

Menescardi, C., Ortega-Benavent, N., Salazar-Cruz, P., Belenguer-Troya, A. y Estevan, I. (2026).
Validación de un instrumento de observación para evaluar la competencia motriz del alumnado.
Cuadernos de Psicología del Deporte, 26(3), 492-505

Validación de un instrumento de observación para evaluar la competencia motriz del alumnado

Validation of an Observation Instrument to Assess Students' Motor Competence

Validação de um instrumento de observação para avaliar a competência motora do alumnado

Menescardi, Cristina¹, Ortega-Benavent, Núria^{1,2}, Salazar-Cruz, Pamela^{1,3}, Belenguer-Troya, Adrián¹, Estevan, Isaac¹

¹*Grupo de investigación AFIPS, Universidad de Valencia;* ²*Florida Universitaria, España;* ³*Universidad de Costa Rica, Costa Rica*

RESUMEN

El objetivo de este estudio fue desarrollar una versión abreviada del test de Desarrollo Motor Grueso-3 (TGMD-3), uno de los test más populares a nivel mundial, y examinar su validez, fiabilidad inter-observador y de la herramienta de observación creada. Se analizaron datos de 217 estudiantes de primaria españoles. Se utilizó un análisis factorial confirmatorio para investigar la validez de reducir el número de habilidades del TGMD-3, así como el índice de correlación intraclase para analizar la fiabilidad de las personas observadoras y la consistencia interna del test. Asimismo, se empleó la Teoría de la Generalizabilidad para valorar tanto la fiabilidad de las personas observadoras como de la herramienta de observación, siendo ambos tipos de análisis complementarios. La versión corta del TGMD-3 con seis habilidades presenta índices adecuados de validez factorial confirmatoria, consistencia interna, fiabilidad inter-observador y calidad de los datos obtenidos con la herramienta de observación que cuenta con categorías exhaustivas y mutuamente excluyentes. A partir de estos datos, las personas investigadoras y profesionales de la educación disponen de una versión corta, válida y fiable del TGMD-3 para la evaluación de la competencia motriz en población española infantil, que posibilita su uso en futuros trabajos con fines de detección niveles de desarrollo motor típico o atípico, así como para medir su evolución a lo largo de la etapa.

Palabras clave: Competencia motriz; Validez; Fiabilidad; Metodología observacional.

ABSTRACT

The aim of this study was to develop a shortened version of the Test of Gross Motor Development-3 (TGMD-3), one of the most widely used assessments of gross motor competence worldwide, and to examine its validity, inter-rater reliability, and the reliability of the observation instrument. Data were collected from 217 Spanish primary school children. Confirmatory factor analysis was conducted to examine the validity of the proposed reduction in the number of TGMD-3 skills. In addition, the intraclass correlation coefficient was used to assess inter-rater

reliability and internal consistency, while Generalizability Theory was applied to evaluate the reliability of both the observers and the observation instrument. These complementary analyses provided evidence of the measurement quality of the proposed short version. The shortened six-skill version of the TGMD-3 demonstrated adequate factorial validity, internal consistency, inter-rater reliability, and high-quality observational data, supported by an observation instrument with exhaustive and mutually exclusive categories. These findings provide researchers and education professionals with a valid and reliable abbreviated version of the TGMD-3 for assessing motor competence in Spanish children, facilitating its use in future studies to identify typical and atypical motor development and to monitor changes in motor competence throughout primary education.

Keywords: Motor competence; Validity; Reliability; Observational methodology.

RESUMO

O objetivo deste estudo foi desenvolver uma versão abreviada do Teste de Desenvolvimento Motor Grosso-3 (TGMD-3), um dos testes mais populares a nível mundial, e examinar sua validade, fiabilidade interavaliador e a fiabilidade do instrumento de observação criado. Foram analisados dados de 217 estudantes do ensino fundamental espanhóis. Utilizou-se análise fatorial confirmatória para investigar a validade da redução do número de habilidades do TGMD-3, bem como o índice de correlação intraclasse para analisar a fiabilidade dos avaliadores e a consistência interna do teste. Além disso, empregou-se a Teoria da Generalizabilidade para avaliar tanto a fiabilidade dos avaliadores quanto do instrumento de observação, sendo ambos os tipos de análise complementares. A versão reduzida do TGMD-3, composta por seis habilidades, apresenta índices adequados de validade fatorial confirmatória, consistência interna, fiabilidade interavaliador e qualidade dos dados obtidos por meio do instrumento de observação, que conta com categorias exaustivas e mutuamente exclusivas. A partir desses resultados, pesquisadores e profissionais da educação dispõem de uma versão reduzida, válida e fiável do TGMD-3 para a avaliação da competência motora em crianças espanholas, possibilitando sua utilização em estudos futuros com fins de identificação de níveis de desenvolvimento motor típico ou atípico, bem como para medir sua evolução ao longo da etapa escolar.

Palavras chave: Competência motora; Validade; Fiabilidade; Metodologia observacional.

INTRODUCCIÓN

Hoy en día, una de las variables con gran relevancia en la práctica y el aprendizaje de la actividad física del alumnado es su competencia motriz real (CMR, Cairney et al., 2019). La CMR se puede definir como el nivel de destreza del estudiantado en la ejecución de habilidades motrices (por ejemplo, de locomoción, de control de objetos y de estabilidad) (Stodden et al., 2008), que se consideran los cimientos de movimientos más avanzados requeridos para participar exitosamente en deportes y juegos (Hulteen et al., 2018; Logan et al., 2018). Una definición más amplia es la propuesta por Bandeira et al. (2025), que sugiere que la CMR es un fenómeno complejo y multidimensional, caracterizado por la capacidad de reunir recursos de diferentes dominios (i.e., físicos, cognitivos, sociales, emocionales y ambientales) para realizar una tarea motriz con facilidad. La CMR se asocia, a su vez, con un peso saludable o una mejor composición corporal (Barnett et al., 2022), una buena condición física (Stodden et al., 2008; Robinson et al., 2015), una mayor percepción de la CMR (Estevan, et al., 2021), así como mejores habilidades cognitivas y académicas (Jylänki et al., 2022; Zeng et al., 2017). A pesar de su relevancia, en los últimos años ha habido una disminución en los niveles de práctica de actividad física y CMR del alumnado (Huotari et al., 2018; John et al., 2024), siendo el nivel de CMR del alumnado bajo (Bardid et al., 2015; Cao et al., 2020; Menescardi et al., 2022) lo que hace necesario seguir investigando qué factores determinan su desarrollo.

Para fomentar que una persona tenga una buena CMR es necesario disponer de test que permitan identificar qué aspectos se deben mejorar. Estos test o pruebas estandarizadas permiten medir el nivel de destreza del alumnado.

Validación de un instrumento de competencia motriz

Dentro del ámbito educativo, para valorar la CMR del alumnado, los métodos existentes, los instrumentos de observación parecen ser los más convenientes por su coste asequible y facilidad de implementación; asimismo, estos proporcionan una retroalimentación más específica sobre la calidad de la ejecución de la habilidad motriz (Bardid et al., 2019). Este tipo de test son fáciles de administrar, requieren de poco instrumental y pueden llevarse a cabo en diversos espacios (e.g., abiertos o cerrados) (Valentini, 2024).

Además, para que sea viable su aplicación dentro del ámbito educativo, es necesario elegir test que sean sencillos y rápidos de implementar (e.g., versiones cortas como la del *Test of Gross Motor Development*, TGMD) (Bandeira et al., 2020; Duncan et al., 2022; Valentini et al., 2018). En este sentido, las versiones cortas de los test permiten una detección más eficiente del desarrollo motor del alumnado en el ámbito escolar, siendo más rentable recopilar datos de grandes muestras (Valentini et al., 2018). Hasta la fecha únicamente hay tres versiones cortas del test TGMD, dos de la segunda versión (TGMD-2, Bandeira et al., 2020; Valentini et al., 2018) y una de la tercera (TGMD-3, Duncan et al., 2022). Estas han sido validadas en población brasileña preescolar (Bandeira et al., 2020, Valentini et al., 2018) e infantil (Valentini et al., 2018) e irlandesa infantil (Duncan et al., 2022). Cabe destacar que, aunque dichas validaciones partan del test TGMD, y aunque coinciden en que las de mayor relevancia dentro de las habilidades de locomoción y control de objetos son el galope y el salto a una pierna, así como el bateo y el lanzamiento, respectivamente, estos trabajos no coinciden completamente en las habilidades que hallaron como más discriminantes. Por ello, es necesario seguir investigando sobre qué habilidades son las más discriminantes en función de la edad y localización de la población objeto de estudio.

Puesto que la versión corta del test TGMD no está validada en población española, se hace necesario analizar la adecuación de la herramienta observacional y la fiabilidad de las personas observadoras y de los datos obtenidos. Debido a los múltiples factores que pueden afectar el comportamiento del alumnado, hay que asegurarse de que las observaciones realizadas son fiables, válidas y precisas. Por ello, el objetivo del presente estudio es validar la versión corta del test TGMD-3 con alumnado español. Para analizar las propiedades psicométricas de la versión abreviada se combinaron dos enfoques complementarios. Por un lado, se realizó un análisis factorial para evaluar la validez estructural del modelo propuesto. Por otro lado, se empleó la Teoría de la Generalizabilidad (TG) que ha sido considerada como una teoría de los errores multifaceta al asumir que cualquier situación de medida posee infinitas fuentes de variación (denominadas facetas) (Cronbach et al., 1972); así como una extensión de la Teoría Clásica de los Test que utiliza los procedimientos del Análisis de la Varianza y de los diseños experimentales (Martínez, 1995). Una de las grandes aportaciones de esta teoría es la unificación de los conceptos de fiabilidad, validez y precisión (Blanco-Villaseñor et al., 2014; Morales-Sánchez et al., 2016).

MATERIAL Y MÉTODOS

Participantes

Participaron en el estudio un total de 217 estudiantes (103 niñas; 44.2%) de primero a cuarto curso de 5 centros en la Comunidad Valenciana, con edades comprendidas entre los 6 y 11 años (media: 7.38 ± 1.83 años). Se obtuvo el consentimiento por escrito de los padres, las madres o tutores de todas las personas participantes antes de su participación. Este estudio fue aprobado por el Comité de Ética de la universidad del primer autor (H1446557620395).

Instrumento

Se empleó el TGMD-3 (Estevan et al., 2017; Ulrich, 2017; Webster y Ulrich, 2017) que es uno de los test más empleados a nivel mundial (Logan et al., 2018; Valentini, 2024), y se puede emplear con grandes rangos de edad (3-11 años, Estevan et al., 2017). Además, este test ofrece información sobre cómo el alumnado realiza las diferentes habilidades, ya que es un test orientado al proceso (Bardid et al., 2019). Dicha evaluación se puede

utilizar como guía para proporcionar información que apoye la enseñanza y la instrucción posteriores (Van Rossum et al., 2021). Esta tercera versión del TGMD comprende 13 habilidades motrices divididas en habilidad de locomoción (i.e., carrera [TGMD1], galope [TGMD2], salto a un pie [TGMD3], *skipping* [TGMD4], salto horizontal [TGMD5], desplazamiento lateral [TGMD6]) y de control de objetos (i.e., bateo a dos manos [TGMD7], golpeo con raqueta [TGMD8], bote de balón [TGMD9], recepción a dos manos [TGMD10], chut a una pared [TGMD11], lanzamiento por encima [TGMD12] y por debajo del hombro [TGMD13]). La elección de este test radica en que está compuesto por habilidades relacionadas con la motricidad gruesa del alumnado y que son las habilidades que suelen desarrollar cuando juegan activamente y practican actividad física en la escuela, lo que apoya la validez ecológica de este test (Valentini, 2024).

Procedimiento

Todas las personas participantes completaron el test TGMD-3 validado al español (Estevan et al., 2017). La implementación del TGMD-3 se realizó en un gimnasio de Educación Física o patio escolar por un asistente de investigación entrenado (último autor). El visionado y codificación se realizó por tres personas observadoras diferentes, un observador experto y dos previamente entrenadas. Para ello, estos últimos realizaron una formación de 2h basada en la familiarización con las pruebas del TGMD-3 y visionado de vídeos para comprobar que su codificación era fiable (acuerdo > 80% con el experto): <https://sites.google.com/view/tgmd-3/videos?authuser=0>. Posteriormente, para comprobar su fiabilidad a lo largo del visionado, se analizó su fiabilidad con el visionado de los intentos de cuatro sujetos. El desempeño de cada niño/a fue grabado en video a 25 Hz con una cámara Lumix TZ7 (Panasonic, Japón ©) para su posterior codificación.

Para analizar las diferentes habilidades, se creó una herramienta de observación acorde a un sistema de categorías, con 13 criterios correspondientes a las diferentes habilidades del test completo del TGMD-3 (i.e., bateo a dos manos, golpeo con raqueta, bote de balón, recepción a dos manos, chute a una pared, lanzamiento por encima y por debajo del hombro, carrera, galope, salto a un pie, *skipping*, salto horizontal, desplazamiento lateral).

Análisis estadístico

La fiabilidad de la codificación en el registro por parte de las personas observadoras se corroboró mediante el cálculo del acuerdo inter-observador y un estudio de generalizabilidad (Blanco-Villaseñor et al., 2014; Morales-Sánchez et al., 2016).

Para validar la versión corta del TGMD-3 se siguió el método propuesto por Duncan et al. (2022) y Valentini et al. (2018) y se realizó un análisis factorial confirmatorio con el modelo completo y, posteriormente, con aquellas habilidades que presentaron mayor carga factorial ($\lambda \geq 0.65$). Para ello, se emplearán las habilidades más representativas del TGMD: salto a una pierna, chute, bote, bateo, desplazamiento lateral y *skipping* (Bandeira et al., 2020; Duncan et al., 2022; Menescardi et al., 2022; Valentini et al., 2018). Además, para evaluar la calidad de los ajustes de los modelos probados, se utilizaron el índice de ajuste comparativo (*Comparative Fit Index*, CFI) y el error cuadrático medio de aproximación (*Root mean squared error of approximation*, RMSEA). Los valores de RMSEA que oscilaron entre 0.05 y 0.08, así como valores de CFI > 0.90 indicaron un buen ajuste (Hu y Bentler, 1999). Para evaluar la calidad del modelo de medida, se calculó la varianza media extraída (*Average Variance Extracted*, AVE) como indicador de validez convergente, considerando adecuados valores ≥ 0.50 (Fornell y Larcker, 1981). Asimismo, se estimó la fiabilidad compuesta (*Composite Reliability*, CR) para evaluar la consistencia interna de cada factor, considerándose aceptables valores ≥ 0.70 (Hair et al., 2019). Finalmente, la validez discriminante entre los factores se examinó mediante el índice Heterotrait–Monotrait (HTMT), aceptándose valores < 0.85 como evidencia de una adecuada diferenciación entre constructos (Henseler et al., 2015).

La invarianza de medición del modelo con la versión corta se probó mediante un análisis multigrupo (niños y niñas). El análisis multigrupo incluyó dos modelos, por un lado, la invarianza configuracional, para analizar si el número de factores e ítems en cada factor era aceptable para niños y niñas y, por otro lado, la invarianza métrica, para analizar si las cargas factoriales eran similares para niños y niñas. Se adoptaron puntos de corte recomendados ($\Delta\text{RMSEA} < 0.015$) para respaldar el supuesto de invarianza (Chen, 2007). Los análisis factoriales se realizaron con el software Mplus. Seguidamente, se utilizó el test completo de 13 habilidades para analizar la validez

Validación de un instrumento de competencia motriz

concurrente y valorar la validación de la versión corta del TGMD-3, a través de un análisis de correlación (e.g., r de Pearson). Posteriormente y para evaluar la consistencia interna, se utilizaron las puntuaciones del alfa de Cronbach, considerándose apropiados valores superiores a 0.70 (Bland y Altman, 1997).

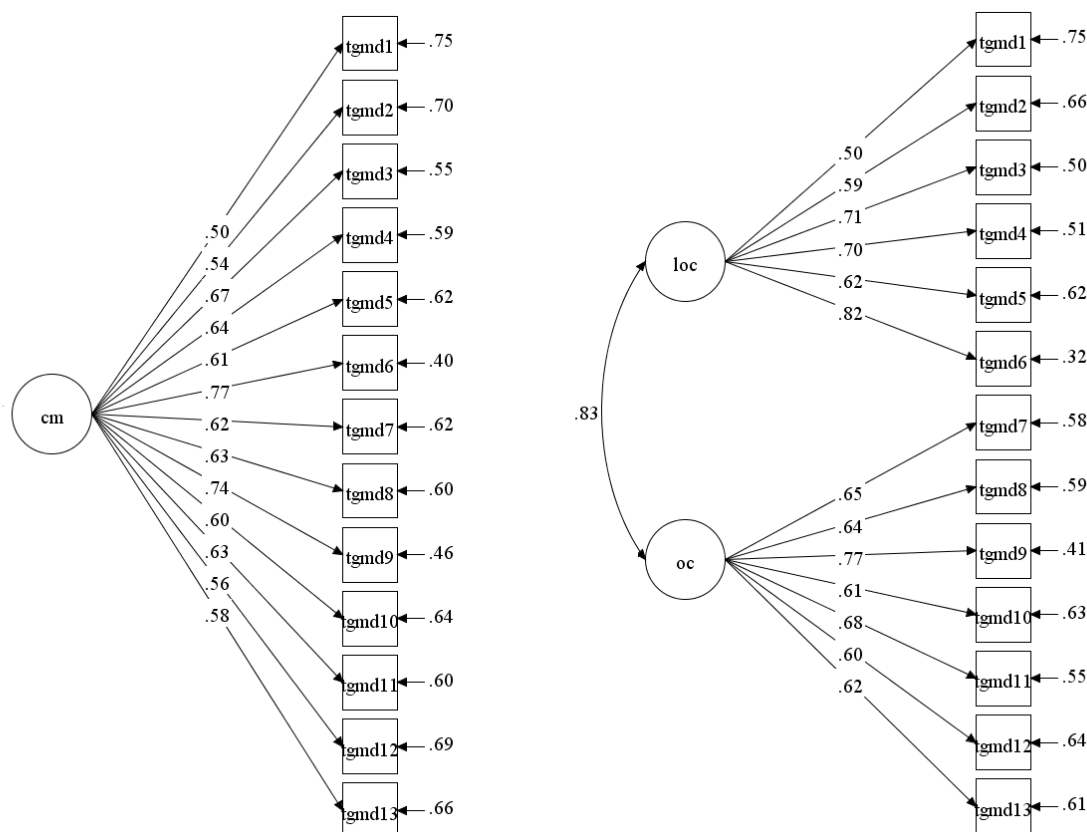
Además, se empleó la Teoría de la Generalizabilidad, que permite conocer si el diseño de medida planteado contiene fuentes de error que estén influyendo en la fiabilidad del proceso de medición y el peso de la misma. Para ello, se consideraron como fuentes de error a las personas observadoras que codifican las grabaciones y la herramienta de observación. El análisis de generalizabilidad se llevó a cabo mediante el software SAGT (Hernández-Mendo et al., 2016), poniendo a prueba el modelo $o \times c$ (siendo 'o' observadores y 'c' categorías). De este modelo se estimaron dos planes de medida, relativos a los objetivos del presente artículo: (1) el plan de medida C/O que permite valorar la concordancia entre observadores para saber si son fiables; y (2) el plan O/C se empleó para comprobar el buen ajuste de las categorías (i.e., si son exhaustivas y mutuamente excluyentes, E/ME), y asegurar que la herramienta observacional permite obtener datos fiables. La exhaustividad se refiere al hecho de incluir en el sistema de categorías cualquier comportamiento del ámbito considerado como objeto de estudio y al cual puede asignársele una de las categorías. Y la mutua exclusividad significa el no solapamiento de las categorías que componen un sistema, por lo que a cada comportamiento se le asignaría una y sólo una categoría (Anguera et al., 2007). En cada uno de estos modelos, el estudio de generalizabilidad proporciona dos coeficientes: el coeficiente de generalizabilidad (e^2) y el coeficiente de fiabilidad (Φ) (Brennan, 2003). Estos coeficientes oscilan entre 0 y 1, donde 1 representa la fiabilidad perfecta mientras que valores entre 0.70 y 0.90 representan fiabilidades óptimas (Lafave y Butterwick, 2014). Valores cercanos a 0.0, indican que los datos no son fiables. Por su parte, la condición de E/ME de las categorías requiere que los valores tiendan a cero (Menescardi et al., 2017).

RESULTADOS

El análisis confirmatorio mostró que el resultado con dos factores presentaba mejor ajuste ($\chi^2 = 125.350$; CFI= 0.94; TLI: 0.93; RMSEA= 0.07; SRMR= 0.05; $AVE_{loc} = 0.44$; $CR_{loc} = 0.82$; $AVE_{oc} = 0.43$; $CR_{oc} = 0.84$) que con un factor ($\chi^2 = 173.783$; CFI= 0.90; TLI: 0.87; RMSEA= 0.09; SRMR= 0.05; $AVE = 0.38$; $CR = 0.89$). Este análisis indicó que los ítems TGMD1, TGMD2 y TGMD5 eran los que menor carga factorial presentaban dentro de las habilidades de locomoción, mientras que los ítems TGMD8, TGMD10, TGMD12 y TGMD13 lo eran dentro de las habilidades de control de objetos (ver Figura 1).

Figura 1

Análisis factorial confirmatorio del modelo de uno (a) y dos factores (b) de la versión completa del TGMD.



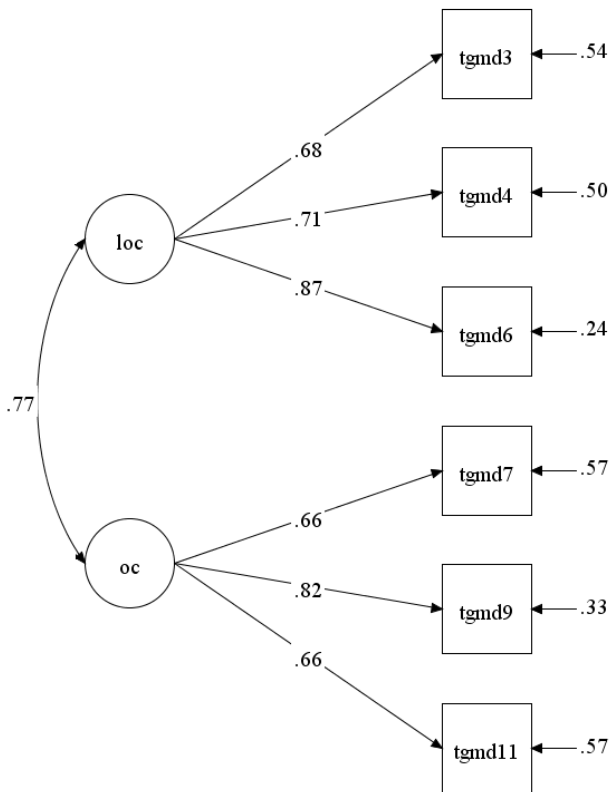
Nota. CM= Competencia Motriz, Oc= Control de objetos, Loc = Locomoción, TGMD1 = carrera, TGMD2 = galope, TGMD3 = salto a un pie, TGMD4 = skipping, TGMD5= salto horizontal, TGMD6 = desplazamiento lateral, TGMD7 = bateo, TGMD8= golpeo con raqueta, TGMD9 = bote, TGMD10= recepción a dos manos; TGMD11 = chute, TGMD12 y TGMD13 = lanzamiento por arriba y por debajo del hombro.

El modelo reducido con seis habilidades mostró un buen ajuste ($\chi^2 = 5.552$; CFI= 1.00; TLI = 1.00; RMSEA= 0.00; SRMR= 0.016; AVE_{loc} = 0.57; CR_{loc} = 0.80; AVE_{oc} = 0.51; CR_{oc} = 0.76; HTMT = 0.78) con cargas factoriales que oscilaron entre 0.66 y 0.87 (ver figura 2). El análisis multigrupo realizado indicó que el modelo sin restricciones demostró invariancia de configuración para el sexo (CFI = 1.00, RMSEA = 0.000), lo que indica que el modelo propuesto inicialmente era aceptable para niños y niñas. No hubo una reducción significativa en los segundos ajustes del modelo en comparación con el primer modelo para niños y niñas (Δ CFI = 0.000; Δ RMSEA = 0.000). Por tanto, los ítems presentados tienen patrones de carga factorial similares para niños y niñas, siendo equivalentes para ambos grupos.

Validación de un instrumento de competencia motriz

Figura 2

Análisis factorial confirmatorio del modelo de dos factores de la versión corta de seis habilidades del test TGMD.



Nota. Loc = habilidades de locomoción, oc = habilidades de control de objetos, TGMD3 = salto a un pie, TGMD4 = skipping, TGMD6 = desplazamiento lateral, TGMD7 = bateo, TGMD9 = bote, TGMD11 = chute.

Los coeficientes de correlación intraclase fueron moderados-excelentes para TGMD completo total (de 13 habilidades: ICC= 0.98), así como para las habilidades de control de objetos (ICC= 0.95) y locomoción (ICC= 0.82). Para la versión corta (Figura 2) también se obtuvieron valores moderados-excelentes de TGMD corta total (de 6 habilidades: ICC= 0.95), control de objetos (ICC = 0.88) y locomoción (ICC = 0.92) (Fleiss, 1981). Asimismo, se obtuvo una correlación de 0.95 entre la versión completa y corta del TGMD-3. Finalmente, en la versión larga, se obtuvieron valores de concordancia interna de 0.89 para las 13 habilidades, 0.82 para las de locomoción y 0.84 para las de control de objetos. En la versión corta, se obtuvo un alfa de Cronbach de 0.83 al considerar las seis habilidades, mientras que para las habilidades de locomoción fue de 0.75 y de 0.70 para las de control de objetos.

Por lo que respecta al análisis de generalizabilidad, tanto al modelo con 13 habilidades (Figura 1) como al obtenido en la versión corta (Figura 2), en ambos el plan de medida O/C arrojó valores de 0.00 en ambos coeficientes (e^2 y Φ), cumpliendo la condición de ser E/ME. Por lo que respecta al plan C/O este dio valores de 0.85, en ambos coeficientes, para el modelo con 13 habilidades y de 0.97 para el modelo con seis habilidades (Figura 2), confirmando la fiabilidad óptima de las personas observadoras, en la línea con los valores de ICC mostrados.

DISCUSIÓN

El objetivo del presente estudio fue validar una herramienta de observación en base a la versión corta del test TGMD-3 en población española que permita a las personas investigadoras llevar a cabo futuros análisis con la adecuada calidad, en términos de fiabilidad y validez. Validar versiones cortas de los test permite una detección más eficiente del desarrollo motor del alumnado y hace más viable la recogida de datos en el ámbito educativo (Valentini et al., 2018). Para ello, se llevó a cabo un análisis factorial, que permite conocer la estructura factorial del test, así como un análisis de validez convergente, discriminante y de consistencia interna. Posteriormente, se llevó a cabo un análisis de generalizabilidad que permite a las personas investigadoras tener garantías de fiabilidad en cuanto a la herramienta de observación utilizada y la fiabilidad de las codificaciones realizadas por las personas observadoras (Blanco-Villaseñor et al., 2014; Cronbach et al., 1972; Wickel y Welk, 2010), siendo ambos análisis complementarios.

Los resultados del presente trabajo muestran evidencias de validez la versión corta del TGMD-3 compuesta por seis habilidades (i.e., salto a un pie, skipping, desplazamiento lateral, bateo, bote y chute), ya que el análisis factorial presenta un ajuste adecuado y es invariante entre niños y niñas. Este modelo reducido, además, presenta una validez convergente adecuada, ya que ambos factores superan el umbral de $AVE \geq 0.50$ (Fornell y Larcker, 1981) y muestran una fiabilidad compuesta satisfactoria ($CR > 0.70$; Hair et al., 2019). También, se obtuvo una adecuada validez discriminante entre factores ($HTMT < 0.85$) (Henseler et al., 2015), que indica que ambos factores representan constructos empíricamente distintos. Además, los resultados de la consistencia interna fueron adecuados ($\alpha > 0.70$) y los de la validez convergente mostraron una alta concordancia entre la versión larga del test (13 habilidades) y la versión corta (6 habilidades) ($r = 0.95$), lo que refuerza la hipótesis de que la versión corta permite realizar mediciones con un coste-tiempo menor (Valentini, 2024). Estos resultados van en la línea de los trabajos previos publicados sobre las versiones cortas del test TGMD (versión 2, Valentini et al., 2018 y versión 3, Duncan et al., 2022) que ofrecen alternativas más eficientes para, por ejemplo, docentes de Educación Física.

Cabe destacar que de las habilidades de la versión corta (Figura 2), tres de ellas (i.e., salto a un pie, bateo y bote) coinciden con las incluidas por Duncan et al. (2022), que es el único estudio donde se ha validado una versión corta de la tercera versión del TGMD en alumnado de primaria. Las pruebas que estos autores consideraron más relevantes fueron las de galope, salto horizontal, salto a un pie, recepción, bote de balón, lanzamiento por encima y bateo), reduciendo las habilidades a siete. Es decir, el modelo propuesto por Duncan et al. (2022) priorizaba el salto horizontal (TGMD5) frente al skipping y desplazamiento lateral (TGMD4 y TGMD6) que en el presente estudio han sido consideradas como las más relevantes. Mientras que, en las habilidades de control de objetos, Duncan et al. (2022) priorizan la recepción de balón (TGMD10) frente al chute (TGMD11). Estas diferencias resultan especialmente interesantes, ya que sugieren que la capacidad discriminativa de determinadas habilidades puede depender del contexto cultural y deportivo en el que se desarrolla la competencia motriz. En el contexto español, el fútbol constituye uno de los deportes con mayor participación durante la infancia y forma parte habitual tanto de las actividades extraescolares como del juego espontáneo (Franco et al., 2022; Isorna et al., 2014), lo que podría favorecer una mayor variabilidad en el rendimiento asociado al chute y, por tanto, incrementar su capacidad para discriminar entre diferentes niveles de competencia motriz. Por el contrario, la recepción con dos manos podría representar una habilidad más básica y ampliamente adquirida en estas edades, reduciendo su capacidad discriminativa dentro de esta población. En cambio, en la muestra irlandesa analizada por Duncan et al. (2022), la recepción mostró un mayor peso dentro de la estructura factorial, posiblemente debido a diferencias en las experiencias motrices predominantes y en la práctica de deportes escolares, donde disciplinas como el fútbol gaélico, el hurling o el rugby implican una mayor utilización de habilidades de recepción y manejo del balón con las manos (Reilly y Collins, 2008; Woods et al., 2023). No obstante, estas interpretaciones deben considerarse con cautela, ya que el presente estudio no evaluó de forma específica la práctica deportiva ni la exposición previa a diferentes actividades físicas. En consecuencia, las diferencias observadas entre ambas versiones requieren ser confirmadas en futuras investigaciones transculturales que analicen la invariancia de la estructura factorial del TGMD-3 y el efecto de la práctica deportiva sobre la relevancia de cada habilidad.

Validación de un instrumento de competencia motriz

A pesar de estas diferencias, la versión abreviada propuesta por Duncan et al. (2022) mostró un ajuste aceptable cuando se aplicó a la muestra española (véase material suplementario, Figura S1), aunque inferior al obtenido con el modelo de seis habilidades propuesto en el presente estudio (Figura 2). Este resultado indica que ambas versiones presentan propiedades psicométricas adecuadas; sin embargo, la versión desarrollada en población española ofrece un mejor ajuste para este contexto específico. Por ello, se recomienda su utilización en población española, ya que presenta un mejor comportamiento psicométrico y una mayor capacidad para representar las características propias de esta población. No obstante, el adecuado ajuste de la versión de Duncan et al. (2022) también sugiere que podría emplearse en estudios internacionales y comparaciones transculturales, favoreciendo la armonización de resultados entre diferentes países.

Asimismo, de las habilidades encontradas como más discriminantes en población española (Figura 2), tres (i.e., salto, bateo y chute) coinciden con la versión corta del TGMD-2 propuesta por Valentini et al. (2018). De acuerdo con la propuesta de Valentini et al. (2018), se observa que los autores consideran importante las habilidades de chute, a diferencia de Duncan et al. (2022), al igual que se ha encontrado en el presente estudio. Del mismo modo, el trabajo presentado por Bandeira et al. (2020) muestra que para los preescolares brasileños, dentro de las habilidades de locomoción, la habilidad de desplazamiento lateral es una de las que más discriminan, tal como se muestra en el presente estudio. Por otro lado, al igual que los datos mostrados en el estudio de Menescardi et al. (2022), en el presente trabajo se mostró que la habilidad de skipping es una de las que más discriminan en alumnado mayor de primaria. Las similitudes encontradas entre la muestra brasileña y la española podrían estar relacionadas con las oportunidades de práctica y de las experiencias motrices adquiridas durante la infancia (e.g., práctica del fútbol), así como al hecho de que los currículos de Educación Física de España y Brasil promueven el desarrollo de habilidades motrices básicas mediante juegos y deportes colectivos, lo que podría contribuir a patrones de desarrollo motor similares. En conjunto, estos resultados sugieren que la relevancia de determinadas habilidades puede variar en función del contexto sociocultural y de las experiencias motrices de la población evaluada. Por ello, la estructura factorial de las versiones abreviadas del TGMD debería confirmarse en cada población antes de su aplicación, ya que no todas las habilidades presentan la misma capacidad discriminativa y esta puede verse influida por las características culturales, deportivas y educativas de cada contexto (Barnett et al., 2016; Newell, 2020; Duncan et al., 2022).

Además del empleo del análisis factorial, que es un enfoque estadístico clásico para evaluar la estructura interna del instrumento, se llevó a cabo un análisis de generalizabilidad considerando dos fuentes de error (observadores 'o' y categorías 'c') sobre las ejecuciones motrices de 217 estudiantes de Educación Primaria. Este análisis pretendió especificar el acuerdo y la fiabilidad interobservador (C/O) y evaluar la precisión de las categorías E/ME presentes en la herramienta observacional para obtener datos fiables (O/C). En cuanto a la fiabilidad de las personas observadoras, el coeficiente de correlación intraclase ($ICC = 0,95$) mostró un nivel de acuerdo excelente entre observadores. Estos resultados fueron consistentes con los obtenidos mediante la Teoría de la Generalizabilidad, donde el plan de estimación C/O arrojó un coeficiente de generalizabilidad ($e^2 = 0,97$), confirmando igualmente una elevada fiabilidad de las observaciones. La convergencia entre ambos enfoques, el estadístico clásico (ICC) y la Teoría de la Generalizabilidad, aporta una evidencia sólida de la consistencia y estabilidad del proceso de observación, reforzando la calidad metodológica de los datos obtenidos. Asimismo, los autores coinciden con estudios previos (McKenzie, 2010; Medina y Delgado, 1999; Menescardi et al., 2017), en la importancia de la formación previa de las personas observadoras para garantizar la objetividad y fiabilidad necesaria de los datos recogidos mediante metodología observacional.

Por último, en relación a la precisión y validez de las categorías de la herramienta de observación (plan de estimación O/C), los valores obtenidos con el modelo tienden a cero, lo que muestra que las categorías son exhaustivas y mutuamente excluyentes (E/ME), es decir, heterogéneas (Blanco-Villaseñor et al., 2014; Menescardi et al., 2017) y nos permiten asumir el buen ajuste de la herramienta de observación, debido a los valores óptimos de los coeficientes de generalizabilidad, dotando de fiabilidad al instrumento de observación (Lafave y Butterwick, 2014). Este resultado constituye una fortaleza metodológica de la herramienta observacional, ya que garantiza que cada conducta observada puede clasificarse de forma inequívoca en una única categoría, minimizando la

ambigüedad durante el proceso de codificación y reduciendo posibles fuentes de error asociadas al sistema de observación. En consecuencia, la estructura categorial empleada proporciona una base sólida para la obtención de datos objetivos y fiables, reforzando la calidad del instrumento de observación y la validez de las evaluaciones realizadas (Lafave y Butterwick, 2014).

CONCLUSIONES

En el presente estudio proporcionamos una herramienta de observación válida y fiable de la versión corta del TGMD-3, a través del empleo de análisis clásicos y el estudio de generalizabilidad, en una muestra de estudiantes de primaria españoles. Esto permite una evaluación más eficiente del desarrollo motor infantil en las escuelas y una forma más rentable de recopilar datos de muestras grandes, ofreciendo las bases para realizar futuros trabajos sobre la CMR del alumnado en primaria.

APLICACIONES PRÁCTICAS

La validación de este instrumento proporcionará al profesorado de Educación Física una herramienta válida, fiable y de fácil aplicación para evaluar la competencia motriz del alumnado en el contexto escolar. Su utilización permitirá identificar necesidades motrices, realizar un seguimiento de la evolución del alumnado, orientar la planificación docente y evaluar la eficacia de programas de intervención dirigidos a promover la competencia motriz y la actividad física. Asimismo, constituirá una herramienta de utilidad para la investigación y para la toma de decisiones educativas basadas en evidencias.

AGRADECIMIENTOS

Este estudio forma parte de un proyecto que ha sido financiado por el ministerio de ciencia, investigación y universidades (PID2024-156274NB-C22 a través del MICIU/AEI/10.13039/501100011033/FEDER, UE). Isaac Estevan ha recibido financiación del Ministerio de Ciencia, Innovación y Universidades, España, de estancias sénior en centros extranjeros [PRX24/00031].

REFERENCIAS

1. Anguera, M. T., Magnusson, M.S., Jonsson, G.K. (2007). Instrumentos no estándar: Planteamiento, desarrollo y posibilidades. *Avances en Medición*, 5, 63-82
2. Bandeira, P. F., Duncan, M., Pessoa, M. L., Soares, Í., da Silva, L., Mota, J., Martins, C. (2020). TGMD-2 short version: evidence of validity and associations with sex, age, and BMI in preschool children. *Journal of Motor Learning and Development*, 8(3), 528-543. <https://doi.org/10.1123/jmld.2019-0040>
3. Bandeira, P. F., Estevan, I., Duncan, M., Lenoir, M., Lemos, L., Romo-Perez, V., Valentini, N., Martins, C. (2025). A Multilayer Network Model for Motor Competence from the View of the Science of Complexity. *Sports Medicine*, 55(2), 245-254. <https://doi.org/10.1007/s40279-024-02164-4>
4. Bardid, F., Rudd, J. R., Lenoir, M., Polman, R., Barnett, L. M. (2015). Cross-cultural comparison of motor competence in children from Australia and Belgium. *Frontiers in Psychology*, 6, 964. <https://doi.org/10.3389/fpsyg.2015.00964>

Validación de un instrumento de competencia motriz

5. Bardid, F., Vannozzi, G., Logan, S. W., Hardy, L. L., Barnett, L. M. (2019). A hitchhiker's guide to assessing young people's motor competence: Deciding what method to use. *Journal of Science and Medicine in Sport*, 22(3), 311-318. <https://doi.org/10.1016/j.jsams.2018.08.007>
6. Barnett, L. M., Stodden, D., Cohen, K. E., Smith, J. J., Lubans, D. R., Lenoir, M., Lander, N. J. (2016). Fundamental movement skills: An important focus. *Journal of Teaching in Physical Education*, 35(3), 219–225. <https://doi.org/10.1123/jtpe.2014-0209>
7. Barnett, L. M., Webster, E. K., Hulteen, R. M., De Meester, A., Valentini, N. C., Lenoir, M., Pesce, C., Getchell, N., Lopes, V. P., Robinson, L. E. (2022). Through the looking glass: A systematic review of longitudinal evidence, providing new insight for motor competence and health. *Sports Medicine*, 52(4), 875-920, <https://doi.org/10.1007/s40279-021-01516-8>
8. Blanco-Villaseñor, A., Castellano, J., Hernández-Mendo, A., Sánchez-López, C. R., Usabiaga, O. (2014). Aplicación de la TG en el deporte para el estudio de la fiabilidad, validez y estimación de la muestra. *Revista de Psicología del Deporte*, 23(1), 131-137.
9. Bland, J., Altman, D. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal*, 314, 572.
10. Brennan, R. L. (2003). *Coefficients and Indexes in Generalizability Theory*. University of Iowa.
11. Cairney, J., Clark, H., Dudley, D., Kriellaars, D. (2019). Physical literacy in children and youth—A construct validation study. *Journal of Teaching in Physical Education*, 38(2), 84-90, <https://doi.org/10.1123/jtpe.2018-0270>
12. Cao, Y., Zhang, C., Guo, R., Zhang, D., Wang, S. (2020). Performances of the Canadian Agility and Movement Skill Assessment (CAMSA), and validity of timing components in comparison with three commonly used agility tests in Chinese boys: An exploratory study. *PeerJ*, 8, e8784. <https://doi.org/10.7717/peerj.8784>
13. Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling: a multidisciplinary journal*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>
14. Cronbach, L. J., Gleser, G.C., Nanda, H., Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. John Wiley and Sons
15. Duncan, M. J., Martins, C., Ribeiro Bandeira, P. F., Issartel, J., Peers, C., Belton, S., O'Connor, N. E., Behan, S. (2022). TGMD-3 short version: Evidence of validity and associations with sex in Irish children. *Journal of Sports Sciences*, 40(2), 138-145. <https://doi.org/10.1080/02640414.2021.1978161>
16. Estevan, I., Menescardi, C., García-Massó, X., Barnett, L. M., Molina-García, J. (2021). Profiling children longitudinally: A three-year follow-up study of perceived and actual motor competence and physical fitness. *Scandinavian Journal of Medicine & Science in Sports*, 31, 35-46. <https://doi.org/10.1111/sms.13731>
17. Estevan, I., Molina-García, J., Queralt, A., Álvarez, O., Castillo, I., Barnett, L. (2017). Validity and Reliability of the Spanish Version of the Test of Gross Motor Development–3. *Journal of Motor Learning and Development*, 5(1), 69-81. <https://doi.org/10.1123/jmld.2016-0045>
18. Fleiss, J.L. (1981) *Statistical method for rates and proportions*. Wiley.
19. Fornell, C., Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39–50.
20. Franco Hidalgo-Chacon, J. P., Rodríguez-Arteche, I. Martínez-Aznar, M. M. (2022). ¿Qué hacen los estudiantes de Educación Primaria españoles fuera del horario académico? Actividades extraescolares. *Revista Complutense de Educación*, 33(3), 459-474. <https://doi.org/10.5209/rced.74490>
21. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.

22. Hernández-Mendo, A., Blanco-Villaseñor, A., Pastrana, J. L., Morales-Sánchez, V., Ramos-Pérez, F. J. (2016). SAGT: Aplicación informática para análisis de generalizabilidad. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, 11(1), 77-89.
23. Henseler, J., Ringle, C. M., Sarstedt, M. (2015). *A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling*. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
24. Hu, L., Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
25. Hulteen, R. M., Barnett, L. M., True, L., Lander, N. J., del Pozo Cruz, B., Lonsdale, C. (2020). Validity and reliability evidence for motor competence assessments in children and adolescents: A systematic review. *Journal of Sports Sciences*, 38(15), 1717-1798. <https://doi.org/10.1080/02640414.2020.1756674>
26. Hulteen, R. M., Morgan, P. J., Barnett, L. M., Stodden, D. F., Lubans, D. R. (2018). Development of foundational movement skills: A conceptual model for physical activity across the lifespan. *Sports Medicine*, 48(7), 1533-1540, <https://doi.org/10.1007/s40279-018-0892-6>
27. Huotari, P., Heikinaro-Johansson, P., Watt, A., Jaakkola, T. (2018). Fundamental movement skills in adolescents: Secular trends from 2003 to 2010 and associations with physical activity and BMI. *Scandinavian Journal of Medicine & Science in Sports*, 28(3), 1121-1129. <https://doi.org/10.1111/sms.13028>
28. Isorna, M., Rial, A., Vaquero-Cristóbal, R. (2014). Motivaciones para la práctica deportiva en escolares federados y no federados. *Retos. Nuevas Tendencias en Educación Física, Deporte y Recreación*, (25), 80-84. <https://doi.org/10.47197/retos.v0i25.34485>
29. John, A., Wainwright, N., Goodway, J. D., Williams, A. (2024). The Relationship between Physical Activity and Motor Competence of Foundation Phase Children in Wales during the School Day. *Children*, 11(6), 629. <https://doi.org/10.3390/children11060629>
30. Jylänki, P., Mbay, T., Hakkarainen, A., Sääkslahti, A., Aunio, P. (2022). The effects of motor skill and physical activity interventions on preschoolers' cognitive and academic skills: A systematic review. *Preventive Medicine*, 155, 106948. <https://doi.org/10.1016/j.ypmed.2021.106948>
31. Lafave M. R., Butterwick, D. J. (2014). Generalizability theory study of athletic taping using the technical skill assessment instrument. *Journal of Athletic Training*, 49(3), 368–372. <https://doi.org/10.4085/1062-6050-49.2.22>
32. Logan, S. W., Ross, S. M., Chee, K., Stodden, D. F., Robinson, L. E. (2018). Fundamental motor skills: A systematic review of terminology. *Journal of sports sciences*, 36(7), 781-796. <https://doi.org/10.1080/02640414.2017.1340660>
33. Martínez, R. (1995). Teoría de la generalizabilidad (I): Conceptos básicos. Estudios G.: Modelos de efectos aleatorios. En R. Martínez. *Psicometría. Teoría de los tests psicológicos y educativos*. Síntesis.
34. McKenzie, T.L. (2010). Observing physical activity and its contexts. *Research Quarterly for Exercise and Sport*, 81(2), 113–122. <https://doi.org/10.1080/02701367.2010.10599656>
35. Medina, J., Delgado, M. A. (1999). Metodología de entrenamiento de observadores para investigaciones sobre E.F. y deporte en las que se utilice como método la observación. *Motricidad*, 5, 69-86.
36. Menescardi, C., Estevan, I., Falcó, C., Hernández-Mendo, A. (2017). Generalizability theory applied to olympic taekwondo combats. *European Journal of Human Movement*, 39, 65-81.

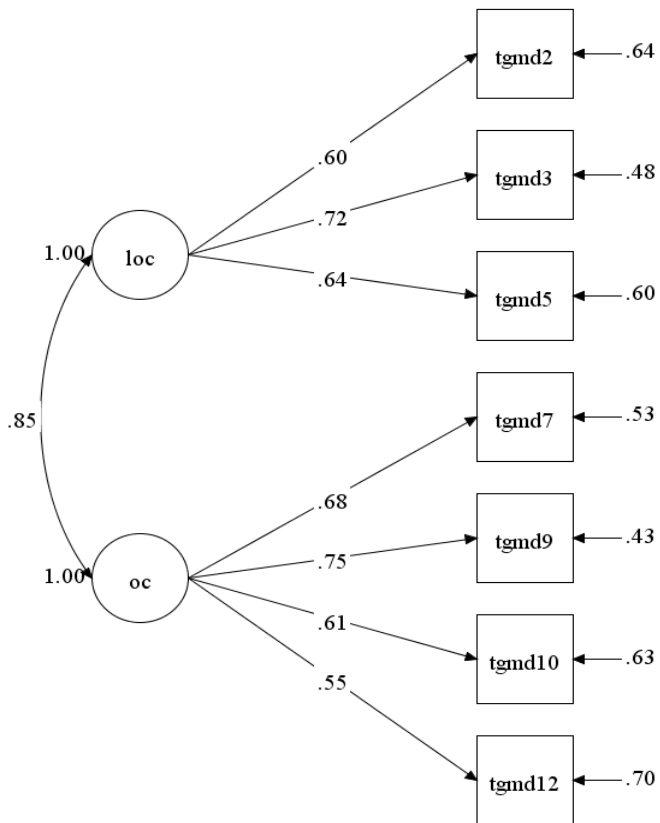
Validación de un instrumento de competencia motriz

37. Menescardi, C., Villarrasa-Sapiña, I., Lander, N., Estevan, I. (2022). Canadian Agility Movement Skill Assessment (CAMSA) in a Spanish Context: Evidences of Reliability and Validity. *Measurement in Physical Education and Exercise Science*, 26(3), 245-255. <https://doi.org/10.1080/1091367X.2021.2020794>
38. Morales-Sánchez, V., Pérez-López, R., Morquecho-Sánchez, R., Hernández-Mendo, A. (2016). Generalizabilidad y gestión deportiva. *Cuadernos de Psicología del Deporte*, 16(1), 161-170.
39. Newell, K. M. (2020). What are fundamental motor skills and what is fundamental about them? *Journal of Motor Learning and Development*, 8(2), 280-314. <https://doi.org/10.1123/jmld.2020-0013>
40. Reilly, T., Collins, K. (2008). Science and the Gaelic sports: Gaelic football and hurling. *European Journal of Sport Science*, 8(5), 231-240. <https://doi.org/10.1080/17461390802251851>
41. Robinson, L. E., Stodden, D. F., Barnett, L. M., Lopes, V. P., Logan, S. W., Rodrigues, L. P., D'Hondt, E. (2015). Motor competence and its effect on positive developmental trajectories of health. *Sports Medicine*, 45(9), 1273-1284. <https://doi.org/10.1007/s40279-015-0351-6>
42. Stodden, D. F., Goodway, J. D., Langendorfer, S. J., Roberton, M. A., Rudisill, M. E., Garcia, C., Garcia, L. E. (2008). A developmental perspective on the role of motor skill competence in physical activity: An emergent relationship. *Quest*, 60(2), 290-306. <https://doi.org/10.1080/00336297.2008.10483582>
43. Ulrich, D. A. (2017). Introduction to the special section: Evaluation of the psychometric properties of the TGMD-3. *Journal of Motor Learning and Development*, 5(1), 1-4.
44. Valentini, N. (2024). Motor Skill Assessment in Children and Adolescents. En A. García-Hermoso (Ed.), *Promotion of Physical Activity and Health in the School Setting* (pp. 133-163). Springer.
45. Valentini, N. C., Rudisill, M. E., Ribeiro-Bandeira, P. F., Hastie, P. A. (2018). The development of a short form of the Test of Gross Motor Development-2 in Brazilian children: Validity and reliability. *Child: Care, Health and Development*, 44(5), 759-765. <https://doi.org/10.1111/cch.12598>
46. Van Rossum, T., Fowweather, L., Hayes, S., Richardson, D., Morley, D. (2021). Expert recommendations for the design of a teacher-oriented movement assessment tool for children aged 4-7 years: A Delphi study. *Measurement in Physical Education and Exercise Science*, 25(4), 283-293. <https://doi.org/10.1080/1091367x.2021.1876070>
47. Webster, E. K., Ulrich, D. A. (2017). Evaluation of the Psychometric Properties of the Test of Gross Motor Development—Third Edition. *Journal of Motor Learning and Development*, 5(1), 45-58. <https://doi.org/10.1123/jmld.2016-0003>
48. Wickel, E.R., Welk, G. J. (2010). Applying generalizability theory to estimate habitual activity levels. *Medicine & Science in Sports & Exercise*, 42(8), 1528-1534. <https://doi.org/10.1249/MSS.0b013e3181d107c4>
49. Woods, C., Ng, K., Britton, U., McClelland, J. F., O'Keeffe, B., Sheikhi, A., ... O'Brien, W. (2023). *Children's Sport Participation and Physical Activity Study 2022*. Physical Activity for Health Research Centre, Department of Physical Education and Sport Sciences, University of Limerick, Limerick, Ireland, Sport Ireland and Healthy Ireland, Dublin, Ireland and Sport Northern Ireland, Belfast, Northern Ireland.
50. Zeng, N., Ayyub, M., Sun, H., Wen, X., Xiang, P., Gao, Z. (2017). Effects of physical activity on motor skills and cognitive development in early childhood: A systematic review. *BioMed Research International*, 2017(1), 2760716. <https://doi.org/10.1155/2017/2760716>

ANEXOS

Figura S1

Análisis factorial confirmatorio del modelo de dos factores de la versión corta de Duncan et al.



Nota. Índices de ajuste aceptables ($X^2 = 32.134$; CFI= 0,96; TLI= 0,93; RMSEA= 0,08 SRMR= 0,04; $AVE_{loc} = 0,43$; $CR_{loc} = 0,69$; $AVE_{oc} = 0,43$; $CR_{oc} = 0,75$).