzaron valores por encima de 0,84, la probabilidad de acuerdo del azar (P_c) para todos los reactivos fue estadísticamente significativo ($p < 0,05$), lo que refuerza la validez interna del proceso.

Tabla 3. Validez de Contenido de los Ítems del Cuestionario REI-Sports mediante el Juicio de Expertos (n=13)

Ítems	Calidad				Relevancia				Representatividad			
	(3 o 4)	I-CVI	P_c	K^*	(3 o 4)	I-CVI	P_c	K^*	(3 o 4)	I-CVI	P_c	K^*
1	13	1	0	1	13	1	0	1	13	1	0	1
2	13	1	0	1	13	1	0	1	13	1	0	1
3	13	1	0	1	13	1	0	0,92	13	1	0	0,92
4	11	0,85	0	1	12	0,92	0	0,92	12	0,92	0	0,92
5	11	0,85	0	1	10	0,77	0,01	0,84	10	0,77	0	0,84
6	13	1	0	1	13	1	0	1	11	0,85	0	1
7	11	0,85	0	1	12	0,92	0	1	12	0,92	0	1
8	12	0,92	0	1	12	0,92	0	0,92	12	0,92	0	0,92
9	11	0,85	0	1	12	0,92	0	0,92	12	0,92	0	0,92
10	12	0,92	0	1	12	0,92	0,01	0,84	12	0,92	0	0,84
11	12	0,92	0	1	12	0,92	0	1	12	0,92	0	1
12	11	0,85	0	1	12	0,92	0	0,92	11	0,85	0	0,92
13	11	0,85	0	1	12	0,92	0	1	11	0,85	0	1
14	13	1	0	1	13	1	0,01	0,84	13	1	0	0,84
15	13	1	0	1	13	1	0	1	13	1	0	1
S-CVI		0,92				0,94				0,92		

Nota. IVC-I = Índice de Validez de Contenido por ítem; P_c = Probabilidad de acuerdo por azar; K^* = Coeficiente Kappa modificado; IVC-G = Promedio de los índices de validez de contenido de la escala. Criterios de interpretación de K^* : > 0,74 (Excelente), 0,60-0,74 (Bueno), 0,40-0,59 (Aceptable)

Estudio de Campo

Se reclutó una muestra piloto por conveniencia de (n= 99) profesionales en formación de las ciencias del deporte, entre los meses de abril a mayo del 2026, con una media de edad de 21,94 ± 2,04 años con un porcentaje de participación de mujeres del 23,6%. Los criterios de inclusión fueron amplios para poder contar con la mayor cantidad de participantes.



Validez factorial

Se realizó la prueba de Kaiser-Meyer-Olkin (KMO) para evaluar si los datos son aptos para el análisis factorial. En este caso, el KMO total fue de 0,705. Se utilizó, la prueba de contraste de Bartlett (X^2 370,885; gl 105,000; $p < ,001$). Se examinó el supuesto de normalidad multivariante mediante el contraste de Mardia. Los resultados indicaron una asimetría multivariante significativa (Mardia-Skewness= 73,71 ($p < ,001$), aunque la curtosis no mostró desviaciones críticas (Mardia-Kurtosis=261,02; $p=0,278$). La asimetría global confirma la ausencia de normalidad multivariante en la distribución de las respuestas del REI-Sports. En consecuencia, y siguiendo las recomendaciones de Li (2016b), se procedió con a realizar el AFE utilizando el estimador de Mínimos Cuadrados Ponderados Diagonalmente Robustos (MCPDR) con rotación oblicua Oblimin, el cual es resistente a las desviaciones de la normalidad y adecuado para los datos de naturaleza ordinal.

En la Tabla No 4, se presenta la carga factorial por cada uno de los respectivos ítems en dos factores. Las cargas factoriales de los 15 ítems, se encontraron entre (0,468-0,939).

Tabla 4. Cargas de los Factores

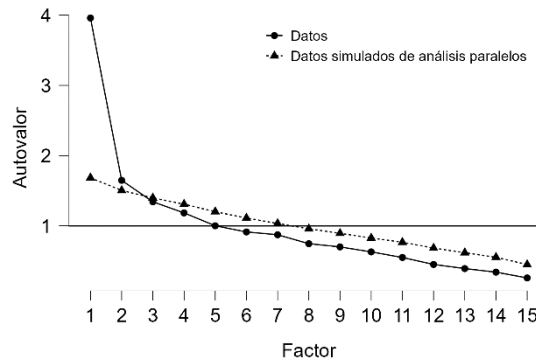
	Factor 1	Factor 2	Unicidad
R6	0,618		0,547
R11	0,558		0,725
R10	0,549		0,718
R13	0,545		0,685
R14	0,532		0,637
R15	0,504		0,665
R9	0,468		0,752
R2		0,591	0,670
R1		0,582	0,625
R4		0,555	0,645
R3			0,832
R5			0,593
R7			0,939
R8			0,865
R12			0,857

Nota. El método de rotación aplicado fue oblimin

Fuente: Elaboración propia

La arquitectura interna del instrumento REI-Sports se determinó mediante la convergencia de múltiples criterios de retención. El análisis paralelo y el criterio de Kaiser (autovalores > 1), coincidieron en sugerir una solución óptima de dos factores diferenciados, la cual explico el 62,4% de la varianza total acumulada, tal como se presenta en el gráfico de sedimentación (Figura 2). El factor 1 denominado de interpretación y razonamiento integral, agrupo 7 reactivos (R6, R9, R10, R11, R13, R14, R1,) con cargas factoriales elegibles que oscilaron entre 0,468 y 0,618, superando el umbral crítico establecido de $> 0,40$. Este dominio refleja la capacidad del futuro profesional para integrar la estadística en la toma de decisiones. Por su parte, el factor 2, titulado de alfabetización y comprensión estadística, quedo constituido por tres reactivos (R1, R2 y R4) con cargas factoriales entre 0,555 y 0,591, representando la competencia cognitiva básica necesaria para la lectura critica de la evidencia científica. Finalmente, los R3, R5, R7, R8 y R12 fueron excluidos del modelo definitivo, debido a la reportaron debilidad psicométrica en las cargas factoriales primarias inferiores a $\lambda 0,32$ y niveles de unicidad elevada ($> 0,80$).

Figura 2. Gráfica de sedimentación



Fuente: Elaboración propia

Esta decisión se fundamenta en la baja representatividad de estos ítems respecto al constructo evaluado. Luego, se realizó un reejecución del AFE con los 10 reactivos restantes resultando un modelo más parsimonioso y conceptualmente más coherente con el modelo de RE propuesto por Garfield & Ben-Zvi (2008). Se ratifica el modelo bifactorial del instrumento.

Análisis de confiabilidad y estabilidad temporal

La consistencia interna global del instrumento fue aceptable, reportando un alfa (α) de Cronbach general de 0,746. Por las dimensiones del instrumento, los coeficientes (α) oscilaron entre 0,701 y 0,750. Respecto a la estabilidad temporal (test-retest), el coeficiente de correlación intraclassa (ICC) fue de 0,748 (IC95%:0,672-0,813) lo que se considera una adecuada confiabilidad temporal. Estos valores confirman que el instrumento posee una precisión psicométrica adecuada y una buena reproducibilidad para la evaluación del RE en las ciencias del deporte.

Discusión

Este estudio tuvo como propósito el diseño, desarrollo y la validación de las propiedades psicométricas del cuestionario REI-Sports para evaluar el RE en una muestra de estudiantes de las ciencias del deporte entre los 19 a 30 años. Los hallazgos confirman que el instrumento es fiable y estable, al igual que posee una estructura interna coherente con los modelos teóricos contemporáneos de alfabetización y RE (Garfield & Ben-Zvi, 2008). El modelo estructural de dos factores: a) Alfabetización y comprensión estadística; b) Interpretación y razonamiento estadístico integral determinado mediante el uso del estimador robusto RDWLS, refuerza la validez del constructo. Esta multidimensionalidad es consistente con el marco teórico propuesto por Wild & Pfannkuch (1999), quienes sostienen que el RE estadístico no es un proceso lineal, sino un ciclo progresivo de conocimiento que mejora partiendo de la base del conocimiento previo, y se fortalece con la interpretación crítica situada. La estabilidad de los 10 reactivos sugiere que el diseño del REI-Sports captura con un nivel suficiente de precisión las competencias de RE básicas para el profesional de las ciencias del deporte. Un hallazgo adicional notorio fue la adecuada estabilidad temporal (0,748 [IC95%:0,672-0,813]). Este valor es superior al reportado en otras validaciones de escalas similares en educación estadística, donde los coeficientes oscilaron entre 0,70 y 0,85 (Garfield et al., 2008). Esta robustez permite que el REI-Sports sea utilizado no solo como una herramienta de diagnóstico, sino también como un indicador de progreso en intervenciones basadas en recursos tecnológicos que faciliten el aprendizaje estadístico (Chance et al., 2007). La consistencia interna general ($\alpha = 0,746$) indica que sus reactivos (ítems) miden de manera coherente y homogénea el constructo de RE.

El instrumento REI-Sports, se alinea parcialmente con lo reportado por Sabbag et al. (2025) quienes sugieren que la alfabetización y el RE tienden a colapsarse en un solo factor de conocimiento general, la estructura bifactorial del instrumento REI-Sports demuestra lo contrario, aunque sí se evidencia un desequilibrio estructural entre los reactivos de cada dimensión. Esto indica que, en estudiantes de niveles avanzados de las ciencias del deporte, estas competencias emergen como constructos diferenciados, por

lo que es plausible pensar que el conocimiento estadístico mejora progresivamente durante el desarrollo del programa de formación. Estos hallazgos también coinciden con las implicaciones prácticas propuestas por Lu (2023) quien enfatiza que el aprendizaje de la estadística en las ciencias del deporte alcanza su mayor efectividad cuando se vincula a proyectos prácticos y análisis cuantitativos aplicados.

Desde una perspectiva aplicada, la legibilidad estándar obtenida mediante el índice de Szigriszt-Pazos (2001) se consideró un elemento determinante para la validez del instrumento REI-Sports. Como señalan Patalay et al. (2018) se recomienda una baja carga cognitiva para asegurar que el desempeño en la prueba refleje el RE real y no la dificultad de la comprensión lectora. Frente a ello el instrumento REI-Sports se posiciona como un instrumento inclusivo y eficiente para la población universitaria hispanohablante.

La validez de constructo del REI-Sports se ve reforzada al contrastar con los desarrollos de otros instrumentos recientes en las ciencias de la salud. En particular, nuestros hallazgos convergen con el trabajo de Gaviria-Bedoya et al. (2025) sobre la evaluación del RE situado en estudiantes de postgrado. Al igual que en esta investigación, REI-Sports evidencia que la evaluación de competencias estadísticas es significativamente más precisa cuando los reactivos se alejan de los problemas abstractos y se sitúan en contextos reales como los escenarios clínicos o los deportivos. Sin embargo, cabe mencionar que el instrumento de Gaviria et al. se centra en los procesos formativos de estadística aplicada en estudiantes de postgrado.

Las fortalezas del proceso de validación del instrumento REI-Sports se sustentan en tres pilares presentados metodológicos y teóricos fundamentales. Primero, el diseño del instrumento situado basado en escenarios auténticos de las ciencias del deporte que potencia la validez ecológica del constructo. Un tratamiento psicométrico avanzado que empleó el contraste de Mardia para identificar la no normalidad multivariante y, en consecuencia, aplicó el estimador DWLS para obtener cargas factoriales precisas y exentas de los sesgos propios de las variables ordinales. Tercero, la evidencia empírica de una solida estabilidad temporal combinada con el control de la legibilidad formal mediante el índice de Szigriszt-Pazos.

A pesar del rigor metodológico empleado en la validación del REI-Sports, se puede identificar algunas limitaciones que deben ser consideradas para la contextualización de nuestros hallazgos. En primer lugar, aunque el tamaño muestral permitió la ejecución del AFE, estudios posteriores deberán confirmar la estructura mediante muestras independientes de mayor tamaño. En segundo lugar, el uso de un muestreo por conveniencia podría limitar la representatividad externa. Sin embargo, la homogeneidad de los participantes en términos de su formación en ciencias del deporte mitiga parcialmente este sesgo.

Futuras investigaciones deberían emplear tamaños muestrales más amplios y multicéntricos que incluyan diversas poblaciones y contextos socioculturales para evaluar la invarianza factorial del instrumento. Se sugiere además realizar nuevos estudios para efectuar análisis factorial confirmatorio (AFC) y otros posibles análisis de validez discriminante y de convergencia del instrumento. Se considera necesario continuar con esta línea de trabajo para considerar incluir otros reactivos dentro del instrumento que permitan fortalecer su estructura factorial. Finalmente, se sugiere realizar estudios experimentales que utilicen el REI-Sports como medida para evaluar el impacto de metodologías activas y el uso de tecnologías como estrategia de apoyo en la enseñanza de la estadística.

Conclusiones

Este estudio proporciona soporte empírico sobre las propiedades psicométricas de validez y confiabilidad del cuestionario REI-Sports necesarias para la evaluación del RE en futuros profesionales de las ciencias del deporte, proyectándose como una herramienta potencialmente útil para la educación superior. La implementación de este instrumento dentro de la enseñanza universitaria permitirá identificar sesgos conceptuales en los estudiantes, además permitirá realizar un monitoreo formativo en el desarrollo del RE como base para el desarrollo de las competencias lógico-matemáticas. Finalmente, la información derivada de su aplicación ofrecerá a las facultades académicas un insumo valioso para el diseño curricular basado en la evidencia.

Agradecimientos

Los autores agradecen la participación de los estudiantes de los cursos de fundamentos de estadística y bioestadística aplicada al deporte y la actividad física del programa de ciencias del deporte de la Universidad de Cundinamarca en el primer semestre del 2026.

Financiación

Este es un producto de proyecto 661 “Efectos del uso de ChatGPT en el rendimiento académico, razonamiento y actitudes hacia la estadística deportiva: un estudio cuasiexperimental”. Convocatoria interna X.

Referencias

- Beatty, P. C., & Willis, G. B. (2007). Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly*, 71(2), 287–311. <https://doi.org/10.1093/poq/nfm006>
- Begum, M., Crossley, J., Strömbäck, F., Akrida, E., Alpizar-Chacon, I., Evans, A., Gross, J. B., Haglund, P., Lonati, V., Satyavolu, C., & Thorgeirsson, S. (2025). A Pedagogical Framework for Developing Abstraction Skills. 2024 *Working Group Reports on Innovation and Technology in Computer Science Education*, 258–299. <https://doi.org/10.1145/3689187.3709613>
- Borg, D. N., Bach, A. J. E., O'Brien, J. L., & Sainani, K. L. (2022). Calculating sample size for reliability studies. *PM&R*, 14(8), 1018–1025. <https://doi.org/10.1002/pmrj.12850>
- Cantón-Chirivella, E., Blázquez-Perera, I., García-Mas, A., Núñez Prats, A., & Peris-Delcampo, D. (2025). Instrumentos de evaluación en psicología del deporte en español: triangulación de métodos en la identificación. *Retos*, 73, 950–972. <https://doi.org/10.47197/retos.v73.117257>
- Carranza-Bautista, D., & Giakoni-Ramírez, F. (2025). Validación psicométrica SECO-3D: un instrumento para la evaluación integral de congresos académicos en ciencias de la actividad física y deporte. *Retos*, 72, 25–36. <https://doi.org/10.47197/retos.v72.116552>
- Ceballos-Bernal, E. A., & Correa-Bautista, J. E. (2025). Efectividad de una intervención educativa basada en escenarios interactivos para mejorar actitudes hacia la estadística deportiva: un estudio cuasi-experimental. *Retos*, 72, 470–481. <https://doi.org/10.47197/retos.v72.116894>
- Chan, S. W., & Ismail, Z. (2014). Developing Statistical Reasoning Assessment Instrument for High School Students in Descriptive Statistics. *Procedia Social and Behavioral Sciences*, 116, 4338–4343. <https://doi.org/10.1016/j.sbspro.2014.01.943>
- Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The Role of Technology in Improving Student Learning of Statistics. *Technology Innovations in Statistics Education*, 1(1). <https://doi.org/10.5070/T511000026>
- Chance, B., del Mas, R., & Garfield, J. (2004). *Reasoning about Sampling Distributions. In The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 295–323). Springer Netherlands. https://doi.org/10.1007/1-4020-2278-6_13
- Conway, B., Gary Martin, W., Strutchens, M., Kraska, M., & Huang, H. (2019). The Statistical Reasoning Learning Environment: A Comparison of Students' Statistical Reasoning Ability. *Journal of Statistics Education*, 27(3), 171–187. <https://doi.org/10.1080/10691898.2019.1647008>
- Corbin, C. B. (2021). Conceptual physical education: A course for the future. *Sport and Health Science*, 10 (3), 308–322. <https://doi.org/10.1016/j.jshs.2020.10.004>
- Dejonckheere, M., & Vaughn, L. M. (2019). Semistructured interviewing in primary care research: a balance of relationship and rigour. *Family Medicine and Community Health*, 7(2). <https://doi.org/10.1136/fmch-2018-000057>
- Delmas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing Student's Conceptual Understanding After a First Course in Statistics. *Statistics Education Research Journal*, 6(2), 28–58. <https://doi.org/10.52041/serj.v6i2.483>
- Dunn, T. J., Baguley, T., & Brunsdon, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>



- Evans, A. D., Roberts, K. P., Price, H. L., & Stefek, C. P. (2010). The use of paraphrasing in investigative interviews. *Child Abuse & Neglect*, 34(8), 585–592. <https://doi.org/10.1016/j.chiabu.2010.01.008>
- Fan, X. (1998). Item Response Theory and Classical Test Theory: An Empirical Comparison of their Item/Person Statistics. *Educational and Psychological Measurement*, 58(3), 357–381. <https://doi.org/10.1177/0013164498058003001>
- Friedrich, S., Antes, G., Behr, S., Binder, H., Brannath, W., Dumpert, F., Ickstadt, K., Kestler, H. A., Lederer, J., Leitgöb, H., Pauly, M., Steland, A., Wilhelm, A., & Friede, T. (2022). Is there a role for statistics in artificial intelligence? *Advances in Data Analysis and Classification*, 16(4), 823–846. <https://doi.org/10.1007/s11634-021-00455-6>
- Gagnier, J. J., Lai, J., Mokkink, L. B., & Terwee, C. B. (2021). COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Quality of Life Research*, 30(8), 2197–2218. <https://doi.org/10.1007/s11136-021-02822-4>
- Galanis, P. (2018). The Delphi method. *Archives of Hellenic Medicine*, 35(4). <https://doi.org/10.4324/9781315728513-10>
- Garfield, J & Ben-Zvi, D. (2008). *Developing Students' Statistical Reasoning*. Springer Netherlands. <https://doi.org/10.1007/978-1-4020-8383-9>
- Garfield, J. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education*, 10(3). <https://doi.org/10.1080/10691898.2002.11910676>
- Garfield, J., & Ahlgren, A. (1988). Difficulties in Learning Basic Concepts in Probability and Statistics: Implications for Research. *Journal for Research in Mathematics Education*, 19(1), 44–63. <https://doi.org/10.5951/jresmetheduc.19.1.0044>
- Garfield, J. B. (2003). Assessing Statistical Reasoning. *Statistics Education Research Journal*, 2(1), 22–38. <https://doi.org/10.52041/serj.v2i1.557>
- Garfield, J., & DelMas, R. (2010). A Web Site That Provides Resources for Assessing Students' Statistical Literacy, Reasoning and Thinking. *Teaching Statistics*, 32(1), 2–7. <https://doi.org/10.1111/j.1467-9639.2009.00373.x>
- Garfield, J., & Gal, I. (1999). Teaching and Assessing Statistical Reasoning". En L. V. Stiff (Ed.), *Developing Mathematical Reasoning in Grade K-12* (pp. 207–219). National Council of Teachers of Mathematics.
- Gaviria-Bedoya, J. A., Villa-Ochoa, J. A., & González-Gómez, D. (2025). Situated statistical reasoning assessment for postgraduate health sciences students: Design and validation. *ZDM - Mathematics Education*, 57(7), 1329–1342. <https://doi.org/10.1007/s11858-025-01753-5>
- He, D., & Lao, H. (2018). Paper-and-pencil assessment. En *The SAGE encyclopedia of educational research, measurement, and evaluation*. SAGE Publications. <https://doi.org/10.4135/9781506326139.n496>
- Heemskerk, L., Norman, G., Chou, S., Mintz, M., Mandin, H., & McLaughlin, K. (2008). The effect of question format and task difficulty on reasoning strategies and diagnostic performance in Internal Medicine residents. *Advances in Health Sciences Education*, 13(4), 453–462. <https://doi.org/10.1007/s10459-006-9057-8>
- Hinkin, T. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21(5), 967–988. [https://doi.org/10.1016/0149-2063\(95\)90050-0](https://doi.org/10.1016/0149-2063(95)90050-0)
- Hussain, M., Zhu, W., Zhang, W., Abidi, S. M. R., & Ali, S. (2019). Using machine learning to predict student difficulties from learning session data. *Artificial Intelligence Review*, 52(1), 381–407. <https://doi.org/10.1007/s10462-018-9620-8>
- Kamp, K., Wyatt, G., Dudley-Brown, S., Brittain, K., & Given, B. (2018). Using cognitive interviewing to improve questionnaires: An exemplar study focusing on individual and condition-specific factors. *Applied Nursing Research*, 43, 121–125. <https://doi.org/10.1016/j.apnr.2018.06.007>
- Kellerhuis, B. E., Jenniskens, K., Kusters, M. P. T., Schuit, E., Hooft, L., Moons, K. G. M., & Reitsma, J. B. (2025). Expert panel as reference standard procedure in diagnostic accuracy studies: a systematic scoping review and methodological guidance. *Diagnostic and Prognostic Research*, 9(1), 12. <https://doi.org/10.1186/s41512-025-00195-7>
- Kincaid, J. P., Fishburne, R.P., Jr., Rogers, R., L., & Chissom, B.S. (1975). *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel*. (Research Branch Report 8-75). Naval technical training command. <https://doi.org/10.21236/ADA006655>



- JASP Team. (2024). *JASP* (Versión 0.19.3) [Software de computación]. <https://jasp-stats.org/>
- Li, C.-H. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- Li, C.-H. (2016b). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, *21*(3), 369–387. <https://doi.org/10.1037/met0000093>
- Lu, H.-F. (2023). Statistical learning in sports education: A case study on improving quantitative analysis skills through project-based learning. *Journal of Hospitality, Leisure, Sport & Tourism Education*, *32*, 100417. <https://doi.org/10.1016/j.jhlste.2023.100417>
- Marco, A. (2023). *A Pen and Paper Introduction to Statistics*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781003398820>
- McGrath, A. L. (2014). Content, Affective, and Behavioral Challenges to Learning: Students' Experiences Learning Statistics. *International Journal for the Scholarship of Teaching and Learning*, *8*(2). <https://doi.org/10.20429/ijstl.2014.080206>
- Nasa, P., Jain, R., & Juneja, D. (2021). Delphi methodology in healthcare research: How to decide its appropriateness. *World Journal of Methodology*, *11*(4), 116–129. <https://doi.org/10.5662/wjm.v11.i4.116>
- Niederberger, M., Schifano, J., Deckert, S., Hirt, J., Homberg, A., Köberich, S., Kuhn, R., Rommel, A., & Sonnberger, M. (2024). Delphi studies in social and health sciences-Recommendations for interdisciplinary standardized reporting (DELPHISTAR). Results of a Delphi study. *Plos One*, *19*(8), e0304651. <https://doi.org/10.1371/journal.pone.0304651>
- Patalay, P., Hayes, D., & Wolpert, M. (2018). Assessing the readability of the self-reported Strengths and Difficulties Questionnaire. *BJPsych Open*, *4*(2), 55–57. <https://doi.org/10.1192/bjo.2017.13>
- Perner, P. (2008). Case-based reasoning and the statistical challenges. *Quality and Reliability Engineering International*, *24*(6), 705–720. <https://doi.org/10.1002/qre.951>
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health*, *29*(5), 489–497. <https://doi.org/10.1002/nur.20147>
- Sabbag, A., Zieffler, A., & Ng, C. (2025). Can We Distinguish Statistical Literacy and Statistical Reasoning? *Statistics Education Research Journal*, *24*(1). <https://doi.org/10.52041/serj.v24i1.587>
- Schwob, M., Duan, Y., Cantoni, B., Flores-Lopez, B., & G. Walker, S. (2025). *Exercises in Statistical Reasoning*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781003493471>
- Sepdanius, E., Binti Sanuddin, N. D., Bin Mohd Nor, M. A., Mohd Sidi, M. A. Bin, Erhan Orhan, B., Aman, M. S., & Afriani, R. (2025). Validity and reliability of a sport participant psychosocial well-being instrument: a preliminary study. *Retos*, *64*, 754–766. <https://doi.org/10.47197/retos.v64.110789>
- Shoozan, A., & Mohamad, M. (2024). Application of Interview Protocol Refinement Framework in Systematically Developing and Refining a Semi-structured Interview Protocol. *SHS Web of Conferences*, *182*. <https://doi.org/10.1051/shsconf/202418204006>
- Silva, P. N., & Sarnecka, B. W. (2025). What Do Your Students Struggle with? A Survey of Statistics Instructors. *Journal of Statistics and Data Science Education*, 1–12. <https://doi.org/10.1080/26939169.2025.2455560>
- Szigriszt, F. (2001). *Sistemas predictivos de legibilidad del mensaje escrito: fórmula de lecturabilidad de la lengua española* [Tesis doctoral, Universidad Complutense de Madrid.]. <https://eprints.ucm.es/id/eprint/4134/>
- Traynor, A., & Christopherson, S. C. (2024). Using Content Relevance and Representativeness Indices in Instrument Revision. *Applied Measurement in Education*, *37*(2), 132–147. <https://doi.org/10.1080/08957347.2024.2347518>
- Wild, C. J., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review*, *67*(3), 223–248. <https://doi.org/10.1111/j.1751-5823.1999.tb00442.x>
- Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What Does Research Suggest About the Teaching and Learning of Introductory Statistics at the College Level? A Review of Literature. *Journal of Statistics Education*, *16*(2), 122. <https://doi.org/10.1080/10691898.2008.11889566>



Datos de los/as autores/as y traductor/a:

María Luisa Zambrano-Rojas
Eliaana Patricia Cuellar-Carvajal
Jorge Enrique Correa-Bautista

mlzambrano@ucundinamarca.edu.co
epcuellar@ucundinamarca.edu.
jorgecorrea@ucundinamarca.edu.co

Autor/a
Autor/a
Autor/a

