

HANDTAC: Validación de un test específico en juego reducido (3×3 + porteras) para la evaluación táctica en jugadoras sub-21 de balonmano

HANDTAC: Validation of a sport-specific small-sided game (3v3 + goalkeepers) test for tactical assessment in U-21 female handball players

Deisy Verónica Mazo Garcés^{1*} , Samuel José Gaviria-Alzate² 

¹Instituto Universitario de Educación Física y Deporte, Universidad de Antioquia, Medellín, Colombia ²Facultad de Educación y Ciencias Sociales, Tecnológico de Antioquia Institución Universitaria, Medellín, Colombia.

* Correspondence: veronica.mazo@udea.edu.co

DOI: 10.17398/1885-7019.22.181

Recibido: 10/07/2025; Aceptado: 17/07/2026; Publicado: 01/07/2026

OPEN ACCESS

Sección / Section:
Handball

 **Editor de Sección / Edited by:**
Sebastián Feu
Universidad de Extremadura,
España

Citación / Citation:
Mazo-Garcés, D.V. & Gaviria-Alzate, S.J. (2026). HANDTAC: Validación de un test específico en juego reducido (3×3 + porteras) para la evaluación táctica en jugadoras sub-21 de balonmano. E-balonmano Com, 22(2), 181-198

Fuentes de Financiación / Funding:
Este trabajo no recibió financiación específica de agencias públicas, comerciales o sin ánimo de lucro. No hubo patrocinadores.

Agradecimientos/ Acknowledgments:
Agradecemos a la Liga Antioqueña de Balonmano (Colombia) por su apoyo logístico y la facilitación de espacios y contactos con entrenadores y jugadoras, así como a las/los evaluadores y participantes que contribuyeron con su tiempo y experiencia al desarrollo del HANDTAC.

Conflicto de intereses / Conflicts of Interest:
Los autores declaran no tener conflictos de intereses. Ninguna entidad pública o privada influyó en el diseño del estudio, la recolección/análisis de los datos, la redacción del manuscrito o la decisión de someterlo a publicación.

Resumen

El objetivo de este estudio fue validar HANDTAC, un test específico para la evaluación cuantitativa del rendimiento táctico en balonmano. Se aplicó un diseño de validación de instrumentos en tres fases: (1) análisis de validez de contenido, (2) sensibilidad para discriminar entre jugadoras con distinta experiencia, y (3) confiabilidad mediante test-retest. El protocolo se desarrolló en situaciones de 3 vs 3 más porteras en un espacio de 15 × 30 m, con dos tiempos de 3 minutos. Participaron 36 jugadoras U-21 (19 con mayor experiencia, 17 con menor experiencia). La validez de contenido fue excelente (CVC = 0.932). El test mostró alta sensibilidad al diferenciar significativamente ambos grupos en el índice de rendimiento táctico ofensivo y defensivo ($p < 0.001$; $d = 1.89$ y 1.41). La confiabilidad fue moderada en jugadoras con mayor experiencia (CCI = 0.50–0.59) y variable en las de menor experiencia (CCI = 0.18–0.64), con dispersión moderada según los gráficos de Bland-Altman. En conclusión, HANDTAC constituye una herramienta válida y sensible para la medición del rendimiento táctico en balonmano, aportando un recurso aplicable a entrenadores e investigadores para optimizar metodologías de entrenamiento y profundizar en la evaluación científica del juego.

Palabras clave: balonmano; evaluación táctica; validez de contenido; confiabilidad; juego reducido.

Abstract

The aim of this study was to validate HANDTAC, a sport-specific test designed to quantitatively assess tactical performance in handball. An instrument validation design was applied in three phases: (1) content validity analysis, (2) sensitivity to discriminate between players with different levels of experience, and (3) reliability through a test-retest approach. The protocol consisted of 3 vs. 3 situations plus goalkeepers in a 15 × 30 m court, played in two 3-minute periods. Thirty-six U-21 female players participated (19 with higher experience, 17 with lower experience). Content validity was excellent (CVC = 0.932). The test demonstrated high sensitivity, showing significant differences between groups in offensive and defensive tactical performance indices ($p < 0.001$; $d = 1.89$ and 1.41). Reliability was moderate in the more experienced group (ICC = 0.50–0.59) and variable in the less experienced group (ICC = 0.18–0.64), with Bland-Altman plots indicating moderate dispersion between repeated measures. In conclusion, HANDTAC is a valid and sensitive instrument for measuring tactical performance in handball. It provides coaches and researchers with an applied tool to optimize training methodologies and advance scientific evaluation of the game.

Keywords: handball; tactical assessment; content validity; reliability; small-sided games.

Introducción

La excelencia en el deporte de alto rendimiento exige un control riguroso y sistemático del proceso de entrenamiento, en el que la evaluación continua y la capacidad de ajuste constituyen factores esenciales para optimizar el rendimiento (Saw et al., 2016). En los deportes de equipo, la complejidad derivada de la interacción dinámica entre compañeros y adversarios, junto con la incertidumbre inherente al entorno de juego (Parlebas, 1988), subraya la necesidad de contar con herramientas de análisis específicas y sensibles a la naturaleza de la práctica (Gabbett et al., 2017). El balonmano, caracterizado por su dinamismo y alta intensidad, comparte estos desafíos. Sin embargo, diversas revisiones sistemáticas han señalado la limitada disponibilidad de instrumentos específicos para evaluar el componente táctico en contextos de entrenamiento, lo que restringe la comprensión integral del rendimiento y dificulta el seguimiento de la progresión táctica de los jugadores (Ferrari et al., 2019; Sarmiento et al., 2022).

La literatura ha destacado de forma consistente la importancia del principio de especificidad en el control del entrenamiento (Viru & Viru, 2003), lo que implica disponer de mediciones precisas de los componentes clave del rendimiento. En el plano táctico, dicha especificidad se traduce en la necesidad de valorar la capacidad de los jugadores para percibir, interpretar y responder en coherencia con los principios propios del balonmano. No obstante, la mayoría de las investigaciones se han centrado en escenarios competitivos. Ejemplo de ello son los trabajos de Aróstica Villa et al. (2016), quienes identificaron indicadores tácticos mediante consulta a expertos; Caldas et al. (2017), que exploraron el conocimiento táctico declarativo a partir de simulaciones; y Quiñones et al. (2019), quienes validaron un sistema de observación del ataque posicional. Si bien estas aproximaciones aportan evidencia valiosa, no logran cubrir la evaluación del rendimiento táctico en el entrenamiento ni incorporan plenamente la especificidad de la lógica interna del juego.

En los últimos años, en balonmano se ha reforzado el interés por construir instrumentos específicos que permitan operacionalizar y observar de forma sistemática los procedimientos tácticos colectivos en situaciones reales de juego, con criterios explícitos de validez y fiabilidad. En esta línea, se han publicado propuestas orientadas al ataque posicional y la identificación de procedimientos colectivos mediante instrumentos observacionales validados, lo que refuerza la pertinencia de desarrollar y validar herramientas que capturen la toma de decisión táctica en contextos representativos. (Zapardiel et al., 2025).

En deportes afines se han desarrollado instrumentos como el GPAI, el TSAP y el FUT-SAT, diseñados para cuantificar principios tácticos en distintos contextos (Gréhaigne et al., 2005). Incluso en balonmano se han propuesto aproximaciones, como las de Greco et al. (2015) y Ribeiro et al. (2023), que introdujeron evaluaciones en formato 3x3. Sin embargo, estas presentan limitaciones significativas al no considerar elementos distintivos del juego, tales como el área de portería o los principios tácticos específicos del deporte. Esta situación evidencia un vacío metodológico: la ausencia de pruebas específicas, objetivas y cuantitativas que permitan evaluar la táctica en entornos de entrenamiento controlados.

El proceso táctico, concebido como un ciclo de percepción, decisión, ejecución y retroalimentación (Lebed & Bar-Eli, 2013), demanda evaluaciones capaces de capturar su complejidad y naturaleza adaptativa. Román (2008) subraya que la variabilidad y la capacidad de actuar inteligentemente frente a la incertidumbre constituyen dimensiones esenciales para impedir la anticipación del rival. Esto refuerza la necesidad de herramientas que valoren no solo el conocimiento declarativo, sino también la ejecución táctica en situaciones representativas del entrenamiento.

El rendimiento táctico en balonmano no puede entenderse únicamente como ejecución técnica, sino como un proceso integrado de percepción, interpretación y decisión bajo presión temporal y con alta variabilidad contextual. Desde aproximaciones metodológicas recientes, se ha insistido en que el diseño de tareas debe contemplar el carácter complejo y sistémico del juego, incorporando incertidumbre, variabilidad y mecanismos de autorregulación que favorezcan decisiones adaptativas. En este marco, la evaluación táctica específica cobra sentido como complemento necesario para orientar el entrenamiento y monitorear el desarrollo del juicio táctico. (Soto-García & Oliver, 2022).

En este contexto, el presente estudio se orienta al diseño y validación de un test específico para el balonmano, con el propósito de determinar su validez de contenido, su concordancia interobservador, su estabilidad temporal mediante test-retest y su sensibilidad. Este instrumento pretende superar las limitaciones de la literatura previa, proporcionando una medida objetiva, específica y aplicable en contextos de formación y alto rendimiento. Su desarrollo busca contribuir a la toma de decisiones basada en evidencia y a la optimización del proceso de entrenamiento táctico en este deporte.

Con base en la literatura sobre evaluación táctica específica y el uso de tareas representativas en deportes de invasión, la presente investigación planteó tres hipótesis: (H1) el contenido del HANDTAC presentará alta validez de contenido según el coeficiente de validez de contenido (CVC); (H2) el test mostrará sensibilidad para discriminar entre jugadoras más experimentadas (ME) y menos experimentadas (mE) en los índices tácticos compuestos; y (H3) la estabilidad temporal (test-retest) será, como mínimo, moderada en los índices compuestos, siendo más variable en principios de baja frecuencia.

Materiales y Métodos

Este estudio adoptó un diseño instrumental de desarrollo y validación de un protocolo de evaluación, con enfoque cuantitativo y fases secuenciales (Ibáñez & Feu, 2026), fundamentado en los lineamientos metodológicos de Pasquali (2010). El proceso se organizó en fases secuenciales para garantizar la robustez psicométrica del test. En primer lugar, se realizó la definición del constructo y la operacionalización de variables; se delimitó el constructo teórico de interés, el rendimiento táctico en balonmano, y se tradujo a indicadores observables y medibles que permitieran su representación empírica de manera fiel y pertinente al dominio de la táctica (Pasquali, 2010).

A continuación, se evaluó la validez de contenido, entendida como el grado en que el instrumento representa adecuadamente el universo del constructo (Hernández et al., 2014). Para ello, se implementaron dos estrategias complementarias. Por un lado, un juicio de expertos conformado por cinco especialistas en balonmano y metodología de la investigación, valoró la relevancia, la representatividad y la claridad de los ítems, asegurando la cobertura integral de los principios tácticos considerados (Hernández et al., 2014). Por otro lado, se condujo una prueba piloto con una muestra reducida de jugadoras U-21, cuyo propósito fue identificar ambigüedades, dificultades de comprensión y potenciales problemas de aplicación; con base en estas evidencias se efectuaron los ajustes pertinentes antes de la aplicación principal (Haynes et al., 1995).

La confiabilidad, concebida como la consistencia y estabilidad de las mediciones, se analizó mediante un diseño test-retest, administrando el instrumento en dos ocasiones a la misma muestra y estimando la estabilidad de las puntuaciones a través del coeficiente de correlación intraclase (CCI), de acuerdo con las recomendaciones para estudios de fiabilidad en contextos de medición repetida (Hernández et al., 2014). Finalmente, se examinó la sensibilidad del test, su capacidad para discriminar entre jugadoras con distinto nivel de experiencia competitiva, comparando las puntuaciones entre grupos mediante la prueba t de Student para muestras independientes y complementando la interpretación con el tamaño del efecto (Cohen's d), de acuerdo con estándares de análisis en ciencias del deporte (Hopkins, 2000).

Participantes

La muestra estuvo integrada por 36 jugadoras U-21, seleccionadas mediante muestreo no probabilístico por conveniencia y disponibilidad (Etikan et al., 2016). La clasificación siguió el Participant Classification Framework (Tiers) (McKay et al., 2022); el grupo de mayor experiencia (ME; n=19) correspondió aproximadamente a Tier 3 (Highly trained/National level), participación en campeonatos de ámbito nacional o pertenencia a equipos representativos regionales/subnacionales y ≥5 sesiones de entrenamiento semanales, mientras que el grupo de menor experiencia (mE; n=17) se alineó con Tier 2 (Trained/Developmental), en proceso formativo sin historial de selecciones regionales o nacionales (Tabla 1).

Tabla 1. Descripción sociodemográfica de la muestra.

Grupo	N	Edad	Talla	Peso	Edad_deportiva
ME	19	20.2 (1.0)*	1.68 (0.14)*	63.02 (7.00)*	9.06 (7.00)*
mE	17	18.1 (0.0)*	1.63 ± (0.04)	57.52 ± (5.04)	4.04 (2.0)*

*Nota: * Mediana y Rango Intercuartil. Edad en años, talla en metros, peso en kilogramos, edad deportiva en años, Grupo de mayor experiencia (ME), Grupo de menor experiencia (mE)*

Criterios de inclusión y exclusión

Para conformar la muestra y salvaguardar la homogeneidad y la validez interna, se definieron criterios de inclusión y exclusión específicos para cada grupo (Vandenbroucke et al., 2014). En el grupo de mayor experiencia (ME) se incluyeron jugadoras U-21 con participación reciente (últimos 12-24 meses) en campeonatos de ámbito nacional o pertenencia a equipos representativos regionales/subnacionales, con ≥5 sesiones de entrenamiento semanales y disponibilidad para todas las fases del estudio. En el grupo de menor experiencia (mE) se incluyeron jugadoras U-21 sin historial de integración en selecciones regionales o nacionales y con disponibilidad total para su participación.

Se excluyeron las deportistas con lesiones o condiciones médicas limitantes, las que no completaran todas las evaluaciones previstas o no otorgaran su consentimiento informado. El tamaño muestral se determinó según la accesibilidad de las participantes y el objetivo de contrastar la sensibilidad del instrumento entre distintos niveles de rendimiento (Etikan et al., 2016; Prinsen et al., 2018).

Control de sesgos y medidas de seguridad

Para maximizar la validez interna y la confiabilidad de los resultados, se implementó un paquete de estrategias metodológicas integrado; se capacitó a los observadores en el protocolo, las definiciones operativas y la práctica de registros hasta alcanzar criterios comunes; se realizó una prueba piloto con jugadoras de características similares a la muestra objetivo para depurar ítems, el protocolo de observación y la logística de recolección de datos (Haynes et al., 1995); cada acción táctica fue codificada de forma independiente por dos observadores, con visionado reiterado cuando fue necesario y resolución por consenso ante discrepancias; y se estandarizaron de manera estricta las condiciones de aplicación (espacios, tiempos y secuencias de juego) para garantizar la replicabilidad.

La concordancia entre observadores se examinó mediante el coeficiente de correlación intraclase (ICC) bajo un modelo de dos vías y acuerdo absoluto (Koo & Li, 2016). Este control se aplicó como parte del aseguramiento de calidad del registro, desde la fase piloto (entrenamiento y estandarización de criterios) hasta el procesamiento final, y las discrepancias fueron resueltas mediante revisión conjunta cuando fue necesario.

Procedimiento

Estudio 1: Diseño del Instrumento, Validación de Contenido y Prueba Piloto

El diseño del instrumento se operativizó en un formato de juego 3 vs 3 más porteras sobre un espacio de 15 × 30 m, con el área de portería delimitada por una línea recta a 5 m (en lugar del semicírculo oficial de 6 m) para simplificar la configuración observacional sin alterar la lógica ofensivo-defensiva del juego. Cada enfrentamiento constó de dos periodos de 3 minutos, separados por un minuto de recuperación, con el objetivo de estandarizar la carga situacional y facilitar la comparación entre intentos (véase Figura 1). Este montaje garantizó un entorno controlado y suficientemente representativo para la valoración de las conductas tácticas y la obtención de registros consistentes destinados,

posteriormente, al juicio de expertos (validación de contenido) y a la prueba piloto orientada a depurar definiciones operativas, planillas y logística de observación

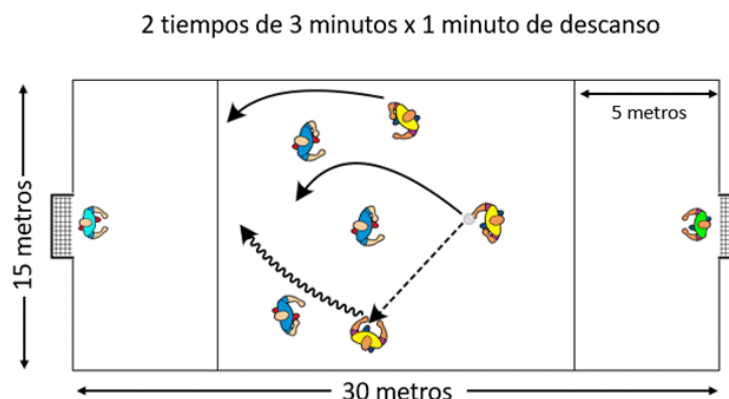


Figura 1. Área y tiempo de juego

Nota. Elaboración propia

Razonamiento del diseño del test

El protocolo se definió con el doble propósito de maximizar la validez ecológica y asegurar la factibilidad de la evaluación táctica en condiciones controladas. En este marco, se optó por un formato de juego reducido que preserve la lógica ofensivo-defensiva del balonmano y, al mismo tiempo, facilita la observación sistemática de las conductas (Araújo et al., 2007; Clemente et al., 2021; Pinder et al., 2011).

En primer lugar, se redujo el espacio de juego a 15 × 30 m para concentrar la atención en los principios tácticos esenciales y simplificar la codificación. La evidencia disponible indica que los juegos reducidos favorecen la identificación del comportamiento táctico en contextos controlados (Clemente et al., 2021) y potencian la toma de decisiones al exigir respuestas más rápidas y eficientes (González-Víllora et al., 2015). Coherente con ello, el formato 3 vs 3 más porterías incrementa el número de interacciones por jugadora y, por ende, la cantidad de acciones observables en un tiempo limitado, lo que resulta pertinente para valorar el rendimiento táctico (Fernández-Espínola et al., 2020).

En segundo lugar, la duración del test se estableció en dos periodos de 3 minutos con un minuto de recuperación entre ellos, buscando un equilibrio entre exigencia y control de la fatiga. Este diseño permite observar la actuación táctica bajo alta demanda sin que la fatiga extrema distorsione los registros (Arslan et al., 2020).

Adicionalmente, el área de portería se adaptó a una línea recta a 5m, en lugar del semicírculo reglamentario de 6m, con el objetivo de estandarizar la marcación y simplificar la observación, manteniendo la coherencia con las disposiciones reglamentarias (International Handball Federation (IHF), 2025). Para preservar la validez táctica del protocolo, se introdujeron ajustes operativos al reglamento oficial; la presencia de un árbitro evitó valorar acciones obtenidas mediante faltas técnicas o disciplinarias; tras cada gol no se efectuó saque de centro (el reinicio al comienzo de cada periodo se realizó mediante saque del árbitro), con el fin de optimizar el tiempo efectivo de observación; en caso de sanciones se registró la falta pero se mantuvo la igualdad numérica para asegurar la comparabilidad de oportunidades entre equipos; y la portera permaneció en su área durante todo el protocolo, evitando alterar el equilibrio táctico previsto. En conjunto, estas decisiones metodológicas incrementan la densidad de eventos relevantes, favorecen la replicabilidad y sostienen la representatividad situacional que exige la evaluación táctica (Clemente et al., 2021).

Principios tácticos en el balonmano

El rendimiento táctico en balonmano se organiza en torno a dos fases nucleares del juego, ofensiva (con balón) y defensiva (sin balón), y su análisis se apoya en marcos consolidados para deportes de invasión y, en particular, para el propio balonmano (Greco et al., 2015; Gréhaigne et al., 2005; Ribeiro et al., 2023). En coherencia con dicha literatura, el test operacionalizó nueve variables clave que condensan las acciones más representativas de cada fase. En fase

ofensiva, se evaluaron: circulación del balón, desmarque, juego 1x1, eficacia de lanzamiento y contraataque; en fase defensiva, se consideraron: marcaje, interceptaciones y blocajes, ayudas defensivas y repliegue.

Estas variables, descritas con detalle en la Tabla 2, actúan como indicadores observables de la capacidad de las jugadoras para percibir, decidir y ejecutar de acuerdo con los principios tácticos del balonmano. Su selección asegura especificidad al deporte, mantiene la representatividad de las situaciones de juego y permite capturar la complejidad inherente del proceso táctico sin diluirlo en métricas exclusivamente técnicas (Porttell et al., 2015; Renshaw et al., 2022; Travassos et al., 2012).

Tabla 2. Subcategorías, variables y definición de las variables tácticas en balonmano

Fases de Juego	Fase ofensiva con posesión de balón	Fase defensiva sin posesión de balón
Principios tácticos fundamentales	Circulación del balón	Marcaje
	El balón se traslada de un lugar a otro con el objetivo de que llegue al jugador que mejores condiciones tenga para crear peligro al rival	Referido a la lucha del 1x1 evitando que su oponente directo logre su objetivo ofensivo
	Desmarque	Interceptaciones y blocajes
	Eludir a los defensores buscando zonas libres SIN balón y generando una opción de pase al compañero con balón	1. Lograr recuperar el balón rompiendo la trayectoria del pase. 2. Bloquear el lanzamiento a portería del equipo rival
	Juego 1x1	Ayuda defensiva
	Utilización de fintas (movimientos de engaño) CON balón o lanzamientos en juego directo con el defensor para superarlo y penetrar a zonas eficaces	Se refiere a la acción de un defensor que abandona temporalmente su posición para ayudar a un compañero que ha sido superado por un atacante o para cubrir un espacio vulnerable en la defensa.
	Eficacia de lanzamiento	Repliegue
	Referido a la efectividad de los lanzamientos desde diferentes posiciones y en situaciones de oposición.	Rapidez y organización del equipo en transición del ataque a la defensa
	Contraataque	
	Transición de defensa a ataque con rapidez y efectividad con no más de 3 pases, midiendo si termina en gol o se mantiene la posesión.	

(Laguna, 2021)

Índices de Rendimiento Táctico (IRT)

Los Índices de Rendimiento Táctico (IRT) se concibieron como conductas observables que caracterizan dimensiones específicas del desempeño y posibilitan su evaluación objetiva (Travassos et al., 2012). Cada variable táctica fue operacionalizada mediante criterios inequívocos y expresada como porcentaje de éxito. Para el registro se empleó una codificación dicotómica: “1” cuando la conducta se realizó correctamente y “0” cuando no se cumplió, de modo que la suma y el porcentaje resultantes reflejan la eficacia de ejecución en cada principio.

La validez de contenido se estableció a través de valoración de expertos siguiendo las recomendaciones de Haynes et al. (1995). Un panel de especialistas juzgó la pertinencia, representatividad y claridad de los ítems incluidos en el instrumento, y los resultados se analizaron mediante el Coeficiente de Validez de Contenido (CVC) de Hernández-Nieto (2002), conforme a los procedimientos descritos por Pedrosa et al. (2014), lo que permitió estimar el grado de acuerdo tanto a nivel de ítem como del instrumento global. En consonancia con la literatura metodológica, se consideró adecuado un número de entre tres y cinco jueces.

Para efectos de esta investigación, se definió como experto a aquel entrenador que cumpliera al menos dos de los siguientes criterios: (i) haber sido jugador de balonmano de alto nivel, con una trayectoria mínima de cinco años en ligas profesionales o selecciones nacionales, experiencia que aporta un capital formativo diferencial para la labor de entrenamiento (Cassidy & Rossi, 2006); (ii) poseer formación profesional en el ámbito deportivo avalada por una

institución de educación superior, que provee sustento académico y metodológico complementario al conocimiento empírico (Feu Molina et al., 2012); y (iii) acreditar ≥ 5 años de experiencia como entrenador en contextos de rendimiento (ligas o selecciones locales, departamentales, nacionales o profesionales), condición que favorece la aplicación y validación práctica de los saberes técnicos (Feu Molina et al., 2012).

Finalmente, se llevó a cabo una prueba piloto orientada a verificar la funcionalidad del sistema de indicadores, depurar el protocolo de registro y estandarizar la observación de las conductas tácticas, asegurando así la consistencia y fiabilidad de los datos en el estudio principal.

Evaluadores

Participaron dos evaluadores con amplia experiencia específica en balonmano, cada uno con más de diez años de trayectoria combinada como entrenadores y jugadores (Chacón-Moscoso et al., 2019). Antes de la recogida de datos, recibieron una formación focalizada en el protocolo, las definiciones operativas de cada conducta y la práctica de registros, lo que garantizó un uso consistente del instrumento y una lectura experta de las acciones tácticas.

Procedimiento prueba piloto

El test 3 vs 3 se aplicó conforme al protocolo establecido, con dos periodos de 3 minutos separados por un minuto de recuperación. Los dos evaluadores observaron y registraron de forma independiente las conductas tácticas de las jugadoras, empleando los IRT previamente definidos. Durante la aplicación, se documentaron observaciones cualitativas sobre la claridad y comprensibilidad de los IRT, la usabilidad del sistema de registro y cualquier ambigüedad o dificultad operativa detectada. Al término de la sesión, se realizó una reunión de debriefing para revisar las notas de campo, discutir eventuales discrepancias y proponer ajustes al protocolo (Chacón-Moscoso et al., 2019; Clemente et al., 2021; Dello Iacono et al., 2021).

Análisis de la prueba piloto

Los registros de ambos evaluadores se utilizaron para estimar la fiabilidad interobservador como evidencia de consistencia del procedimiento. Cuando fue posible, se calculó el coeficiente de correlación intraclase (ICC) bajo un modelo de dos vías y acuerdo absoluto, con sus intervalos de confianza del 95%, en línea con las recomendaciones internacionales (Koo & Li, 2016). A partir de estos resultados y de las incidencias observadas, se introdujeron ajustes específicos en el protocolo y en los criterios de observación; además, se realizó una sesión adicional de capacitación para afinar la concordancia entre evaluadores y asegurar la consistencia en la recolección de datos del estudio principal.

Consideraciones éticas y de confidencialidad

La participación en la prueba piloto fue voluntaria, con posibilidad de retiro en cualquier momento sin repercusiones. Se garantizó la confidencialidad de la información y el anonimato en el informe y la difusión de resultados, en estricto cumplimiento de los principios éticos para investigación con seres humanos establecidos por la Declaración de Helsinki (World Medical Association, 2013). El protocolo fue evaluado y aprobado por el Comité de Ética en Investigación del Instituto Universitario de Educación Física y Deporte de la Universidad de Antioquia (CEI-EDUFID; aval ACEI 67-2025, Acta n.º 145; 11 de abril de 2025).

Estudio 2: Sensibilidad del test (validez por grupos conocidos)

Se examinó la capacidad discriminativa del HANDTAC entre deportistas con diferente nivel de pericia táctica, contrastando dos grupos U-21 definidos a priori según su trayectoria competitiva: un grupo de mayor experiencia (ME; participación reciente en campeonatos de ámbito nacional o pertenencia a equipos representativos regionales/subnacionales, ≥ 5 sesiones de entrenamiento semanales) y un grupo de menor experiencia (mE; en proceso formativo, sin historial de selecciones regionales o nacionales). Este enfoque responde al principio de validez por grupos conocidos en la evaluación de constructos, en el que se formulan hipótesis direccionales ($ME > mE$) y se contrastan

mediante diferencias entre grupos (Mokkink et al., 2010). A efectos inferenciales, se verificaron supuestos de normalidad (Shapiro-Wilk) y se aplicó t de Student para muestras independientes, reportando tamaño del efecto (Cohen's d) con IC95% e interpretaciones estandarizadas para magnitudes prácticas (Hopkins, 2000).

Estudio 3: Confiabilidad del test (test-retest)

La estabilidad temporal del HANDTAC se evaluó mediante un diseño test-retest con un intervalo de ocho días entre mediciones sobre las mismas participantes. Este intervalo de ocho días entre mediciones se estableció con el fin de reducir el efecto de memoria y el aprendizaje inmediato, manteniendo al mismo tiempo una ventana temporal lo suficientemente corta como para asumir estabilidad razonable del constructo táctico y del contexto de entrenamiento. Intervalos cercanos a una semana han sido utilizados en estudios de fiabilidad con tareas específicas en balonmano. La consistencia se estimó con el coeficiente de correlación intraclass (ICC) bajo un modelo de dos vías y acuerdo absoluto (ICC(2,1) o ICC(2,k), según agregación), con intervalos de confianza del 95%, de acuerdo con las guías metodológicas para estudios de fiabilidad (Koo & Li, 2016). La fiabilidad absoluta se describió mediante análisis de Bland-Altman (sesgo y límites de acuerdo) (Bland & Altman, 1986, 1999).

Análisis estadístico

Los análisis se realizaron en Jamovi v2.3.28. Se efectuó una caracterización descriptiva de las variables (media $\pm$ DE cuando la distribución fue aproximadamente normal; mediana y RIC en caso contrario) y se verificaron los supuestos de normalidad mediante Shapiro-Wilk. Todas las contrastaciones fueron bilaterales con un umbral de significación de $\alpha = 0,05$.

Para el Estudio 1 (validez de contenido), las valoraciones del panel se sintetizaron mediante el Coeficiente de Validez de Contenido (CVC) de Hernández-Nieto (2002), calculado por ítem y global, con corrección del error y criterio de retención habitual $CVC \geq 0,80$ (Pedrosa et al., 2014; Sánchez Sánchez, 2021). Este procedimiento permitió estimar la pertinencia, representatividad y claridad de cada indicador y, en conjunto, del instrumento.

Para el Estudio 2 (sensibilidad/validez por grupos conocidos), se evaluó la capacidad discriminativa del HANDTAC entre deportistas de mayor y menor experiencia. Cuando se cumplieron los supuestos paramétricos, se aplicó la t de Student para muestras independientes; en caso contrario, se recurrió al test U de Mann-Whitney como alternativa no paramétrica. En todos los contrastes se reportó el tamaño del efecto (Cohen's d) con su interpretación en términos prácticos, conforme a criterios habituales en ciencias del deporte (Hopkins, 2000), junto con los IC95% cuando correspondía.

Para el Estudio 3 (confiabilidad test-retest), la fiabilidad se abordó mediante un enfoque test-retest, estimando la estabilidad temporal de las puntuaciones y su consistencia. Este tipo de aproximación es habitual en pruebas específicas del balonmano y se ha utilizado para valorar estabilidad del gesto con indicadores como el coeficiente de correlación intraclass (ICC) y medidas de variabilidad (Koo & Li, 2016), contemplando además la posible influencia de la familiarización entre sesiones. (Andrades-Ramírez et al., 2022). La fiabilidad absoluta se describió mediante gráficos de Bland-Altman, informando el sesgo medio y los límites de acuerdo al 95% (Bland & Altman, 1986, 1999). Todos los análisis y gráficos se realizaron en Python (v. 3.11), ejecutados en Visual Studio Code (v. 1.93). Se emplearon numpy (v. 1.26), pandas (v. 2.2), scipy (v. 1.11), statsmodels (v. 0.14) y matplotlib (v. 3.8).

Resultados

Estudio 1: Coeficiente de Validez de Contenido

El panel de cinco expertos (Tabla 3) evaluó la pertinencia, representatividad, claridad conceptual, redacción y terminología de las nueve variables tácticas del HANDTAC a partir de una síntesis del protocolo y sus definiciones operativas.

Tabla 3. Caracterización de expertos

Nº Experto	Años de experiencia como entrenador	Años de experiencia como jugador profesional	Estudios académicos en el ámbito deportivo	Títulos más relevantes en su carrera profesional
1	Más de 10 años	Más de 10 años	Maestría	2 medallas de bronce Olímpicas y 5 Campeonatos de Europa con la selección española. "Mejor jugador del Mundo".
2	Más de 10 años	Más de 10 años	Doctorado	Campeón del Mundo y 3 medallas olímpicas con la Selección Española de Balonmano
3	Más de 10 años	0 años	Maestría	Subcampeón de la Copa del Rey (en España) Subcampeón Juvenil de Europa (Selección Española) Subcampeón de los Juegos Mundiales Juveniles (Selección Española)
4	Más de 10 años	Más de 10 años	Especialización	Oro en juegos Europeos de la juventud y liga. Competiciones como liga Asobal, Champions y copa EHF, internacional absoluto y categorías inferiores con la selección española.
5	Más de 10 años	0 años	Maestría	Campeón de la Copa Italia. Campeón del Torneo Centro y Sur Americano IHF países emergentes como seleccionador de Colombia. Campeón Super Copa de Emiratos Árabes.

Sobre la base de estas valoraciones se estimó el Coeficiente de Validez de Contenido de Hernández-Nieto (2002) por ítem y global, aplicando la corrección por error ($Pei=0,00032$ para $j=5$), y se obtuvo un CVC global de 0,93 (Tabla 4). Los CVCtc por ítem oscilaron entre 0,89 (juego 1x1) y 0,99 (repliegue), con valores $\geq 0,90$ en todos los restantes principios (circulación del balón, desmarque, eficacia de lanzamiento, contraataque, marcaje, interceptaciones y blocajes, ayudas defensivas).

Tabla 4. Coeficiente de validez de contenido Hernández-Nieto (2002)

VARIABLES	Sxi	Mx	CVCi	Pei	CVCtc
Circulación del balón	96	4.80	0.96	0.00032	0.96
Desmarque	90	4.50	0.90	0.00032	0.90
Juego 1x1	89	4.45	0.89	0.00032	0.89
Eficacia de lanzamiento	92	4.60	0.92	0.00032	0.92
Contraataque	93	4.65	0.93	0.00032	0.93
Marcaje	95	4.75	0.95	0.00032	0.95
Interceptaciones y blocajes	93	4.65	0.93	0.00032	0.93
Ayuda defensiva	92	4.60	0.92	0.00032	0.92
Repliegue	99	4.95	0.99	0.00032	0.99
PROMEDIO					0.93

Nota. Sxi: suma de las puntuaciones asignadas por los expertos a cada variable; Mx: media de las puntuaciones de la variable; CVCi: coeficiente de validez de contenido inicial ($Mx/Vmáx$); Pei: probabilidad de error; CVCtc: coeficiente de validez de contenido corregido ($CVCi - Pei$).

Consistencia interobservador

En la prueba piloto, dos evaluadores registraron de forma independiente los conteos por principio. El ICC interobservador fue 0.91 para los principios del HANDTAC, excepto en el principio de circulación del balón ($ICC = 0.87$).

Estudio 2: Sensibilidad del Test

El supuesto de normalidad se verificó mediante Shapiro-Wilk para los índices posttest (IRT_Defensivo: $p=0.81$; IRT_Ofensivo: $p=0.90$). Con base en ello, se aplicó la t de Student para muestras independientes. Tal como se resume

en la Tabla 5, el grupo de mayor experiencia superó al de menor experiencia en ambos índices, con diferencias estadísticamente significativas y tamaños del efecto grandes ($d > 0.80$).

Tabla 5. Prueba T para muestras independientes

		Estadístico	gl	p	Tamaño del efecto	
IRT_Def_Post	T de Student	4.22	34.0	<.001	d de Cohen	1.41
IRT_Ofen_Post	T de Student	5.66	34.0	<.001	d de Cohen	1.89

Nota: IRT_Def_Post: Índice de Rendimiento Táctico Defensivo Posttest. IRT_Ofen_Post: Índice de Rendimiento Táctico Ofensivo Posttest.

Estudio 3: Confiabilidad del Test

La estabilidad temporal del HANDTAC se estimó mediante el coeficiente de correlación intraclass (ICC, modelo de dos vías con acuerdo absoluto) para cada principio y para los índices compuestos, a partir de dos aplicaciones separadas por ocho días (Koo & Li, 2016). Como sintetiza la Tabla 6, los índices compuestos mostraron fiabilidad de pobre a moderada: en el grupo de mayor experiencia (ME), IRT_Defensivo=0.50 (IC95%: 0.08–0.77) e IRT_Ofensivo=0.59 (IC95%: 0.21–0.82); en el grupo de menor experiencia (mE), IRT_Defensivo=0.64 (IC95%: 0.25–0.85).

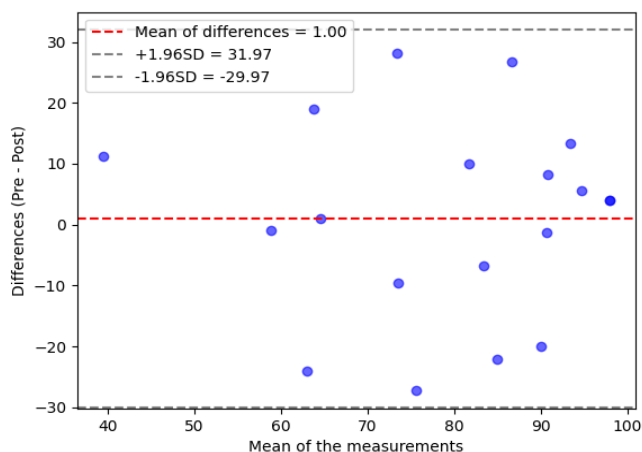
Tabla 6. Estadísticos descriptivos pre y posttest y Coeficiente de Correlación Intraclass de ME y mE.

Mayor Experiencia								
Principio	Media/ Mediana pre	DE/RIC pre	Media/ Mediana post	DE/RIC post	ICC	p	95% CI lower	Upper
IR_CirBal	100	9.72	100	9.55	0.33	0.07	-0.11	0.67
IR_Des	100	29.17	100	15.56	0.24	0.14	-0.21	0.61
IR_1vs1	100	33.33	100	41.67	0.42	0.02	-0.00	0.73
IR_EfiLanz	100	36.67	66.70	23.33	0.36	0.05	-0.08	0.69
IR_Contra	100	100	100	100	0.46	0.01	0.04	0.75
IR_Marc	66.70	42.86	71.50*	22.55*	0.25	0.13	-0.20	0.62
IR_Intercep	0.00	70.83	60.00	100	-0.07	0.62	-0.49	0.37
IR_AyudaDef	0.00	87.50	0.00	75	-0.03	0.55	-0.46	0.41
IR_Replie	83.30	48.57	57.90*	31.6*	0.42	0.03	-0.01	0.72
IRT_Def	51.90*	22.90*	53.80*	24.33*	0.50	0.01	0.08	0.77
IRT_Ofen	79.70*	17.68*	78.70*	17.67*	0.59	p<.001	0.21	0.82
Menor Experiencia								
Principio	Media/ Mediana pre	DE/RIC pre	Media/ Mediana post	DE/RIC post	ICC	p	95% CI lower	Upper
IR_CirBal	92.30	9.02	88.90	13.04	0.22	0.17	-0.26	0.62
IR_Des	42.90	75.00	44.16*	34.20*	0.04	0.43	-0.42	0.49
IR_1vs1	0.00	50.00	0.00	14.29	0.24	0.16	-0.24	0.63
IR_EfiLanz	0.00	50.00	0.00	50	0.24	0.15	-0.23	0.64
IR_Contra	0.00	0.00	0.00	0.00	-0.16	0.74	-0.57	0.32
IR_Marc	0.00	50.00	38.82*	33.10*	-0.18	0.77	-0.59	0.30
IR_Intercep	0.00	0.00	0.00	0.00	0.64	0.00	0.26	0.85
IR_AyudaDef	0.00	0.00	0.00	0.00	-0.07	0.61	-0.51	0.40
IR_Replie	11.10	28.60	22.20	60.00	0.42	0.03	-0.03	0.74
IRT_Def	12.50	22.20	22.43*	19.60*	0.64	0.00	0.25	0.85
IRT_Ofen	39.35*	16.80*	41.28*	21.90*	0.18	0.23	-0.30	0.59

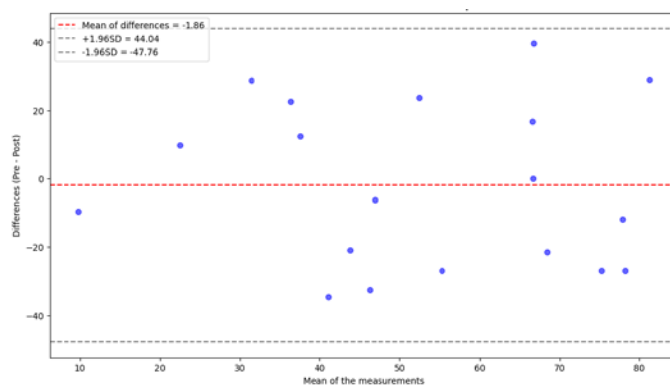
Nota. IR: índice de rendimiento. IR_CirBal: IR circulación del balón. IR_Des: IR desmarque. IR_1vs1: IR juego 1x1. IR_EfiLanz: IR eficacia del lanzamiento. IR_Contra: IR contraataque. IR_Marc: IR marcaje. IR_Intercep: IR interceptaciones y blocajes. IR_AyudaDef: IR ayudas defensivas. IR_Replie: IR repliegue. IRT: índice de Rendimiento Táctico. IRT_Def: IRT defensivo. IRT_Ofen: IRT ofensivo. * Mediana y Rango Intercuartil (RIC).

La fiabilidad absoluta se examinó mediante diagramas de Bland-Altman para cada combinación de fase táctica y nivel de experiencia (Figuras 2A-2D). En mayor experiencia (ME), el IRT ofensivo presentó una media de las diferencias de +1.00 puntos y límites de acuerdo (LoA) de -29.97 a +31.97 (Figura 2A). El IRT defensivo mostró una media de -1.86 con LoA de -47.76 a +44.04 (Figura 2B). En menor experiencia (mE), el IRT ofensivo registró una media de -1.94 y LoA de -50.26 a +46.39 (Figura 2C), mientras que el IRT defensivo presentó una media de -6.45 y LoA de -34.43 a +21.54 (Figura 2D). En todos los casos, los puntos se distribuyeron alrededor de la media de las diferencias sin un patrón evidente a lo largo del rango de las medias.

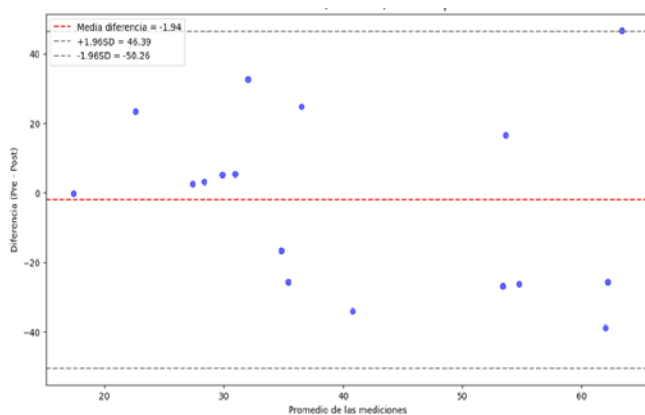
(A) IRT Ofensivo mayor experiencia



(B) IRT Defensivo mayor experiencia



(C) IRT Ofensivo menor experiencia



(D) IRT Defensivo menor experiencia

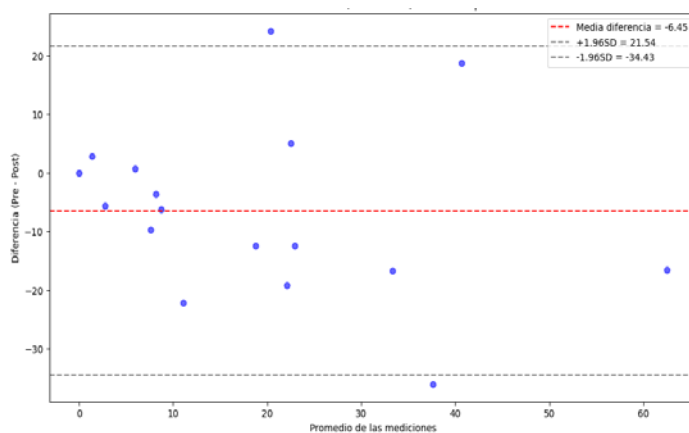


Figura 2. Diagramas de Bland–Altman del HANDTAC por fase táctica (ofensiva/defensiva) y nivel de experiencia (mayor/menor), paneles 2A–2D.

Cada panel muestra las diferencias Pre–Post frente al promedio de ambas mediciones; la línea roja discontinua indica la media de las diferencias y las líneas grises discontinuas los límites de acuerdo ($\pm 1.96 \cdot DE$).

2A (Ofensivo, mayor experiencia): media = +1.00; LoA = -29.97 a +31.97.

2B (Defensivo, mayor experiencia): media = -1.86; LoA = -47.76 a +44.04.

2C (Ofensivo, menor experiencia): media = -1.94; LoA = -50.26 a +46.39.

2D (Defensivo, menor experiencia): media = -6.45; LoA = -34.43 a +21.54.

Discusión

Este estudio diseñó, validó y examinó la sensibilidad del HANDTAC para evaluar principios tácticos ofensivos y defensivos en balonmano, procurando una operacionalización aplicable en cancha y con propiedades de medición explícitas. La propuesta se inscribe en la línea de instrumentos previos que han estructurado el dominio táctico a partir de principios de juego y han documentado su calidad metrológica (Costa et al., 2011; Gaviria-Alzate & Ramírez-González, 2023; Greco et al., 2015; Morales et al., 2012; Sánchez-López et al., 2021; Serra-Olivares & García-López, 2016). En ese sentido, el HANDTAC aporta: (i) validez de contenido alta, (ii) capacidad discriminativa por nivel de experiencia, y (iii) reporte integrado de fiabilidad relativa (ICC) y fiabilidad absoluta (Bland–Altman), conforme a guías de reporte (Bland & Altman, 1986, 1999; Koo & Li, 2016).

La elección del formato de evaluación en juego reducido resulta congruente con evidencia que muestra que estas tareas modifican demandas atencionales y el esfuerzo percibido, sugiriendo que el entorno del SSG intensifica la necesidad de seleccionar información relevante y decidir con rapidez. Bajo esta lógica, un test como HANDTAC se beneficia de un contexto donde el número de interacciones y decisiones por unidad de tiempo aumenta, favoreciendo la sensibilidad de la medición táctica. (Reigal et al., 2023).

Validez de contenido: coherencia conceptual y aplicabilidad

Los CVC $\geq .90$ por ítem y el CVC global sostienen la pertinencia, claridad y representatividad de los principios del HANDTAC cuando son juzgados por entrenadores expertos con formación de posgrado, lo que converge con los hallazgos del FUT-SAT (Costa et al., 2011), con las evaluaciones del TCTP (Greco et al., 2015; Morales et al., 2012) y con la validación Delphi/CVC de FOCOS (Sánchez-López et al., 2021). Además, dialoga directamente con la experiencia acumulada en contextos no convencionales de invasión como el rugby subacuático, donde Gaviria-Alzate & Ramírez-González (2023) reportaron CVC altos en el RUSTAC y mostraron que un marco de principios facilita tanto la codificación como la decisión práctica del entrenador. En términos del argumento de validez (de Vet et al., 2011), esta evidencia respalda de forma consistente la base de contenido y los procesos de respuesta del instrumento.

Sensibilidad y gradientes de pericia

El HANDTAC discriminó entre mayor y menor experiencia en ofensiva y defensiva ($p < .001$; d grandes), un patrón convergente con la literatura que documenta gradientes de pericia cuando los principios están bien definidos y el entorno de medición preserva la información funcional de juego (Gaviria-Alzate & Ramírez-González, 2023; Sánchez-López et al., 2021; Serra-Olivares & García-López, 2016). Este resultado es consistente con la idea, también desarrollada en Gaviria Alzate et al. (2025), de que el rendimiento táctico emerge de la adaptación situada a tareas condicionadas, oponente y contexto, y que la pericia modula la selección y el encadenamiento de principios en tiempo real.

Fiabilidad relativa (ICC) y estructura por niveles

La especificación del modelo ICC (dos vías, acuerdo absoluto; ICC(2,1)/ICC(2,k)) con IC95% cumple estándares de transparencia (Koo & Li, 2016). En paralelo, la concordancia interobservador documentada en la prueba piloto (ICC = 0.91; circulación del balón: ICC = 0.87) respalda la consistencia del registro; por tanto, la menor estabilidad test-retest observada en algunos principios es más compatible con variabilidad situada y restricción de rango que con errores sistemáticos de codificación. En principios específicos se observaron ICC $< .50$ y, en casos de baja ocurrencia (p. ej., interceptaciones y ayudas defensivas), ICC negativos. Este comportamiento está bien descrito en medición aplicada al deporte: restricción de rango, efectos techo/suelo y frecuencias bajas reducen la varianza entre unidades y deprimen los ICC (Atkinson & Nevill, 1998; Liljequist et al., 2019). Por contraste, los índices compuestos (IRT) amortiguan esa inestabilidad al agregar conductas, lo que los vuelve especialmente útiles para comparaciones entre grupos o para seguimiento programático (Gaviria-Alzate & Ramírez-González, 2023).

Fiabilidad absoluta (Bland–Altman) y márgenes esperados

Los Bland-Altman mostraron sesgos medios pequeños con límites de acuerdo amplios en las combinaciones ofensiva/defensiva × experiencia. Este patrón, estabilidad promedio con variabilidad intraindividual, es habitual en variables de comportamiento en juego y complementa a los ICC al ofrecer la reproducibilidad en unidades de la medida (Atkinson & Nevill, 1998; Bland & Altman, 1986, 1999). En estudios tácticos previos (Gaviria Alzate et al., 2025), la coexistencia de estabilidad grupal y fluctuación individual se interpreta como rasgo funcional de un sistema dinámico sujeto a información y restricciones cambiantes.

Naturaleza situada de la táctica y representatividad de la prueba

Siguiendo marcos de dinámica ecológica y pedagogías no lineales, la táctica no es estacionaria ni lineal; emerge de acoplamientos percepción-acción bajo restricciones (Chow et al., 2015; Davids et al., 2013). De ahí que la expectativa de consistencia alta para principios raros pueda ser metodológicamente irrealista. Esta perspectiva, que ya se discute en Gaviria Alzate et al. (2025), sugiere leer la fiabilidad en dos escalas: (a) principios para diagnóstico fino y (b) compuestos para monitorización robusta. Asimismo, la literatura de análisis del rendimiento aconseja múltiples sesiones o ventanas de observación más extensas para mejorar la precisión de estimadores de fiabilidad en contextos situados (Hopkins, 2000; O'Donoghue, 2009).

Limitaciones

Entre las principales limitaciones se encuentra el tamaño muestral acotado (propio de un contexto competitivo sub-21), el muestreo por conveniencia y la procedencia de un único entorno de entrenamiento, lo que limita la generalización a otros niveles, sexos y estructuras competitivas. Además, al tratarse de una tarea en juego reducido, los principios tácticos pueden presentar frecuencia baja y alta dependencia del contexto, lo que tiende a reducir la estabilidad temporal de algunos indicadores. Finalmente, aunque el protocolo incluyó estandarización y entrenamiento de observadores, el estudio se centra en validez de contenido, sensibilidad y estabilidad temporal; aunque se estimó la concordancia interobservador en la prueba piloto, futuros estudios deberían ampliar la doble codificación a una muestra predefinida y suficiente e informar intervalos de confianza para estas estimaciones.

Fortalezas, alcances y proyección

El estudio ofrece: (i) validez de contenido explícita (CVC alto) alineada con la tradición de FUT-SAT, TCTP, FOCOS y RUSTAC (Costa et al., 2011; Gaviria-Alzate & Ramírez-González, 2023; Greco et al., 2015; Morales et al., 2012; Sánchez-López et al., 2021) (ii) sensibilidad con efectos grandes; (iii) reporte estándar de ICC y Bland-Altman; y (iv) transparencia computacional (Python/VS Code), que facilita la replicabilidad. Se plantea como agenda de investigación: (1) incrementar la densidad/ventana de observación para principios de baja ocurrencia; (2) incorporar modelos multinivel (posesiones anidadas en jugadoras/equipos) o Teoría de la Generalizabilidad; y (3) explorar validez de constructo (estructura interna y relaciones con desempeño objetivo) en distintos contextos competitivos. Asimismo, se espera transferir y contrastar el HANDTAC en población masculina y en otras categorías etarias y competitivas, con el fin de evaluar la invariancia del instrumento y su aplicabilidad transversal en el balonmano.

Conclusiones

El HANDTAC representa adecuadamente el dominio táctico del balonmano, discrimina niveles de experiencia y ofrece un panorama realista de fiabilidad; el HANDTAC mostró alta validez de contenido y capacidad discriminativa, pero su estabilidad temporal osciló de pobre a moderada según el índice y el nivel de experiencia; por ello, los índices compuestos parecen más adecuados para comparaciones grupales que para interpretar cambios individuales. Al conservar la arquitectura de referencias del campo, este estudio se posiciona en continuidad con la literatura táctica latinoamericana e internacional y traza una ruta clara para consolidar su validación y transferencia a la práctica.

Aplicaciones prácticas

Las aplicaciones prácticas del HANDTAC se orientan a la toma de decisiones tácticas en contextos de formación y rendimiento. En primer lugar, los índices compuestos (IRT ofensivo y defensivo) permiten la monitorización grupal y el benchmarking entre categorías o niveles de pericia, facilitando además el seguimiento de tendencias a lo largo de los mesociclos de entrenamiento. Complementariamente, los reportes desagregados por principio habilitan una retroalimentación específica para jugadoras y entrenadores, útil para diseñar tareas representativas, como la creación de superioridades, las ayudas y las coberturas, y para articular la planificación con una evaluación formativa que combine resultados pre-post y resúmenes por fase (ofensiva/defensiva), priorizando así contenidos y comprobando si las intervenciones generan los ajustes esperados en la toma de decisiones. En términos de uso, conviene privilegiar los índices compuestos cuando el objetivo es decidir a nivel de programa o de equipo, y recurrir a los principios específicos, con ventanas de observación suficientes, cuando se busca un diagnóstico fino. La interpretación de cambios individuales debe considerar los márgenes de variación evidenciados por los límites de acuerdo y, cuando sea posible, complementarse con métricas como el cambio mínimo detectable (MDC) y el cambio mínimamente importante (SWC).

Estas aplicaciones requieren estandarizar el protocolo (mismas consignas, tiempos, criterios de codificación y formación de evaluadores) y se benefician de la integración sistemática del videoanálisis, que mejora la consistencia de la codificación, promueve el aprendizaje de los evaluadores y genera material pedagógico para la sesión siguiente. En conjunto, el instrumento es aplicable tanto en procesos de formación como de alto rendimiento, favoreciendo la triangulación con métricas físicas y técnicas y guiando la selección de contenidos tácticos prioritarios.

Author Contributions: Conceptualización, D.V.M.G. y S.J.G.A.; metodología, D.V.M.G. y S.J.G.A.; software, S.J.G.A.; validación, D.V.M.G. y S.J.G.A.; análisis estadísticos, S.J.G.A.; investigación, D.V.M.G. y S.J.G.A.; recursos, D.V.M.G.; preparación de datos, D.V.M.G. y S.J.G.A.; preparación del manuscrito (redacción - borrador original), D.V.M.G. y S.J.G.A.; redacción - revisión y edición, D.V.M.G. y S.J.G.A.; visualización, S.J.G.A.; supervisión, S.J.G.A.; administración del proyecto, D.V.M.G. y S.J.G.A. Todos los autores han leído y aceptado la versión publicada del manuscrito.

Referencias

- Andrades-Ramírez, O. A., Alfaro-Castillo, B. S., Martínez-García, D., Cuevas-Aburto, J., Ulloa-Díaz, D. L., & Chirisa-Ríos, L. J. (2022). Fiabilidad y estabilidad del gesto de diferentes variantes de lanzamiento de balonmano. *MHSalud*, 19(2), 1–12. <https://doi.org/10.15359/mhs.19-2.5>
- Araújo, D., Davids, K., & Passos, P. (2007). Ecological Validity, Representative Design, and Correspondence Between Experimental Task Constraints and Behavioral Setting: Comment on Rogers, Kadar, and Costall (2005). *Ecological Psychology*, 19(1), 69–78. <https://doi.org/10.1080/10407410709336951>
- Aróstica Villa, O., Hurtado Rodríguez, J., & Rafael Sebrango Rodríguez, C. (2016). Indicadores para el control y la evaluación de la preparación técnico-táctica en el balonmano. *E-Balonmano.Com: Journal of Sport Science*, 12.

- Arslan, E., Ozer, G. E., & Clemente, F. M. (2020). Running-based high-intensity interval training vs. small-sided game training programs: Effects on the physical performance, psychophysiological responses and technical skills in young soccer players. *Biology of Sport*, 37(2), 165–173. <https://doi.org/10.5114/BIOLOSPORT.2020.94237>
- Atkinson, G., & Nevill, A. M. (1998). Statistical Methods For Assessing Measurement Error (Reliability) in Variables Relevant to Sports Medicine. *Sports Medicine*, 26(4), 217–238. <https://doi.org/10.2165/00007256-199826040-00002>
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England)*, 1(8476), 307–310.
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135–160. <https://doi.org/10.1177/096228029900800204>
- Caldas, I., Tavares, M., & Botelho, E. (2017). Aplicativo para avaliar o nível de conhecimento tático declarativo no handebol. *E-Balónmano: Revista de Ciencias Del Deporte*, 13(2), 77–84.
- Cassidy, T., & Rossi, T. (2006). Situating learning:(Re) examining the notion of apprenticeship in coach education. *International Journal of Sports Science & Coaching*, 1(3), 235–246. <https://doi.org/10.1260/174795406778604591>
- Chacón-MoscOSO, S., Anguera, M., Sanduvete-Chaves, S., Losada, J., Lozano-Lozano, J., & Portell, M. (2019). Methodological quality checklist for studies based on observational methodology (MQCOM). *Psychothema*, 31(4), 458–464. <https://doi.org/10.7334/psychothema2019.116>
- Chow, J. Y., Davids, K., Button, C., & Renshaw, I. (2015). *Nonlinear Pedagogy in Skill Acquisition*. Routledge. <https://doi.org/10.4324/9781315813042>
- Clemente, F., Sarmiento, H., Aquino, R., Loureiro, M., Martins, F., & Mendes, R. (2021). Small-sided games as a methodological resource for team sports: umbrella review of systematic reviews. *PLOS ONE*, 16(2).
- Costa, I. T., Garganta, J., Greco, P. J., Mesquita, I., & Maia, J. (2011). Motricidade Sistema de avaliação tática no Futebol (FUT-SAT): Desenvolvimento e validação preliminar. *Motricidade*, 7(1), 69–84.
- Davids, K., Araújo, D., Vilar, L., Renshaw, I., & Pinder, R. (2013). An Ecological Dynamics Approach to Skill Acquisition: Implications for Development of Talent in Sport Theoretical Principles of Ecological Dynamics. In *An Ecological Dynamics Approach to Skill Acquisition 21 Talent Development & Excellence* (Vol. 5, Issue 1). <http://www.iratde.org>
- de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in Medicine*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511996214>
- Dello Iacono, A., Vigotsky, A. D., Laver, L., & Halperin, I. (2021). Beneficial Effects of Small-Sided Games as a Conclusive Part of Warm-up Routines in Young Elite Handball Players. *Journal of Strength & Conditioning Research*, 35(6), 1724–1731. <https://doi.org/10.1519/JSC.0000000000002983>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Fernández-Espínola, C., Robles, M. T. A., & Fuentes-Guerra, F. J. G. (2020). Small-sided games as a methodological resource for team sports teaching: A systematic review. *International Journal of Environmental Research and Public Health*, 17(6). <https://doi.org/10.3390/ijerph17061884>
- Ferrari, W. R., Sarmiento, H., & Vaz, V. (2019). Match Analysis in Handball: A Systematic Review. *Montenegrin Journal of Sports Science and Medicine*, 8(2), 63–76. <https://doi.org/10.26773/mjssm.190909>
- Feu Molina, S., Godoy Ibáñez, S., Lorenzo Calvo, A., Jiménez Saiz, S., & Cañadas Alonso, M. (2012). El conocimiento profesional adquirido por el entrenador de balonmano: experiencias y formación. *Revista de Psicología Del Deporte*, 21(1), 107–115. <http://www.redalyc.org/articulo.oa?id=235124455014>
- Gabbett, T. J., Nassis, G. P., Oetter, E., Pretorius, J., Johnston, N., Medina, D., Rodas, G., Myslinski, T., Howells, D., Beard, A., & Ryan, A. (2017). The athlete monitoring cycle: a practical guide to interpreting and applying training monitoring data. *British Journal of Sports Medicine*, 51(20), 1451–1452. <https://doi.org/10.1136/bjsports-2016-097298>
- Gaviria Alzate, S. J. O., Valencia Sánchez, W., & Arias Arias, E. A. (2025). Evaluating the Tactical Critical Thinking Program (TPCT) in elite underwater rugby: Effects on tactical efficiency index and critical thinking skills. *International Journal of Sports Science & Coaching*, 20(5), 1958–1971. <https://doi.org/10.1177/17479541251362146>

- Gaviria-Alzate, S., & Ramírez-González, A. (2023). Validación de un test para evaluar la táctica en rugby subacuático. *Viref: Revista de Educación Física*, 12(3).
- González-Víllora, S., García-López, L. M., & Contreras-Jordán, O. R. (2015). Decision making and skill development in youth football players. *International Journal of Medicine & Science of Physical Activity & Sport*, 15(59). <https://doi.org/10.15366/rimcafd2015.59.005>
- Greco, P. J., Pérez Morales, J. C., Aburachid, L. M. C., & Silva, S. R. da. (2015). Evidência de validade do teste de conhecimento tático processual para orientação esportiva - TCTP: OE. *Revista Brasileira de Educação Física e Esporte*, 29(2), 313–324. <https://doi.org/10.1590/1807-55092015000200313>
- Gréhaigne, J. F., Wallian, N., & Godbout, P. (2005). Tactical-decision learning model and students' practices. *Physical Education & Sport Pedagogy*, 10(3), 255–269. <https://doi.org/10.1080/17408980500340869>
- Haynes, S., Richard, D., & Kubany, E. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238–247.
- Hernández, R., Fernández, C., & Baptista, M. (2014). *Metodología de la investigación* (6ta ed.). McGraw-Hill.
- Hernández-Nieto, R. A. (2002). *Contributions to Statistical Analysis*. Universidad de Los Andes.
- Hopkins, W. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, 30(1), 1–15.
- Ibáñez, S. J., & Feu, S. (2026). Research designs in Sports Science. *E-Balonmano Com*, 22(1), 1–44. <https://doi.org/10.17398/1885-7019.22.1>
- International Handball Federation (IHF). (2025). *Reglas de Juego*. Real Federación Española de Balonmano.
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Laguna, M. (2021). *Curso Integral de Balonmano*. . Asociación de Entrenadores de Balonmano de la República Argentina (AEBRA).
- Lebed, F., & Bar-Eli, M. (2013). *Complexity and Control in Team Sports. Dialectics in Contesting Human Systems* (First). Routledge.
- Liljequist, D., Elfving, B., & Skavberg Roaldsen, K. (2019). Intraclass correlation – A discussion and demonstration of basic features. *PLOS ONE*, 14(7), e0219854. <https://doi.org/10.1371/journal.pone.0219854>
- McKay, A. K. A., Stellingwerff, T., Smith, E. S., Martin, D. T., Mujika, I., Goosey-Tolfrey, V. L., Sheppard, J., & Burke, L. M. (2022). Defining Training and Performance Caliber: A Participant Classification Framework. *International Journal of Sports Physiology and Performance*, 17(2), 317–331. <https://doi.org/10.1123/ijsspp.2021-0451>
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research*, 19(4), 539–549. <https://doi.org/10.1007/s11136-010-9606-8>
- Morales, J., Greco, P., & Andrade, R. (2012). Validade de Conteúdo do Instrumento para Avaliação do Conhecimento Tático Processual no Basquetebol. *Cuadernos de Psicología Del Deporte*, 12, 31–36. <http://revistas.um.es/cpd>
- O'Donoghue, P. (2009). *Research Methods for Sports Performance Analysis*. Routledge. <https://doi.org/10.4324/9780203878309>
- Parlebas, P. (1988). *Elementos de sociología del deporte*. Unisport.
- Pasquali, L. (2010). Testes referentes a construto: teoria e modelo de construção. *Instrumentação Psicológica: Fundamentos e Práticas*, 167–198.

- Pedrosa, I., Suárez-Álvarez, J., & García-Cueto, E. (2014). Evidencias sobre la Validez de Contenido: Avances Teóricos y Métodos para su Estimación. *Acción Psicológica*, 10(2), 3–20. <https://doi.org/10.5944/ap.10.2.11820>
- Pinder, R. A., Davids, K., Renshaw, I., & Araújo, D. (2011). Representative Learning Design and Functionality of Research and Practice in Sport. *Journal of Sport and Exercise Psychology*, 33(1), 146–155. <https://doi.org/10.1123/jsep.33.1.146>
- Portell, M., Anguera, M., Chacón-Moscoso, S., & Sanduvete-Chaves, S. (2015). Guidelines for reporting evaluations based on observational methodology. *Psychothema*, 27(3), 283–289. <https://doi.org/10.7334/psychothema2014.276>
- Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1147–1157. <https://doi.org/10.1007/s11136-018-1798-3>
- Quiñones, Y., Morillo, J., Reigal, R., Morales, V., Vázquez, J., & Hernández, A. (2019). El ataque posicional en balonmano: validación de un sistema de observación. *Cuadernos de Psicología Del Deporte*, 19(3), 114–124. <http://revistas.um.es/cpd>
- Reigal, R. E., Delgado-García, G., Melero-Cañas, D., Rishenko, A., & Gamboa-Morales, A. (2023). The effect of the small-sided games based on handball on attention and perceived effort in adolescent students. *Cuadernos de Psicología del Deporte*, 23(2), 107–122. <https://doi.org/10.6018/cpd.563001>
- Renshaw, I., Davids, K., O'Sullivan, M., Maloney, M. A., Crowther, R., & McCosker, C. (2022). An ecological dynamics approach to motor learning in practice: Reframing the learning and performing relationship in high performance sport. *Asian Journal of Sport and Exercise Psychology*, 2(1), 18–26. <https://doi.org/10.1016/j.ajsep.2022.04.003>
- Ribeiro, L., Praça, G., Figueiredo, L., Bredt, S., Torres, J., & Greco, P. (2023). Effects of an implicit-explicit hybrid learning model on handball tactical knowledge. *Human Movement*, 24(2), 25–34. <https://doi.org/10.5114/hm.2023.114911>
- Román, J. (2008). Táctica colectiva grupal en ataque: los modelos en el balonmano español. *E-Balonmano.Com: Revista de Ciencias Del Deporte*, 4(2), 29–51.
- Sánchez Sánchez, R. (2021). El tema de validez de contenido en la educación y la propuesta de Hernández-Nieto. *Am. J. Phys. Educ*, 15(3). <http://www.lajpe.org>
- Sánchez-López, R., Echeazarra, I., & Castellano, J. (2021). Validation of a football competence observation system (Focos), linked to procedural tactical knowledge. *Sustainability (Switzerland)*, 13(12). <https://doi.org/10.3390/su13126780>
- Sarmiento, H., Clemente, F. M., Afonso, J., Araújo, D., Fachada, M., Nobre, P., & Davids, K. (2022). Match Analysis in Team Ball Sports: An Umbrella Review of Systematic Reviews and Meta-Analyses. *Sports Medicine - Open*, 8(1), 66. <https://doi.org/10.1186/s40798-022-00454-7>
- Saw, A. E., Main, L. C., & Gastin, P. B. (2016). Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review. *British Journal of Sports Medicine*, 50(5), 281–291. <https://doi.org/10.1136/bjsports-2015-094758>
- Serra-Olivares, J., & García-López, L. M. (2016). DISEÑO Y VALIDACIÓN DEL TEST DE CONOCIMIENTO TÁCTICO OFENSIVO EN FÚTBOL (TCTOF). *Revista Internacional de Medicina y Ciencias de La Actividad Física y Del Deporte*, 16(63), 521–536. <https://doi.org/10.15366/rimcafd2016.63.008>
- Soto-García, D., & Oliver, J. F. (2022). Aplicación del conocimiento neurocientífico a un modelo sistémico de entrenamiento en balonmano. Una aproximación metodológica. *E-Balonmano.com: Revista de Ciencias del Deporte*, 18(3), 201–210.
- Travassos, B., Duarte, R., Vilar, L., Davids, K., & Araújo, D. (2012). Practice task design in team sports: Representativeness enhanced by increasing opportunities for action. *Journal of Sports Sciences*, 30(13), 1447–1454. <https://doi.org/10.1080/02640414.2012.712716>
- Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., Poole, C., Schlesselman, J. J., & Egger, M. (2014). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *International Journal of Surgery*, 12(12), 1500–1524. <https://doi.org/10.1016/j.ijsu.2014.07.014>
- Viru, A., & Viru, M. (2003). *Análisis y control del rendimiento deportivo* (1ra ed.). Paidotribo.

- World Medical Association. (2013). Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*, 310(20), 2191. <https://doi.org/10.1001/jama.2013.281053>
- Zapardiel, J. C., Rodríguez-Hernández, M. L., González-Moreno, P. M., & Sánchez-Sánchez, F. (2025). Development of an instrument for the observation of collective tactical procedures of positional attack in beach handball. *E-Balonmano Com*, 21(4), 495–514. <https://doi.org/10.17398/1885-7019.21.495>